Universiteit Leiden
The Netherlands

Opleiding Informatica

Collaborative web-based tool support for IAMA assessments on public sector algorithms.

Koen Bron

Supervisors:
Prof.dr.ir. J.M.W. Visser Dr. D.M. Pelt

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

04/07/2024

Abstract

With the rapid developments made in the AI landscape also comes the need to make sure that these models adhere to certain ethical and qualitative standards. Recently the Dutch government has passed legislation that will require IAMA assessments to be performed on high-risk public sector algorithms they have in use or development. This thesis researches whether web-based tool support for IAMA is desired and effective. To this end we designed a proof of concept Django application called IAMA Checker that covers phases 1 & 4 of IAMA with a number of added functionalities to improve on the original IAMA document. Functionalities such as user collaboration, the production of an assessment summary as a PDF file and tracking the completion status of IAMA's, alongside content to further assist in answering the IAMA questions.

Through a demonstration with six government employees using the Thinking-aloud methodology combined with a questionnaire utilising aspects of the TAM, SUS and NPS evaluation methods, we evaluated the user reception of IAMA Checker and determined whether IAMA Checker was perceived as a significant improvement on the original IAMA. We found that employees appreciated what new functionalities IAMA Checker brought to the table and that such a web-based support tool for IAMA is mainly seen as a desired improvement on the currently available IAMA document, although it is important for that tool to have a user interface that is as unambiguous and intuitive as possible to accommodate for the large and diverse audience.

Contents

1 Introduction

The ever increasing capabilities of artificial intelligence (AI from here on) has resulted in an increased reliance by the Dutch government on AI systems [1] to assist them in governing the population. This, of course, creates the need for these AI systems to be as reliable and consistent as possible, seeing as they have more and more influence over the lives of civilians than ever before [1]. The *Toeslagen Affaire* [2] that came to light in 2019 in the Netherlands poses as a powerful example of how algorithms in use by the government can become gruesomely detrimental to people's lives should they not perform as intended.

As a means to push for more responsible AI use, algorithm auditing tools were created. These are meant to start a discussion around an algorithm on whether it is lawful and feasible to pursue or continue development and continuously assess the algorithm at various stages of its life cycle. They can be instrumental in deciding on AI policies by virtue of them being broad, operationalizable, flexible, iterative, guided and participatory [3]. One such tool is the Impact Assessment Mensenrechten en Algoritmes (translated: impact assessment human rights and algorithms, from here on IAMA) [4] with a focus on determining the legality of using an algorithm that infringes on human rights.

In this bachelor's thesis we develop a web-application called IAMA Checker[1]to serve as a proof of concept that aims to optimise and improve the process of performing a IAMA while offering additional support for people inexperienced with IAMA, thus making IAMA's more accessible. Furthermore, we produce evaluations of this proof of concept on whether it accomplishes these goals. From processing these evaluations into quantitative and qualitative data it can be determined whether further development of this tool or a likewise product is worthwhile.

1.1 Motivation & Research Question

The Dutch government as of late is making efforts to be more transparent about their algorithms by including them in their algorithm register [5], which lists the algorithms in use by the government along with several characteristics to explain the algorithm's function and serves provide more insight on how they can affect civilians. As of now there are unfortunately a number of issues with the register, such as algorithm registration not being mandatory and the algorithms that are registered often lack crucial information like the outcome of an impact assessment such as a IAMA [6, 7].

Legislation has since passed which will, in the near future, make it required by law to register high-risk algorithms that are currently in use or still in development at any level of government [8]. Furthermore, legislation has passed which will require a IAMA to be executed on these algorithms [9] and the results of the subsequent assessments are required to be published in the register or a different publicly accessible platform [10].

Completing IAMA's for the backlog of algorithms in use by different divisions on all levels of the government will prove to be daunting task. Especially seeing as the current form of IAMA is already pushing the limits of how effective and efficient a 95 page collaborative PDF file can be in

[1]Source code available: https://github.com/KoenBron/IAMA-checker

a setting that requires a thorough and lengthy discussion amongst a team that consists of as much as sixteen different professions. So a pilot program, created by the creators of IAMA at Universiteit Utrecht and the Rijks ICT Gilde (RIG from here on)[2], was put in place to assess the experiences of multiple teams across different divisions of government whilst performing a IAMA under supervised circumstances. From these 8 documented test cases it was concluded that the current form of IAMA comes with a number of issues that hinder both its effectiveness and efficiency[11], though the general response to IAMA was remarkably positive. There was a lot of feedback given on various aspects of IAMA, but here are the most notable issues relevant to the scope of this bachelor thesis in no particular order:

- Requiring a large group of people to be available at the same time, which can lead to scheduling issues.
- Confusion on what constitutes as a good and complete answer to the questions posed in IAMA.
- Not enough context given to the questions, which lead to confusion on the relevance of some questions.
- The assessment taking a long time to complete.
- Uncertainty about the goal of performing a IAMA.

With this feedback in mind we found a need for a software solution that would aid its users in performing a IAMA, thus creating the research question:

Research Question Does collaborative, web-based tool support improve IAMA assessments performed on public sector algorithms?

The answer to this question can be produced through the answering of the following sub questions:

1. Does web-based tool support for IAMA provide improved support for individuals in producing answers of noticeably better quality than they otherwise would?
2. Does web-based tool support make the process of performing a IAMA assessment on public sector algorithms be perceived as more accessible and efficient?
3. Do the additional functionalities that are made possible by the web-based format significantly improve the experience of individuals when performing a IAMA assessment on public sector algorithms?

1.2 Research Approach

Van Schaik performed similar research in his bachelor thesis that has inspired the research approach we decided to take in this thesis. He created a software tool called CompAI that streamlines the usage of CapAI [12]. CapAI is a compilation of multiple metrics and a tool that perform a conformity assessment with the AI act that recently came into effect by the EU [13]. He produced Comp AI that both guides its users through performing the assessment and provides all the necessary information and insights needed to correctly complete the assessment.

[2] https://www.rijksorganisatieodi.nl/rijks-ict-gilde

To design IAMA Checker we employed the design method outlined in the Design Science Research Methodology (from here on DSRM) by Peffers et al. [14]. This process started with analysing the original IAMA document and determining what parts should be covered by IAMA Checker. As this document is 95 pages and implementing all of it would take too much time, we decided to incorporate phases 1 and 4 of IAMA into IAMA Checker. This way it can still serve as a viable proof of concept that demonstrates the potential of a web-based IAMA software solution. Through the DSRM iterations we continuously improved the functionalities and presentation of IAMA Checker until we found that it sufficed as a proof of concept for this thesis.

To measure how well the final version of IAMA Checker achieves its goals stated in Section 1.1, we demonstrated the product to six different government employees that were at some level familiar with IAMA. We employed the thinking-aloud protocol during these demonstrations as to gather information on their experience and general sentiment towards IAMA Checker. These six individuals were handed a questionnaire afterwards to further evaluate their experience. This questionnaire combined aspects of TAM (Technology Acceptance Model) [15], SUS (System Usability Scale) [16] and the NPS (Net Promoter Score) [17]. It also includes several open-ended questions to gather additional insights on the feature reception and other qualitative data. After combining all of the feedback gathered we interpret the results and consequently draw conclusions concering the research questions stated in Section 1.1.

1.3 Thesis overview

This bachelor thesis is done at the Leiden Institute of Advanced Computer Science (LIACS) of Leiden University under the supervision of Prof. J. M.W. Visser and Dr. D.M. Visser. From this point on we start with introducing the necessary background information in Section 2 concerning IAMA. Then in Section 3 we will go into detail on the methods that we used to design IAMA Checker and the methods we employed to evaluate the success of IAMA Checker.

Afterwards we discuss the overall design and implementation of IAMA Checker in Section 4. The results of the filled out questionnaires are listed in Section 5. We will then discuss the results and draw conclusions from them about IAMA Checker supported by the gathered verbal feedback in Section 6 and lastly, we will conclude this paper in Section 7.

2 Background

This background section goes over the necessary background information relevant to understanding this thesis meant to provide context on the current AI landscape in the Netherlands.

2.1 AI Act & Algorithm Risks

Now that AI has become so powerful and versatile, it is only natural that we would see a rapid increase in adoption across many fields and disciplines. Unfortunately, the rate of adoption is significantly higher than the rate at which we develop risk-management systems, due to the processes of devising new legislation and product standards around AI/algorithms taking longer than creating new software [1]. Though we do see the European Union making major strides by introducing the

new AI act, which serves as an EU wide standard for regulating all manor of AI and algorithms based on their perceived function [18].

The AI act regulates algorithms by dividing them into 4 separate risk-classes that will determine what regulations apply to the algorithm, these categories are listed below with some examples to clarify the concept [19]:

- Unacceptable (Social scoring systems, manipulative AI)
- High (Managing employees, AI infringing on constitutional rights)
- Limited (AI systems that need transparency about their functioning)
- Minimal (AI in video games, spam filters)

In compliance with the AI act and due to recent motions that have passed the House of Representatives [9, 8], it will soon be required by law to perform a IAMA on these high-risk algorithms and certain other entities that provide public services [20].

2.2 Algorithm Register

The algorithm register, called algoritmeregister in Dutch [5], is an initiative launched by the Dutch government in December of 2022. It was conceived as a means to be more transparent and informative about the (high-risk) algorithms that Dutch government employs after the events of the "Toeslagen Affaire" had greatly diminished the public's trust in how the government utilises algorithms [21].

The register allows different branches of government to document the algorithms they have in use. This documentation does not only contain information about the algorithms function, but also provides more meta-data like who is responsible for the algorithm and whether it's in use or not. Furthermore, it allows for documentation on the potential impact the algorithm has on the population.

The register is still in its beginning stages and is continuously improved [22]. This means that as of now it is still optional for government branches to publish their algorithms in the register and there is no requirement for the registration to be complete [7]. Though, with efforts from the Dutch government to create legislation around the register [23, 8, 21] and legislation found in the AI act created by the EU [18], it will soon be required by law that the Dutch government registers its high-risk algorithms in the register.

