Master's Thesis

User Interface for a Rehabilitation Training Decision Support System

Andersen, Frederik

201508251

Hansen, Theis Egsgaard

201506770

Supervisor

Christian Fischer Pedersen, cfp@ece.au.dk

May 31, 2021

Preface

This master's thesis has been made to fulfill the requirements for the Master of Science degree in Computer Engineering at the Department of Engineering at Aarhus University.

We would first and foremost like to thank our supervisor, Christian Fischer Pedersen, for providing guidance and knowledge during the entirety of this thesis. Furthermore, would we like to thank Christian Marius Lillelund for the continued feedback on the report and technical support. Additionally, we would like to thank Kasper Løvborg Jensen for acting as a temporary supervisor during the paternity leave of Christian Fischer Pedersen.

In truth, none of the results achieved in this thesis would have been possible without the immense cooperation by Aalborg Municipality. Thank you so much for providing the required participants for the experiments.

Lastly, we would like to thank our friends and family for their support.

Frederik Andersen

Theis Egsgaard Hansen

Abstract

Case workers in Denmark's municipalities are by law required to provide citizens with rehabilitative training. As a result, case workers have to assess the citizens to determine if they can complete such training. This assessment is currently based on the case worker's subjective opinions and experiences, as no decision support currently exists to assist them.

This thesis seeks to develop a web-based prototype capable of supporting Aalborg municipality's case workers' decision by providing an objective evaluation from a machine learning model developed by Aarhus University. Furthermore, the prototype is focusing on the user experience and is developed in collaborating with potential end users by including them in the design process.

To achieve this, three experiments were performed with the end users to evaluate the prototype's user experience. The experiments included both quantitative and qualitative data gathering techniques to improve the prototype through several iterations.

The results showed potential as the prototype achieved a high user experience through usability and utility tests. In the usability test, the prototype reached 95, 8 out of 100 in a Usability Metric for User Experience questionnaire. The utility test was made as a User Acceptance Test, where the end users acknowledged that the prototype fulfilled the requirements.

Furthermore, the end users stated they felt more confident in their decision by using the prototype. For future work, the prototype should be field-tested to justify its continued development and high user experience.

Resume

Contents

Chapter 1

Introduction

1.1 Background

On December 27th 2014, a law change was made to §83 in "Serviceloven" [1] that required the danish municipalities to offer rehabilitation training. The training is meant to be offered to citizens, e.g., having trouble walking unassisted, if it could result in improved bodily functionality. The rehabilitation training is short-term and time-limited, meaning that the municipality will help the citizens get started and ensure progress before they are asked to continue by themselves. This law becomes more relevant each year as the number of elderly citizens has increased by 31% from 2009Q4 to 2020Q4 [2].

Furthermore, the change in §83 also caused an addition called §83a, forcing the municipalities to decide if a citizen could benefit from rehabilitation training. Nevertheless, it is common knowledge that training regularly is considered healthy both physically and mentally [3, 4]. This is corroborated by S. Zampieri et al. [5], who found that senior citizens who train have higher muscle retention than those who are sedentary.

Having lower muscle retention can result in the citizen becoming weaker and requiring an assistive aid or home help to complete the same tasks. While receiving assistive aid can be beneficial, it can also be a disabler [6], as it reduces stress on the body, thereby removing indirect training, such as standing up from a toilet or a chair. Reducing the indirect training can cause even lower muscle retention, thereby further weakening the citizen, requiring them to acquire additional assistive aids and home help, creating a vicious spiral.

This is supported by G. Häggblom-Kronlöf and U. Sonn [6], who conducted a longitudinal study with participants at the age of 76 and then again at 86. The study found that at the age of 86, 34% of the previously assistive aid independent citizens now required an assistive aid. It also showed that 35% were permanent users, meaning they had an assistive aid at both screenings, with only seven percent transitions to being independent of assistive aids. Furthermore, 81% of the citizens requiring home help also had an assistive aid.

Combining the knowledge of training to increase muscle retention, and that with age and home help citizens are more likely to have assistive aids, it can be beneficial to train to postpone this. While it could be beneficial to assign training to every eligible citizen in the municipality, it is not feasible from a monetary perspective. One reason is that some citizens have other physical or mental limitations, such as paralysis or dementia. Taking this into account increases the difficulty of successfully assigning training to citizens capable of completing rehabilitation training.

1.2 Motivation

Assessing a citizen is done by a municipality case worker to see if rehabilitation training can be completed. In such an assessment, many variables are taken into account such as the citizens current assistive aids, mental health, diagnoses and current rehabilitation plans from the hospital. When a case worker performs an assessment it is based on their previous experience, the variables mentioned above and communication with the citizen, which results in completely subjective assessment. While such an assessment is not necessarily bad, it does not mean it is good either, as it could result in not assigning training to a citizen even if the citizen was perfectly capable.

To provide the case workers with additional support, the AI Rehabilitering (AIR) project has been commenced. The AIR project is one out of 15 signature projects, initiated by the danish government to test the ability of machine learning to achieve better society welfare [7]. AIR is centered around creating a decision support system (DSS), a computer system that provides support and recommendations for a specific use case. The goal for AIR is to provide the case workers with an objective evalutation based on machine learning [8].

This DSS's goal is two fold, as it is meant as way to help the case workers identify citizens who can complete a rehabilitation courses, and those in need of fall prevention rehabilitative training. By focusing on these two goals it can help the case workers make a better decision because they have an objective measure of what the citizen needs, and if the citizen is capable of completing a course and thereby reaping the benefits that persistent training offers. A benefit of providing the case workers with an objective evalutation is that they might assign more rehabilitation courses to eligible citizen, and consequently reducing the amount of home help needed, which is otherwise constantly increasing. Another benefit might be that the case workers assign rehabilitation courses to those capable of completing them, and thus reducing the amount of wasted resources spend on a citizen not able or willing to complete the rehabilitation training course.

1.3 Problem formulation

The case workers in Aalborg municipality often experience a lack of proper decision support when deciding whether or not a citizen should receive rehabilitation training or fall preventive training. The lack of support has lead to decision-making based on the case worker's own knowledge and experience, resulting in fewer assigned rehabilitation training courses to the citizens in need.

A research group at Aarhus University has developed a machine learning model capable of evaluating raw citizen data. The model is focused on two cases. Case 1 predicts a citizen's probability for completing rehabilitation training and, case 2, the risk of falling within the next three months. Because of the fall risk, it can be beneficial for a municipality to assign fall preventive training to the citizen. This thesis aims to establish a prototype that combines data from the case workers current system and the probability data from the machine learning model and translates it to an unambiguous and useful user interface. To ensure a successful user experience, a pilot group from Aalborg municipality has been used to evaluate and optimize the prototype to their specific needs and requirements.

Therefore the goals of this project are to:

1. Develop a useful prototype that displays information from both the machine learning model and the case workers currently used system.

2. Optimize the pilot group's user experience with the prototype.

3. Increase the objectivity of a case worker's assessment when assigning rehabilitation training.

4. Increase the objectivity of a case worker's assessment when assigning fall preventive training.

This thesis presents the methods and choices that have been made throughout the development of the prototype. It also presents the methods and results obtained from the experiments conducted with the pilot group and how the results have optimized the prototype.

Chapter 2

State of the art

This chapter outlines the state of the art of various decision support systems focusing on applicability in general and in the healthcare sector. Furthermore has the concepts behind explainable AI as well as user interface principles been presented. The researched guidelines and recommendations have been encapsulated in a summary at the end of this chapter to provide an overview of the chosen improvements that impact the thesis goals.

2.1 Decision Support Systems in General

A Decision Support System (DSS) helps people make decisions by providing an assessment or recommendation based on data [9], and has been in development and studied for many decades [10]. It consists of three components: the data receiving *database* fed into a *model* to produce a presentable output through a *user interface* [11].

Over the years, various types of Decision Support Systems have emerged to cover many different sectors such as finance and healthcare. In 2004 a method called the expanded framework was developed by D. J. Power to clarify how to distinguish the different DSSs [12]. The following section focuses on the Decision Support Systems used in the healthcare sector.

2.2 Decision Support Systems in Healthcare

Introduction

Focusing on the healthcare sector, a study by R. T. Sutton et al. [13] has done extensive research and compiled papers from university libraries, through hand searching, and from the MEDLINE database from 1980 - 2018. The study describes a specific type of DSS called a Clinical Decision Support System (CDSS), which supports physicians and other clinical personnel by assisting the decision-making with guidelines and advice. CDSSs are commonly classified as either a knowledge or non-knowledge base type of CDSS [13].

Knowledge-based CDSSs usually contains three major components, a knowledge base, an inference engine, and a communication mechanism between the CDSS and the end user [13]. A knowledge base is a compiled base of expert knowledge, mostly in IF-THEN statements that are translated into rules. To produce an output, clinical data is fed into the inference engine that applies these rules and presents it through the communication mechanism, from where the user can interact with it [13]. Non-knowledge base based CDSSs use machine learning or AI algorithms to look for patterns in the clinical data instead of using a knowledge base to produce the output [13, 14, p. 17, 15]. Regardless of the CDSS type, the output consists of guidelines and advice needed to assist the user in the decision-making.

The guidelines and advice can be obtained by the end users in two ways depending on the system. If the CDSS is active the guidelines and advice gets “pushed” to the end users, while they for a passive CDSS gets “pulled” by the end users [15].

Advantages and Usages

E. S. Berner et al. [14, pp. 17–20] researched the effect on healthcare quality caused by a CDSS and concluded using systematic reviews that a CDSS has a positive effect on healthcare quality. This is corroborated by a study made by R. S. Castillo and A. Kelemen [16], that also concludes that an increase in healthcare quality occurs with the use of a CDSS. Furthermore, they emphasize the beneficial attributes to expect from a CDSS, such as a significant reduction in errors made by the clinical staff and a more consistent patient care quality [16].

As for suppliers of CDSSs, multiple companies provide solutions all around the world. According to a market search report, [17] Wolters Kluwer from the Netherlands [18] and Zynx Health from the US [19], were the top two leading companies providing CDSS solutions. Both companies provide CDSSs with an integrated knowledge base [18, 19], which according to market research is the most popular with a 59% market share [20].

Looking further into the two companies, Wolters Kluwer [18] has multiple solution sectors, ranging from finance to healthcare, with the healthcare department alone providing 79 different solutions [18]. One of these solutions is a CDSS called "Lippincott Advisor" [21], that provides the leading evidence-based CDSS for clinical personnel in familiar and unfamiliar situations [21]. According to Wolters Kluwer, the usage of Lippincott Advisor resulted in an improved patient outcome and a reduced number of errors made by the personnel [21].

ZynxHealth [19] has a primary focus on providing healthcare solutions, with one of the areas being the post-acute care sector. In this sector, they have a solution called "ZynxHealth for Rehabilitation" [22] which surrounds the patient during rehabilitation [23].

Pitfalls

While many companies leverage the benefits of providing a good CDSS, these can also worsen lives if certain pitfalls are not avoided. One of such pitfalls described by R. T. Sutton et al. [13] is to avoid alert fatigue, which occurs when a user receives a large number of alerts. They state that "studies have found up to 95% of CDSS alerts are inconsequential"[13], meaning that the user can end up distrusting and ignoring the alerts. The user receiving excessive alerts is deemed a more prominent problem in active CDSSs than in passive CDSSs. The same problem is also mentioned in a study by J. Horsky et al. [15], where they state that "interruptions ... may quickly become an irritant, hazardous disruption when used inappropriately for frequent alerts about minimally important events" [15].

Another example of a significant pitfall is that some users might not have a high technical proficiency when interacting with a CDSS. As this pitfall is very user-specific, it can slow down the interaction with the system for some users, requiring a larger cognitive effort to complete tasks. The study by R. T. Sutton et al. [13] proposed two solutions, with the first being more training of the personnel and the second being a less complex design of the

CDSS.

A CDSS can also impact users skills to be self-reliant [13]. Over time, some users would not be able to make their own decision without relying on the advice and guidelines given by the CDSS. Being too reliant can cause the user to trust the CDSS more than necessary, prompting them to make errors. While this is a concern, a study by M. Laka et al. [24] found that CDSS users were 47% less likely to consider the CDSS a threat to their professional autonomy, and by extension, their self-reliance. Users might also distrust the CDSS to such a degree that the guidance is rendered useless.

2.3 Explainable AI in Healthcare

In non-knowledge based CDSSs, machine learning can generate the output, which can be beneficial to explain, as it can provide the user with context behind the output. Machine learning is the concept of training a model on a dataset to be able to make decisions.

An article done by R. Bhardwaj et al. [25] made a study about machine learning in the healthcare sector. Through their research, they discovered a positive effect using machine learning, as an increasing amount of healthcare data had become available over the years, improving the accuracy and precision of the decision made by the machine learning model. The study also addressed that harnessing large quantities of healthcare data could result in a more or near-perfect patient diagnosis [25].

One drawback to machine learning models is called the black box phenomenon [26, 27]. This phenomenon states that machine learning models tend not to explain what has impacted the decision, leading to distrust from the end users [13]. This issue was studied by D. Dave et al. [26], who found negative impacts using machine learning or AI in the healthcare sector when a models decision could not be explained.

A study by A. B. Arrieta et al. [28] researched the concepts of explainable AI (XAI). XAI provides techniques to machine learning models that otherwise would be incapable of being explained on their own [27, 28]. As stated in section 1.3, this thesis uses a machine learning model developed by Aarhus University. This model utilizes an XAI technique called SHAP, which is a feature relevance explanation technique [28]. This technique was studied by D. Dave et al. [26], where they compared SHAP with another technique called LIME. The study concluded that using SHAP could be used to avoid an otherwise black-box machine learning model.

2.4 User Interface in a CDSS

A user interface in any DSS is one of the three major components as mentioned in section 2.1. It facilitates the communication between the user and the rest of the DSS to deliver the guidelines needed to support the decision-making.

Process

J. Horsky et al. [15] stated that "poor usability continues to be one of the leading obstacles to CDSS adoption and a deterrent to routine use in clinical practice." Poor usability is considered unwanted in a CDSS and in general, as it could lead to a non-utilized system, even though the functionality might be beneficial for the intended users. To combat poor usability, the study highlighted several ways to make sufficient usability improvements. One approach which both J. Horsky et al. and R. Castillo and A. Kelemen [16] found was to "adopt design practices that include user-centered, iterative design and common standards based on human–computer interaction" [15]. Furthermore, these suggestions should be adhered to by involving the potential end users in the design process [15, 16, 29].

Applying an iterative design has been done by A. M. Kanstrup et al. [30, pp. 65–73]. The study's goal was to create a prototype, using the iterative design process, that could minimize medication errors and "place the response from the decision support module in focus of the healthcare professional" [30, pp. 65–73]. The process used by A. M. Kanstrup et al. [30, pp. 65–73] consisted of having a workshop first to generate a low-fidelity paper-based design, which would then be given to the end users to be improved and tailored more specifically to their needs. After this, a high-fidelity prototype was developed and tested using an authentic simulation environment, with the feedback from this resulting in a final prototype. An important thing the study discovered was the possibility to "place the response from the decision support module in focus of the healthcare professional" [30, pp. 65–73] by "placing the decision support in the central part of the application panel and placing other groups of information ... around it" [30, pp. 65–73].

Another study by E. Kilsdonk et al. [31] focused on applying the user-centered approach as they investigated how to design a CDSS. They performed a series of experiments during the study using both a think-aloud technique, semi-structured interviews, and questionnaires. The process was divided into three steps, with the first being an analysis of the participants current system through the use of the semi-structured interview and the think-aloud technique. During the second step, they developed a prototype based on the information gathered throughout the first step. The third and last step was to evaluate the usefulness of the prototype through another think-aloud session, semi-structured interview, and a questionnaire. The results from the study showed that the produced CDSS prototype had a "fairly high usability" [31]. Another discovery made by the study [31] was that the participants had varying opinions during the interview and that the results from the think-aloud session contradicted these opinions. The conclusion to this particular problem is that "healthcare practitioners may not (always) be able to verbalize essential information needs" [31].

Design

A narrative review paper by K. Miller et al. [32] researched design criteria for a CDSS and found 14 papers from January 1, 2000, to December 31, 2016. These 14 papers identified 42 unique design-related recommendations, which could be divided into three main categories, with one of the main takeaways being improved usability for the end users [32]. Each category represents an arbitrary list of recommendations associated with the user interface, the information presented by the user interface, and how the interface should respond and function [32].

The first category regarding the user interface emphasizes the urge to provide simplicity, enhance overall readability, proper use of information placement techniques, and not only rely on presenting information using text [32]. Suggested techniques to achieve some of these includes consistent terminology, colors, font sizes, and the use of tables, graphs, and buttons. The second category involves the information, including terminology and language deemed appropriate in the user domain using the CDSS [32]. As for the third category, the focus has shifted to involve interaction elements, such as minimizing the amount of typing and mouse-clicking, thereby reducing the cognitive load of the end user.

2.5 Summary

A DSS is a system that is capable of providing an assessment or recommendation to people, with a CDSS doing this in a clinical setting. There are different ways to categorize these systems with the most common being if a CDSS is passive or active and if it has a knowledge base or uses techniques such as machine learning. Because a CDSS can be related to peoples health, there is extensive research on some of the benefits and pitfalls of using such a system. While the research is mainly centered around the system, there is also research regarding how to construct such a system, with what process can be used and design and conventions elements to be aware of. Since some CDSS's use machine learning for their inference engine, it is also important to be able to explain the results e.g. using SHAP.

Using the findings from the sections above classifies the prototype in this thesis as a passive non-knowledge based CDSS. While the prototype is not legally a CDSS because the goal is to implement a resulting product in a more administrative setting outside normal clinical use, it will still be considered a CDSS like system, as the data used by the prototype relates to peoples health and healthcare. Categorizing the prototype as a CDSS can also make it easier to enter a new market in the clinical sector, should the future prove there is basis for this.

Chapter 3

Methodology

This chapter contains the methods and techniques used to achieve the thesis goals, including the reasoning for why the given method has been used, and which benefits it brings to the thesis. The chapter is centered around methods related to the user and the user experience, as this has been one of the core areas of research. During the chapter different methods for including the user in the development is present, along with how to design a prototype and test the user experience of it.

3.1 What is User Experience?

Many versions of what the term "user experience" means, exists [33]. One definition by the NN group says that the ""user experience" encompasses all aspects of the end-user's interaction with the company, its services, and its products" [34]. Another definition from a book called "Measuring the User Experience" [35] by T. Tullis and B. Albert says that the "user experience takes a broader view, looking at the individual's entire interaction with the thing, as well as the thoughts, feelings, and perceptions that result from that interaction" [35, p. 5]. Each definition almost had the same meaning, and that the user experience is related to how the user interacts with a product and how content they are.

One article made by the Interaction Design Foundation were able to capture the entire user experience phrase using four distinct levels, namely usability, utility, desirability, and brand experience [36]. Usability is making a product easy to use with a nice feel and good design, seen from a users perspective

[37, 36]. Utility are the desired functionality for the product in development [37, 36]. Desirability is what differentiates a product from similar products that provides the same amount of utility and usability [36]. Brand experience is associated with marketing and the overall branding of the product [36], which in this thesis is not a focus, and will therefore be disregarded. The first three levels have been visualized on figure 3.1.

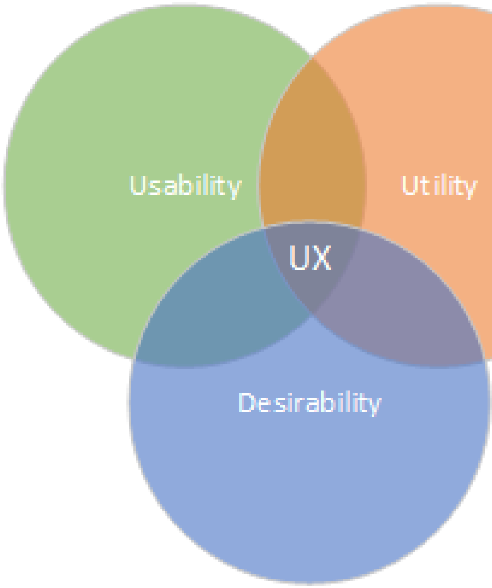

Figure 3.1: Definition of User Experience

In the problem formulation it is stated that the goal is to "Optimize the ... user experience ...", which makes the focus on user experience a key aspect of the project. Having a focus on user experience, and by optimizing it helps achieve another thesis goal, namely "Develop a useful prototype ...". This is because "useful" can be defined as the sum of the usability and utility, according to NN group [37]. By developing a useful prototype it automatically creates a focus on the user experience during the process. The desirability of the "user experience" definition cannot be mapped directly to a thesis goal, as it more related to comparing a product against another, which is difficult in a space with no competitors. Regarding the desirability, it is assumed that it is present, as this thesis is related to the AIR project [38] by Aarhus University, which has received funding before the start of this thesis.

Hvorfor kan User Experience være med til at løse det vi gerne vil i projektet?

3.2 User Centered Approach

When developing a new product or prototype it is important to choose a fitting approach, as not having one can cause the project to lack direction. Choosing an approach can be a difficult task, as it determines how the development should be done, how empirical evidence should be gathered and how users should be involved. As seen in chapter 2, J. Horsky et al. [15] emphasised an approach called the user-centered approach, when developing a CDSS. This approach puts the user at the center of attention throughout the entire development process, and thus has an intentional focus on the user experience [39]. The user centered approached was made by J. Gould and C. Lewis in 1985 [40, pp. 47–49], and is defined using three principles.

DOMÆNE VIDEN!!!

The first is "Early focus on users and task", and focuses on gaining an understanding of the users cognitive and behavioral characteristics early on [40, p. 48]. This is achieved by observing the users performing tasks in their daily work environment and involving them in the design process. For this thesis the principle has mainly manifested itself through the involvement of the users in the design process, as it has provided valuable domain knowledge and preferences into what information is expected to be where, and which functionalities should be present under what conditions. Observing the users have not been possible due to Covid-19, which limits the readily available information on their cognitive and behavioral characteristics. An alternative approach to this problem has been taken by talking to the users during the experiments performed in this thesis. Talking to the users have enabled the group members to obtain expert domain knowledge, and get better understanding of what user goals and user needs should be in focus.

The second is "Empirical measurement" [40, p. 48], stating that empirical evidence of the users reactions and performance should be gathered. Gathering different measurements provides an indication of the systems user experience in terms of usability, utility, and desirability. During this thesis, empirical evidence has been the result of the performed experiments, and helped determine whether the thesis goals were achieved or not.

The third principle is "Iterative design" [40, p. 48], and states that the whole product should not be created all at once, but rather in smaller chunks called iterations. Doing so makes it easier to make corrections and improvements during the development process, and thus cheaper both in terms of time and money, as the duration is reduced between validations. It is important to note that both of the previous principles is included through each iteration to ensure that the project direction is kept, the user is still involved, empirical evidence is gathered, and the product of the iteration is validated.

Iteration Process

The prototype has been through five iterations as seen on figure 3.2, with each iteration building upon the knowledge obtained during the previous iteration. The first for this master's thesis is iteration 3, which is a continuation of the group's previous work done in two R&D projects [41, 42], illustrated as iterations 1 and 2. Iterations 3, 4, and 5 each map to one of three experiments performed in this thesis.

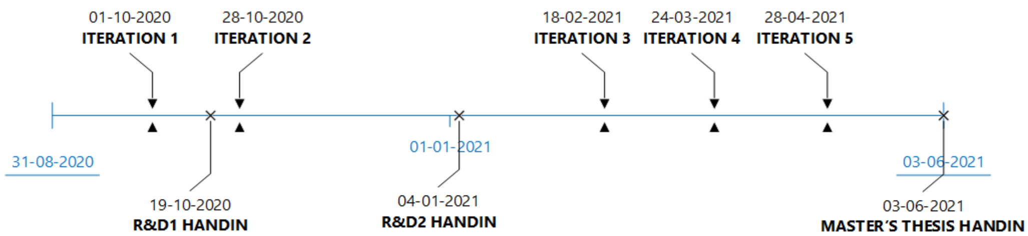

Figure 3.2: Iteration timeline over the course of the process, showing five iterations beginning from R&D1 till the end of this master's thesis.