2.3 IAMA

Impact assessment Mensenrechten en Algoritmes, or IAMA [4], is an algorithm auditing tool that assesses whether an algorithm is compliant with generally accepted best practices and it has a special focus on assessing the impact an algorithm can have on constitutional rights. It was commissioned by the Ministry of Interior and Kingdom Relations of the Dutch government and developed by Utrecht University in 2020-2021 [24].

It is important to state that IAMA is not a checklist that guarantees an algorithm is compliant when filled out correctly. IAMA is meant to lay many questions out in the open concerning ethical development and use of an algorithm at the different stages of an algorithm's life cycle. With these questions a discussion can be started that aims to let parties responsible for the algorithm assess whether their development and/or use of the algorithm is both ethical and lawful. The answers to these questions will then be recorded in the IAMA and can serve as a recollection of the decisions made around the algorithm. It can then possibly offer guidance in the later stages of development and provide accountability for the decision processes around an algorithm.

While performing the IAMA it may be concluded that the algorithm can absolutely not be used and that development must promptly be halted. It can also conclude that there are certain risks involved due to the algorithm for example infringing on constitutional rights, this however does not mean that the algorithm cannot be put to use. Only by looking at countermeasures that mitigate or eliminate the risks can it be determined whether the algorithm is allowed to be operational.

The IAMA consist of four phases, those are in order:

1. **Why**, concerning the reason the algorithms should be used and what its goals are.
2. **What**, concerning the data that comes in the algorithm and the way the algorithms works with the data.
3. **How**, concerning output the algorithm produces and how the output will be used.
4. **Human rights**, a special section for determining whether certain constitutional rights are infringed upon and how the accompanying risks can be minimised.

The first three phases go over more general questions concerning ethics, much like questions that can found in other auditing tools, as IAMA is closely related to a number of impact assessments like the DPIA [25] and other guidelines on responsible use of algorithms. This causes some answers that are required in the IAMA to be interchangeable with answers given in other impact assessments or likewise audits.

The fourth phase is what makes IAMA truly unique and is the main reason it was commissioned by the Dutch government. It aims to find all the constitutional rights that might be infringed upon by the algorithm and then opens a discussion on how the algorithm can still responsibly be put to use.

3 Methods

3.1 Design Science Research Methodology

Design Science Research Methodology (DSRM from here on) serves as a commonly accepted framework for design science processes in the world of information systems. Created by Peffers et al. [14] to formally introduce a standard methodology for the information systems discipline, as they stated that "Design Science is of importance in a discipline oriented towards creating successful artefacts" [14, p. 46].

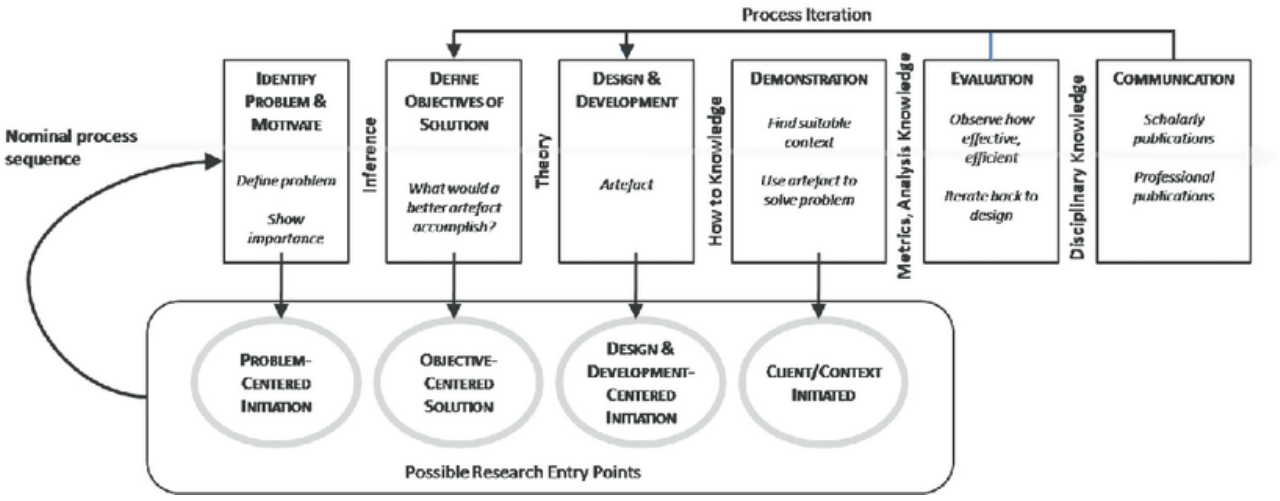


Figure 1: Overview of the Design Science Research Methodology inserted from Peffers et al. [14]

In Figure 1 we see DSRM illustrated as a process consisting of six steps with the later steps looping back to earlier ones, thus making the design process iterative and implementing a way to incorporate intermediate feedback into subsequent designs decisions.

We first considered using DSRM as we saw that it had served C. van Schaik well in the creation of his web-application CompAI [12]. From there we found that what DSRM offered aligned well with what we needed from a design methodology. It was concluded that it neatly formalises an approach to designing IAMA Checker in which we could intimately involve people likely to use a product like IAMA Checker in the design process. This sentiment was further solidified as the tooling conceptualised by IAMA Checker is meant to be adapted to frequent daily usage and we require only one optimal design methodology in order to design an objective improvement over the original IAMA document [26, p. 9].

From here on is shown how the DSRM steps shown in Figure 1 were utilised to design IAMA Checker and thus create a software solution that can help answer our research questions outlined in Section 1.1. Though the first two steps (identify problem and motivate & derive objectives of solution) are covered in Section 1.1, so these are not presented here. Furthermore, we applied DSRM throughout three of the iterations as is outlined by the 'Process Iterations' arrow in Figure 1 going back from step 6 to step 3. The following subsections report on what these steps entailed on a per iteration basis.

3.1.1 Design & development

Iteration 1 First came the condensing all the content relevant to phases 1 & 4 into a compact version with essential information necessary to perform a IAMA, from there the places where additional context or information was needed were identified and provided with that content. This way the original intentions of IAMA were not lost while users were facilitated through short but concise information. Following this, we designed and developed IAMA Checkers initial features based on the initial grievances we acquired through struggling with the collaborative PDF format of the original IAMA document.

Iteration 2 & 3 We implemented additional features based on the feedback gathered from previous iterations and simultaneously rearranged development priorities, as to create the most complete version of IAMA Checker whilst remaining within the allotted time window and inside the scope of this thesis.

3.1.2 Demonstration

Iteration 1 We demonstrated the first prototype of IAMA Checker along with our intentions for the final version to a member of the RIG who co-lead the government pilot programs with IAMA mentioned in Section 1.1. This person was also partly responsible for reporting on these pilot programs in a report, which was shared with us and provided insight into what improvements IAMA Checker should focus on [11].

Iteration 2 We demonstrated an intermediate version of IAMA Checker to a group of 3 RIG members that all have experience in working with IAMA and performing IAMAs. Here we mostly showcased the tool while working through a fictional use-case, which allowed IAMA Checker to show all of its implemented functionalities while the design decisions were thoroughly explained. After the demonstration the participants were further informed of our design plans for the final version.

Iteration 3 We demonstrated the final version of IAMA Checker to five members of the RIG and one member of the Dutch Tax Administration, all of which had varying levels of experience with IAMA. These subjects had full access to the final version of IAMA Checker with minimal interference, although intervention was inevitable when a subject seemed to get stuck on user interface elements or if they didn't organically explore certain features of IAMA Checker.
We did, for the sake of time, explain how to use the different features in IAMA, as is otherwise explained through the included tutorial, for it would have required a rather long time to understand this information relative to the time we had allotted for the demonstrations. These demonstrations took place during work-hours on workdays, which unfortunately limited the available time per demonstration to 1-1.5.

3.1.3 Evaluation & Communication

Iteration 1 & 2 We had the respective participants comment on their experience during their demonstration and any remarks deemed useful and within the scope of this thesis were collected. Even more so, feedback that could possibly be turned into future work was also gathered. Alongside the presented feedback on the state of IAMA Checker as it was presented then and there, we engaged the participants in conversation about their personal experiences with IAMA as a means to encourage them to think of the functionalities they would like to see in a web-based software tool such as IAMA Checker.

Iteration 3 Here is presented a high-level overview of the evaluation methodologies that were employed to produce the evaluation results that can be viewed in Section 5, they will be elaborated

on further in the subsequent subsections.

During the demonstrations we employed the thinking-aloud methodology[27] to gather as much information of any constructive nature about the final implementation of IAMA-Checker. After the demonstrations, the six subjects were handed a questionnaire to assess both the usefulness and the ease-of-use of IAMA Checker in the form of quantifiable results, along with six open ended questions meant to gather additional qualitative data. This evaluation questionnaire is elaborated on further in section 3.2.2 and it was made using Google Forms[3]. Lastly, all respondents answered within 24 hours upon concluding their respective demonstrations, as the time constraints in some cases didn't allow for thequestionnaire to be filled out upon finishing the demonstration.

3.2 Evaluation Methods

3.2.1 Thinking-aloud

Thinking-aloud is a usability testing method where the subjects quite literally think aloud continuously while performing certain tasks. This way they verbalise their thoughts on the matter at hand [27, 28] which can be converted into usable feedback on the user experience. We decided to employ this method during the final demonstrations as thinking-aloud is found to consistently provide valuable insights into both the minor and major problems of design implementations during software usability testing [29, 30, 31], while also proving to be a very robust and easy to implement methodology [28].

3.2.2 Questionnaire

The questionnaire is meant to gather concrete data on how its respondents experienced using IAMA Checker as an alternative to traditional way of performing IAMA's and in doing it gathers the data necessary to answer the research questions from Section 1.1.

We were inspired by the original Technology Acceptance Model (TAM from here on) [15] when structuring the questionnaire. It proposes that the intention a person has of using a certain product is influenced more by the attitude a persons has towards that product rather than directly by the features that the product bolsters, which in turn influences the acceptance and adoption rate of that product [15, 32]. TAM states that there are two factors that determine this attitude: perceived usefulness and perceived ease-of-use, as can be seen in Figure 2.
Since predicting the adoption rate of our prototype is not entirely relevant to the aims of this thesis, it was quickly concluded that utilising TAM in its entirety was unnecessary. Instead we decided to take the concepts of estimating the perceived usefulness and perceived ease-of-use to quantify the overall perceived IAMA Checker experience. These concepts are thus integrated into two separate sections of the questionnaire.

The original version of TAM is criticised for being incapable of recognising the role that product innovation plays in the attitude held towards that product [33], which is vital information to gather for this thesis. Therefore a separate section of the questionnaire contains six open ended questions guided towards comparing IAMA Checker to the original document along with the NPS

[3]Google Forms: https://www.google.nl/intl/nl/forms/about/

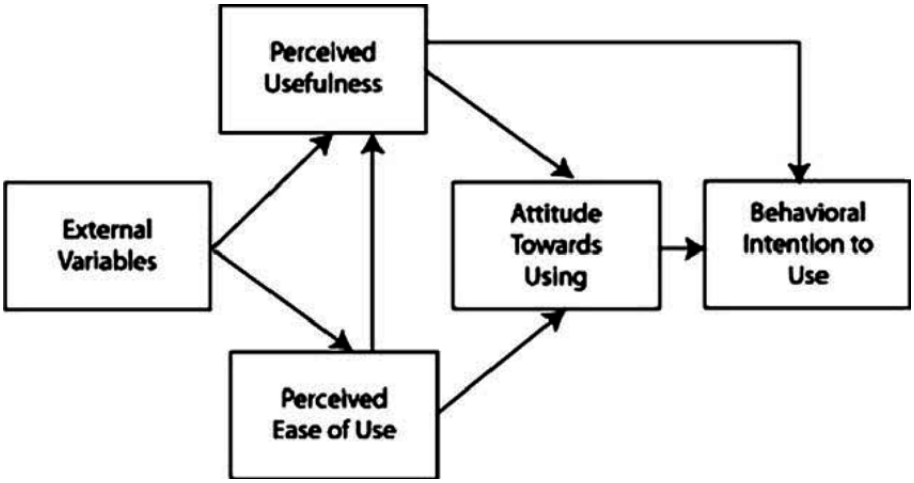


Figure 2: The TAM model inserted from Davis et al. [15]

score[17], which acts as a summarising question on the overall impression IAMA Checker leaves. The methodologies used in the three sections of the questionnaire are further explained below.

Perceived Usefulness The six questions proposed in the original TAM paper[15] were used to assess the perceived usefulness, as the simplicity allows for few questions to still offer us significant insights on how useful the target audience perceives IAMA Checker. The questions have been adapted to the context of performing a IAMA with IAMA Checker and the answers are given on a Likert scale[34] ranging from 1 (strongly disagree) to 5 (strongly agree).

1. Using this product at work would help me complete IAMA related tasks faster.
2. Using this product would improve my IAMA related job performance.
3. Using this product would improve my productivity when working with IAMA.
4. Using this product would increase my effectiveness at working with IAMA.
5. Using this product would make it easier to do my job when working with IAMA.
6. I would find this product useful when working with IAMA.

Perceived Ease-of-use The SUS was created to measure a system's usability depending on the context it is used in [16], with the answers to its ten questions measured on the same Likert scale used in TAM. Though the original paper labels SUS as "A quick and dirty usability scale"[16], it has proven itself on multiple occasions to be a very robust and reliable tool [35, 36]. Attrakdiff 2 was considered but ultimately not used, as its focus on gauging user stimulation and users' intent to continue using a product [37] exceeds the scope of estimating the ease-of-use. Furthermore, UMUX [38] and UMUX-LITE[39] have been proposed as viable replacements for SUS and one that consists of only 4 (or 2 in UMUX-Lite). [40] As our small sample does not allow for statistically significant results, it was decided that the discourse on whether or not SUS and UMUX(-LITE) produce very similar results [40, 41] would not influence the ease-of-use method selection process.

So, UMUX(-LITE) was ultimately passed on as the larger number of questions in SUS allowed for the extraction of more detailed information from the results, while not taking so long that the questionnaire would become cumbersome or repetitive to answer. The SUS questions posed in the questionnaire are the following:

1. I think I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I though the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought the was too much inconsistency in the system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident in using the system.
10. I needed to learn a lot of things before I could get going with this system.

Open questions The open questions are partly inspired by the response types to system features used in the Kano model [42]. Due to its focus on wording how the user experiences the features present in a product, which is critical information that cannot be gathered solely through the TAM and SUS questions. Lastly, the NPS question is meant to conclude the questionnaire by giving a score on how likely the participant is to recommend IAMA Checker to a colleague [17] on a Likert scale ranging from 1-10, summarising the respondent's general stance on IAMA Checker. Then questions of this section are the following:

1. Which must-have features are missing in this product?
2. Which must-have features are present in this product?
3. Which nice-to-have features are missing in this product?
4. Which nice-to-have features are present in this product?
5. In which manner do you perceive this product as an improvement on the original IAMA document?
6. In which manner do you perceive this product as an degradation on the original IAMA document?
7. On a scale from 1 to 10, how likely are you to recommend this product to a colleague?

4 Design

4.1 Django Implementation

Django[4] was the web-framework used to create the back-end of IAMA Checker and Django templating[5] was used as the front-end. Django was chosen for its quick development time so the development focus could be directed towards implementing new features, as a lot of the basic functionalities needed in a web-application, such as user authentication and session handling, are covered by Django right out of the box.

The back-end architecture was implemented with a focus on long term development due to IAMA Checker posing as proof of concept that would need further development to be turned into a complete product. This means that dynamic content, such as the questions and their context, is stored in the database following a standardised JSON format. As a consequence, adding the remaining phases and their questions mostly boils down converting the information into the aforementioned JSON format and loading it into the database. For more information see the documentation on our GitHub page: https://github.com/KoenBron/IAMA-checker.

The free version of the Material Dashboard Django by Creative Tim[6] combined with Bootstrap[7] formed the basis for the front-end as its components were customised to fit the needs of our user interface designs.

4.2 Feature Design

IAMA Checker boasts a number of features that leverage the advantages of the web-application format against the original IAMA document. The most important of these features are subdivided amongst the following sections and further explained.

4.2.1 User Collaboration

Arguably the most important feature of IAMA Checker is the editors system that grants other registered user accounts editorial privileges for an assessment. Performing a IAMA can require a group as large as sixteen people to work on it by virtue of the different professions needed to weigh in on the discussions preceding the answers to the questions in IAMA. The editors system relieves its users of the necessity to be present in the same space at the same time.

Actively tracking the parties responsible for answer contributions safeguards the integrity of a IAMA and serves as a recollection of accountability on the decision made throughout the IAMA. IAMA Checker implements this with its contributor system, see Figure 3, where a user can register a contributor by name, organisation name and profession to each question. This registration is to be done manually for each question, as to not require every registered contributor to have a IAMA Checker account. Though the burden of repetitive registration of the same collaborator is alleviated through a shorthand for adding previously register collaborators as can be seen in Figure 3.

[4]Django: https://www.djangoproject.com
[5]Django templating: https://docs.djangoproject.com/en/5.0/ref/templates/
[6]Material Dashboard Django: https://www.creative-tim.com/product/material-dashboard-django
[7]Bootstrap 5.3: https://getbootstrap.com/docs/5.3/getting-started/introduction/

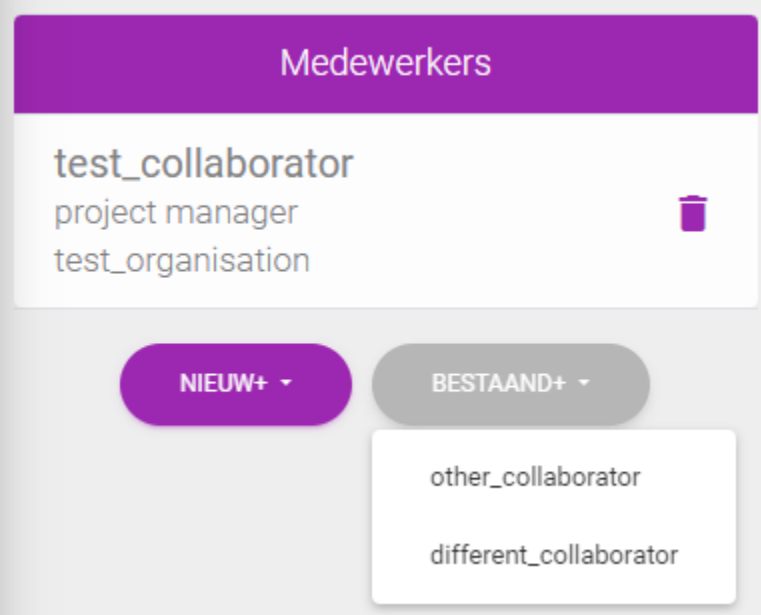


Figure 3: Image of the collaborators system as implemented in IAMA Checker (in Dutch).