The process for all iterations consists of four steps: analysis, specification, development, and test, as illustrated on figure 3.3. The analysis is based on the empirical evidence gathered during the previous iteration. The specification refers to the obtained requirements as a result of the analysis phase, whereas the development phase refines the prototype based on the specified requirements. The test phase is the performed experiment to gather empirical evidence, which then are used for the next iterations analysis phase.

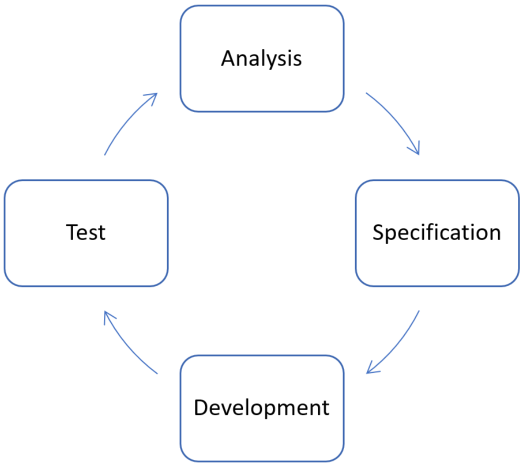

Figure 3.3: Illustration of the processes for each iteration.

3.3 Target Audience

Using a user-centered approach requires a focus on the users for the product being developed. The users for this master's thesis are the case workers, who are responsible for assigning rehabilitation and fall preventive training to citizens in Aalborg Municipality, Denmark.

In an article by K. Le she states that "All users are humans, but not all humans will be your users" [43]. This is true, as the case workers managers will not use the product the same amount as the case workers. It is therefore important to define the users of the product, and other stakeholders such that the correct involvement and communication can be initiated with the different groups.

For this thesis each stakeholder have been identified using the basic stakeholder analysis technique [44, pp. 10–11] and placed in a power versus interest grid [44, p. 11], as seen on figure 3.4. The grid is a two by two matrix, designed by C. Eden and F. Ackermann in 1998 [45] and visualizes the stakeholders with the least and highest amount of both power and interest within a project. The most important stakeholders are those with both the highest amount of power and interest [44, p. 12].

As seen on figure 3.4, the case worker have been identified as a stakeholder with both high power and interest. This is because the case worker is the end user, and have the power to make the product obsolete by not using it, and have interest in creating the best possible product for themselves.

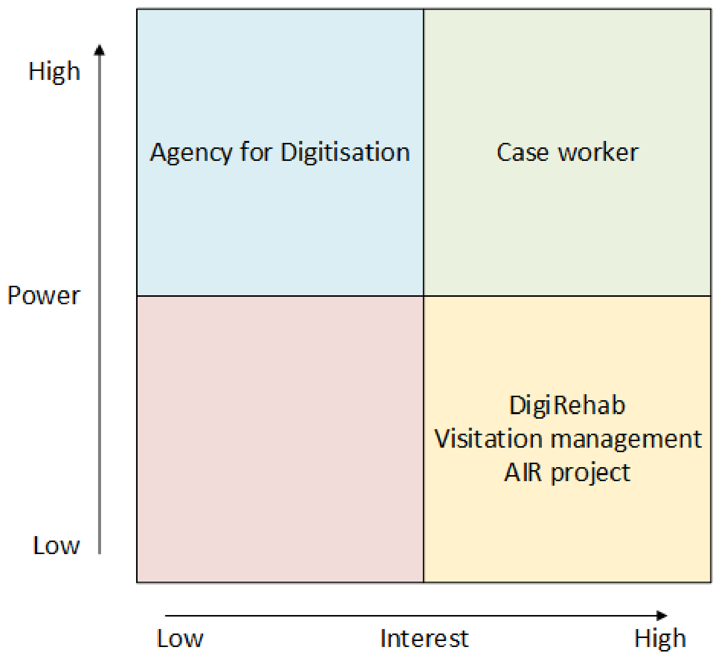

Figure 3.4: Stakeholder analysis

A stakeholder with the same amount of interest as the case workers are their managers, as they would like the case workers to be happy with their solution stack. The managers have lower power as they are not going to use the product. It can be argued that the management has a medium to high power instead of a low power, since the management might have a larger vote in which system the case workers have to use, and therefore have the possible potential to stop the system from being used. The Agency for Digitisation has a high power, since they fund the AIR project [7], and low to medium interest, as they are focused on the projects results.

3.3.1 User Goals

Developing a product requires goals such as user goals, which are important statements of what the user wants to achieve by using the product, which ensures the project direction. These goals have been discovered through R&D1 [41], R&D2 [42], and interviews during this thesis, and can be seen in table 3.1.

User goal 1 and 2 can be respectively mapped directly to the thesis goals "Increase objectivity of case workers assessment when assigning rehabilitation

Id	User goal
1	Increased objectivity when assigning rehabilitation training.
2	Increased objectivity when assigning fall preventive training.
3	An overview based on information from their currently used system.

Table 3.1: User Goals

training" and "Increase objectivity of case workers assessment when assigning fall preventive training". This means that one of the core areas of focus for the developed prototype should be the case workers feeling of receiving useful objective support.

User goal 3 stems from the users need to put the objective support measure in the context of the citizens data. To avoid the users needing to switch between the prototype and their current system, the data needs to be implemented and shown to minimize the users cognitive workload, thereby making it easier to see things in the context of each other. This goal can be partly mapped to the thesis goal "Develop a useful prototype which displays information from ... the case workers currently used system", as the overview required by the case workers comes from displaying the information from their currently used system.

To achieve the user goals, a list of user requirements has been generated (see section 4.2.1), which represents the prototypes specification seen from the users perspective, and have defined the prototypes utility. The user requirements does not concern itself with low level implementation detail, such as if a number should be an integer or not, but rather on the general utility of being able to see the number.

3.3.2 User Expectations

"Users spend most of their time on other sites. This means that users prefer your site to work the same way as all the other sites they already know." [46] is known as Jakob's law and put forth by Jakob Nielsen. Using the law means when developing a product to either replace or support another product, the developed product should work the same way, which for the thesis' prototype should make it look like the case workers current system Cura [47]. While the case workers might not consciously think about this as problem, the users always have expectations of how things works. Meeting some of the expectations can lessen the cognitive effort required by the users to get to know a new system, because it looks and functions the same way the old did. User expectations can therefore be defined as "what a user expects from a product".

User Centered Approach har også en indflydelse på user expectations?

3.4 User Interface

To develop a prototype, a user interface has to be created. A user interface is an interface that provides user control and is everything a user can see and interact with, and it coexists with the user experience. Looking at the available material online it has not been possible to find a precise definition of how to create a good user interface, but looking at the subjects discussed it has been possible to isolate areas that should be focused on. These areas include the layout, UI elements, color, and terminology [48], which is presented in the following sections, along with the users involvement of the prototype design.

3.4.1 User Involvement

To develop a user interface, the users have been involved according to the first principle of the user-centered approach. This means that the user has been part of the design process for each iteration during this thesis. The same approach was employed during R&D1 and R&D2, which means that users were also involved in the design during these two projects. The users' involvement was for the iterations in R&D1 and R&D2, more focused on discovering the utility they needed and its placement. This changed slightly during this thesis, as the need to discover new functionality was minimal, which meant the focus became to improve the utility and gaining feedback on the design of the current iteration.

Despite the user's involvement in the design, there has been a large degree of freedom in designing the prototype, as the style and precise element placement have all been determined by the group. The reason behind this is that the users have been vague about the exact desires, which is corroborated by E. Kilsdonk et al. [31] stating "healthcare practitioners may not (always) be able to verbalize essential information needs." This means that the group has been responsible for determining the things like the color, placement, and style of the presented elements.

3.4.2 Layout Style

Layout represents the overall structure, and is used to provide slots for occupation of UI elements for the user to interact with. E. N. McKay [48, pp. 136–152] presents different examples of how the layout can affect a user when examining a user interface to find the necessary information quickly. Among these examples she states that the user tends to scan the user interface rather than thoroughly reading it [48, p. 147]. This statement is based on the findings done by S. Krug [49, pp. 22–23], who proposed a scanning pattern as shown on figure 3.5. The pattern works by dividing the user interface into four quadrants, with the initial focal area being the top left and then moving towards the lower right corner. This division enables the user interface to show the most prominent information to the user first by placing this in the “primary optical area”.

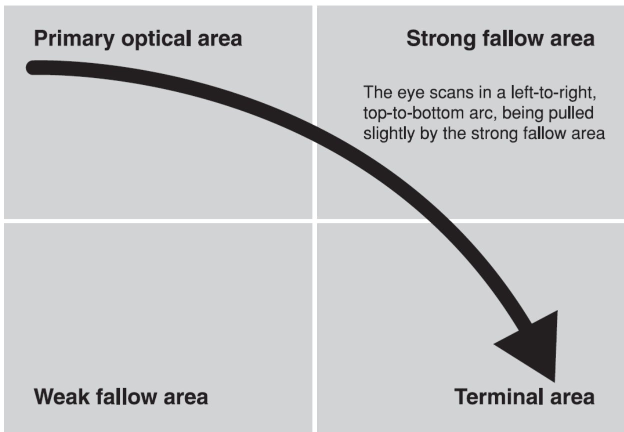

Figure 3.5: Scanning Technique [48, p. 137]

Looking at the study by K. Miller et al. [32] as mentioned in section 2.4, one major design recommendation associated with the layout has been identified. This recommendation, namely “Placement and Positioning” [32] is directly linked to the proposed scanning pattern by S. Krug [49, pp. 21–23], since it is about the placement of information. Details regarding the application of the scanning pattern layout can be seen in section 4.4.2.

3.4.3 The Use of UI Elements

A user interface without UI elements would just be a blank interface. Using UI elements is therefore important to capture the user's attention, with a variation being key as K. Miller et al. state [32], "Avoid using only text". To avoid using only text it is important to be aware of the four main categories of UI elements [50, 51, 52] as seen on figure 3.6.

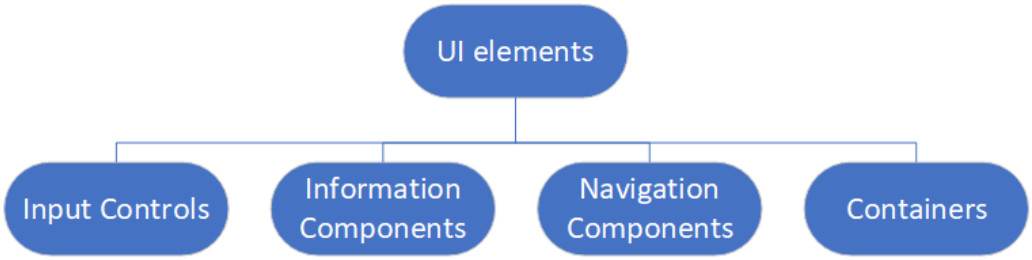

Figure 3.6: Derived UI element categories

Input controls describe elements from where a user can interact with the system through things such as clicking a button or writing in a text field. Informational components share information with the user such as text, statistical information, tooltips, or notifications. Navigation components include different navigational options that transport the users to different pages like a search bar or an image with a hyperlink attached. Containers are used when relatable UI elements should be in close proximity to each other.

Input Controls
Whenever a user is required to input some information by either typing or clicking on an element such as a button or an input field, a type of input control is utilized. These controls are used every day when e.g. there is a need to enter credentials on a login screen, or clicking on the "Google Search" button. Requiring the user to use input controls has in the prototype been minimized as much as possible, as it increases the cognitive workload [32]. This minimization also decreases the amount of time needed to complete a task, as well as the number of mouse clicks and keystrokes required [32].

Informational Components
Information has to be presented in such a way, that it is interpretable and unambiguous [32], and can represent anything from text and numbers to tables, graphs, and diagrams. Including a variety of different informational

components can reduce the amount of time needed to understand complex information, and increase the interpretability of the system [32].

Through the development of the prototype, different informational components such as plain text, tables, and graphs, have been utilized. Using tables and graphs has made it possible to visualize the same data in different ways, thereby decreasing the user's need to spend time and cognitive effort on performing a mental visualization.

To improve the readability it is important to have a fitting font size. This was studied by L. Rello et al. [53] as they performed an experiment using eye tracking technology on 104 participants to establish a baseline between font size and readability. They concluded that in the font size range of 12 to 22, the objectively readability would continuously increase, and they recommended a minimum font size on 14 for adults [53].

While a fitting font size can improve readability, the selection of the appropriate font family also have an effect. J. Horsky [15] stated that "The use of sans serif font families (Arial, Helvetica and other) is recommended". Both the font size and font family recommendations have been used in the prototype to both inform and to create awareness and to indicate importance [49, p. 34].

Navigation Components

Navigation components are used when new information should be presented that is not currently shown on the user interface. Two subcategories of navigation has been used, namely global and local navigation. Global navigation is also referred to as main navigation, and is used to lead the user into a completely new page within a system, thereby allowing them to switch topic [54, pp. 86–87]. Local navigation can be referred to as page-level navigation and allows the user to navigate between UI elements within the same page [54, p. 89].

In the prototype, global navigation taking the form of a navigation drawer, as it resembles the main navigation already used by the case workers in their current system, Cura [47]. This enables the user to switch between case 1, rehabilitation training, and case 2, fall prevention. Within both of these cases, local navigation has been employed to allow the user to switch between information relevant for the given case.

Containers
A container does not bring new functionality like input controls or navigation components. Containers are able to wrap UI elements together thereby creating a connection, which can help the end user to more easily identify elements coupled together thus reducing the search time [32]. An example of use for containers could be gathering citizen information such as name, age, and social security number (SSN) beside each other, and have the container determine the stylistic properties associated with the informational components.

3.4.4 Color Schemes

Colors are a difficult matter as it is subjective what a person defines as pretty and what makes a good composition [48, p. 158]. Because colors are so subjective it is also important to choose the right ones since they can have a significant impact on any user interface. Choosing an inappropriate color or contrast scheme drastically reduces the readability.

Because choosing the right colors is a difficult task, and even more so as the users associated with the project require it to support people suffering from red-green colorblindness (see UR14). A thesis by T. M. Sparks [55] studied the effects of choosing an appropriate color scheme to support people with normal color sight as well as people suffering from red-green colorblindness. Red-green color blindness is the most widespread condition [56] and has two variations: protanopia [57] and deuteranopia [58] as illustrated on figure 3.7. She concludes that using colors like beige, black and white along with colors like blue had a highly positive effect on the participants. The findings by T. M. Sparks [55] form the basis of the color scheme used in the prototype as the primary colors are blue and blue-grey.

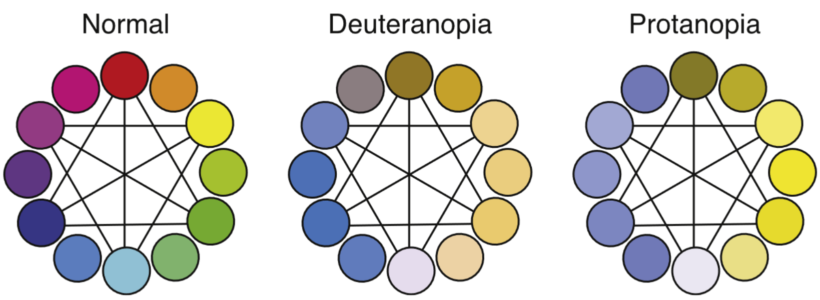

Figure 3.7: Color schemes that display the difference between color perception of normal, red-green and green-red colorblindness [48, p. 160]

Manipulating the color scheme in low light or dark environments is also a user requirement (see UR12) as it otherwise might induce eye fatigue [59]. A way to reduce the fatigue is to use dark mode, which involves changing the contrast polarity between the background and text. A negative polarity is what is defined as dark mode which is a dark on light contrast scheme, whereas a positive polarity or light mode is a light on dark contrast scheme [59].

The usage of dark mode and its influence on the user experience has been studied by H. Eisfeld and F. Kristallovich [59]. Their conclusion was that dark mode can objectively improve the user experience, and should be included as an option.

3.4.5 Terminology

K. Miller [32] proposed that all information should be presented as clear, concise and the terminology should be standardized. Doing so would help the end users to identify and recognize functionalities within the user interface to do their work more effectively and efficiently [32].

For the prototype the terminology has been standardized to fit the end users current system, by gathering the necessary information in the experiments performed in this thesis. It has not been possible to acquire the correct expressions from their system, which means that the prototype is only synchronized to the point of users stating something is wrong. The AI does not have a standardized terminology for reasons explained in section 3.5, but provides clear and concise information.

3.5 Objective Support

As mentioned in section 1.3, machine learning is used to provide the probability for a citizen being able to complete rehabilitation training or them falling within the next three months. Machine learning is the concept of using data to train a mathematical model to either predict or classify new data. Looking at the data used in machine learning each data sample consists of one or more features. An example of a data sample could be a citizen and the features could be age and number of assistive aids. The model then uses the features as an input to output a prediction or classification, based on all data samples the model has been trained on.

In this thesis, a binary classification model [60], developed by Aarhus University has been used. This model will be referred to as "the project model" throughout the thesis. The project model is capable of outputting a citizen's classification along with the probability of the project model's certainty for the citizen belonging to a class 0 or class 1. While the project model can classify between these two classes, the output will always yield class 1 and the associated probability of belonging to that class. This means the project models output always states a citizen's probability of completing rehabilitation training or falling within the next three months. Inversely, a low probability for belonging to class 1 results in a high probability for the citizen NOT being able to complete rehabilitation training or NOT falling within the next three months.

The project model has helped ensure both the thesis and user goals regarding increasing objectivity, as it is able to provide the case workers with a probability to support their decision in whether or not training should be assigned.

During the R&D2 [42] the pilot group from Aalborg Municipality stated a set of requirements, needed for them to support their common understanding of the probability. This includes providing textual arguments explaining the probability, and differentiate between the positive and negative arguments influencing the probability. The requirements was stated by the pilot group to ensure a greater insight into the project model's reasoning. This is corroborated by state of the art studies section 2.3 that states a prediction by itself can lead to distrust. To solve the issues presented by the requirements SHAP was used, as it was the interpretability method provided by Aarhus University, who have created the project model.

3.5.1 SHAP (SHapley Additive exPlanations)

SHAP is an interpretability method used to express how much a feature has contributed to the prediction in either a positive or negative direction. These contributions are called a SHAP value (SV), where a feature with a positive SV equals a positive contribution to the given probability and vice versa. The range of the SVs is bound between -1 and 1 as expressed in equation (3.1), because the range of the project model's probability is between 0 to 1. In a case where only a single feature exists, the feature must represent the entire contribution of the probability by itself. This means that the sum of the SVs is identical to the probability as expressed in equation (3.2).

$$-1 \leq \text{SV} \leq 1 \tag{3.1}$$

$$\sum_{i=1}^{n} \text{SV} = \text{Probability} \tag{3.2}$$

Figure 3.8 displays an illustration of the inner workings of SHAP, and how it interprets the features given to the project model. These features are acquired from the users current system, Cura, where four features has been used in this thesis. These includes: Number of Assistive Aids, Age, Gender and the Loan Period of the assistive aids, which corresponds to four different SV. The red bars represent a feature's positive contribution, whereas the blue bars represent the negative contribution.

Figure 3.8: SHAP used in the project model (Inspired by [61]).

As stated earlier, SHAP is an interpretability method, which though providing explainability, also has its limitations. One of the limitations is that SHAP cannot explain and justify the reason behind the contribution for a given feature. This results in SHAP not being able to provide sufficient arguments required by the users. To acquire sufficient arguments, the SHAP values was then related to their corresponding input feature from which a textual argument could be generated.

An example based on the process: Feature → SHAP Value → Argument was made on the feature "age" as shown in figure 3.9. Each generated argument has been written in a neutral language to avoid any misinterpretations. This means the statements in itself does not reflect any positive or negative influences on the probability, but rather the context of the statements decides the intended influence.

[1]Danish Translation: Personen er 64 år

The person is 64 years

Figure 3.9: Text example[1]based on the age input.

Figure 3.10 shows how SHAP treats each of the four inputs as either positive or negative, respectively. Input types being positive belongs in the *Arguments For* category, and the rest in the *Arguments Against* category. This, combined with the textual arguments shown on figure 3.9 fulfills the user's requirement by both justifying the probability and separating each argument through the abilities of SHAP.

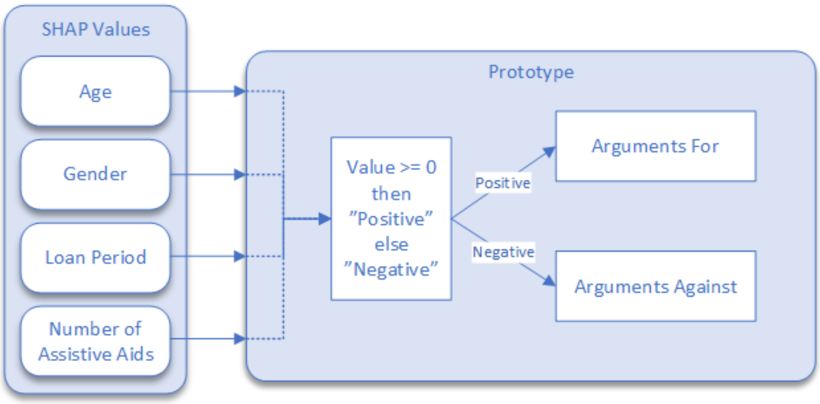

Figure 3.10: SHAP interpretation based on all four inputs.

As stated earlier the users required a text-based argument. This contradicts the statement by K. Miller et al. [32] who states that "avoid using only text and use graphs ... to ensure that the density of information is appropriate" This combined with "healthcare practitioners may not (always) be able to verbalize essential information needs", by E. Kilsdonk et al. [31] which indicates that an alternative to the textual arguments should be developed. Two different design setups, where one includes only textual arguments and the other a combination of both arguments and visual illustrations were tested as described in section 5.4. The experiment showed that the users requirement for a text-only argument, provided the best user experience, and was therefore kept.

3.6 Gathering of Empirical Evidence

The user centered approach's second principle, mentioned in section 3.2, states that empirical evidence has to be gathered and used for improving and validating a system. This section describes the utilization of the different empirical evidence-gathering techniques [40, pp. 259–260] used in this thesis, along with the different kinds of collectable evidence.

Qualitative and Quantitative Data
Two main categories exist during data gathering: quantitative data and qualitative data. Quantitative data can be statistically represented by either a number or value, such that answers can be counted and compared. Qualitative data are usually expressed in words and focuses on the reason behind an action or an opinion.

3.6.1 Interview

Interviews are comparable to conversations and are an efficient way to get information regarding a certain topic and either take place as a one-on-one or as a focus group interview. One-on-one interviews have three categories: unstructured, semi-structured, and structured. Unstructured interviews have open-ended questions without a pre-planned question list for each participant. Structured interviews are the opposite, with a strict list of questions, and in between is the semi-structured interview.

Interviews have been used to gather qualitative feedback during the experiments, as the interview questions have been related to how the users experienced the prototype, and what they felt. This has enabled the group to gather feedback related to the experiment while also getting an in-depth explanation of the user's thinking process. This is important as it ensures a common frame of reference for the received feedback. As a pre-planned set of both open and closed-ended questions has been necessary to ask each participant, the semi-structured interview has been used [40, p. 272].