4.2.2 Summarising Document

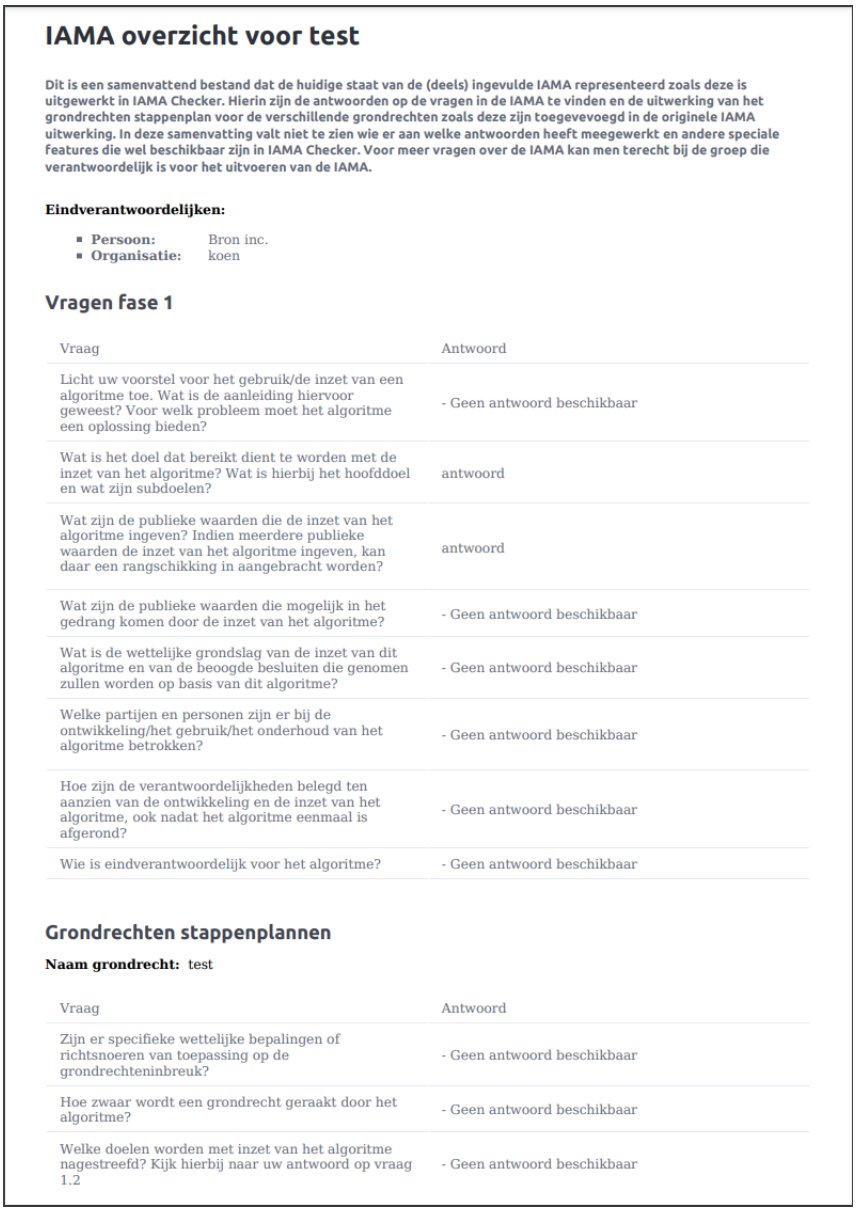
IAMA Checker generates a downloadable summarising PDF document with the given answers to all questions included in IAMA Checker. The document consists of an introductory text, the listing of parties that are ultimately responsible for the IAMA and the given answers, an example is presented in Figure 4.

4.2.3 Assessment Status Tracking

IAMA Checker tracks the completion status of an assessment. Furthermore, it tracks the completion status of all questions in the assessment. A question has one of three statuses: *unanswered*, *answered* or *reviewed*. Unanswered and answered are self-explanatory and kept track of automatically, but a question is considered to be reviewed only when all the people involved in answering the question sign off on the final answer. Meaning the reviewed status needs to be set manually and an assessment is in turn considered complete only when all the questions have been marked as reviewed. Every submitted answer to a question is stored and accessible in the that question's answer history. Stored is the content of the answers, the time and date of submission and the username of the author. The answer history is meant to preserve the discussion happening during the answering of the question, to this end it is made explicitly clear in the tutorial that submitting intermittent answers containing comments or questions for colleagues is considered a best practice when using IAMA Checker.

4.2.4 Assisting Features

Each question page in IAMA Checker consists of: the question as originally posed in IAMA, optional additional context and explanation on the nature of the question and brief instructios on what content is expected from a good answer. The context to a question is made optional as to avoid



IAMA overzicht voor test

Dit is een samenvattend bestand dat de huidige staat van de (deels) ingevulde IAMA representeerd zoals deze is uitgewerkt in IAMA Checker. Hierin zijn de antwoorden op de vragen in de IAMA te vinden en de uitwerking van het grondrechten stappenplan voor de verschillende grondrechten zoals deze zijn toegevevoegd in de originele IAMA uitwerking. In deze samenvatting valt niet te zien wie er aan welke antwoorden heeft meegewerkt en andere speciale features die wel beschikbaar zijn in IAMA Checker. Voor meer vragen over de IAMA kan men terecht bij de groep die verantwoordelijk is voor het uitvoeren van de IAMA.

Eindverantwoordelijken:

- **Persoon:** Bron inc.
- **Organisatie:** koen

Vragen fase 1

Vraag	Antwoord
Licht uw voorstel voor het gebruik/de inzet van een algoritme toe. Wat is de aanleiding hiervoor geweest? Voor welk probleem moet het algoritme een oplossing bieden?	- Geen antwoord beschikbaar
Wat is het doel dat bereikt dient te worden met de inzet van het algoritme? Wat is hierbij het hoofddoel en wat zijn subdoelen?	antwoord
Wat zijn de publieke waarden die de inzet van het algoritme ingeven? Indien meerdere publieke waarden de inzet van het algoritme ingeven, kan daar een rangschikking in aangebracht worden?	antwoord
Wat zijn de publieke waarden die mogelijk in het gedrang komen door de inzet van het algoritme?	- Geen antwoord beschikbaar
Wat is de wettelijke grondslag van de inzet van dit algoritme en van de beoogde besluiten die genomen zullen worden op basis van dit algoritme?	- Geen antwoord beschikbaar
Welke partijen en personen zijn er bij de ontwikkeling/het gebruik/het onderhoud van het algoritme betrokken?	- Geen antwoord beschikbaar
Hoe zijn de verantwoordelijkheden belegd ten aanzien van de ontwikkeling en de inzet van het algoritme, ook nadat het algoritme eenmaal is afgerond?	- Geen antwoord beschikbaar
Wie is eindverantwoordelijk voor het algoritme?	- Geen antwoord beschikbaar

Grondrechten stappenplannen

Naam grondrecht: test

Vraag	Antwoord
Zijn er specifieke wettelijke bepalingen of richtsnoeren van toepassing op de grondrechteninbreuk?	- Geen antwoord beschikbaar
Hoe zwaar wordt een grondrecht geraakt door het algoritme?	- Geen antwoord beschikbaar
Welke doelen worden met inzet van het algoritme nagestreefd? Kijk hierbij naar uw antwoord op vraag 1.2	- Geen antwoord beschikbaar

Figure 4: Example of the generated summarising document, produced by IAMA Checker to present the answers given in an assessment as a comprehensive pdf file (in Dutch).

displaying large amounts of text that could deter users from reading any of the presented information and to avoid disrupting the users' workflow with excessive scrolling. The user is further provided with links to external resources presented in the original IAMA on a per question basis and they are accompanied by instructions on how to utilise them. The appendix information included in the back of IAMA is in IAMA Checker now readily available for the questions that require them.

Each question page presents its users with the professions required to produce a well informed answer, along with the disciplines that are only recommended to include in the answering process. The phase introduction pages also includes an overview of these professions per question, so the required individuals needed for that phase can be anticipated before starting on that phase. An example of this table is shown in Figure 5.

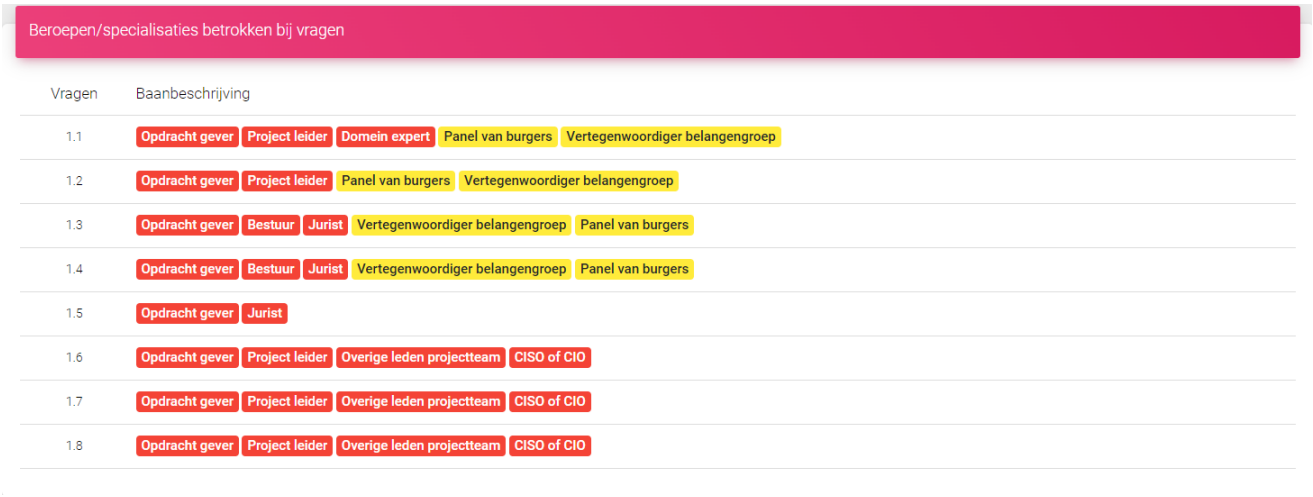


Figure 5: A table containing an overview of the disciplines required (red buttons) and recommended (yellow buttons) for each question of phase 1 of IAMA (in Dutch).