3.6.2 Questionnaire

A questionnaire consists of a structured series of questions and is an easy and fast way to gather data from multiple people. Questionnaires in this thesis are quantitative in nature, such that it is possible to compare the results between users and across experiments. The quantitative nature is beneficial when an iterative development process is used, as any improvement can be clearly analyzed and measured between each iteration.

None of the used questionnaires have been made from scratch but selected from a set of well-defined questionnaires that have been used and researched through decades. Because the users participating in the experiments have been Danish along with the authors, it was decided that each questionnaire should be presented in that language. This means that each questionnaire has been translated to the best of the group's abilities while retaining the original meaning of the psychometric qualified text. To get a sense of how well the translation had become, each question was translated back into English, as recommended by a user experience pioneer, J. Sauro [62], though pre-testing the questionnaire has not been possible. Questionnaires have been used to evaluate the usability as part of optimizing the user experience, and can therefore be read about in greater detail in section 3.7.3 and 3.7.4.

3.6.3 Recordings

Data recording can capture important feedback which would otherwise be missed during a later stage in a development cycle. Three types of data recording techniques exist: note-taking, audio recording, and video recording [40, p. 266].

All three types have been used, with note-taking being the primary technique during interviews. Audio and video recording have been used to capture the sessions with each user during each experiment, with the video being explicitly used for computer screen-capturing. Capturing the computer screen has enabled review of the users performance during the sessions and has enabled analysis of the users behavior. Audio recording has been used to record the user's voice, enabling further evaluation and review of potentially missed segments during the note-taking process.

3.7 Usability Testing

Usability testing is the gathering of empirical evidence to identify problems in the design of a system, uncover opportunities to improve the system, and to learn about the target user's behavior and preference [63]. This is done by observing the users behaviour while they are performing tasks and listening to their feedback [63]. It is here important to note that the users performance and feedback can change according to e.g. the tasks, the participants environment, and the participants state of mind [35, p. 114]. Using the empirical evidence gathered in a usability test can improve the user experience, as problems and inconveniences is solved. The improvement of the user experience helps achieve the thesis goals of optimizing the user experience, and developing a useful prototype. The optimization of the user experience comes by repeating the usability test for each experiment performed in this thesis. The optimization also helps make the prototype useful, as the term "useful" is part of the definition for user experience. A usability test consists of three core elements as illustrated on figure 3.11.

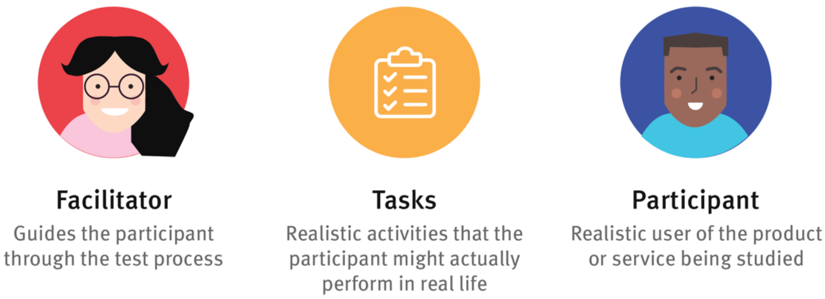

Figure 3.11: Core elements of usability testing [63]

The facilitator is the developer or researcher, with the intention of collecting empirical evidence from the participant, through the completion of a set of tasks. The participant is a person, and preferably an end user, solving the set of tasks given by the facilitator. A set of tasks reflects the users requirements, as it is something the user want to achieve using the system in real life [63].

For this master's thesis, the facilitators are the authors, and by extension the developers of the prototype. The participants have been the case workers provided by Aalborg Municipality. The groups members have created tasks around the user requirements found in section 4.2.1, such that the participants can test the functionality provided by the prototype. Each task has been defined and described in accordance to the format recommended by the NN group to avoid ambiguous task descriptions [64].

The following sections describe four utilized usability metrics, used to gather empirical evidence during the usability tests, based on the book "Measuring the User Experience" by T. Tullis and B. Albert [35]. These includes performance, usability issues, post-task questionnaires, and post-session questionnaires.

3.7.1 Performance

Performance metrics are a beneficial asset to quantitatively determine how well a user performs any given set of tasks in a system [35, p. 64]. Each metric measures the users success rate of a task, amount of time spent on a task, amount of errors, efficiency, and ability to learn. These metrics can help identify if, e.g., a user spends a long time completing a task or if any errors have been made. Therefore, they can provide statistics of the users average performance, showing possible inconsistencies between the tasks. Each of the following segments describes one of these five metrics.

Task Success
Task success is defined by if users can complete a task or not, which is extendable to state if a user needed help during the tasks. *Levels of success* have been used to achieve this definition, as it can define the conditions for a user's success. Using the levels of success has resulted in two measurements for the project. One stating if a participant completed a task, marked with a ✓for success or ✗for failure, and one stating if a participant required any assistance, marked with a ✓for assistance received and ✗for no assistance received.

This configuration has been employed to achieve a greater granularity compared to a success or failure state. It can indicate if a task might not be intuitive, if a significant number of users require assistance.

Task Time
Task time is a measurement for the amount of time a participant spends on a task [35, p. 74, 65, p. 14]. The measurement can be divided into three categories: *task completion time*, *task time till failure*, and *total time on task*.

The task completion time is measured using the definition of task success and is the amount of time spent from the start of a task till the task reaches the success criteria. The total time on task is measured from when a participant starts a task until the participant starts another task, whether or not the task was completed. Task time till failure is identical to the total time on task for the cases where a participant does not manage to reach the success criteria. Out of the three measurements, only two can be represented as total time on task is a constant, and since task completion time and task time till failure depend on whether or not the participant completed the task.

The task time measurements are used to identify time-consuming tasks, where the user spends a lot of time, which is primarily seen in the task completion time, and the total time on task. This knowledge can indicate harder tasks and that additional cognitive effort is required by the participant, which could be lessened by increasing the prototype's usability.

Errors
An error is a mistake or omission a user performs during a task that results in some kind of path deviation from completing the task and could potentially result in task failure [35, p. 82]. Errors are counted when the participant makes an action not in the set of actions for completing the task. Each error is categorized by its cause to identify how many of each type of error was made and where the error made it. This is shown in table A.1.

Categorizing and counting the errors has given an insight into what actions the participants took during the usability test. This insight has then been used to identify areas of the prototype the participant found difficult to use, which could then be improved.

Efficiency
Efficiency is used to determine the amount of physical effort the participant has exerted to complete a task, and is measured as the total number of mouse clicks and keyboard strokes a participant performs during a task. The measured efficiency has been compared to an expected efficiency, describing the least amount of effort required to solve a task. Comparing the participants efficiency to the expected efficiency enables analysis of the required task effort. It can indicate if a task is challenging to complete, and if the participants *explored* the prototype and its functionalities.

There is a causation effect between errors and efficiency, as an error causes an efficiency increase, but not vice versa. This means that a high efficiency count can symbolize a high error count, but a high error count will always symbolize a high efficiency count.

Learnability
Learnability expresses the amount of learning required to use a system [35, p. 92], and is measured over the course of multiple experiments containing

usability tests. Measuring learnability is done using any of the previous performance metrics to see an improvement of the participant's task performance over time. The learnability for this thesis is an aggregated value based on the participants mean task completion time, the total number of errors, and mean efficiency. These metrics have all been chosen as they reflect the participants performance during the task execution.

3.7.2 Usability Issues

A usability issue is used to indicate a problem with either the utility or usability of a system. These issues can include but are not limited to the participant's behavior, facial expression, or verbal comments [35, p. 100]. The majority of the experiments have been conducted remotely, due to Covid-19, limiting the possibility for gathering usability issues, using, e.g., facial expression. Usability issues have therefore only been gathered via verbal comments from the participants. This has been achieved using a think-aloud technique, to allow the participant to verbalize their thought during the tasks [35, pp. 102–103], which is influenced by the participant's state of mind.

Each identified issue has to be prioritized to understand its importance and impact on the user experience. To prioritize the issues, a technique called severity rating has been used [35, p. 106]. These rating techniques have been discussed in an article by J. Sauro [66], where he compares six different types. He concludes that the preferred rating scale consists of three items, as seen on figure 3.12, which is the rating technique used in this master's thesis. Once an issue is prioritized, the source of the issue has been discussed, along with possible solutions to be implemented for the next iteration.

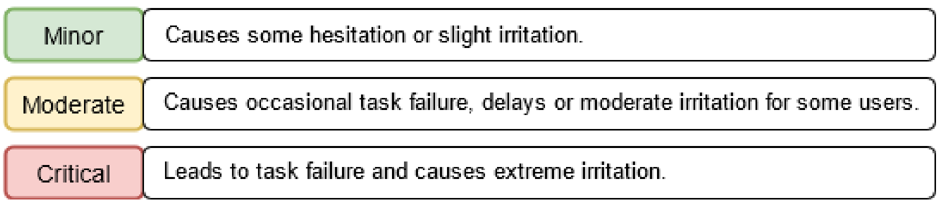

Figure 3.12: Severity rating levels and their description based on [66]

3.7.3 Post-Task

Post-task questionnaires gather data regarding participants perceived task difficulty. In contrast, the performance metrics indicate the participants difficulty level during the task execution. This means a participant can perceive a task as easy, whereas the gathered performance data reveals a failed task. Therefore, the post-task questionnaire can complement the performance metrics to gain insight into what the participants feel, thus indicating where to improve the prototype.

It is important to note that social desirability bias might influence the results when performing a post-task questionnaire [67]. The social desirability bias means that a participant might submit a higher rating in the presence of the facilitators.

SEQ

As mentioned, rating a task after completion provides guidance as to which tasks were perceived as difficult and which tasks need improvement. Figure 3.13 shows the accuracy of the different types of post-task questionnaires according to the number of participants. During each conducted experiment, the number of participant where five case workers from Aalborg Municipality. The most accurate post-task rating using five people was according to T. Tullis and B. Albert, the single ease question (SEQ)[2] [35, p. 136]. The simple question allows the participant to focus on answering rather than the meaning behind the question.

SEQ provides a single question asking “Overall, this task was” on a seven-point scale, ranging from “1: Very difficult” to “7: Very easy” as seen on figure 3.14.

Processing the responses received by the participants is done using equation (3.3). An average SEQ score is calculated for each task based on the number of participants. The average indicates the perceived task difficulty level for the average participant, making it easier to identify the more difficult tasks needing improvement.

$$\text{average} = \frac{1}{n} * \sum_{i=1}^{n} \text{SEQ score} \tag{3.3}$$

[2] “1: Easy/Difficult Task” seen in figure 3.13

[3] Adapted version by B. Albert and T. Tullis [35, p. 136]

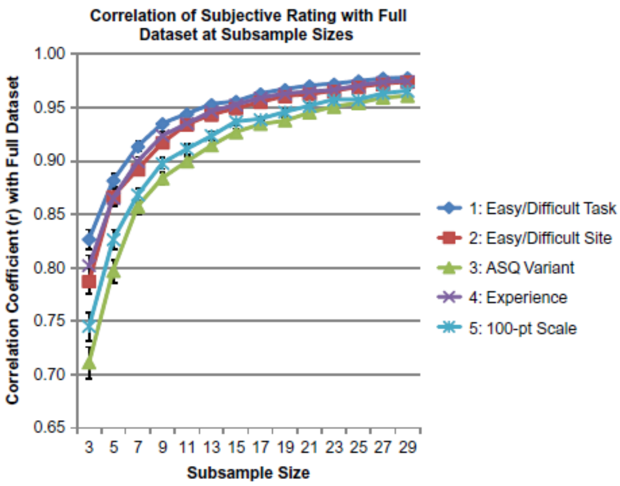

Figure 3.13: Post task rating comparisons[3]

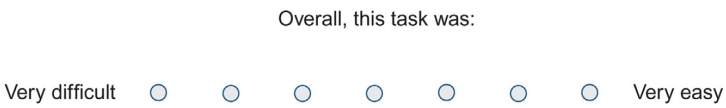

Figure 3.14: SEQ rating scale [65, p. 218]

3.7.4 Post-Session

Post-session questionnaires measures the perceived usability of the entire system [35, p. 137]. Compared to post-task, which are used after each task, post-session questionnaires are used after the entire usability test has ended. As with the post-task questionnaires, post-session questionnaires are also affected by social desirability bias, which is important to have in mind when looking at the results.

CSUQ

The CSUQ provides 16 statements on a seven-point scale ranging from “1: Strongly Agree” to “7: Strongly Disagree” with an additional “NA” option if the participant does not find the statement relevant. The CSUQ statements is divided into four categories, one overall score and three sub-scales: system quality, information quality, and interface quality. These categories provide an evaluation of different parts of the system, giving a broader image of the users perceived usability.

The CSUQ version 3 [65, p. 231] has been chosen as the post-session questionnaire for this thesis, because of its ability to provide different classes of information while retaining a high accuracy compared to other post-session questionnaires. CSUQ also offers high reliability [68], and does not have an inherent gender bias [65, p. 195, 69].

The high reliability stated by J. R. Lewis [68] is corroborated in a study by T. Tullis and J. N. Stetson, [35] who compared the CSUQ to four other questionnaires. The results of that study, seen on figure 3.15, show that have the best performing post-session questionnaire with a low number of participants was the CSUQ, which continues to be reliable even as the number of participants grows.

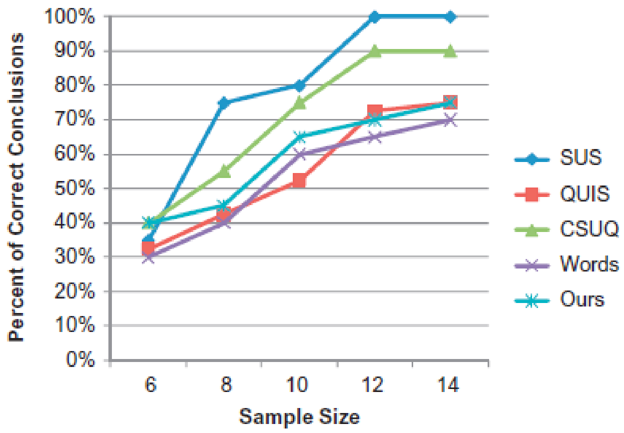

Figure 3.15: Post session comparisons [35, p. 145]

Processing the responses from each participants has been done using the rules shown in table 3.2.

Score	Score description
Overall	Average of the response across all 16 statements
System quality	Average of the response across statements 1-6
Information quality	Average of the response across statements 7-12
Interface quality	Average of the response across statements 13-15

Table 3.2: Score description of the CSUQ scores [68]

SUS

System Usability Scale (SUS) is a post-session questionnaire, developed in 1986 by John Brooke [35, p. 137]. The SUS consists of ten statements on a five-point scale ranging from “1: Strongly disagree” to “5: Strongly agree”. The SUS exists in two versions, with the standard having half of the statements positively worded and the other half negatively worded [35, p. 137, 65, p. 198]. All ten statements have a positive wording in the second version. During the usability tests, the original version of SUS is used, as stated by J. Sauro and J. Lewis [70] “Researchers who use the standard SUS have no need to change to the all positive version provided that they verify the proper coding of scores”.

SUS has been chosen as it a popular and reliable questionnaire [68], with usability as its only focal point. Due to its focal point, it has been used to compare the two designs suggestions mentioned in section 3.5, and to obtain a final usability score for the entire prototype, during the third and final experiment. An important thing to mention in the context of comparing designs, a typical method to achieve the same is to conduct an A/B test [35, p. 216]. I an A/B test, the participant pool is split in two, where each group only evaluates one of the two design variants. However, such a test has not been possible, as a sample size of five participants would potentially make the uncertainties of the results too high.

The data processing of the SUS is different compared to the previous questionnaires because half of its statements are positive, and half is negative. Calculating the score for each statement is done using equation (3.4) on the odd numbered, and equation (3.4) on the evenly numbered. The entire SUS score is calculated and shown in equation (3.6).

$$x_i - 1 = \text{odd} \tag{3.4}$$

$$5 - x_i = \text{even} \tag{3.5}$$

$$\text{SUS Score} = 2,5 * \left(\sum_{i=1,3,5,7,9} x_i - 1 + \sum_{i=2,4,6,8,10} 5 - x_i \right) \tag{3.6}$$

UMUX

The Usability Metric for User Experience (UMUX), developed by K. Finstad [71] in 2010, is a new and shorter questionnaire, and used as it aims to provide the same reliable usability score as the SUS [68]. UMUX has only four statements compared to SUS, ranging from “1: Strongly disagree” to

"7: Strongly agree", but with a broader ranging scale of seven points instead of five. The calculated score is also made the same way as SUS, but only with a change to the constants (see equations (3.7) to (3.9)).

The UMUX is used in the same cases as SUS, as it is according to research not recommended to use the UMUX without using SUS [72, 73]. This stems from the fact that the UMUX is a relatively new questionnaire, and as such not much time has been spend researching the UMUX compared to SUS.

$$x_i - 1 = \text{even} \tag{3.7}$$

$$7 - x_i = \text{odd} \tag{3.8}$$

$$\text{UMUX Score} = \frac{100}{24} * \left(\sum_{i=2,4} x_i - 1 + \sum_{i=1,3} 7 - x_i \right) \tag{3.9}$$

3.8 Utility Testing

Utility refers to whether a system provides the desired functionality, requested by the user. This mean it is part of the thesis goal "Develop a useful prototype ...", as "useful" is defined as usability + utility.

Utility is as usability, testable, using the requirements set by the users (see section 4.2.1). Utility testing can be done using User Acceptance Testing (UAT) and Functional Testing. These differ in execution, as the developers performs the functional tests and the users performs the UAT. Functional testing ensures the functionality works as intended, and UAT makes sure the functionality matches the requirements.

As the focus has been the users and meeting their requirements, only UAT is performed in this thesis. The UAT has been made using the book "User Acceptance Testing" by B. Hambling and P. Goethem [74], as one of the examples can validate all user requirements. This example is called test scenarios [74, pp. 140–141].

Test scenarios describe a sequence of actions necessary for the user to complete the scenario, with an expected result for each action. The expected result is then either confirmed or disconfirmed, resulting in the action and consequently the task either passing or failing. No matter the outcome, the user then proceeds to the next task. Each test scenario performed in this thesis is shown in chapter B. UAT is performed at the end of the development stage, and as such performed during the third and final experiment.

Chapter 4

System Design and Implementation

This chapter describes how the methodology presented in chapter 3 is used to define the prototype. All requirements are presented along with a described architecture to showcase how the prototype has been connected with external endpoints. The groups prototype design from before and at the end of the thesis is presented, with the latter encapsulating the requirements. Lastly are some implementational details presented regarding the establishments between the prototype and the endpoints.

4.1 Data Description

This section describes the information presented in the prototype, which can be seen in table 4.1. Each type of information is represented along with a description.

Begrundelse for hvorfor læseren skal læse om Data Description

It should be noted that *Assistive Aids*, *Citizen Information*, *Diagnoses and Motivation*, *Registered Falls*, and *Training Plans* are retrievable from Cura.

Table 4.1: Data presented in the prototype

Data	Description
	Arguments
Argument	Plain text describing the argument.
	Assistive aids
Name	The name of the assistive aid based on the HMI number.
HMI-number	An assistive aids identification number.
Grant type	Under which law has the citizen received the assistive aid. The type can either be APV [75], §83a [76] or §86 [77]
Grant reason	Plain text describing the reason as to why the citizen has received the assistive aid.
Date received	Date for when the citizen received the assistive aid.
Hand in date	Date for when the citizen returned the assistive aid.
	Citizen Information
Name	Plain text stating the citizens full name.
Age	A number representing the citizen's age.
Home help hours	The number of home help hours the citizen has received.
SSN[1]	Plain text stating the citizens SSN.
	Diagnoses and Motivation
Diagnoses	Plain text describing the citizen's diagnoses.
Motivation	Plain text describing the citizen's motivation.
	Probability
Probability	Number stating the probability for either case 1 or case 2.
	Registered Falls
Cause	Plain text describing the cause of the fall.
Date	Date for when the fall occurred.
	Training Plans

Continued on next page

Table 4.1 – continued from previous page

Data	Description
Active	Describing whether a training plan is active, finished, or not yet started.
Type	The type of training plan the citizen received.
Goal	Plain text describing the end goal for the training plan.
Start date	Date for when the training plan was started.
End date	Date for when the training plan ends.
End status	Plain text describing the citizen's status at the end of the training plan.

4.2 Requirements

4.2.1 User Requirements

The user requirements is defined with the purpose of highlighting a users interaction with the system [78, 79]. Each of these requirements is used to produce system requirements, which describes the prototypes functional and non-functional requirements.

Table 4.2: User Requirements

Req. Id	Description
UR1	The case worker shall be able to log into the prototype.
UR2	The case worker shall be able to see the citizens probability for completing rehabilitation training.
UR3	The case worker shall be able to see the citizens probability for falling within the following three months.
UR4	The case worker shall be able to see both the positive and negative arguments for the citizens probability.
UR5	The case worker shall be able to see the citizens name, age, and SSN.

Continued on next page

[1]Social Security Number

Table 4.2 – continued from previous page

Req. Id	Description
UR6	The case worker shall be able to see the citizens diagnoses and motivation.
UR7	The case worker shall be able to see the citizens training plans.
UR8	The case worker shall be able to see the citizens assistive aids.
UR9	The case worker shall be able to see a graphical representation of the assistive aids over time.
UR10	The case worker shall be able to see a graphical representation of the home help hours received over time.
UR11	The case worker shall be able to see how many registered falls a citizen has.
UR12	The case worker shall be able to use dark mode.
UR13	The case worker shall be able to log out of the prototype.
UR14	The case worker shall be able to use the prototype, while suffering from red-green colorblindness.

4.2.2 System Requirements

The following system requirements reflects the user requirements to provide a more detailed requirement description. Each system requirement has been written using the guidelines proposed by I. Hooks [80], and a book by D. D. Walden et al. [81, p. 61].

Functional Requirements

The functional requirements is a detailed description of the functionality provided by the system. The following requirements does not state how the functionality is provided, but only what is provided.

Table 4.3: Functional Requirements

Req. Id	Description
FR1	The prototype shall provide a login screen.
FR2	The login screen shall provide a field for entering a username.
FR3	The login screen shall provide a field for entering a password.

Continued on next page

Table 4.3 – continued from previous page

Req. Id	Description
FR4	The login screen shall autofill the username.
FR5	The login screen shall be able to show the entered password characters as asterisks.
FR6	The login screen shall be able to show the entered password characters in plain text.
FR7	The prototype will enable a case worker to log in.
FR8	The prototype will enable a case worker to log out.
FR9	The prototype shall be able to change the contrast polarity between the background and text.
FR10	The prototype will be able to interface and acquire data from subsystem 1[2].
FR11	The prototype will be able to interface and acquire data from subsystem 2[3].
FR12	The prototype will show the citizens probability for completing rehabilitative training.
FR13	The prototype will show the citizens probability for falling within the next three months.
FR14	The prototype will show the citizens probability as an integer.
FR15	The prototype shall be able to show the positive factors of the citizens probability.
FR16	The prototype shall be able to show the negative factors of the citizens probability.
FR17	The prototype shall display the citizens SSN.
FR18	The prototype shall display the citizens name.
FR19	The prototype shall display the citizens age.
FR20	The prototype shall display the citizens active GOP[4].
FR21	The prototype shall display the citizens assistive aids.
FR22	The prototype shall plot the citizens assistive aids over time.
FR23	The prototype shall plot the citizens home help hours over time.
FR24	The prototype shall be able to toggle the visibility of the plot from FR22 and FR23.
FR25	The prototype shall display the citizens diagnoses.