4.3 User Experience

In the appendix are included a number of screenshots of IAMA Checker to provide an idea of the IAMA Checker user interface a user is prestented with when utilising the application. Furthermore, below in Figure 6 is a flowchart included, made with lucidchart[8], representing the rough user experience when using IAMA Checker to perform one or more IAMA's.

The chart is dived into two sections, where the top part covers the navigation bar that is at all times available to the user and the bottom part depicts the core user experience. The rounded buttons represent entry points, the squares represent actions or (sections of) screens available in IAMA Checker and the diamonds are decision points. For the decision diamonds only the "yes" paths are graphed as the "no" allows users to interact with navigation bar or question index in order to move through the tool.

[8]Lucidchart: https://www.lucidchart.com/blog/how-to-make-a-user-flow-diagram

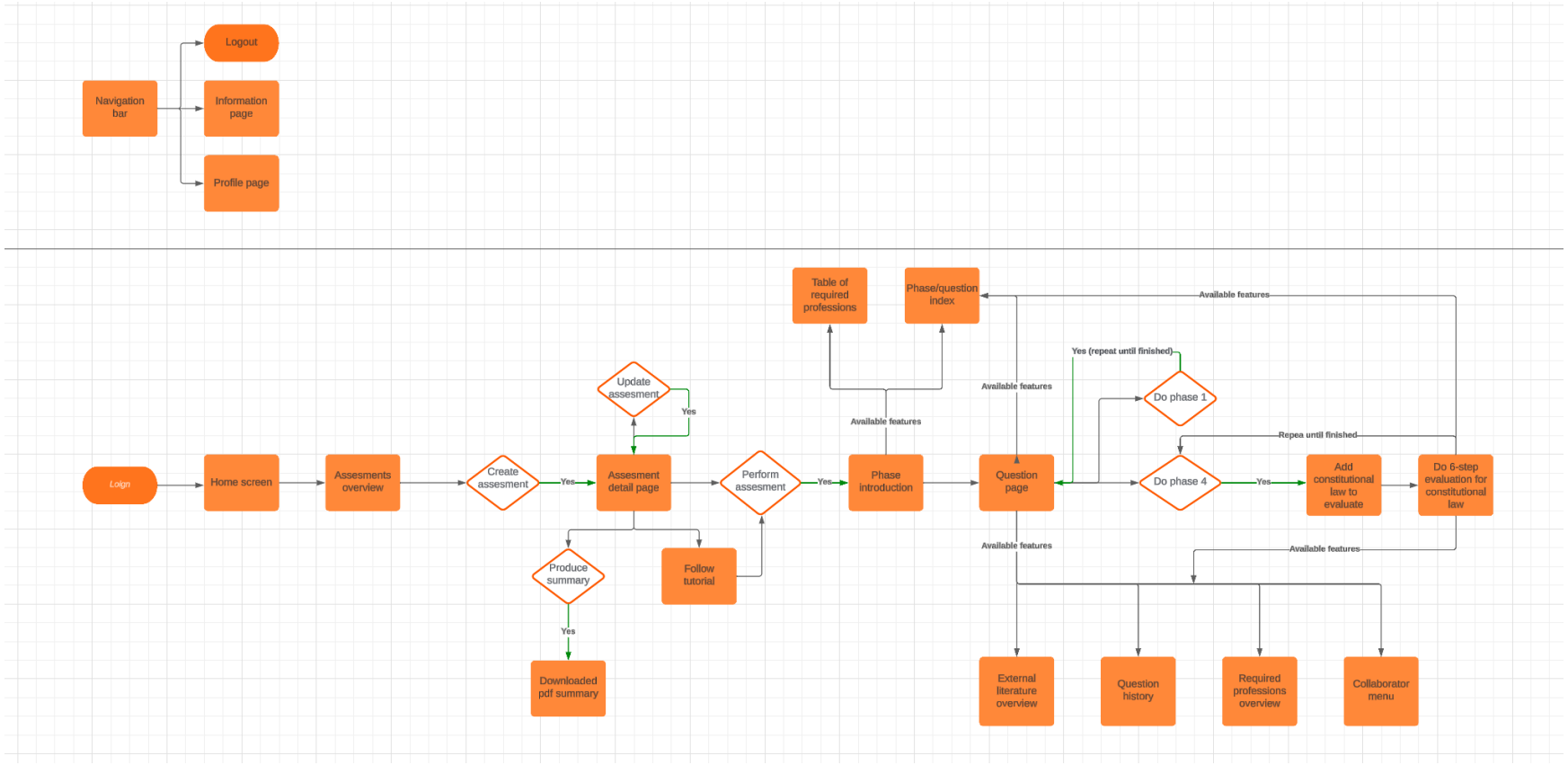


Figure 6: User flowchart representing the user experience when working on IAMA assessment with IAMA Checker.

5 Evaluation Results

This section presents the results gathered through the evaluation methodologies presented in Section 3.2 . The results for the multiple choice questions regarding usefulness and ease-of-use are respectively presented Tables 1 & Table 2 along with the NPS results in Table 3. In Section 5.2 we provide a summary of the results to the open questions of the questionnaire.

5.1 Quantifiable Results

In Table 1 are the results to the TAM questions we employed to evaluate usefulness, rated on a Likert [34] scale ranging from 1 (strongly disagree) to 5 (strongly agree). The answers are mostly grouped around a rating of 4 (agree) with only questions 2 & 3 partially diverting to a rating of 3.

In Table 2 are the results of the SUS questions for evaluating the ease-of-use, again rated on a Likert scale ranging from 1-5. Here we do not show such a decisively positive reaction as can be observed in Table 1.

#	Question	1 (strongly disagree)	2 (disagree)	3 (neutral)	4 (agree)	5 (strongly agree)
1	Using this product at work would help me complete IAMA related tasks faster.	(0) 0%	(0) 0%	(0) 0%	**(6) 100%**	(0) 0%
2	Using this product would improve my IAMA related job performance.	(0) 0%	(0) 0%	**(4) 66.7%**	**(2) 33.3%**	(0) 0%
3	Using this product would improve my productivity when working with IAMA.	(0) 0%	(0) 0%	(0) 0%	**(6) 100%**	(0) 0%
4	Using this product would increase my effectiveness at working with IAMA.	(0) 0%	(0) 0%	**(1) 16.7%**	**(5) 83.3%**	(0) 0%
5	Using this product would make it easier to do my job when working with IAMA.	(0) 0%	(0) 0%	(0) 0%	**(5) 83.3%**	**(1) 16.7%**
6	I would find this product useful when working with IAMA.	(0) 0%	(0) 0%	(0) 0%	**(4) 66.7%**	**(2) 33.3%**

Table 1: Results from the six IAMA Checker evaluations to the TAM [15] questions used to determine the usefulness of IAMA Checker, rated on a Likert scale [34] from 1 (strongly disagree) to 5 (strongly agree). Results presented in the format: (quantity) percentage%.

#	Question	1 (strongly disagree)	2 (disagree)	3 (neutral)	4 (agree)	5 (strongly agree)
1	I think I would like to use this system frequently.	(0) 0%	**(2) 33.3%**	**(2) 33.3%**	**(2) 33.3%**	(0) 0%
2	I found the system unnecessarily complex.	**(2) 33.3%**	**(2) 33.3%**	**(2) 33.3%**	(0) 0%	(0) 0%
3	I though the system was easy to use.	(0) 0%	**(1) 16.7%**	**(1) 16.7%**	**(4) 66.7%**	(0) 0%
4	I thik that I would need the support of a technical person to be able to use this system.	**(3) 50%**	**(2) 33.3%**	(0) 0%	**(1) 16.7%**	(0) 0%
5	I found the various functions in this system were well integrated.	(0) 0%	(0) 0%	**(1) 16.7%**	**(4) 66.7%**	**(1) 16.7%**
6	I thought there was too much inconsistency in the system.	**(3) 50%**	**(2) 33.3%**	**(1) 16.7%**	(0) 0%	(0) 0%
7	I would imagine that most people would learn to use this system very quickly.	(0) 0%	(0) 0%	**(1) 16.7%**	**(3) 50%**	**(2) 33.3%**
8	I found the system very cumbersome to use.	**(2) 33.3%**	**(4) 66.7%**	(0) 0%	(0) 0%	(0) 0%
9	I felt very confident in using the system.	(0) 0%	(0) 0%	**(3) 50%**	**(3) 50%**	(0) 0%
10	I need to learn a lot of things before I could get going with this system.	**(3) 50%**	**(3) 50%**	(0) 0%	(0) 0%	(0) 0%

Table 2: Results from the 6 IAMA Checker evaluations to the SUS [16] questions,used to determine the ease-of-use and rated on a Likert scale [34] from 1 (strongly disagree) to 5 (strongly agree). Results presented in the format: (quantity) percentage%.

On a scale from 1 to 10, how likely are you to recommend this product to a colleague?									
1	2	3	4	5	6	7	8	9	10
(0) 0%	(0) 0%	(0) 0%	(0) 0%	(0) 0%	(0) 0%	**(2) 33.3%**	**(3) 50.0%**	**(1) 16.7%**	(0) 0%

Table 3: Results from the 6 IAMA Checker evaluations to the NPS [17] question. Answers given on a Likert scale [34] rating from 1 to 10, in the format: (quantity) percentage%.

5.2 Results Open Question

Below is summarised the most relevant and useful feedback gathered through the six open questions.

1. Which must-have features are missing in this product? Additional collaborative features, most suggested of which is the ability to leave comments for other users on highlighted pieces of text. This would greatly facilitate in the discussion element pertinent to IAMA. Also, the ability to have multiple user accounts working on an answer at the same time is suggested, as opposed to the current implementation where only the last saved answer is shown and a refresh of the page is required to update it.

2. Which must-have features are present in this product? All of the features named in Section 4.2 were explicitly mentioned in the accumulated answers to this question.