Continued on next page

[2]The users current system, Cura
[3]The project model
[4]Rehabilitation plan issued by a hospital

Table 4.3 – continued from previous page

Req. Id	Description
FR26	The prototype shall display the citizens motivation.
FR27	The prototype shall display the citizens training plans.
FR28	The prototype shall display the citizens registered falls.

Non-Functional Requirements

Non-functional requirements describes the quality attributes of the prototype, reflecting the prototypes performance and how it provides the functionality.

Table 4.4: Non-functional Requirements

Req. Id	Description
NFR1	The prototype will use a responsive design.
NFR2	The prototype shall be compatible with color schemes associated with people suffering from deuteranopia [58] and protanopia [57].
NFR3	The prototype shall be compatible with all modern browsers [82].
NFR4	The case worker shall be able to navigate to any place using a maximum of three clicks [83].

4.3 System Architecture

The system architecture is a conceptual model that depicts the product and its interfaces. During this section, the framework from which the prototype has been made will be justified, along with a description of external endpoints interfacing with the prototype.

Figure 4.1 shows the system architecture for the prototype and its interfaces to two endpoints. Connecting the prototype to Cura makes it possible to achieve the user goal of having an overview as it consists of data from Cura and increasing the case worker's objective support. The objective support is because the data from Cura is, as mentioned in section 3.5 necessary to

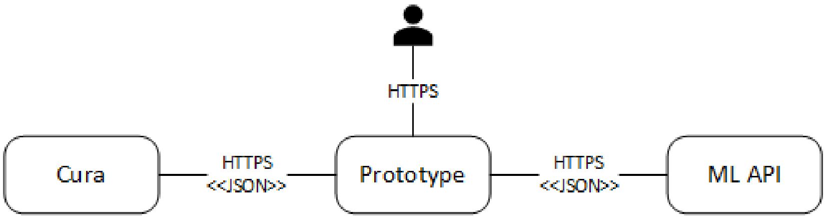

Figure 4.1: System architecture diagram, showing the prototype, and its interfaces to both the endpoint of the users current system, Cura, and the ML API.

provide both a probability and to make the requested textual arguments. By achieving the user goals, also helps achieve the thesis goals of the same name.

4.3.1 Web Framework

Throughout R&D1 and R&D2, three of the most popular front-end web frameworks were analyzed, Angular, React, and Vue, from which the chosen framework was Vue, version 2.6.12 [42, pp. 7–10]. Vue is a front-end web framework that has become a more popular candidate throughout the years, even though it has been around for fewer years than the other two. One of Vue's key advantages is its flatter learning curve compared to its competitors [84, 85, 86], which is relevant, as the group members do not have much experience with web development. This, combined with Vue's overall fast performance [87] in a small app size with increasing popularity, made it an ideal candidate [84].

The latest version of Vue [5] is Vue 3.0.7, which is smaller and faster compared to Vue 2.6.12. This version, however, has not been used by the group, as the chosen UI libraries (see section 4.4.3) do not support Vue 3. However, Vue has provided a comprehensive guide [88] for migrating from Vue 2 to Vue 3 when the library gets supported.

4.3.2 Cura API

Systematic develops Cura [89], and is, as mentioned, the existing system used by the case workers, as information needed by a case worker is located within the same system. Cura is accessed through a protocol called Fast

[5] As of the 1st of March 2021

Healthcare Interoperability Resources (FHIR), developed by Health Level Seven International (HL7)[90], and is built on top of a REST protocol [91]. Using FHIR made it possible for Systematic to generate several profiles, representing some of the information existing in Cura [92]. Each profile consists of several accessible resources. Through the use of one out of seven available REST operations in the REST protocol [92], a resource can be read, deleted, updated, created, etc. In the case of the prototype, a resource is only necessary to read, as it makes it possible to display the equivalent information for what the resource represents in Cura. Shown on listing 4.1 is an example on a read operation through the REST protocol. The *id* represents the specific citizen from which the resource should be read [92].

```
GET https://basepath/{resourceType}/{id}
```

Listing 4.1: Resource Example of a Read Operation to Cura's API through REST

However, acquiring a resource is only possible when authenticated by Cura or by one of their trusted external providers. The authentication consists of providing a username and password, from which the user gets validated and receives a token if successful [92]. A login screen has been implemented in the prototype to simulate the requirement for a username and password. However, due to limited resources at Systematic, the authentication process has not been possible to integrate. Nevertheless, a proof has been conceptualized to show how it could be done (see section 4.6.1).

While it has not been possible to retrieve data directly from Cura, the data and associated descriptions presented in section 4.1 has been developed in cooperation with the case workers. This means that they have validated that the presented fields in the prototype are also present in Cura, thereby enabling the desired data to be shown.

4.3.3 ML API

To summarise the project model, it is the machine learning model capable of predicting the probability for completing a rehabilitation training course for any given citizen and if a citizen is at risk of falling within the next three months. It also provides additional information such as SHAP values which have been described and discussed in section 3.5. Communication is done using an API that uses the same RESTful concepts as mentioned in section 4.3.2. The endpoints for the API can be found in the documentation [8].

The API consists of two endpoints from where data can be retrieved. The first endpoint is *basepath/predict_complete*[6], whereas the second is *basepath/predict_fall*[7]. Each of these endpoints is responsible for returning a prediction for whether or not a citizen could complete rehabilitation training and returning a prediction for whether or not a citizen falls within the next three months.

Receiving a prediction from either endpoint is done using the same body in a *POST* request. An example of a request body is seen on listing 4.2. Looking at the example shows the four data inputs as discussed in section 3.5, which is needed to receive a prediction. This information includes the citizens gender[8], birth year, average loan period of all assistive aids, and a list of the assistive aids, represented by its HMI-number in chronological order.

```json
{
    "Gender": 1,
    "BirthYear": 29,
    "LoanPeriod": 42,
    "Ats": "123456, 654321"
}
```

Listing 4.2: Example JSON request body for *basepath/predict_complete* and *basepath/predict_fall* endpoints

The response body is as the request body, identical for both endpoints. An identical response body makes it easier to process the returned data in the prototype, although the processed data is stored at different locations to differentiate between the two endpoints. The response body is shown on listing 4.3, and consists of three primary values, which are dividable into two categories.

The first category consists of the *Prediction* and *Probability*, which are the result from the project model, stating whether or not the citizen can complete rehabilitation training or falls within the next three months. The second

[6]Basepath represents https://air-fastapi.azurewebsites.net
[7]Basepath represents https://air-fastapi.azurewebsites.net
[8]'0' means female, and '1' means male

category is the SHAP values which, as mentioned in section 3.5 provides an interpretation of the feature's positive or negative impact on the probability.

```
{
    "Prediction": 1,
    "Probability": 0.5896564722061157,
    "ShapValues": {
        "Gender": 0.087,
        "BirthYear": -0.01,
        "LoanPeriod": 0.028,
        "NumberAts": 0.075,
    }
}
```

Listing 4.3: Example JSON response body for *basepath/predict_complete* and *basepath/predict_fall* endpoints

4.4 UI Design

This section presents both the initial design on which this thesis' prototype is based upon, as well as the prototype's final design iteration. Furthermore, are the UI libraries presented.

4.4.1 Initial Design

As mentioned in section 3.2, two prototype iterations have already been completed during R&D1 [41] and R&D2 [42]. The design shown on figure 4.2 is the design made at the end of R&D2. The initial design has been included to show the starting point for this thesis prototype design and serve as a frame of reference for the changes made to the final design.

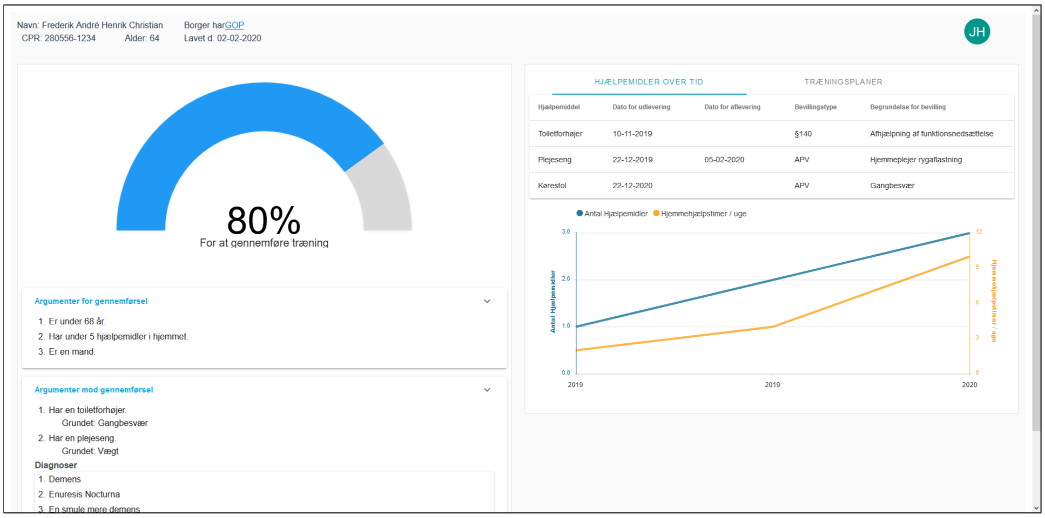

Figure 4.2: Resulting design of R&D2 [42], and used as the initial design for this thesis.

4.4.2 Final Design

The design described in this section is considered the final version of the developed prototype in this thesis. It incorporates all the requirements, presented methodologies from section 3.4 and changes based on the feedback gathered from the pilot group during all three experiments. The final design is presented for both case 1, can a citizen complete rehabilitation training, and case 2, will a citizen fall within the next three months. Each case differentiates in its functionality, which will be described and clarified as well. Each of the two designs are shown on figures 4.3 and 4.4, respectively. Furthermore has dark mode been enabled on figure 4.4 to showcase how the prototype is intended to be used in low light environments. Additional screenshots of the final design can be found in appendix section A.4.2.

As mentioned in section 3.4.2, the layout is developed by applying the principles of the scanning technique. To summarise, the scanning technique enables the user to scan an application and find the necessary information quickly [48, p. 137]. According to the technique, the users gaze starts in the top left corner called the "Primary Optical Area", traveling right through the "Strong Fallow Area", and ends in the bottom right called the "Terminal Area" [48, p. 137]. The prototype shown on figure 4.3 is divided into four groups, outlined using red borders. All groups are wrapped in separate containers, as the individual elements within each group have high coherence. Each of these groups will be presented and described separately in greater detail during this section.

Fix overgang

Regarding the terminology, both group #2 and #4 display data directly from Cura, and as such, these groups have been validated by the case workers. This keeps the prototype close with their current systems terminology, which is relevant as described in section 3.4.5. Group #1 and #3 tries to mimic the terminology used by the case workers, by being clear and concise.

Fix overgang

Colors have only been used in a restricted amount, with the primary colors being blue, blue-grey, and orange, as shown in figures 4.3 and 4.4. Each of these colors has been selected based on the findings presented in section 3.4.4, ensuring that users with color blindness such as protanopia and deuteranopia are still able to use the prototype. For simplicity, white, black, and grey are used for the background, fonts, and minor highlighting [32].

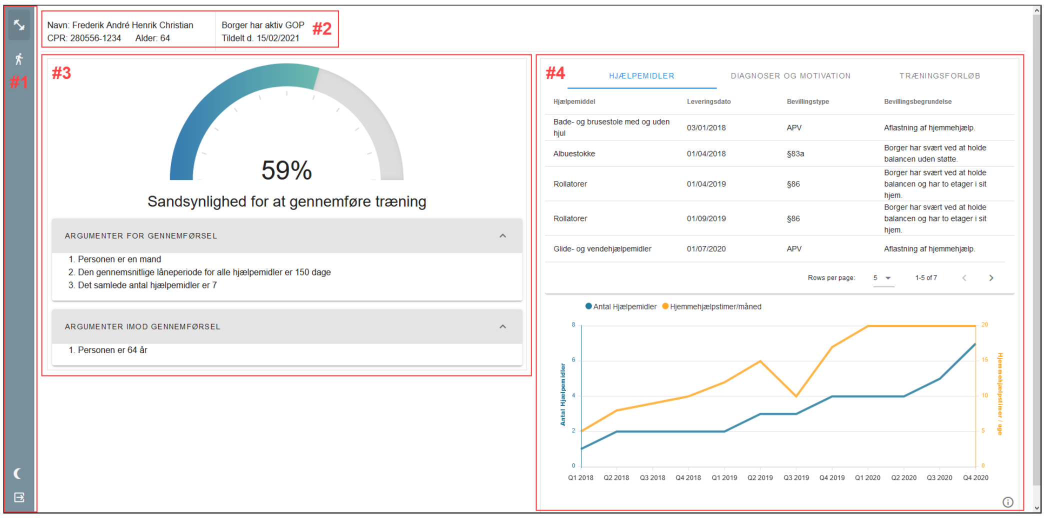

Figure 4.3: Final Prototype Design - Case 1 (Dark mode: Disabled)

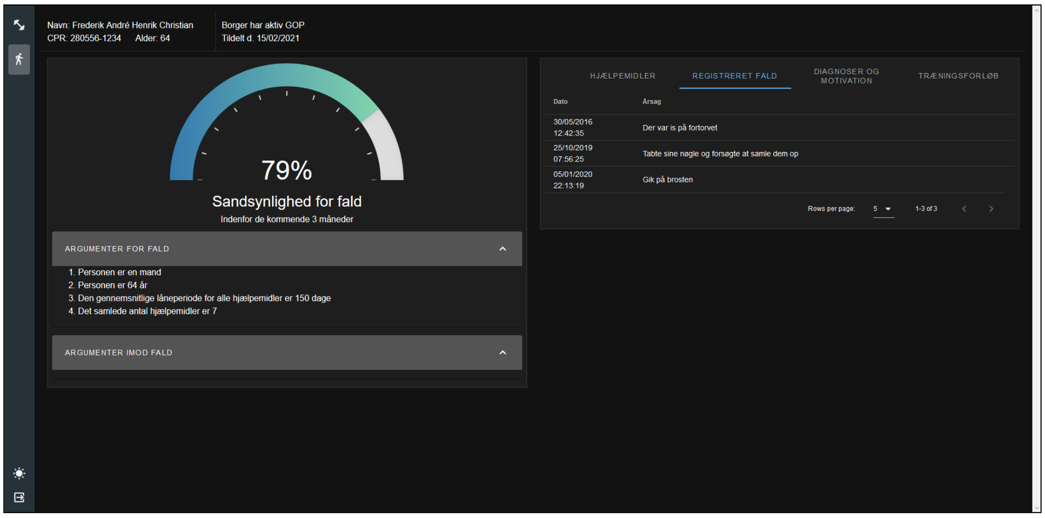

Figure 4.4: Final Prototype Design - Case 2 (Dark mode: Enabled)

Group #1 - Navigation Drawer

This group consists of a navigation drawer seen on figure 4.5 which, as mentioned in figure 3.6 enables the user to navigate to a different location in the prototype. The navigation drawer is global and thus always accessible to the user.

Functionality located in the navigation drawer from top to bottom enables the case workers to switch between case 1 and case 2, change from light to dark mode and vice versa, and sign out of the prototype. Dark mode and log out are each located in the "weak fallow area", as they are assumed less used, and thus less important than switching between cases. The navigation drawer can expand upon a mouse over as shown in figure A.37, to show a full description of the functionality the user is about to interact with. If a user then clicks on the inactive case, group #3 and group #4 change accordingly. Furthermore, the active case is marked with a variant of the navigation drawer color.

At the time of developing the initial design (see figure 4.2), a navigation drawer was not considered necessary as only case 1 existed. After the inclusion of case 2, a navigation drawer became a valuable

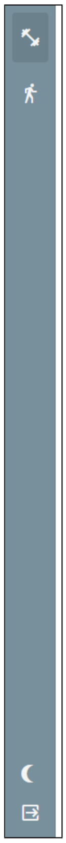

Figure 4.5: Final Prototype Design -

addition to make the end users capable of changing between cases at will.

Group #2 - Header

This group as seen on figure 4.6 displays the citizen's information such as name, age, and SSN, along with the citizen's rehabilitation training plan supplied by a hospital (GOP) if available.

The group itself is comprised purely of text, and due to its location in the "Primary Optical Area" it is easy for the case workers to see and verify that the correct citizen is being assessed, before consulting the probability located in group #3.

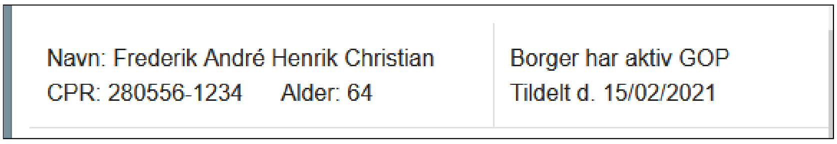

Figure 4.6: Final Prototype Design - Group #2 - Header

Compared to the initial design, the number of changes to this group has been minimal, with only minor text adjustments and alignments.

Group #3 - Probability and Arguments

This group contains as shown on figure 4.7 the probability from the project model, and the arguments described in section 3.5. Both are located in the "Primary Optical Area", as the main purpose of the prototype is to provide the case workers with objective decision support. The probability is visualized as a numeric percentage value and as a gauge. The gauge is a visual aid to emphasize the numeric probability value, providing greater context for the case workers. Additionally, the inside of the gauge contains indicators, each corresponding to 10% of the probability, making the gauge easier to read. The probability description text states what the probability indicates, depending on case 1 and case 2.

For quick interpretation of the probability, four arguments have been created as described in section 3.5. Each argument is found in one of two expandable containers called "ARGUMENTER FOR GENNEMFØRSEL" and "ARGU-

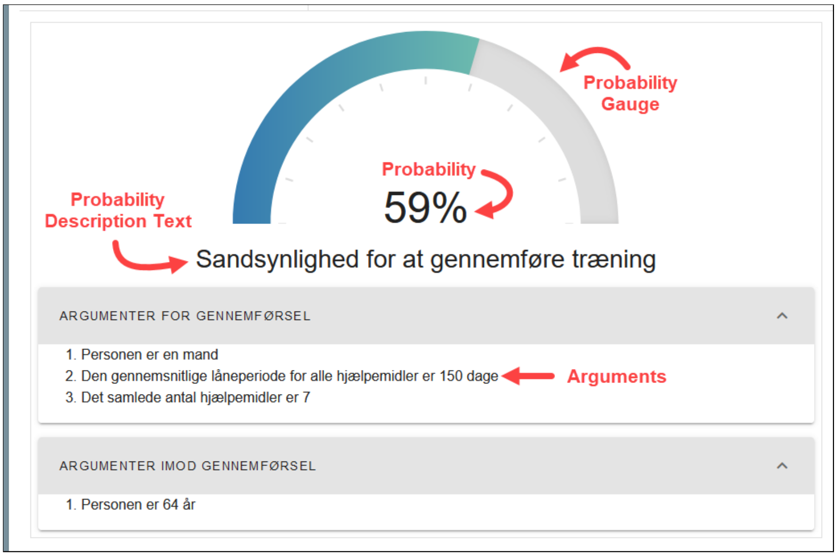

Figure 4.7: Final Prototype Design - Group #3 - Probability and Arguments

MENTER IMOD GENNEMFØRSEL". The "ARGUMENTER FOR GENNEMFØRSEL" shows the positive influences while "ARGUMENTER IMOD GENNEMFØRSEL" shows the negative influences. Both containers are by default closed (see figure 5.26) as it decreases the amount of information, thereby decreasing the cognitive load of the case workers.

Comparing the group to the initial design, the UI elements resemblance and placement have not changed, although size and color adjustments have been made. The probability gauge in particular have been changed to be less dominant in the UI space, with the addition of indicators. To keep the presentation consistent [32] both the font size and color on the expandable containers matches the headers in group #4.

Group #4 - Cura Information Overview

Group #4 is located in the "Strong Fallow Area", and shows the information from Cura, including the citizen's assistive aids, training plans, and diagnoses and motivation. These three information types are separated into sections called tabs, as they are easily interpretable and hard to miss [49, pp. 80–81]. Each tab is shared for both case 1 and case 2, while case 2 includes an additional tab, containing the citizens registered falls (see figure 4.4). All tabs are considered a part of local navigation as they do not redirect to a new page.

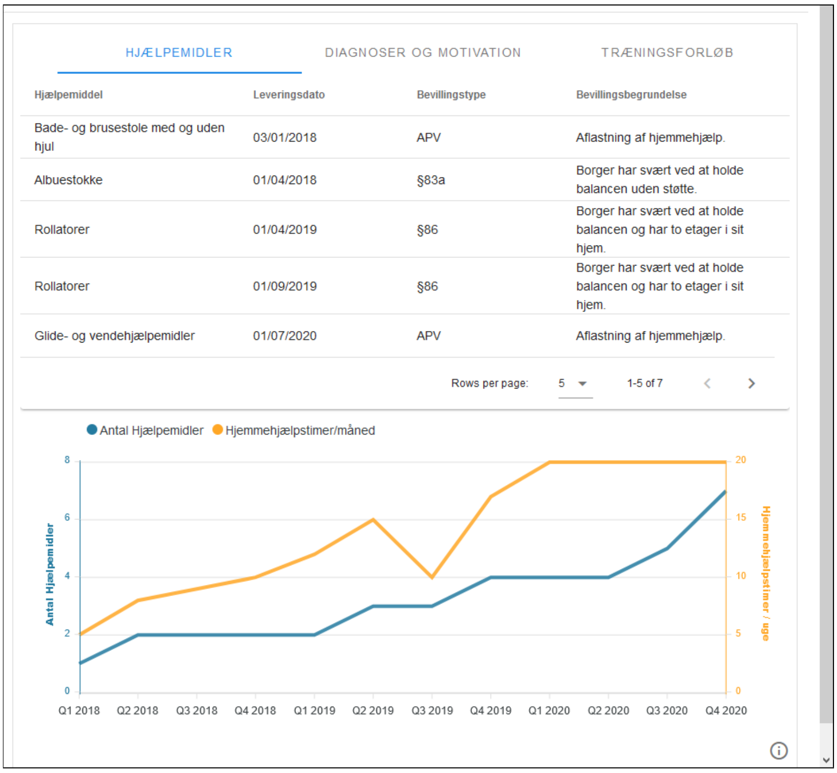

Figure 4.8: Final Prototype Design - Group #4 - Cura Information Overview

Placing group #4 in the "Strong Fallow Area" has the advantage of drawing the case workers' attention after having scanned group #3. The location allows the caseworker to establish greater context for the probability based on the information in group 4, which serves to fulfill the user and thesis goals seen in sections 1.3 and 3.3.1.

The tab shown on figure 4.8 represents the citizen's assistive aids using both a table and a graph for easy interpretation. It is corroborated by K. Miller et al. [32] stating "avoid using only text and use tables, graphs ... to ensure that the density of information is appropriate". While the statement by K. Miller et al. also applies to the "TRÆNINGSFORLØB" and "REGISTRERET FALD", it should be noted that no graph is present in these two tabs. As for the "DIAGNOSER OG MOTIVATION", the information is represented in plain-text as this is consistent with Cura.

Looking at the initial design (see figure 4.2), only two tabs were present as "DIAGNOSER OG MOTIVATION" was located in the argument section. The information was moved to a separate tab in group #4, as it did not influence the probability. Another reason was that the information originated from Cura, which means that it would be possible to gather all Cura-related data in a single group.

4.4.3 UI Libraries

Making all the UI elements from scratch can be a long process and requires expertise and hundreds of lines of code. As many developers deal with the same struggle, many different solutions exist to solve the issue. Throughout this thesis, three UI libraries have been used to avoid using resources on already solved issues.

The UI library Vuetify [93] has been responsible for the majority of the UI elements, such as the tables, navigation drawer, icons, and containers. Vuetify [93] was chosen due to its high and increasing popularity over the competition [94] and diverse UI element catalog [93].