3. Which nice-to-have features are missing in this product? Here the ability to create two different types of summarising documents was proposed, one document that contained all the names of contributors to each answer that can be used as an internal document for keeping track of responsibilities and one document to be posted publicly that contains no names as to protect the privacy of the participants.

Furthermore, as of now a user is to explicitly save their answer before moving to another page in the application. Meaning that it is very possible that after producing an answer, the user could lose that answer by accidentally moving to another page or refreshing the browser due to IAMA Checker acting on page refreshes. So the proposed ability to automatically save your temporary answers would prove to be a nice quality-of-life feature.

Last is suggestion to use commonly seen icons as a means to intuitively draw user attention towards certain actions they might want to take, to make the user interface more intuitive.

4. Which nice-to-have features are present in this product? The tracking of completion statuses for each of the constitutional rights covered in phase 4 of IAMA was appreciated along with the fact the fact that viewing context to a question is optional.

5. In which manner do you perceive this product as an improvement on the original IAMA document? First and foremost is the improvement on collaboration, allowing teams to perform a IAMA in small groups that are not required to be present at the same location or available at the same time. It was found to make performing a IAMA an iterative process, one that aided in tackling a IAMA step by step while it can also be resumed at any point during the assessing process. The web-application format was found to make it noticeably faster to navigate the required pages and present the required information on a per question basis, whereas a PDF file requires you to scroll back and forth between pages to revisit its content.

Lastly, producing an easy to style PDF file containing all the content of the assessment is greatly appreciated and seems to be regarded as one of the main selling points for IAMA Checker, seeing as PDF manipulation is notoriously difficult and this feature makes it significantly easier to communicate about IAMA's internally as well as externally.

6. In which manner do you perceive this product as a degradation on the original IAMA document? Though it is faster to navigate to the page you are looking for in IAMA Checker that holds true if you know where to go, there is a consensus among the responses to this question that it is too easy to 'get lost' in IAMA Checker and lose overview. The subsequent lack of providing users with an intuitive way to get back to where they were or clear assistence on how to go to their desired page is remarked as a quality that proves detrimental to the user experience.

6 Discussion

Before discussing the results, it is first and foremost important to reflect on the fact that IAMA Checker serves as a proof-of-concept and by virtue of that fact it is an incomplete product. It does not include phases 2 & 3, which makes up roughly half of IAMA's original content. Although we found these phases very similar in structure to phase 1 and concluded that only phases 1 & 4 would suffice in demonstrating the potential of IAMA Checker, we can't be 100% sure certain that it does until a version of IAMA Checker is evaluated where this missing content is implemented. Even more so, features that the evaluation participants would not interact with during the IAMA Checker demonstration, like inviting users to edit an assessment or user authentication, are primitively implemented. This was intentional as the development setting of this thesis does not require these features to be implemented in a way that would suffice in deployment. As a consequence a lot of development time is saved on parts of the application that the participants to an evaluation aren't likely to interact with. Thus the implementation of these features was simplified in a manner such that the end result doesn't accurately represent the time and effort it would take to implement them when adhering to the quality and security standards that are expected of a tool in use by the government.

Concerning the results, we acknowledge that our sample size of six lends no statistical significance to the produced results. Though it does hold true that useful insights into the reception of IAMA Checker can still be gained from closely analysing the results presented in Section 5. This is the reason that we decided against computing the SUS score [16] and the NPS score [17], as their lack of statistical significance in this case makes it that these scores don't provide more insights into the reception of IAMA Checker than is readily apparent when looking at Tables 1 2 3. Future works should increase the sample size, with a focus on including employees from local governments in the evaluation as to not only demonstrate/pilot IAMA Checker to employees of the central government. Seeing as requiring IAMA's to be performed on algorithms at every level of government would mostly include the algorithms in use by the numerous cities in the Netherlands. Including more participants from these sample pools in the evaluation process would solidify the results so that statistical analyses can be performed on them whilst also making the results more representative IAMA Checker's intended user base. We further found that the experiment setup we chose, where we sat down with the evaluation participants during work hours, made it that the evaluations sometimes were rushed, as each participant worked at their own page through the application. We mentioned earlier that it also led us to explain the tutorial presented in IAMA Checker instead of letting the user organically read and apply it. Future evaluations would do well to carve out anywhere between 1,5-2 hours for the evaluations to make sure as little intervention by the evaluators is needed.

First we discuss the percieved usefulness results in Table 1. It is apparent that IAMA Checker's functionalities are generally perceived as useful as the ratings mostly lean towards a rating of 4 (agree) or higher. This claim is further supported by the answers to open question 2 where clear appreciation is shown for the features discussed in Section 4.2 with a special focus on the collaborative features, administrative features and the summarising document. These features make for a positive influence on the accessibility and effectiveness of IAMA Checker, see question 1, 3 & 5. During the evaluations it seemed that the ability to generate a summarising PDF was an especially welcome surprise, as is shown in the answers to open question 2. Furthermore, as is suggest the results of open question 3, the upgrade to producing two variants of the documents, one for internal use and the other for external, is expected to yield great practical applications. This practical improvement is a logical next step for IAMA Checker and would be rather simple to implement due to the HTML to PDF pipeline we implemented.

Question 2 of Table 1 shows us that IAMA is not really thought to significantly improve IAMA related job performance, which is to be expected as users that are quite experienced with IAMA aren't likely to leverage IAMA Checker's added functionalities as a means to give measurably better answers than they otherwise would have given. This is also further supported through the verbal feedback gathered during the final demonstrations, where comments were made with the implication that the additional information provided for questions does not yield noticeable benefits for users that know what is expected of them. Rather, it has the possibility of hindering the users workflow by getting in the way of quickly answering a question, though small user interface changes like making the viewing of all additional information optional could remedy this. Conversely, the additional information presented by IAMA Checker can support less experienced users in answering the IAMA questions, which is supported by the score to question 4. In any case, the absence of 1 and 2 ratings to all questions gives the impressions that merits of IAMA Checker functionalities are recognised by the results.

We find through the answers to open questions 1 & 3 a further need for focusing on collaborative functionalities. IAMA Checker already improves in this aspect compared to the original IAMA PDF, though it is not that big of a step up. Using JavaScript to combat the static nature of the templated Django implementation would allow for development of the desired features, which are commonly found in other collaboration tools such as Google Docs[9] such as real-time collaborative editing and leaving comments for other collaborators. The desire for these features is reflected in the answers to open questions 1 & 3.

Now we look at Table 2, where we are presented with some negative results in questions 1, 3 & 4 in contrast to the results in Table 1. These scores imply that users experienced some troubles when working their way through the tool. This is further reflected in the verbal feedback gathered through the demonstrations, where intervention was sometimes needed to get the participants back on track when they got stuck. It mostly comes down to the user interface not communicating enough on where you are in the tool and not indicating in an intuitive matter what actions are available to the user to proceed. This conclusion is further supported by the answers given to open question 6. It became clear through the demonstrations that the user interface must be as unambiguous as possible by not leaving anything to the users interpretation. This is because in

[9]Google docs: https://www.google.nl/intl/nl/docs/about/

practice, users need all the hand holding they can get to move intuitively through the tool, seeing as the pool of employees that are the intended users of IAMA Checker is very diverse in terms of experience, technical background and age. Suggestions like the implementation of commonly seen symbols proposed in the an answer to open question 3 can help combat this.
On the other hand, these are user interface issues that are not particularly difficult to resolve and IAMA Checker's ease-of-use was in general positively perceived. Complemented by the fact that IAMA Checker's accessibility is well regarded, see questions 4, 7 & 10. We can draw the conclusion that user interface appears to enable its users in effectively using IAMA Checker without too much difficulty, though it is certainly not without its issues that can hold it back from being considered for very frequent use, see question 1.

Currently, IAMA is becoming more and more widespread as its practice within all levels of algorithm governance becomes required by law. As a consequence, all the troubles that the current IAMA implementation entails will also be multiplied by this new adoption rate. All in all it means that IAMA isn't going away anytime soon and that the desire for an improved manner of performing IAMA's will only become more pressing. The NPS results in Table 3 shows us that the concept that IAMA Checker proposes is well received by our participants which is further embellished considering the NPS results indicate that this tool promises enough additional benefits to have it be rather seriously considered for recommendations to other people that work with IAMA. It appears to succeed in providing a collaborative web-based experience that supports its users in many different ways when performing a IAMA, though the user experience is not without its problems due to poor user interface design choices. It shows us that there is a future for web-based tooling support for IAMA, where these tools can make a meaningful difference in the Dutch AI landscape.

7 Conclusions and Further Research

In conclusion, we set out to research whether web-based tool support improves the process of performing a IAMA assessment on public sector algorithms and to that end we presented the implementation and subsequent evaluation of IAMA Checker as a prove of concept meant to answer this question. In regards to improvements on the answer quality, the results of our work led us to conclude that IAMA Checker doesn't enable individuals who are already quite experienced with IAMA to produce answers of a noticeably better quality. Although the additional resources presented by IAMA Checker for answering the questions do support less experienced individuals in understanding what is expected from their answers and can aid them in producing well informed answers.

Our results lead us to conclude the following facts on the perceived accessibility and efficiency of performing web-based, tool supported IAMA's. We can confidently state that the web-based environment allows for IAMA assessments to become more accessible by virtue of them being able to guide individuals significantly better than the original pdf format of IAMA can. Its ease-of-use is regarded positively, making it so that the advantages of utilising IAMA Checker are not difficult to leverage into performing IAMA's more efficiently.

The additional functionalities proposed by IAMA Checker through its implemented features are well received. They have shown to assist users by keeping them well informed of the state of

assessments and enabling them to convey information that is vital to preserving the answering processes happening throughout the IAMA questions, all the while not diminishing the legitimacy of the IAMA performed using this tool. The possibility to generate a PDF summary that reflects the state of an assessment was especially well regarded when working with public sector algorithms as it allows for convenient and concise communication to the public about the algorithms they have in use. Nevertheless, the collaborative environment created by IAMA Checker remains the main selling point when comparing with the IAMA document, even though it is quite primitively implemented.

The feedback to IAMA Checker made clear that future works should focus on improving the user experience by making the user interface less ambiguous and more intuitive, as to make it an even more accessible and intuitive experience that doesn't become frustrating through frequent usage. Second to that would be devoting more development time into the collaborative aspects of IAMA Checker, in an effort to make the tool more responsive. Additionally, the current evaluation was done with just six participants and further research would do well to increase the sample size by gathering participants from all levels of government that have public sector algorithms which are subject to a IAMA under the new legislation. That way statistical significance can be lent to the results.

All in all, this thesis has shown that there is a need and a want for web-based IAMA tooling and that such a tool can be genuinely proficient at improving the process of performing a IAMA on public sector algorithms, so further development of IAMA Checker, or a comparable tool, into a complete product would in accordance with to our findings be a worthwhile investment of time and resources.

References

[1] D. C. Algoritmes, "Rapportage ai & algoritmerisico's nederland," Authoriteit Persoonsgegevens, Tech. Rep., 2023, page 5.

[2] P. O. Kinderopvangtoeslag, "Ongekend onrecht," December 2020.

[3] D. Schiff, B. Rakova, A. Ayesh, A. Fanti, and M. Lennon, "Principles to practices for responsible ai: Closing the gap," 06 2020.

[4] J. Gerards, M. T. Schafer, I. Muis, and A. Vankan, "Impact assessment mensenrechten en algoritmes," July 2021, published by: Ministerie van Binnenlandse Zaken en Koninkrijksrelaties and Universiteit Utrecht. [Online]. Available: https://www.rijksoverheid.nl/documenten/rapporten/2021/02/25/impact-assessment-mensenrechten-en-algoritmes

[5] Rijkshoverheid, "Algoritmeregister," December 2022. [Online]. Available: https://algoritmes.overheid.nl/nl

[6] H. Raba, "Nauwelijks zicht op 'zwarte zoemende dozen' van overheid: 'algoritmeregister wassen neus'," *NOS*, June 2023.

[7] M. Janssen, "Data quality analysis for the dutch algorithm register," Bachelors thesis, Leiden University, August 2023.

[8] H. Dekker-Abdulaziz, "Motie van het lid dekker-abdulaziz over alle nieuwe hoogrisicoalgoritmes opnemen in het algoritmeregister en die onderwerpen aan een mensenrechtentoets," Tweede Kamer motie, June 2023.

[9] K. Bouchallikh and H. Dekker-Abdulaziz, "Motie van de leden bouchallikh en dekker-abdulaziz over verplichte impactassessments voorafgaand aan het inzetten van algoritmen voor evaluaties van of beslissingen over mensen download," Tweede Kamer motie, March 2022.

[10] C. voor Digitaliseren, "Toezegging publicatie mensenrechtentoetsen ai-gebruik (cxlvii)toezegging publicatie mensenrechtentoetsen ai-gebruik (cxlvii)," Eerste Kamer brief, March 2023.

[11] M. et al., "Het iama: best practices voor het toepassen van het impact assessment mensenrechten en algoritmes." Unpublished, Universiteit Utrecht, Rijks ICT Gilde, Tech. Rep., 2024.

[12] C. van Schaik, "Compai: A user-friendly software tool to support the capai procedure for ai act conformity assessment," Bachelor's thesis, Leiden University, 2023.

[13] F. et al., "capai - a procedure for conducting conformity assessment of ai systems in line with the eu artificial intelligence act," March 2022.

[14] P. et al., "A design science research methodology for information systems research," *Journal of Management Information Systems*, 2007.

[15] F. Davis, *User Acceptance of Information Systems: The Technology Acceptance Model (TAM)*, 1987. [Online]. Available: https://books.google.nl/books?id=Fzmw33E6YHQC

[16] J. Brooke, "Sus: A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, November 1995.

[17] F. Reichfeld, "The one number you need to grow," *Harvard Buiseness Review*, Decemember 2003, accessed on May 2024.

[18] E. Comission, "Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts," March 2024. [Online]. Available: https://artificialintelligenceact.eu/the-act/

[19] ——, "Ai act," Published as a high-level article on the AI act., accessed on May 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[20] J. van der Wijst, "Overeenstemming over de ai act," December 2023. [Online]. Available: https://bg.legal/overeenstemming-over-de-ai-act/

[21] "Kamerstukken ii 2022/23, 26643, nr. 924," Published as letter to the Tweede Kamer, October 2022.

[22] Rijksoverheid, "Algoritmeregister voor de overheid." [Online]. Available: https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/algoritmes/algoritmeregister/

[23] A. van Huffelen, Published as letter to the Tweede Kamer, December 2022. [Online]. Available: https://www.rijksoverheid.nl/documenten/kamerstukken/2022/12/21/kamerbrief-over-het-algoritmeregister

[24] U. U. Data School, "Impact assessment mensenrechten en algoritmes," accessed on May 2024. [Online]. Available: https://dataschool.nl/iama/

[25] A. Persoonsgegevens, "Data protection impact assessment," accessed on May 2024. [Online]. Available: https://www.autoriteitpersoonsgegevens.nl/themas/basis-avg/praktisch-avg/data-protection-impact-assessment-dpia

[26] J. Venable, John R.; Pries-Heje and R. L. Baskerville, "Choosing a design science research methodology," ser. ACIS 2017 Proceedings, no. 112, 2017. [Online]. Available: https://aisel.aisnet.org/acis2017/112

[27] K. Ericsson and H. Simon, *Protocol Analysis: Verbal Reports as Data*. MIT Press, 1984.

[28] J. Nielsen, "Thinking aloud: The #1 usability tool," *Nielsen Norman Group*, January 2012, accessed on June 2024.

[29] T. Z. S. McDonald, H. E. Edwards, "Exploring think-alouds in usability testing: An international survey," in *IEEE Transactions on Professional Communication*, vol. 55, March 2012, pp. 2,19, issue 1. [Online]. Available: https://ieeexplore.ieee.org/document/6151862/citations#citations

[30] O. Alhadreti and P. J. Mayhew, ""thinking about thinking aloud": An investigation of think-aloud methods in usability testing." BCS Learning & Development, July 2016.

[31] S. Denning, D. Hoiem, M. Simpson, and K. Sullivan, "The value of thinking-aloud protocols in industry: A case study at microsoft corporation," *Proc. Hum. Factors Soc. Annu. Meet.*, vol. 34, no. 17, pp. 1285–1289, Oct. 1990.

[32] P. Surendran, "Technology acceptance model: A survey of literature," *International Journal of Business and Social Research*, vol. 2, no. 4, pp. 175–178, August 2012.

[33] T. Z. William Ratjeana Malatji, Rene Van Eck, "Understanding the usage, Modifications, Limitations and Criticisms of Technology Acceptance Model (TAM)," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 6, pp. 113–117, 2020.

[34] R. Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, vol. 22, 1932.

[35] Bangor, Aaron, Kortum, P. T., Miller, and J. T., "The system usability scale (sus): an empirical evaluation," *International Journal of Human-Computer Interaction*, vol. 24, pp. 574–, 08 2008.

[36] J. Lewis, "The system usability scale: Past, present, and future," *International Journal of Human-Computer Interaction*, pp. 1–14, 03 2018.

[37] M. Hassenzahl, M. Burmester, and F. Koller, "Attrakdiff: A questionnaire to measure perceived hedonic and pragmatic quality," in *Proceedings of the Mensch & Computer 2003.* Stuttgart, Leipzig: Teubner, 2003, pp. 187–196.

[38] K. Finstad, "The usability metric for user experience," *Interacting with Computers*, vol. 22, pp. 323–327, 09 2010.

[39] J. Lewis, B. Utesch, and D. Maher, "Umux-lite: when there's no time for the sus," 04 2013, pp. 2099–2102.

[40] M. Schrepp, J. Kollmorgen, and J. Thomaschewski, *A Comparison of SUS, UMUX-LITE, and UEQ-S*, 02 2023, pp. Vol. 18, Issue 2, February 2023 pp. 86–104.

[41] S. Borsci, S. Federici, S. Bacci, M. Gnaldi, and F. Bartolucci, "Assessing user satisfaction in the era of user experience: Comparison of the sus, umux, and umux-lite as a function of product experience," *International Journal of Human–Computer Interaction*, vol. 31, no. 8, pp. 484–495, 2015.

[42] E. Sauerwein, F. Bailom, K. Matzler, and H. Hinterhuber, "The kano model: How to delight your customers," *International Working Seminar on Production Economics*, vol. 1, 01 1996.

A Appendix

Here you find a number of screenshots of different pages a user encounters all through their experience in working with IAMA Checker. The text in the screenshots is in Dutch considering that IAMA Checker is an application made with Dutch government officials in mind as its intend user base.

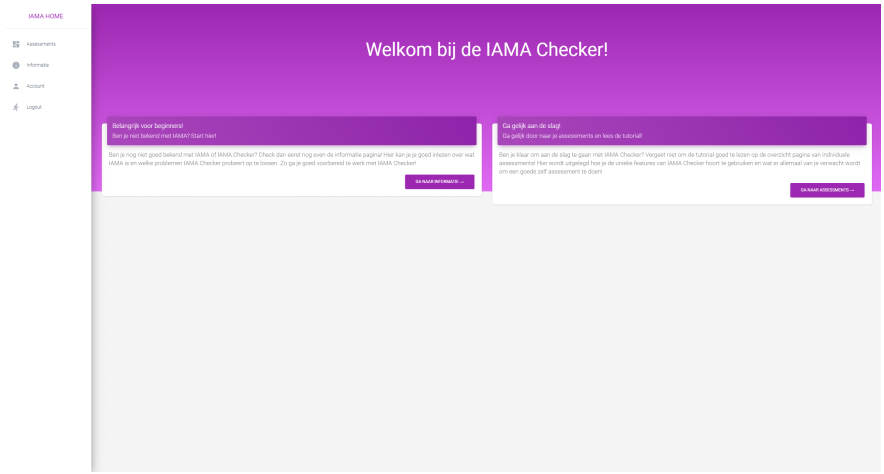


Figure 7: The home screen, meant to greet logged in users and providing them with the option for more information on IAMA or going straight to their assessments.