While Vuetify offers many different types of UI elements, it does not provide a multi-axis graph like the one shown on figure 4.8. For this purpose, the library ApexCharts were used, as no other library seemed to offer multi-axis graph functionality. ApexCharts was also initially used to make the probability gauge seen on figure 4.2, but was replaced by vue-svg-gauge [95] due to customization limitations.

4.5 System Design

This section seeks to explain the structure of the prototype using the class diagram shown on figure 4.9. Looking at figure 4.9, the prototype consists of several components, which are reusable packages containing both design and functionality. These components can also consist of other components, making the prototype easily expandable for further additions.

The components are divided into three groups *VueX*, *Login*, and *Home*, and provide the basis of all interactions a user can perform using the prototype. While VueX is regarded as its own group, it does not contain any design as it is a state manager. The state manager is a globally distributed data store and is only interactable through other components. Having a global state manager enables all other components to access and manipulate the data regardless of their relationships with other components. Furthermore, this

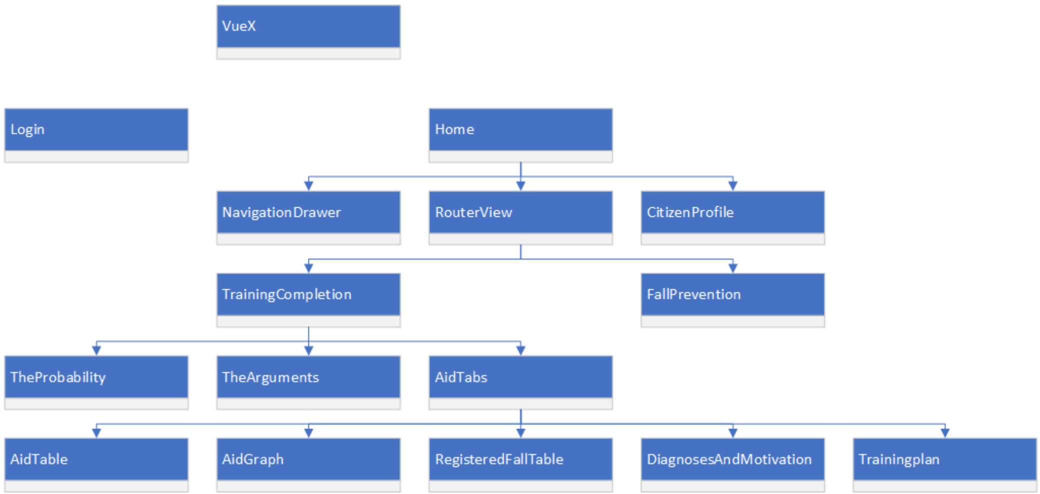

Figure 4.9: A class diagram of the prototype, based on components.

enables easy data transmission from one component to another, as the state manager is a single point of access.

The second group is the *Login*, responsible for validating and redirecting the user to the *Home* group upon successful login. *Home* is the main application and contains the components that constitutes the designs seen on figures 4.3 and 4.4. *Home* consists of three components, *NavigationDrawer*, *RouterView*, and *CitizenProfile*. Each of these can be mapped to the groups shown on figure 4.3.

NavigationDrawer represents group #1, *CitizenProfile* group #2, and *RouterView* group #3 and #4 as it consists of *TrainingCompletion* (TC) or *FallPrevention* (FP). However, the *RouterView* itself does not contain the components TC and FP but can interchange them at any given time, as it is a router component. This means that the *RouterView* does not have any design or functionality besides presenting other components.

It should be noted that FP does not seem to consist of any components compared to TC. This is not the case as they both use the same components, with the only difference being that FP shows the *RegisteredFallTable* component while TC hides it. However, the components contained within FP have been left out of the class diagram on figure 4.9 to save space.

4.6 Implementation

This section describes the implementation of the established connections and data acquisitions from Cura and the ML API. During the development, it was made clear that establishing a connection to Cura was not possible, as mentioned in section 4.3.2. This has resulted in the prototype only having an established connection to the ML API. However, a proof is provided, showing a plausible integration with Cura given the chosen technology stack. This includes how the data described in section 4.1 can be accessed and thus used in the prototype on a conceptual level.

4.6.1 Cura API

For proving an established connection to Cura, a diagram was provided by Systematic (see figure C.1). The diagram has been used as a basis for creating figure 4.10, which includes the components utilized in the prototype.

A user first needs to be authenticated to access the data, starting by entering a UserId[9] and Password. An AuthenticationURL can then be generated and thus acquired based on the user's UserId. The AuthenticationURL provides access to an authentication server called OpenId Connect [96]. OpenId Connect uses both the provided UserId and Password to verify the user, which in turn returns an AccessToken[10]. This token grants access to a protected resource consisting of a list of organizations and roles displayed using the Login component. The user then selects an organization and role validated against the FHIR Server, which returns a SessionToken to be stored in the VueX component. The stored SessionToken can then be used to access data through the REST operations [97] mentioned in section 4.3.2.

[9]Username
[10]JWT Token [92]

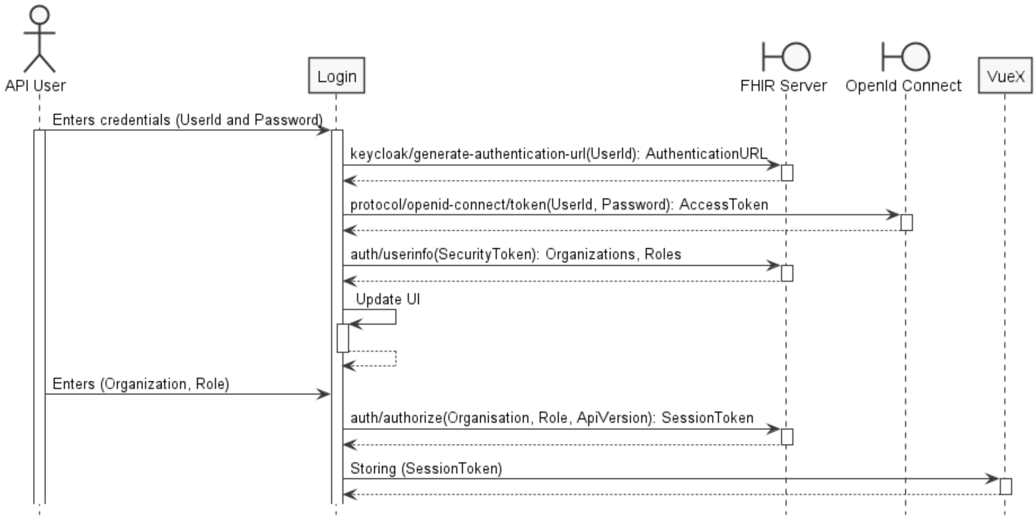

Figure 4.10: Sequence Diagram: Cura authentication of third party users.

4.6.2 ML API

This section shows how the connection between the prototype and the ML API has been established. Figure 4.11 shows a sequence diagram with the steps necessary to store data using the ML API. The sequence is based on the assumption that the user has been logged in successfully and has access to the actual Cura data based on the proof shown in section 4.6.1. Because it has not been possible to retrieve actual data from Cura, as previously mentioned, the prototype uses mocked data to use the ML API.

The Home component first needs to be created, after which a call to the ML API using REST is automatically made, as shown on listing 4.2. It should be noted that the sequence is based on case 1, as the sequence for case 2 is identically performed. The ML API returns both a probability and SHAP values as shown on listing 4.3, which are stored in the VueX component. The probability can thereupon be displayed, while the SHAP values is used to show the positive and negative influenced arguments as described in section 3.5.

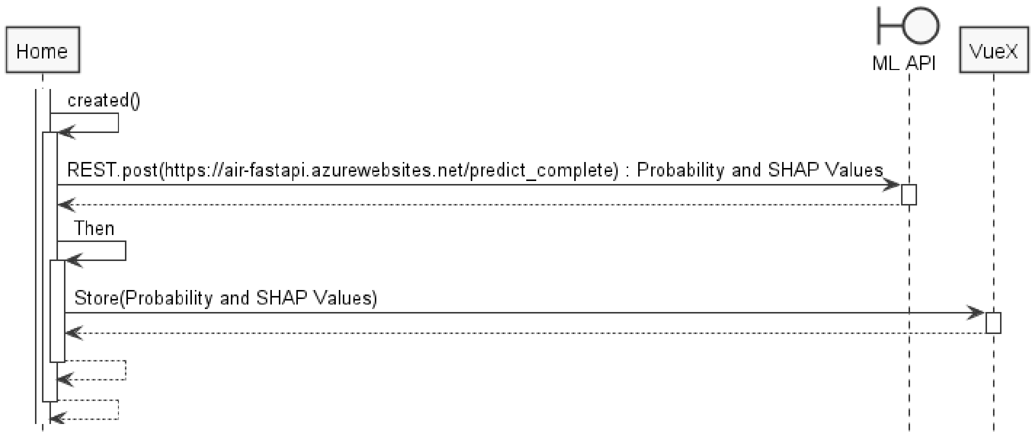

Figure 4.11: Sequence Diagram: Retrieving and storing information using the ML API

Chapter 5

Experiments and Results

This chapter introduces the three experiments that have been conducted over the course of the thesis. An experiment has required possible end users as participants to evaluate, validate and provide feedback using the methodology described in sections 3.6 to 3.8. Each of the three experiments presents its procedure, the state of the prototype at the time of the experiment, results, and a discussion. Data gathered from an experiment is used to redefine the prototype until tested in the next. All the data from which the results are based on can be found in chapter A.

5.1 The Participants

Every experiment has been performed using the recommended minimum of five participants [98], except for the third experiment, where only four participants could attend. The participants have been part of the prototype development from the beginning, as mentioned in section 3.4.1. The criteria under which the participants were chosen is as follows [41, pp. 5–6]:

- A case worker who is preferably over the age of 55.
- A case worker who is preferably under 30.
- A case worker with limited experience as a case worker.
- A case worker with few technical skills.

The four criteria target two types of participants. The first type is middle-aged case workers, with possibly many years of experience and presumably less technically inclined. The second type is younger case workers, with possibly fewer years of experience and presumably more technically inclined.

These two groups have been chosen as they complement each other. The middle-aged case worker's presumable lack of technical skills can cause them to struggle using the prototype. Ensuring high usability for this group could result in a larger degree of usability when applied to younger case workers. On the other hand, the younger case workers could rethink certain work processes and how these could be improved in the prototype [41, p. 6]. All criteria have been fulfilled, as the participants ages range from 36 to 63 with an experience level from 6 to 20 years.

5.2 Statistical Distribution

The results represented by different figures throughout each experiment uses a t-distribution. A t-distribution is similar in shape compared to a normal distribution but adjusted to account for smaller sample sizes, and thus ideal given the project's sample size of five [99]. Using a sample size of five, could potentially lead to a high the margin of error.

The used confidence interval is 95%, with an upper and lower boundary unable to exceed the maximum and minimum possible score. This was suggested by P. B. Stark [100], who stated "If the upper endpoint of a confidence interval for a population percentage is greater than 100%, it is legitimate to replace the upper endpoint by 100%. The confidence level remains the same." [100]. It should be noted that some of the results do not have a confidence interval, as no variation between the samples occurred.

5.3 Experiment 1

5.3.1 Purpose

The purpose of the experiment was to test the changes made since the last prototype iteration during R&D2 [42]. The experiment focused on establishing a baseline regarding the prototype's user experience using quantitative and qualitative data gathering techniques.

5.3.2 Experiment Design

The experiment was conducted remotely using Microsoft Teams [101] on the 18th of February 2021 with five participants, as illustrated on figure 5.1. It should be noted that the experiment was conducted five times, each with a different participant and that the participant's peripherals, as well as their computer screen size, was unknown while gathering the results during this experiment.

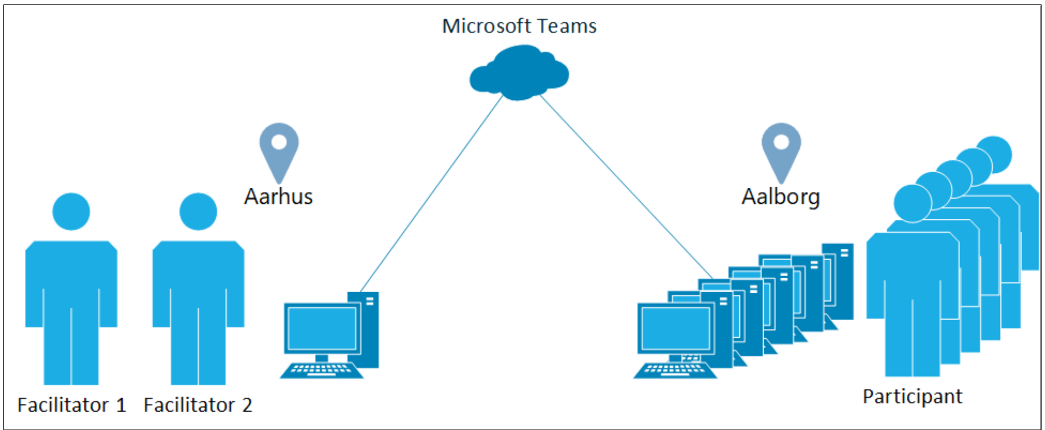

Figure 5.1: Experiment Setup for Experiment 1..

Table 5.1 shows the included data gathering methods described in chapter 3. A 75 minutes slot was reserved for each participant to make sure all necessary measurements were gathered. The experiment consisted of a usability test with a semi-structured interview regarding the prototype's design, first impressions, and functionality.

Table 5.1: Methods applied for Experiment 1.

Method	**Types**
Data Recording	Notes, Audio, Video
Interview	Semi-Structured
Usability Test	Performance, Usability Issues, SEQ, CSUQ

Prototype design

For this experiment, the prototype design has been developed using the initial design shown on figure 4.2. The prototype design for this experiment can be seen on figures 5.2 and 5.3, with the figures showing case 1 and case 2, respectively. Additional screenshots can be found in section A.2.2. Changes between the initial design and this experiments design is presented in table 5.2, with a complete changelog located in section A.2.1.

Table 5.2: Significant changes between the initial design and the design used throughout Experiment 1.

Id	Change description	Reason
1	Added a page for case 2.	At the time of developing the initial design (see figure 4.2), case 2 did not exist. When it did, it was considered relevant for the project and thus included.
2	Added a Navigation Drawer.	To be able to access both case 1 and case 2 and additionally functionality if needed.
3	Added a User Profile.	To see who has logged into the prototype.
4	Added a tab for the citizen's Diagnoses and Motivation.	To get a more clean separation between the information provided by the ML API and Cura. The pilot group approved this change (see id 38 in table A.16).
5	Added an SSN Search Field.	For the case workers to search for a citizen to obtain a result from both the ML API and Cura.
6	Added a Dark mode.	A function located in the navigation drawer for the end users to change the color contrast polarity to reduce eye fatigue.
7	Added a Color blind mode.	A function located in the navigation drawer for the end users to change the color scheme to enable support for red-green color blindness.

Continued on next page

Table 5.2 – continued from previous page

Id	Change description	Reason
8	Added a Login page.	A safety measure to prevent unauthorized users from gaining access to the prototype. It also acts as a verification measure for Cura as mentioned in section 4.3.2.

Case 1 - Rehabilitation Training
The prototype design for case 1 is presented on figure 5.2. The figure shows a probability of 77% for completing rehabilitation training and the positive and negative arguments that influence the probability. Among the changes described in the changelog is the SSN search field in the top left corner. Furthermore are the citizens *TRÆNINGSPLANER* shown on the right side.

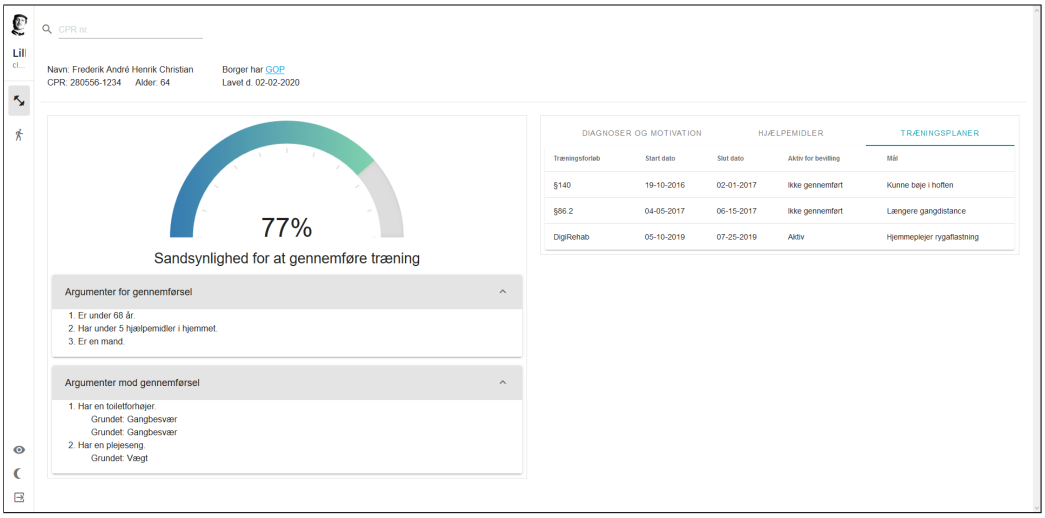

Figure 5.2: Prototype Design - Case 1 for Experiment 1.

Case 2 - Fall Preventive Training
The prototype design for case 2 is presented on figure 5.3. The figure shows a probability of 77% for falling within the next three months and the positive and negative arguments that influence the probability. Among the changes described in the changelog is the SSN search field in the top left corner. Furthermore are the citizens *HJÆLPEMIDLER* shown on the right side.

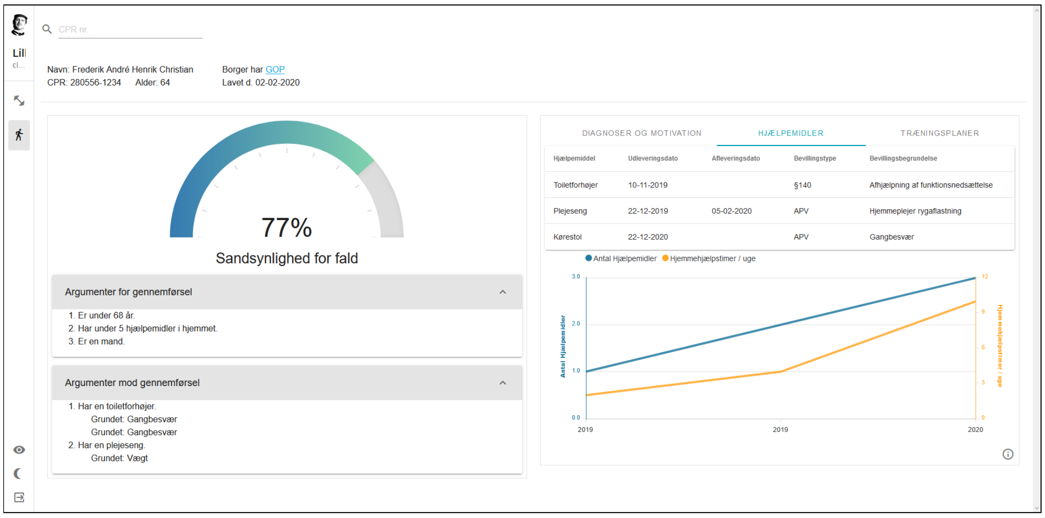

Figure 5.3: Prototype Design - Case 2 for Experiment 1.

5.3.3 Experiment Procedure

Before the experiment began, an introductory meeting was held with the participants to describe the changes and additions made since the last experiment during R&D2 [42]. The meeting also included an agenda explaining how the experiment would be conducted.

The experiment itself consisted of a remote usability test with 12 tasks shown on table 5.3, in the order of execution. An SEQ was answered after each task to identify the perceived task difficulty among the participants.

Table 5.3: English version of the SEQ used during Experiment 1. (Danish version is found in table A.4).
Status: A=Added, M=Modified, R=Removed, U=Unmodified

Task #	Identifier	Task description	Status
1	Log In	Log on the website. The username and password is the same as in Cura.	A
2	SSN Search	Search for a citizen, that has the SSN: 010151-0101.	A
3	Colorblind	Activate colorblind mode on the website.	A

Continued on next page

Table 5.3 – continued from previous page

Task #	Identifier	Task description	Status
4	Arguments For	Find and show the arguments for the citizen being able to complete training.	A
5	Arguments Against	Find and show the arguments against the citizen being able to complete training.	A
6	Hide Plot	Hide one of the plots in the graph.	A
7	Training Plans	Find and show the citizens training plans.	A
8	Assistive Aids	Find one the citizens assistive aids that has been issued in 2020.	A
9	GOP	Find and show the citizen's rehabilitation training plan.	A
10	Dark Mode	Activate dark mode on the website.	A
11	Fall Prevention	Find and show the page for fall prevention.	A
12	Log Out	Log out of the website.	A

The SEQ questionnaire followed each task to identify the perceived task difficulty, while both performance and usability issues were gathered. After all tasks were completed, the participants answered the CSUQ, followed by the semi-structured interview. The whole experiment session was recorded through audio and screen-capture, with note-taking during the semi-structured interview.

5.3.4 Results

SEQ

Figure 5.4 shows the mean Single Ease Question (SEQ) results for each task, with 1 being the worst score and 7 being the best. Tasks Arguments For, Arguments Against, and Log Out all received a perfect score from the participants. The worst scoring tasks were Hide Plot and Colorblind, as they respectively received 4,2 and 5,2. These two tasks also have the largest confidence interval, showing that the participants have mixed opinions. This is not the case compared to the tasks Assistive Aids, GOP, and Dark Mode which have the lowest confidence interval, thereby indicating larger agreement between the participants.

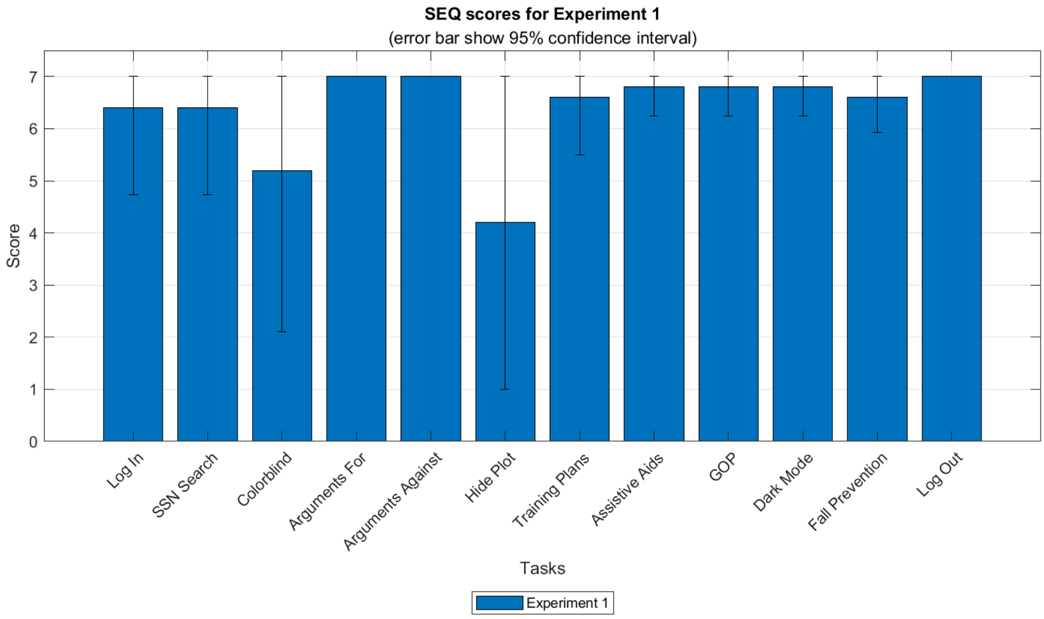

Figure 5.4: Mean SEQ scores for the tasks performed for Experiment 1.