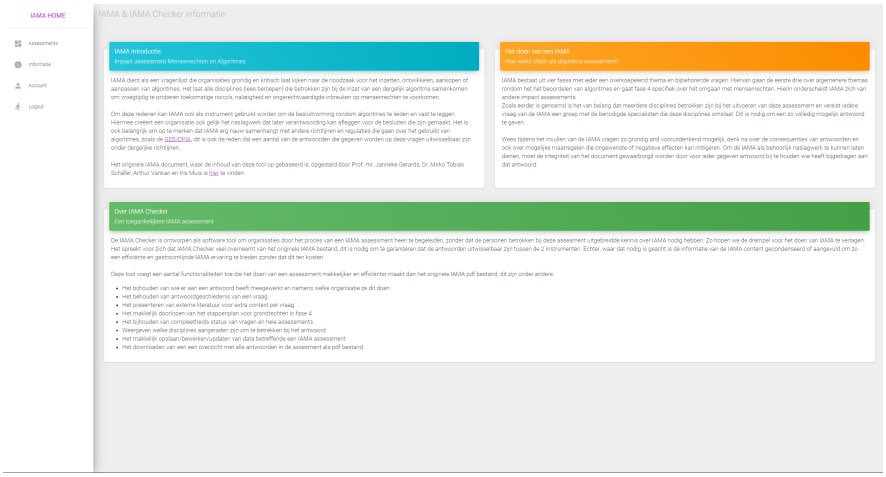


Figure 8: The information screen that presents background information on IAMA and the introduces IAMA Checker.

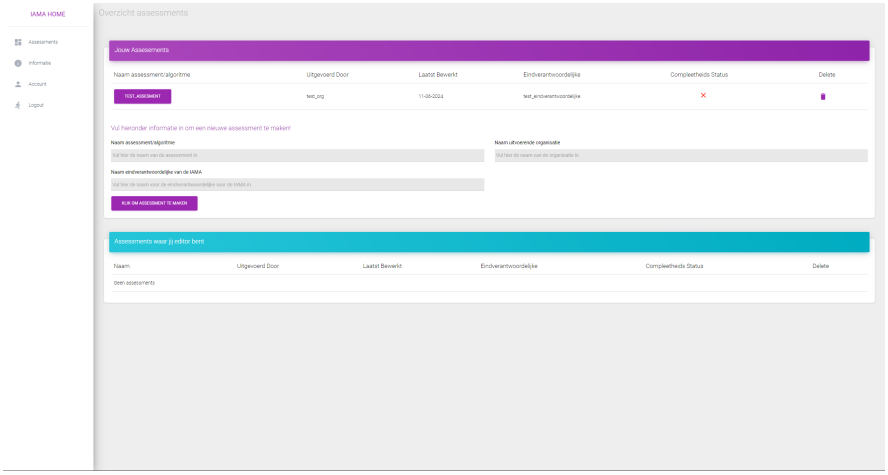


Figure 9: The screen that lists the created assessments and allows for the creation of assessments.

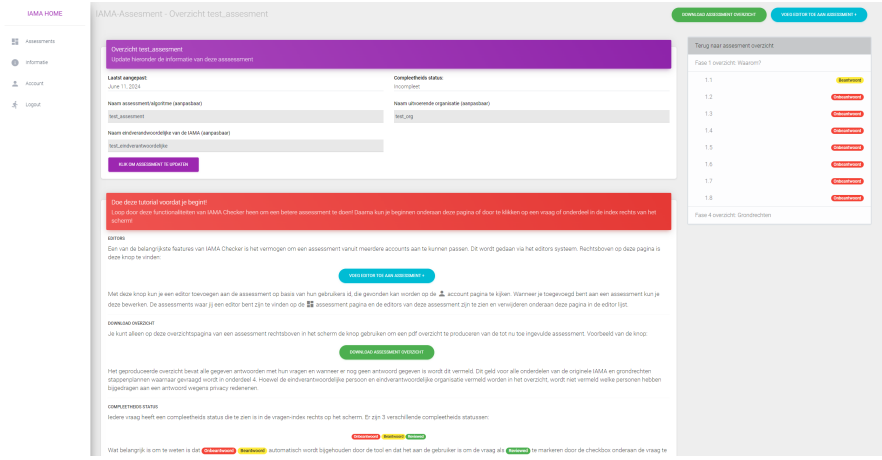


Figure 10: The screen that presents the assessment details and the features tutorial explaining how to utilise its functionalities.

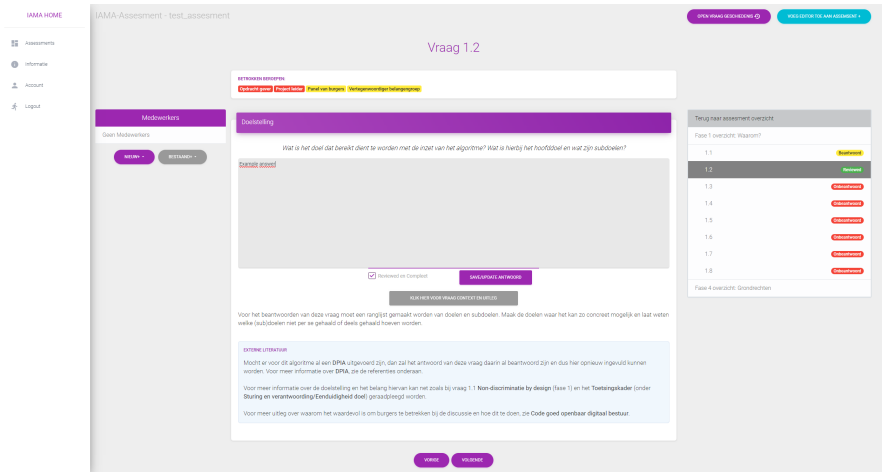


Figure 11: A phase 1 question screen.

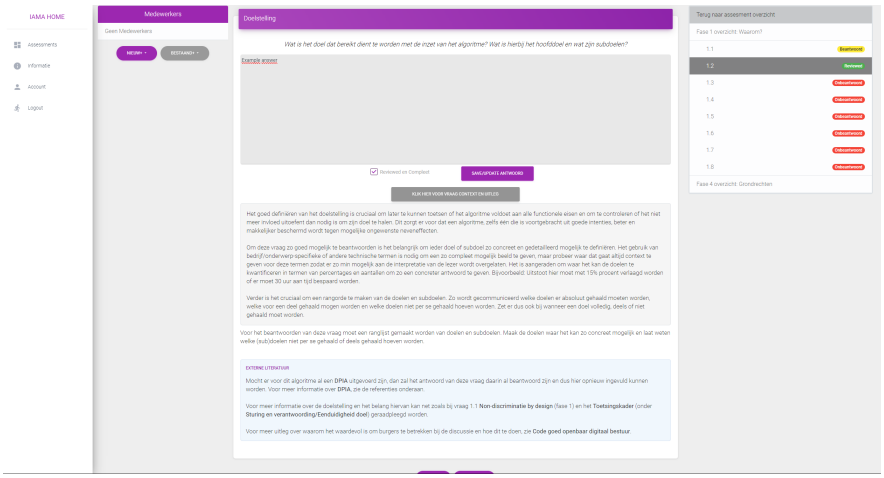


Figure 12: A phase 1 question screen, with the context extended.

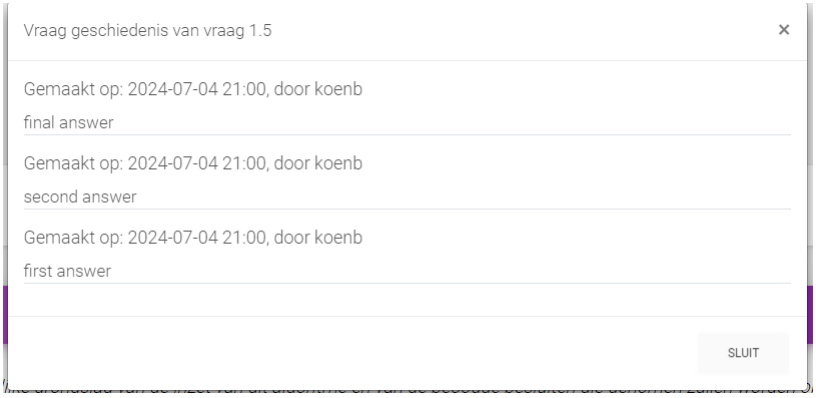


Figure 13: Image of how the answer history is presented to its users.

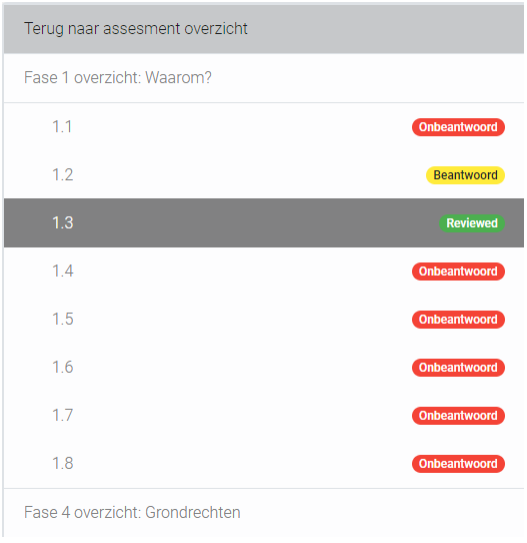


Figure 14: The question index used to straightforwardly navigate between question pages and phase introductions and present the answer statuses of each question.