CSUQ

In figure 5.5 the results from the Computer System Usability Questionnaire (CSUQ) are shown, with 1 and 7 being the best and worst scores, respectively. As shown in the figure, the overall score for the prototype is 1,40, which is approaching a perfect score of 1,00. The worst of the four scores is the InfoQual with a score of 1,63, which also has the largest confidence interval. Since all scores are close to 1,00, the participants consider the prototype to have good usability across the board.

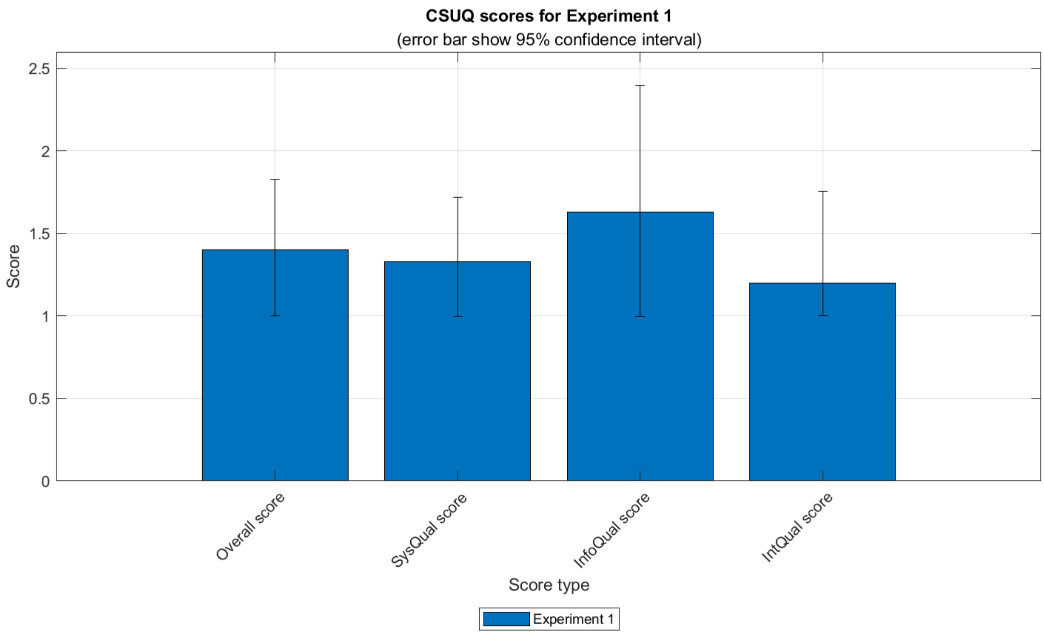

Figure 5.5: Mean CSUQ scores for Experiment 1.

Performance: Task success

The graph on figure 5.6 shows the percentage of participants that were able to complete the tasks without assistance, with assistance, and failed to complete the task. Figure 5.6 shows that the participants generally were able to complete the tasks, even if it required assistance. The GOP task is the most significant outlier as 80% of the participants were unable to complete the task, with only one participant to complete with assistance.

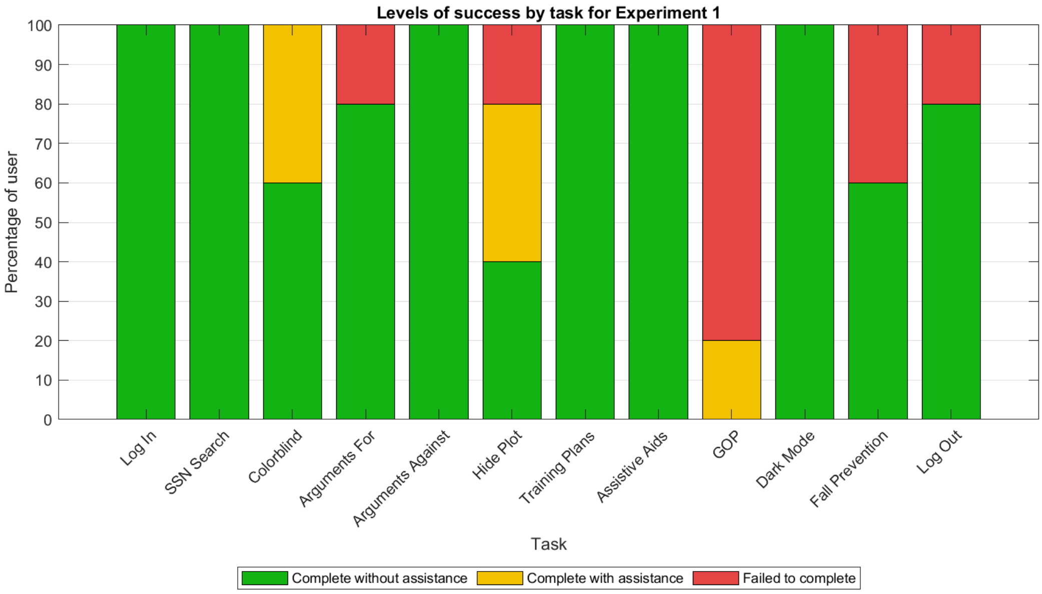

Figure 5.6: Percentage of participants for each task that could complete the tasks without assistance, with assistance, and failed to complete the task for Experiment 1.

Performance: Task time

Figure 5.7 shows the mean values for task completion time and total time on task. Comparing the task completion time and total time on task, the biggest difference is found in the task Arguments For, as the participants on average spent 2,82 as much time on the task as they did completing it. While the Arguments For task has the largest difference between the two mean values, other tasks such as Arguments Against, Hide Plot, and Fall Prevention also experienced more than a doubling between the two mean values.

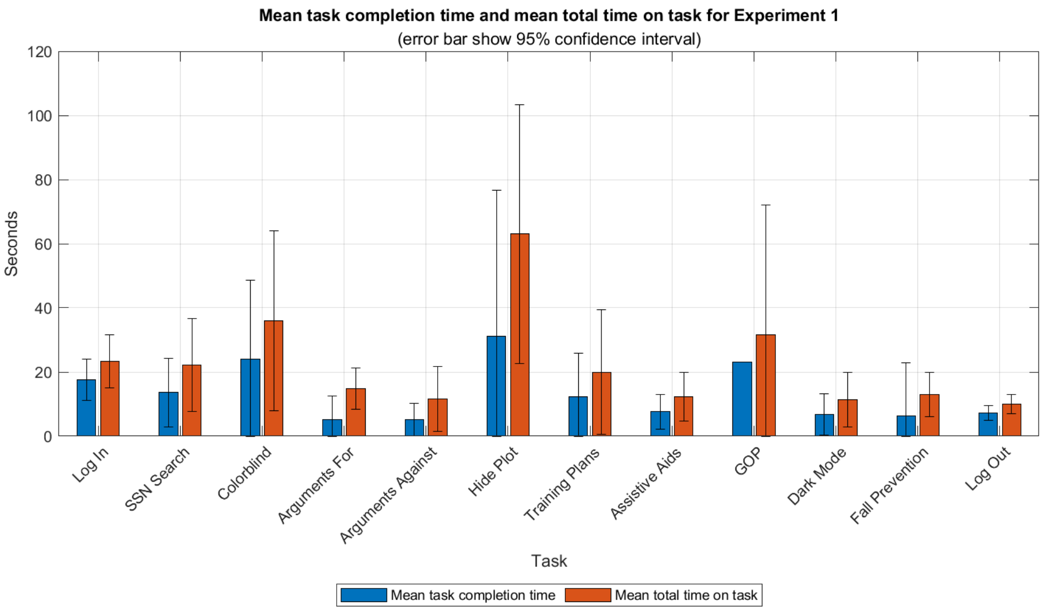

Figure 5.7: Mean task completion time and mean total task time for Experiment 1.

Performance: Errors

Looking at figure 5.8 it is worth noting that the stacked bars each represent their part of the total number of errors. E.g., task Arguments Against shows a total of five errors, with four errors being *Random click error* and one being *Manipulated argumentbox error*. Additional information regarding the error types can be found in section A.1.

Figure 5.8 shows that the same type of error typically is repeated in successive tasks and that the tasks Hide Plot and GOP have the largest amount of total errors. The Hide Plot task has a total of 22 errors, where the GOP task has less than half of that at only ten errors. Among all the tasks, Hide Plot also has the largest amount of different error types as there are seven, with the second-largest being three different error types. The three tasks Log In, SSN Search, and Log Out does not show any errors, as no errors where made during these tasks.

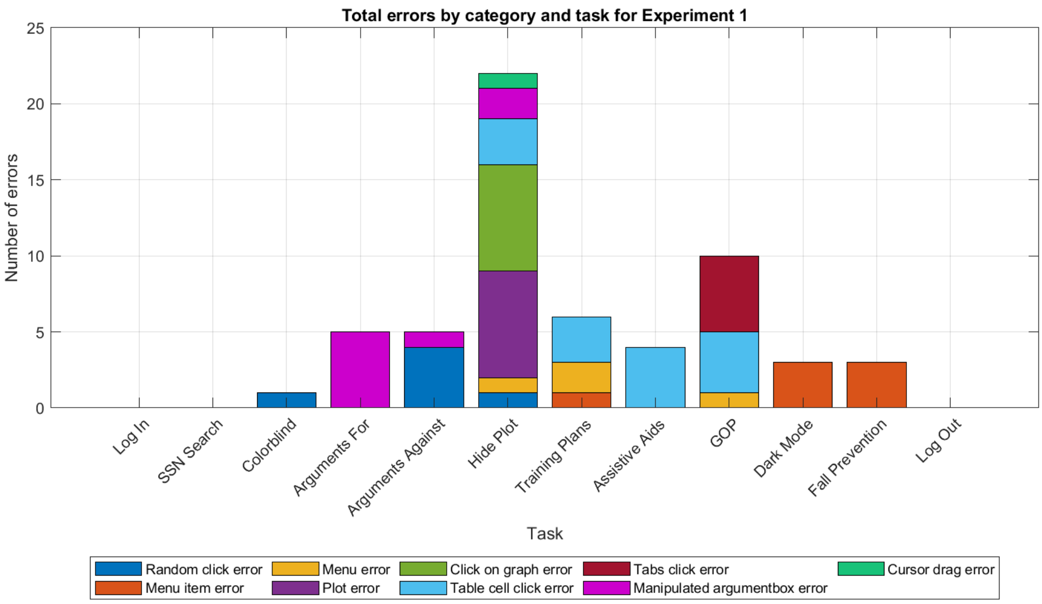

Figure 5.8: Categorised number of errors and total number of errors for each task for Experiment 1.

Performance: Efficiency

As described in section 3.7 efficiency consists of counting both mouse clicks and keyboard strokes in the prototype. The mouse clicks and keyboard strokes have also been counted after a task has reached the criteria for being completed, and the count is thus only stopped when the participant moves on to the next task.

The tasks Log In and SSN Search in figure 5.9 does not show a confidence interval, as it is not possible as no variance exists. Another thing to notice is that task Log Out has a lower *mean task efficiency* than the *expected efficiency*, which should not be possible. This can only be achieved if the participant does not perform the task correctly and thus failed to complete the task. All other tasks except for the three aforementioned show a greater mean efficiency compared to the expected efficiency, with the largest difference being the Hide Plot task with 3,3 times larger.

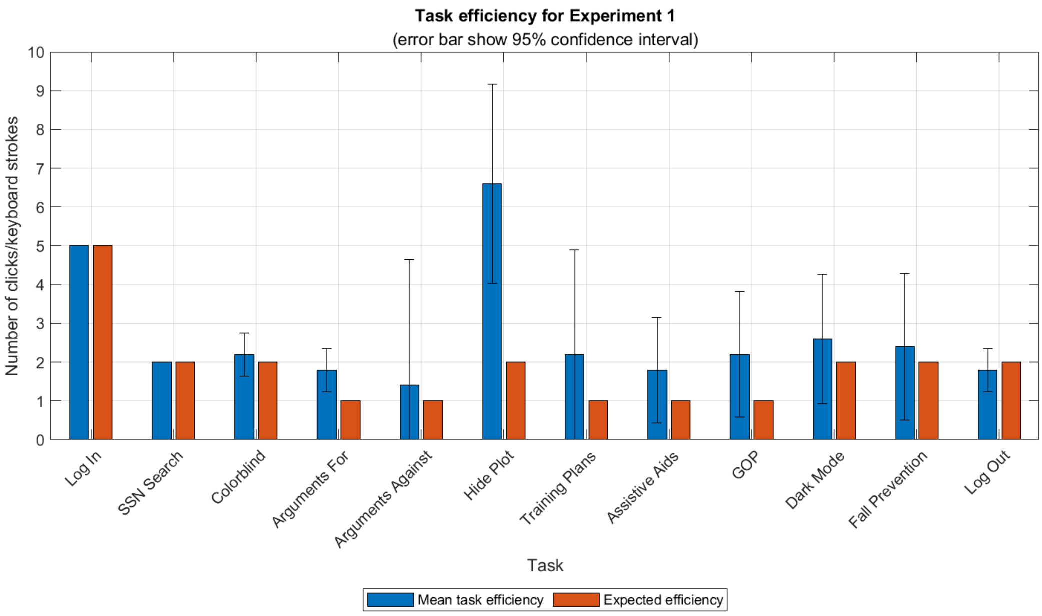

Figure 5.9: Mean task and expected efficiency for each task for Experiment 1.

Usability Issues

Each identified usability issue has been translated from table A.15 and shown on table 5.4 along with an identification of how many participants mentioned the same issue and an associated severity rating. Furthermore, a fix has been discussed for each usability issue as described in section 3.7.

Table 5.4: Experiment 1: List of usability issues and how many participants drew attention to the issue.

Id	Usability issue description	Mentioned by number of participants	Severity Rating
1	It is not obvious that the argument-boxes are clickable.	1	Minor
2	The contrast in the text is too high and the text size is generally to small.	2	Minor
3	It is not obvious that there is a menu in the system.	1	Minor

Continued on next page

Table 5.4 – continued from previous page

Id	Usability issue description	Mentioned by number of participants	Severity Rating
4	The inactive tabs have to be sought after to find.	2	Minor
5	The inactive tabs text is too pale and almost unreadable.	1	Minor
6	Some of the text in dark mode is not clear.	1	Moderate
7	The GOP should only be shown if there is an active GOP.	5	Moderate

Semi-structured interview

The five items seen in table 5.5 is a translated subset of the gathered feedback from table A.16. The table shows that the feedback is mostly positive and directed towards the UI as either compliments or minor changes. While item 4 can also be argued to contain some changes to the UI, it is mainly concerned with the prototype's functionality.

Table 5.5: Subset of translated feedback from semi-structured interview for Experiment 1.

Id	Feedback
1	The overall design is pleasant, and the blue colors are pleasing for the eyes.
2	The initial amount of information is good, as there is not too much or too little.
3	The color of the probability chart could be red-yellow-green.
4	Only active assistive aids should be shown in the assistive aids tab.
5	The menu should be placed at the top.

Objectivity: Probability and Arguments

During the semi-structured interview, some of the feedback was related to the probability and arguments and increasing the case workers' objectivity. This is shown in table 5.6, and has been translated from the danish version found in table A.16. Since this experiment has not been focused on the objective support, the feedback are limited. The feedback gathered however associates with the arguments being too sparse, or that the probability is the main focal point.

Table 5.6: Subset of translated feedback from the semi-structured interview related to the provision of objective support for Experiment 1.

Id	Feedback
1	I only look at the number that has been presented.
2	The arguments was a bit deficient.

5.3.5 Discussion

SEQ

Looking at the results and the mean values for the SEQ questionnaire in figure 5.4, the participants generally find the tasks very easy to perform. This is most likely due to the participants in the experiment has been part of the design process. Because of this previous knowledge, they potentially remember how the prototype works, causing them to find the tasks more straightforward and trivial.

Focusing on the task Hide Plot, which as mentioned has the worst mean SEQ score of all, as the participants possibly did not understand the task. Another factor could be that the participants used time and effort to explore the prototype before requesting assistance, after which they succeeded in completing the task. The effects of the exploration are also reflected in the measurements for the total time on task, errors, and efficiency (see figures 5.7 to 5.9), where the task has the worst performance among all the task. A way to improve the results would be to improve the task description by making it unambiguous.

Shifting the focus to the Colorblind task, which has the second-worst SEQ score, all the participants managed to complete the task while 40% needed assistance as shown in figure 5.6. The main reason could be found among the identified usability issues as some of the participants stated that the menu was hard to see (third usability issue in table 5.4).

CSUQ

The scores from the CSUQ, shown in figure 5.5, generally tell that the prototype is usable in all four measured categories, where the mean overall score is 1,40, which is close to a perfect score (1,00). A drawback of having an almost perfect score indicates that little to no further improvements should be made to the prototype. The near-perfect score could be explained by the participant's involvement in the design process, causing them to be biased towards liking the prototype more than a person using the prototype for the first time. The worst score across all four categories is InfoQual as shown on table A.7, which is as expected as item 7 in a CSUQ usually generally produces a worse score compared to the rest [65, p. 196].

Performance: Task success

While most of the participants could complete the tasks without and with assistance, at least 20% failed to complete five tasks. This could indicate that the actions necessary to complete the tasks were unintuitive.

Looking specifically at the GOP in figure A.10 it is possible to see that 80% of the participants failed to complete the task while 20% completed it with assistance. Knowing that 80% of the participants failed to complete the task, while the task received the second-highest SEQ score, shows that the participants did not understand what the task entailed.

Focusing on the Fall Prevention task, which has the second-highest failure rate at 40%. This could be explained by the participants misunderstanding the task and not remembering the menu was used during the Colorblind task. A usability issue also emphasized the issue with the participants not remembering the menu (see the third usability issue in table 5.4). While the usability issue can partly explain why the participants had trouble completing the Fall Prevention task, it should be noted that the menu is an added functionality as seen in table 5.2.

Performance: Task time

At the beginning of the experiment, the difference between the task completion time and total time on task was more prominent than at the end of the experiment, as seen on figure 5.7. This is most likely because they had not interacted with the prototype since R&D2 and had memorized the location of the different functionalities at the end. Studying the video recordings of the participant's performance during the usability test also shows uncertainty in interacting with the prototype. Another reason could be that the participants were uncomfortable with the experiment setup due to the experiment being conducted remotely.

Looking at the Hide Plot in figure 5.7, a significant difference is seen between task completion time and total time on task. The average participant spends 63 seconds on this task, which could be an explanatory factor to some of the errors produced (see figure 5.8) since the participants continued to explore the prototype after reaching the task completion criteria. The exploration has also influenced the efficiency results (see figure 5.9) since they, on average, had more than three times the amount of the expected efficiency.

Performance: Errors

Figure 5.8 reveals that the same type of errors is usually made in adjacent tasks. This is, in all likelihood, due to the construction of the task sequence in the usability test. Also shown is the variety of different error types, which are kept to one, two, or three for all tasks except for Hide Plot.

Studying these error shows that seven out of the 22 errors is categorised as the Plot error, which is a combination of the two error types Hide plot error and Show plot error (see table A.2). This means that after a participant completed the task, they continued experimenting, as it would not be possible to make these error types without completing the task. The continued experimentation is also reflected in the total time on task and efficiency, as stated earlier, and in the participants' usability test recording.

The GOP task, which has the second-highest number of errors, consists of three error types, with Tabs click error being responsible for half of these errors. This type of error could be due to a previous task, where the participant was to find the table containing a citizen's training plans. During this task, a GOP was shown to the participants, which primed them to believe that they should find the same GOP again during the GOP task. Instead, the correct action was to click on the citizens GOP, located at the top of the prototype (see figure 5.2).

Because the participants located the wrong GOP and the GOP task stated that they should "show the citizen's rehabilitation training plan", this influenced them to click on the wrong GOP, which explains the Table cell click error. When a participant clicks in the table, nothing happens, which is likely the cause for one participant requesting assistance with the task and thereby completing it (see figure 5.6).

Performance: Efficiency

The efficiency results as seen on figure 5.9 show the participants, on average, performs the correct sequence of actions without many additional actions. As seen on the figure 5.9, some tasks have a mean task efficiency matching the expected efficiency and no confidence interval, meaning no errors were made. This is due to the causation effect between errors and efficiency as mentioned in section 3.7, as an error causes an efficiency increase, but not vice versa.

An example of this is seen in the Log Out task, which has a lower mean task efficiency (1,8) compared to the expected efficiency on (2,0). This is due to 20% of the participants failed the task, and no errors were made, which have caused the task to have a lower mean efficiency than expected.

Usability Issues

Looking at the usability issues, it is essential to remember that the experiment was conducted remotely through screen sharing and remote control using Microsoft Teams [101]. This means that it is impossible to know whether or not the participants had the screen shared in fullscreen or if the internet connection caused pixelation, distorting the prototype's look.

The seven usability issues shown on table 5.4 is divided into three groups where the solution to the usability issues applies to the whole group. The first group consists of issue 2 and 4 which is related to the text readability and discoverability, where the solution would be to increase the font size. The individual end user can do this by employing the browser built-in zoom functionality, as only two participants experienced these issues. Furthermore, the validity of the issue can be questioned as there are conflicting statements regarding the font size (see table A.16), and the screen size for the participants was unknown.

The second group consists of usability issues 1, 3, 5, and 6, which relate to text color and color in other parts of the prototype. For issue 3, where one participant stated the menu was not obvious to find, the solution was to color the navigation drawer to attract more attention. This solution should also reduce the task completion time, total time on task, errors, and efficiency while increasing the task success rate without the participants needing assistance or failing when the menu should be used. For the remaining usability issues 1, 5, and 6, the solution could also be to add more color, but this should be done carefully so that the prototype's general coherence is intact.

Usability issue 7 was mentioned by all five participants, indicating the importance of finding a solution. The issue can be solved by adding functionality that only shows the GOP when it is active.

Semi-structured interview

The five points of feedback in table 5.5 is an indication of what the participants think of the prototype. Most of the feedback is positive, which is also reflected by the very positive overall CSUQ score. The complete feedback list can be found in table A.16.

Looking at the feedback, the participants generally seem satisfied with the design and layout of the prototype, while there is an opportunity for improvements. A possible improvement could be to move the navigation drawer to the top of the prototype. While this is a valid improvement, it will not be done as Cura have their navigation on the left, and matching this creates greater coherence between the applications.

Some of the other feedback seen in table 5.5 have to do with the color, or lack thereof, which is also expressed in the usability issues (see table 5.4). Even though color can have a considerable impact, it can be applied with relative ease from a programmatic standpoint, where the difficulty could lie in finding the "correct color." Using these five items as a guide shows how the feedback relates to the prototype design, presented information, and styling.

Objectivity: Probability and Arguments

Because the focus of this experiment has not been on measuring an increase in the participants objectivity, the gathered data has been sparse. The data stems from the feedback given during the semi-structured interview, with only two points of feedback. Both feedback points have a neutral to negative tone. Some participants stated they do not look at the arguments but only at the probability, and other participants stated that the arguments were deficient. This feedback shows an overall trend of the participants not being satisfied with the arguments.

5.3.6 Summary

The results from the SEQ questionnaires show that the participants found nearly all tasks, except for two, easy to complete. These exceptions indicate a problem with the prototype's usability, as the participants have been part of the design process, and therefore should have an idea of how to operate the prototype.

The CSUQ shows that the participants found the prototype very usable. This usability is proved through the near-perfect scores, indicating that the participants suffer from a bias because they have been part of the design process. Another downside is that they leave little to no room for improving the prototype.

Shifting the focus to the performance metric results, the tasks requiring the most time is also the tasks that have the lowest SEQ scores. The metrics also reflect causation between the total time on task, the amount of errors produces, and the efficiency. This correlation is seen most strongly for the Hide Plot and GOP tasks.

The usability issues and semi-structured interview feedback were mainly centered around the design. Not all of the feedback was usable, as Experiment 1 was performed remotely, resulting in no knowledge about how the participants actually experienced the prototype. While the objective support is present in the prototype, the usefulness can be questioned due to the participants finding the arguments deficient.

5.4 Experiment 2

5.4.1 Purpose

The second experiment was much like the first, as it used the same methods to allow for a comparison to detect a potentially optimized user experience. Furthermore was two argument-design suggestions evaluated to identify the participants preferences regarding how the arguments should be presented as mentioned in section 3.5.

5.4.2 Experiment Design

Experiment 2 was conducted remotely using Microsoft Teams [101] on the 24th of March, 2021 with five participants. This is illustrated on figure 5.10. It should be noted that the experiment was conducted five times, each with a different participant and that the participant's peripherals, as well as their computer screen size, were unknown while gathering the results during this experiment.

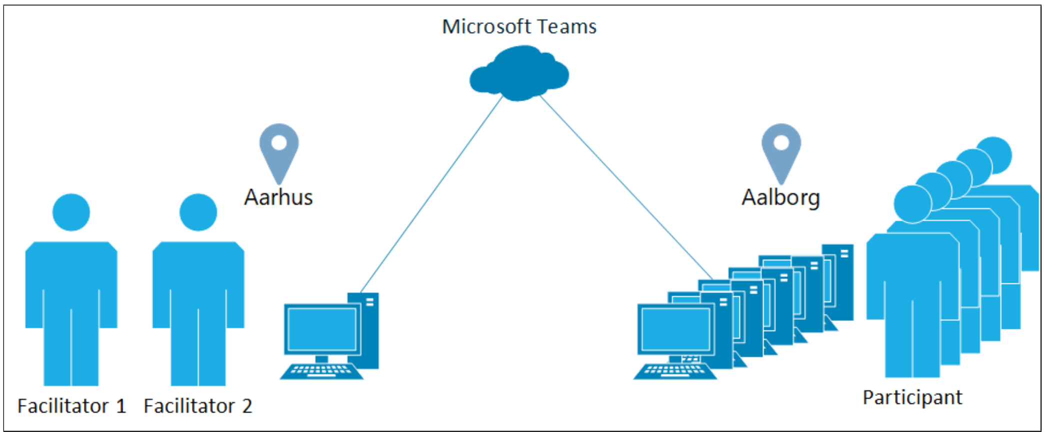

Figure 5.10: Experiment Setup for Experiment 2.

Table 5.7 shows the included data gathering methods described in chapter 3. The experiment consisted of a usability test and a semi-structured interview, with each experiment having a duration of 60 minutes for each of the five participants. As for the included methodology, the only difference compared to Experiment 1 was adding the SUS and UMUX to evaluate the aforementioned design suggestions, and the inclusion of the learnability performance metric.

Table 5.7: Methods applied for Experiment 2.

Method	Types
Data Recording	Notes, Audio, Video
Interview	Semi-Structured
Usability Test	Performance, Usability Issues, SEQ, CSUQ, SUS, UMUX

Prototype design

For this experiment, the prototype design has been developed using the design from Experiment 1 shown on figures 5.2 and 5.3. The prototype design for this experiment is shown on figures 5.11 and 5.12, and includes changes compared to the design during Experiment 1. The most significant changes are shown and justified in table 5.8, with the complete changelog located in section A.3.1.

Table 5.8: Significant changes between Experiment 1 and 2

Id	Change description	Reason
1	Added an established connection between the prototype and the ML API.	To show a non-hardcoded probability for both case 1 and case 2.
2	Changed the color on the navigation drawer.	Resolved usability issue found during Experiment 1 (see number 3 on table 5.4).
3	Changed the terminology to be closer to the information used in Cura.	The pilot group saw a discrepancy between the terminology used in the prototype and Cura during the semi-structured interview (see number 49 and 50 in table A.16).
4	Removed the SSN Search Field.	The search field was no longer necessary due to the reasons explained in section 4.4.2.
5	Removed the User Profile.	It was mentioned during the semi-structured interview as unnecessary (see number 74 in table A.16).
6	Removed the explanation content from the argument containers.	After the connection to the ML API was made, it became clear that a justification behind each argument was not possible due to the current state of the project model.

Case 1 - Rehabilitation Training

Shows a citizen's 59% probability to complete rehabilitation training, which originates from the ML API. The menu gained a light-grey color, the search field was removed, and more assistive aids were added.

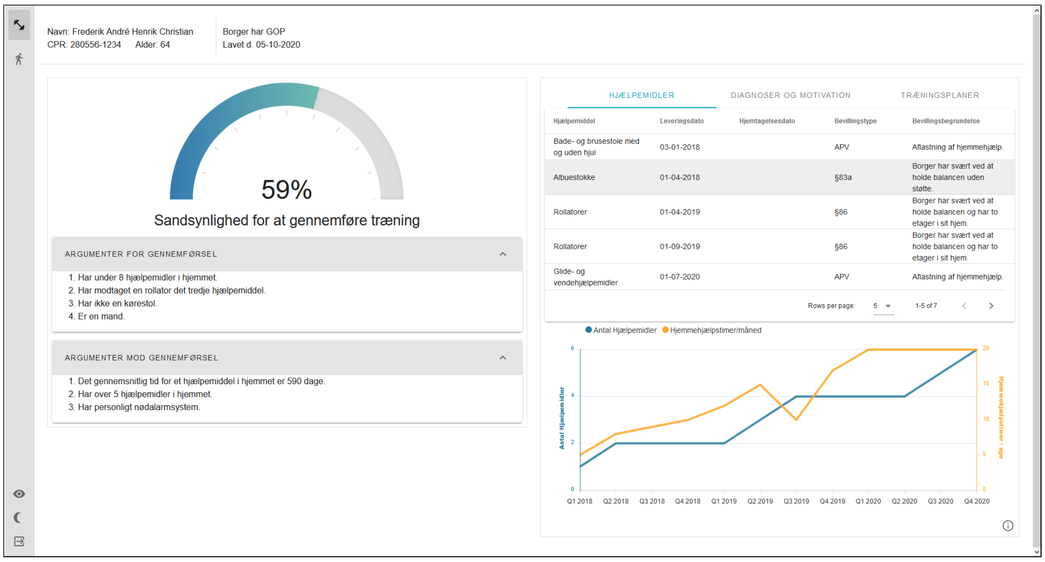

Figure 5.11: Prototype Design - Case 1: Argument-Design 1 for Experiment 2.

Case 2 - Fall Preventive Training

Just like case 1, the 79% probability for falling within the next three months originates from the ML API. As mentioned in section 3.5, two argument-designs were made. The prototype on figure 5.12 shows the alternative argument-design.

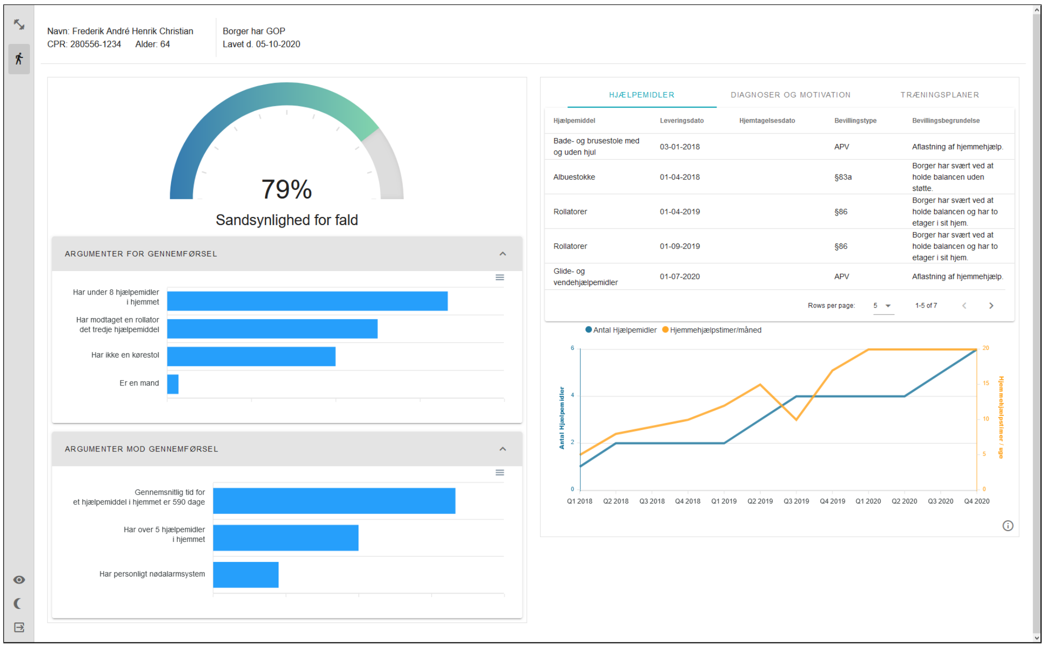

Figure 5.12: Prototype Design - Case 2: Argument-Design 2 for Experiment 2.

5.4.3 Experiment Procedure

Experiment 2 was structured similar to Experiment 1, as the objective had not changed except for the two argument-design comparisons. Before the experiment began, an introductory meeting was held with the participants to describe the changes and additions made since Experiment 1. The meeting also included an agenda explaining how the experiment would be conducted.

The experiment itself consisted of a remote usability test, with 11 tasks to complete shown on table 5.9, in the order of execution. Due to some of the changes shown in table 5.8, two tasks have been removed, and one added. Furthermore, some of the task descriptions were changed to increase the task's understandability. The tasks and task descriptions are shown on table 5.9.

Table 5.9: English version of the SEQ used during Experiment 2. (Danish Version is found in table A.18).
Status: A=Added, M=Modified, R=Removed, U=Unmodified

Task #	Identifier	Task description	Status
1	Log In	Log on the website. Use the pre-filled username and input a random password.	M
-	SSN Search	Search for a citizen, that has the SSN: 010151-0101.	R
2	Colorblind	Activate colorblind mode on the website.	U
3	Arguments For	Find an argument that states the citizen is able to complete training.	M
4	Arguments Against	Find an argument that states the citizen is unable to complete training.	M
5	Hide Plot	Hide one of the plots that show the assistive aids and homehelp hours over time.	M
6	Training Plans	Find the citizens training plans.	M
7	Assistive Aids	Find the citizens assistive aids.	M
-	GOP	Find and show the citizen's rehabilitation training plan.	R
8	Diagnoses	Find the citizens diagnoses.	A
9	Dark Mode	Activate dark mode on the website.	U
10	Fall Prevention	Find the probability that shows if a citizen should receive fall preventive training.	M
11	Log Out	Log out of the website.	U

The SEQ questionnaire followed each task to identify the perceived task difficulty, while both performance and usability issues were gathered. After all tasks were completed, the participants answered the CSUQ. The participant was then presented with the two argument-designs. After the participant spent some time with the first design, the SUS and UMUX questionnaires were answered. The participant was then presented with the second design, followed by another set of SUS and UMUX questionnaires.

At the end of the experiment, the semi-structured interview was conducted. The whole experiment session was recorded through audio and screen-capture, with note-taking during the semi-structured interview.

5.4.4 Results

SEQ

Figure 5.13 shows the mean Single Ease Question (SEQ) results for Experiment 1 and 2, with 1 being the worst score and 7 being the best. The graph shows that the amount of tasks with the highest possible score has increased from three tasks in Experiment 1 to eight in Experiment 2.

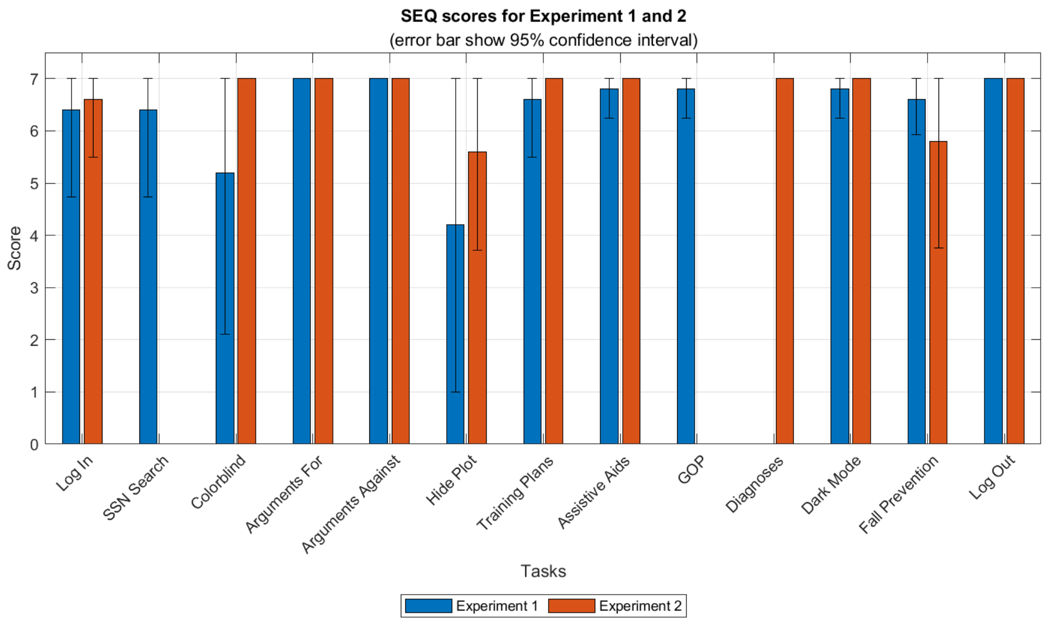

Figure 5.13: Mean SEQ scores for the tasks performed for Experiment 1 and 2.

The task Hide Plot was still the lowest scoring task with a score of 5,8, up from 4,2, with a more narrow confidence interval, meaning a more significant agreement of the task's difficulty among the participants. The Fall Prevention task closely follows the Hide Plot task, whose score has decreased from 6,6 to 5,8. Except for the Fall Prevention task, all other reoccurring tasks have seen an improvement, compared to the results from Experiment 1.

CSUQ

Figure 5.14 shows the mean Computer System Usability Questionnaire (CSUQ) scores for Experiment 1 and 2, with 1 and 7 being the best and worst scores, respectively. The graph shows a worsening in the results since Experiment 1. While the SysQual scores are identical, the Overall and InfoQual score have only worsened slightly compared to the IntQual score.

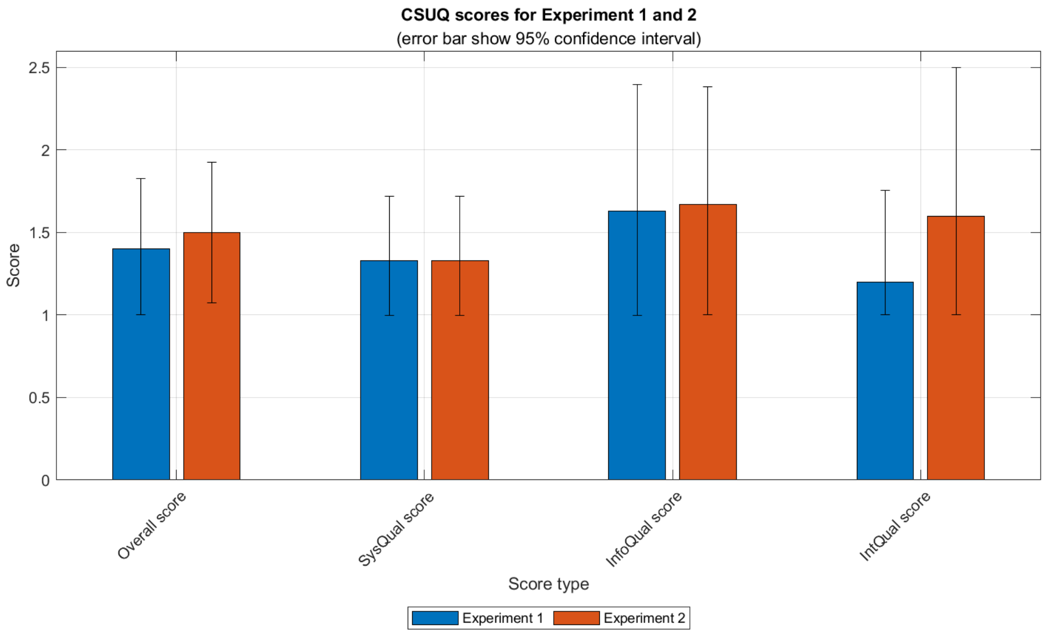

Figure 5.14: Mean CSUQ scores for Experiment 1 and 2.

SUS & UMUX

Figure 5.15 shows the mean System Usability Score (SUS) and mean Usability Metric User Experience (UMUX) result for the different argument-designs. *Design 1* refers to the participants requested textual argument-design seen on figure 5.11 while *Design 2* refers to the alternative argument-design seen on figure 5.12.

Looking at figure 5.15 shows an overall better UMUX score compared to SUS and that design 1 has a significantly higher score compared to design 2. A thing to notice is the confidence intervals between the two designs, as a larger agreement is indicated for design 1, due to the significantly narrower confidence interval.

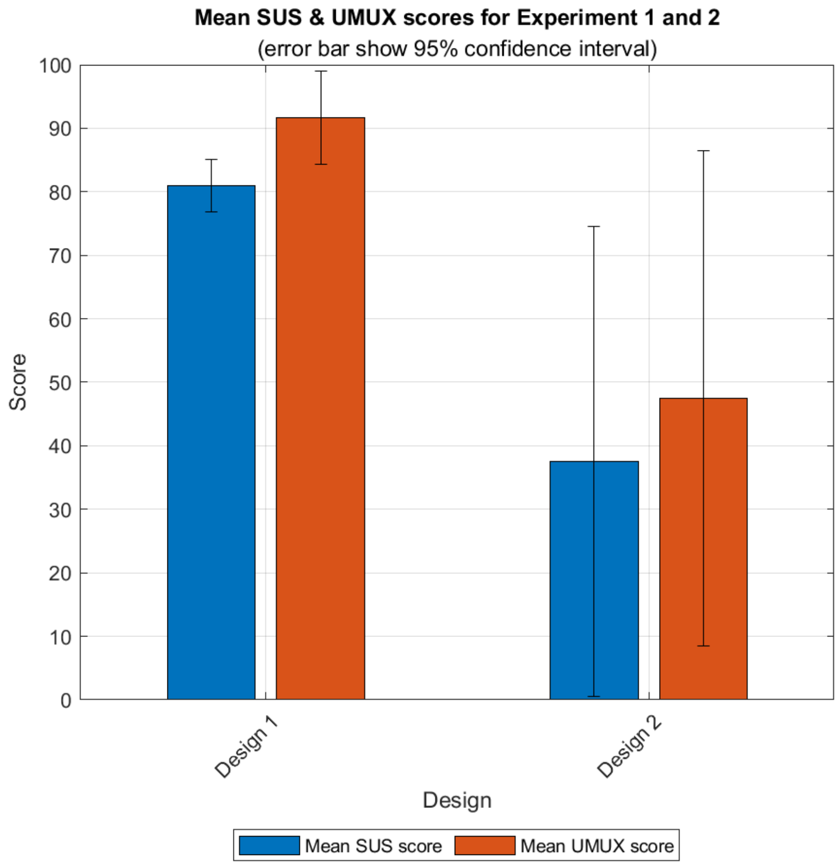

Figure 5.15: Mean SUS and UMUX scores for design 1 and design 2 for Experiment 2.

Performance: Task success

The levels of success for each task performed during Experiment 2 is shown on figure 5.16. As shown, no participants required any assistance to complete a task. All participants completed four out of the eleven tasks, meaning that at least one participant failed the remaining seven tasks.

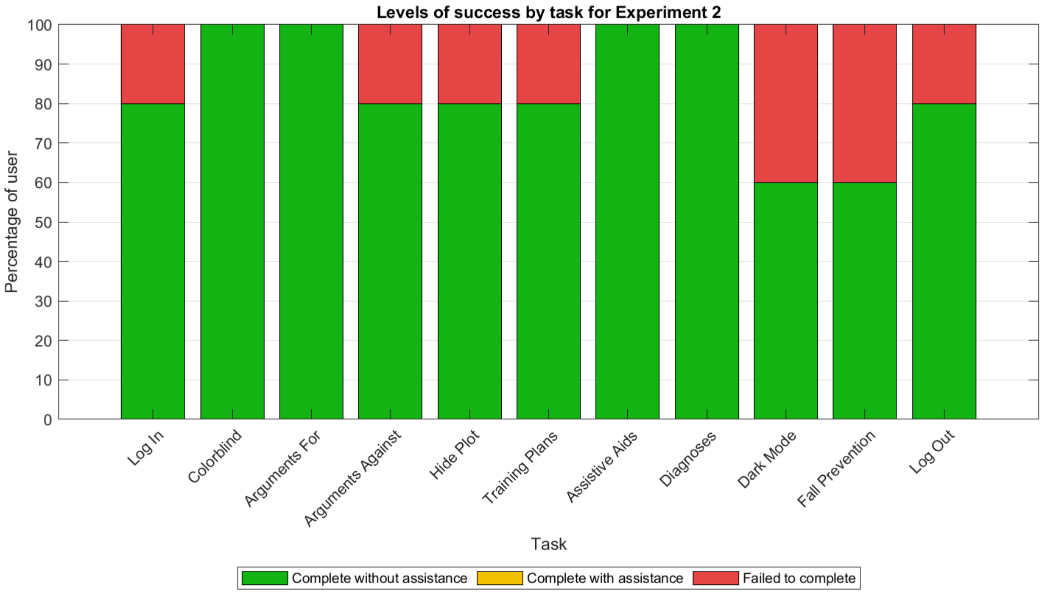

Figure 5.16: Percentage of participants for each task that could complete the tasks without assistance, with assistance, and failed to complete the task for Experiment 2.

Comparing the completed number of tasks between Experiment 1 and Experiment 2, as seen on figure 5.17, shows that only the Arguments For task has an increased completion rate. Five kept the completion rate out of the remaining tasks, and four saw a decrease in their completion rate. Among the tasks with the decreased completion rate, the most significant difference was found in the Dark Mode task, where 40% failed to complete the task.

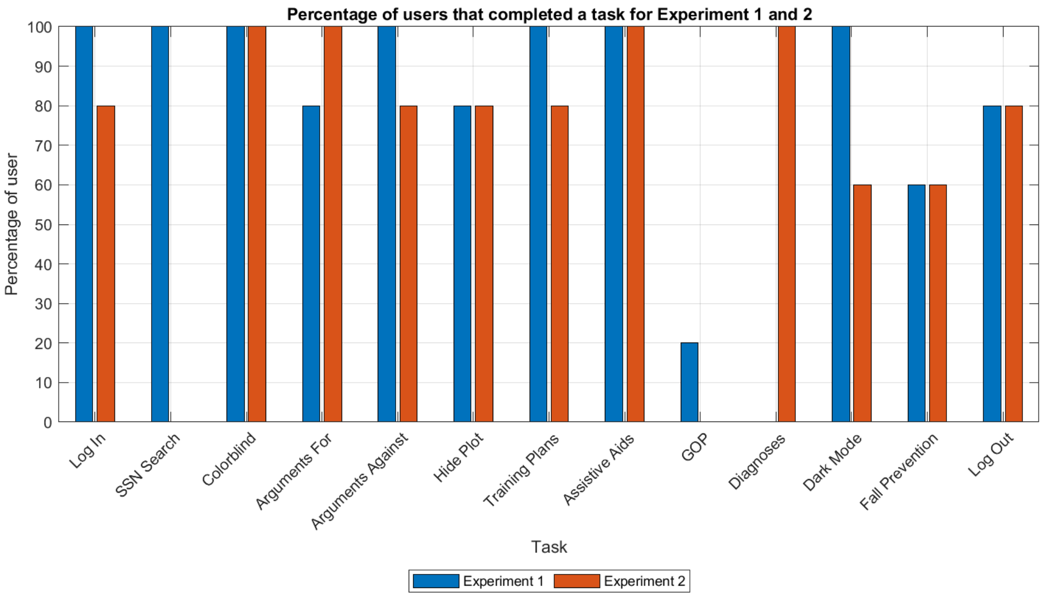

Figure 5.17: Percentage of participants that completed each task for Experiment 1 and 2.

Performance: Task time

Figure 5.18 shows the mean value for the task completion time and the total time on task for each task performed during Experiment 2. The Fall Prevention has the highest task completion time and the smallest difference between the completion time and the total time on task.

Comparing Experiment 1 and 2 the participants have, on average, gotten faster at completing the tasks, as seen on figure 5.19, with the exception of the Arguments For and Fall Prevention tasks. Total time on task seen on figure 5.20 shows the same trend, with the same two exceptions.

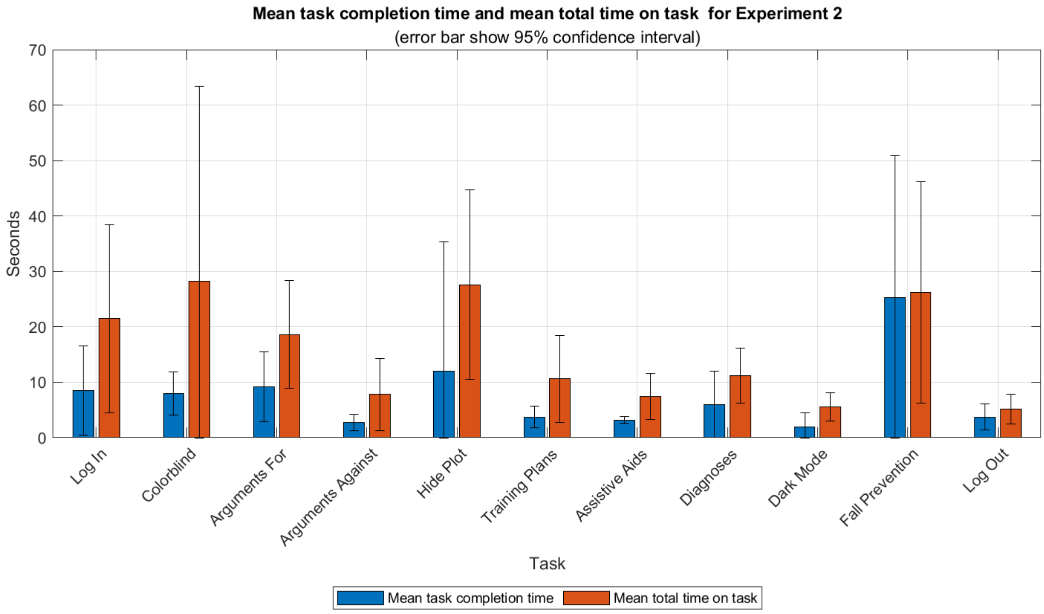

Figure 5.18: Mean task completion time and mean total task time on task for Experiment 2.

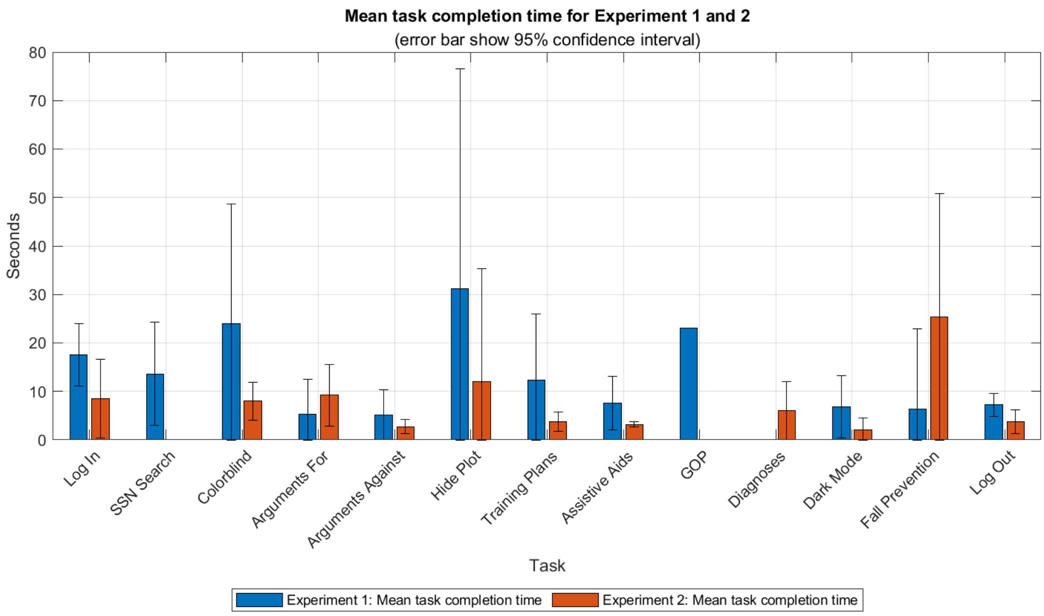

Figure 5.19: Mean task completion time for Experiment 1 and 2.

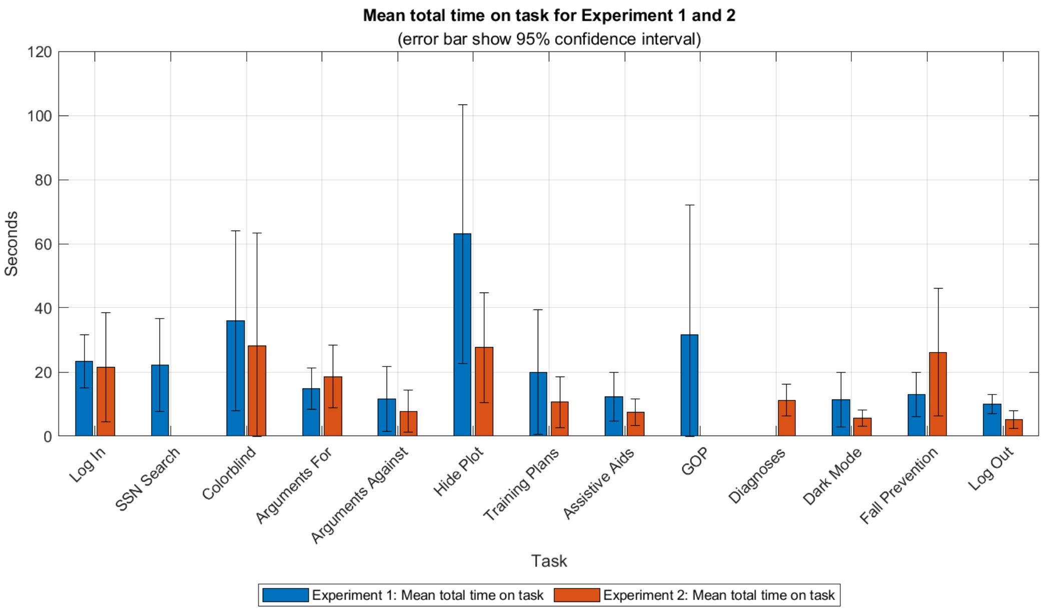

Figure 5.20: Mean total time on task for Experiment 1 and 2.

Performance: Errors

Looking at figure 5.21 it is worth noting that the stacked bars each represent their part of the total number of errors. E.g., task Arguments For shows a total of two errors, with one error being *Random click error* and another *Menu error*. Additional information regarding the error types can be found in section A.1.

The largest amount of errors are made by the two tasks Hide Plot and Fall Prevention, which combined contributes 80% of the total number of errors. During all other tasks, either zero, one, or two errors were made in total.

Figure 5.22 shows the total amount of errors made during Experiment 1 and 2. As seen on the graph, the total amount of errors for Experiment 2 were almost halved the amount of those made during Experiment 1.

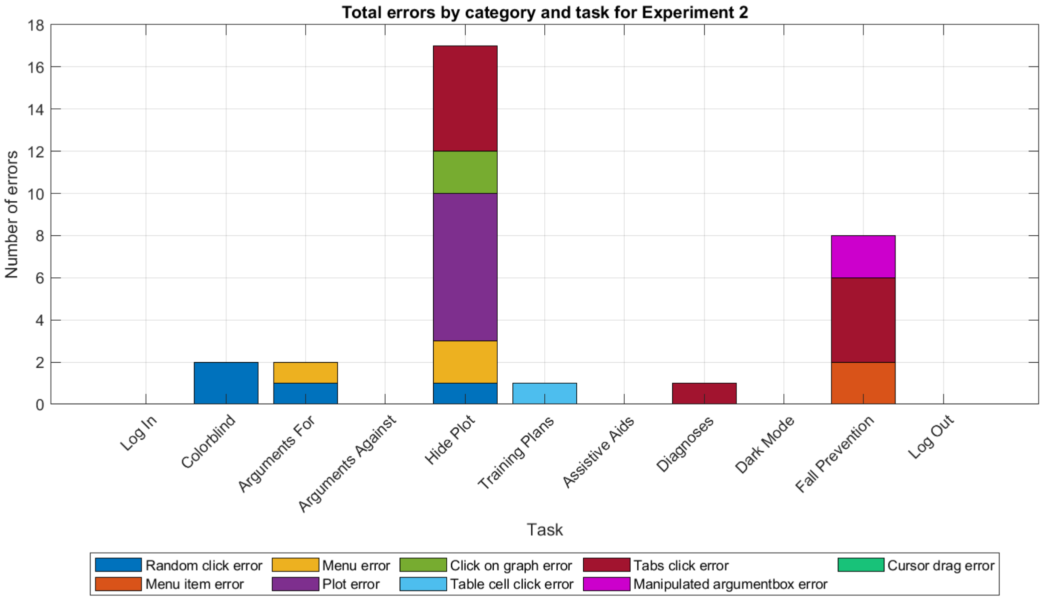

Figure 5.21: Categorised number of errors and total number of errors for each task for Experiment 2.

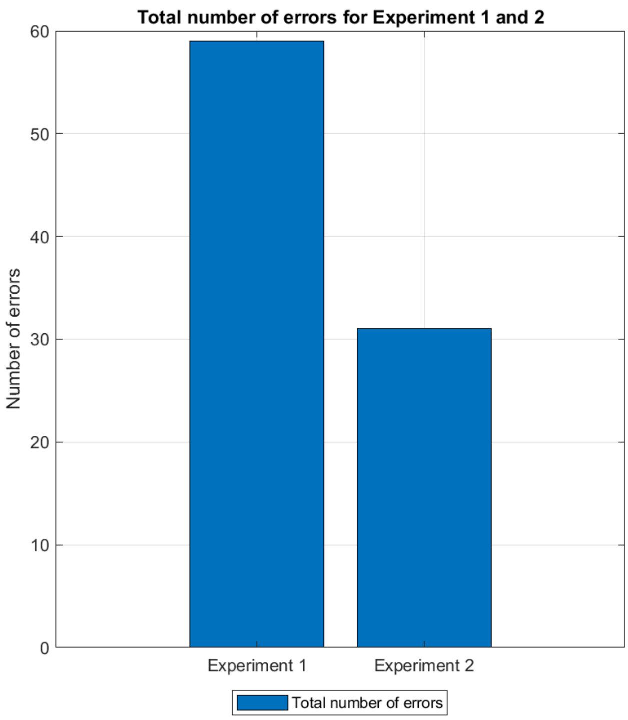

Figure 5.22: The total amount of errors for Experiment 1 and 2.

Performance: Efficiency

Figure 5.23 shows the mean efficiency for each task performed during Experiment 1 and 2. It is here important to notice that for Experiment 2, the expected efficiency for task Log In has changed from five to two (see number 4 and 9 in table A.17).

The tasks Colorblind and Fall Prevention, have a higher efficiency score for Experiment 2 compared to Experiment 1. As seen on the graph, the tasks Arguments Against, Training Plans, Assistive Aids, Dark Mode, and Log Out all have a lower mean task efficiency than the expected efficiency. This can only be achieved if at least one participant has failed to complete the tasks. Hide Plot has the highest efficiency in both Experiment 1 and 2.

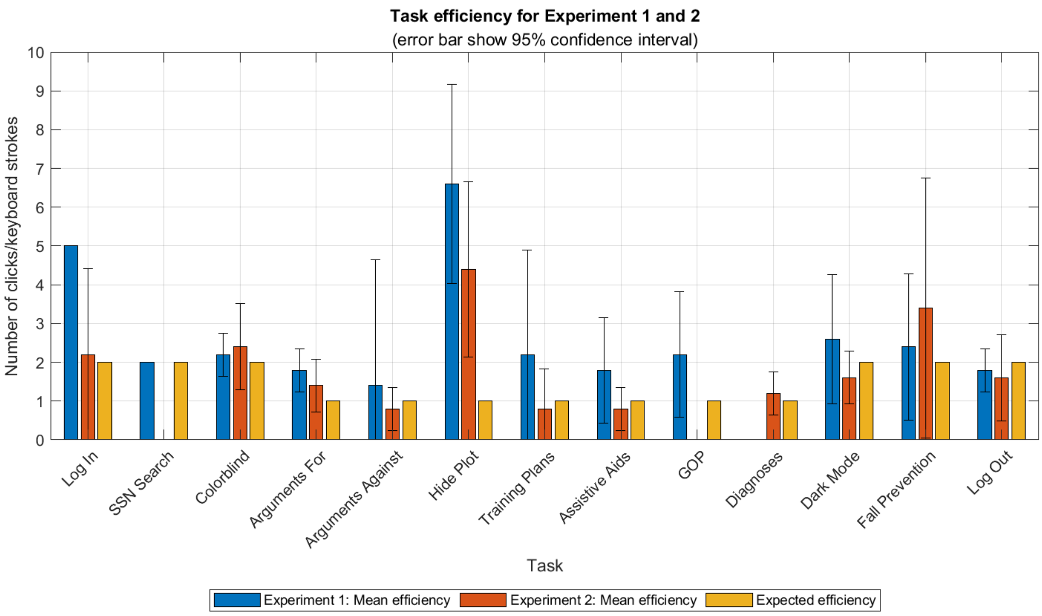

Figure 5.23: Mean efficiency scores for the different tasks for Experiment 1 and 2.

Performance: Learnability

Figure 5.24 shows the learnability score for Experiment 1 and 2. The learnability score is an aggregated score from total time on task, the total number of errors, and the mean task efficiency score, where a lower score is better.

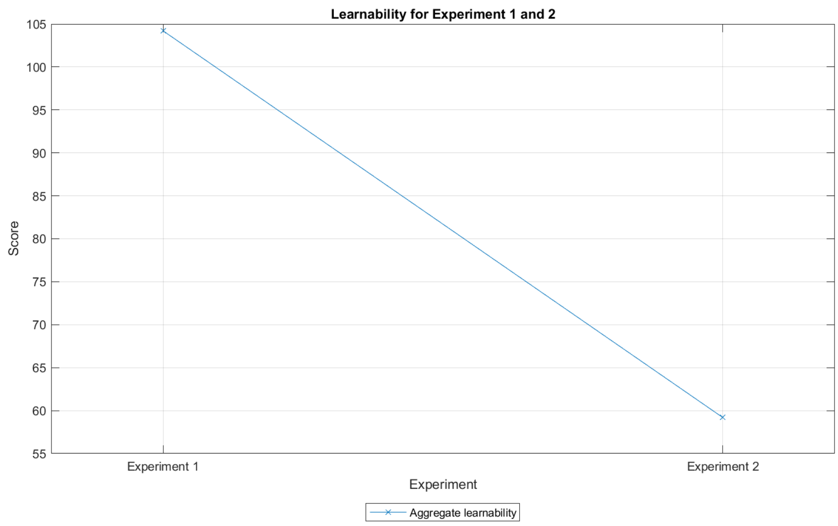

Figure 5.24: Aggregated learnability score for Experiment 1 and 2.

Usability issues

During the usability test, the participants did not verbalize any usability issues. While this does not mean that no usability issues exist, it only means that the participants did not verbalize any issues during the test.

Semi-structured interview

The five items seen in table 5.10 is a translated subset of feedback from table A.34. The five items provides an overview of the general perception of the prototype, as there is some feedback regarding the different designs, the UI, and new functionality. Items 2 and 4 show that some of the improvements made to the prototype after Experiment 1 were well received, and some of the items like 3 and 5 indicated that the prototype lacked functionality.

Table 5.10: Subset of translated feedback from semi-structured interview for Experiment 2.

Id	Feedback
1	The new bar graph for the arguments are confusing to look at, and it takes longer to extract the relevant information.
2	I did not miss the user profile as the less to confuse the better.
3	Being able to filter assistive aids in the table would be good.
4	I love the blue-grey color of the menu because it highlights without being intrusive.
5	I would like to have an overview of the citizens registered falls like in Cura.

Objectivity: Probability and Arguments

During the semi-structured interview, some of the feedback was related to the probability and arguments, regarding the increasing of the case workers' objectivity. This is shown in table 5.11, and has been translated from the danish version found in table A.34. The feedback are mainly comments regarding either the arguments or the trustworthiness of the probability.

Table 5.11: Subset of translated feedback from the semi-structured interview related to the provision of objective support for Experiment 2.

Id	Feedback
1	The arguments should be short and precise.
2	It is annoying if the arguments take up more than one line.
3	I will be able to use the probability and the arguments in my decision-making process as long as they are trustworthy.
4	There is still a lack of context as to why the arguments are as they are.
5	There must be more arguments having an influence than the ones currently shown.

5.4.5 Discussion

SEQ

Comparing the SEQ scores between Experiment 1 and 2, as seen on figure 5.13, shows an overall improvement, except for the worsened score in the Fall Prevention task. The worsening is reflected in the performance metrics as the participants, on average, spent more time completing the task and made additional errors compared to Experiment 1. A possible explanation for this worsening could be the reformulation of the task description for Fall Prevention as seen in table 5.9. Another explanation could be the time that had passed since the participants last interacted with the prototype, and therefore had forgotten how to accomplish the task. The worsening could also be explained by the participant's need to use the navigation drawer to complete the Fall Prevention task. However, since one of the changes as seen in table 5.8 to the navigation drawer has been to add more color, it would presumably have a positive effect on the Fall Prevention task.

Looking at the Hide Plot task in figure 5.13, an improvement is seen compared to Experiment 1 in figure 5.4. This improvement shows that the participants found the task easier to perform, and the narrower confidence interval shows they agreed on the score to a larger extent. The improvement is also reflected in the number of participants completing the task without assistance, which increased from 40% to 80%.

CSUQ

While the SEQ scores indicate that the participants find the tasks easier, and thereby greater usability in the prototype, the scores from the CSUQ contradict this as seen in figure 5.14. All CSUQ scores have either stayed the same or worsened, which could be due to the participants being more critical as they from Experiment 1 expect how the prototype works. Assuming that the participants are more critical, the only slightly worsened scores still indicate high usability.

The relatively large worsening of the IntQual score cannot be explained, as it would be expected to be near-identical to the equivalent score from Experiment 1. Because there have been no changes to the external interface for Experiment 2, it is reasonable to assume that the worsening comes from the participant's understanding of the word "interface".

SUS & UMUX

The results from the SUS and UMUX regarding the argument-design show a preference for design 1 (5.11) compared to design 2 (5.12) as seen in figure 5.15. The participants preferring design 1 is also reflected in the confidence interval, as it is narrower, meaning a less variance across the participants scores. During the semi-structured interview, the participants explained as to why they preferred design 1. They meant that design 2 was too intrusive, and too much information needed to be interpreted (see table A.34).

Performance: Task success

Comparing the participants task success rate between Experiment 1 and 2 shows either an identical or a worse completion rate in all tasks except for the Arguments For task as seen in figure 5.17. The most notable decrease in completion rate is the Dark Mode task, for which the explanation is unclear. Studying the participant's recordings of the task shows a correct identification from where completing the task can be completed by activating dark mode, but a refusal to activate it, resulting in a failed task. The way the participant performs this task makes it look like they know what the result is, and therefore do not complete it. This is pure conjecture as the participants did not say anything during the task or the semi-structured interview.

Looking at the failures in figure 5.17 it is seen that 20% of the participant failed the tasks Arguments Against, Hide Plot, Training Plans, Dark Mode, Fall Prevention, and Log Out. The recording of the participant's usability test shows that during the tasks Arguments Against, Training Plans, Fall Prevention, and Log Out the participant just looked at the prototype for a few seconds then returned to the SEQ questionnaire. This interaction pattern also affected the other performance results which is especially visible in the efficiency, seen on figure 5.23, and total time on task, seen on figure 5.20. An effect of this particular usability test is that it is possible to question if the results should be included. For this thesis, the results have not been excluded, as it is possible to explain the source of the failures and low efficiencies. Furthermore, reducing the participant pool from five to four makes it more difficult to perform statistical analysis.

While there have been many failed tasks for Experiment 2, none of the participants asked for any help, which indicates that the participants have learned how to operate the prototype. Assuming this is true, it would also explain why the performance metrics, in general, have improved.

Performance: Task time

In general, when comparing Experiment 1 and 2 as seen on figures 5.19 and 5.20, the participants have gotten faster. This is seen in all tasks for both task completion time and total time on task, except for the two tasks Arguments For and Fall Prevention. The reason for the participants being slower is unclear, but a possible explanation could be found in the tasks description change, as seen in table 5.9. While the rewording of the Arguments For task caused the participants to be slower, the same rewording of the Arguments Against task caused them to be faster. This discrepancy indicates that there is another reason.

A rewording for the Fall Prevention task was also made as seen in table 5.9, which might have had a significantly larger impact, when comparing the task completion time results for Experiment 1 and 2 shown on figure 5.19. The rewording of this task might have required additional mental effort by the participants to figure out what the task entailed. Furthermore, the difference between the task completion time and the total time on task for the Fall Prevention task, is only $0,9$ seconds. This difference is due to one of the participants who failed the task, only spent 3 seconds in total, which is included in the calculation of the average total time on task, thereby lowering the result.

Looking at the figure 5.18, it might look like the Log Out task suffers from the same issue as the Fall Prevention task, with a small difference between the task completion time and total time on task. This is not deemed to be the case, as the Log Out task redirects the participant away from the main part of the prototype. The redirect would cause the participants to quickly realize the usability test is over, thereby decreasing the time spend looking at the prototype.

Performance: Errors

Comparing the number of errors made during Experiment 1 and 2, the results as shown in figure 5.21 indicate the participants have learned how to interact with the prototype without the need to explore. Looking at figure 5.21, which shows the errors made in Experiment 2 they are, like Experiment 1, concentrated around the Hide Plot task. Comparing the tasks errors from Experiment 2 to Experiment 1, the participants made five errors less.

Studying the errors in the Hide Plot task on figure 5.21, they mainly stem from two error types, the Plot error and Tabs click error. The Plot error is to be expected as some of the participants would inadvertently continue to explore the prototype after the task was completed. The concerning error is the Tabs click error, which indicates that the participants were unable to find the plots, despite already having the plots shown due to the execution order of the tasks.

The Tabs click error is also found in the Fall Prevention task, where it is the main contributor. As with the Hide Plot task, it indicates that the participants were unable to find the correct sequence of actions to complete the task. This could also explain the increase in the task completion time and total time and efficiency.

Performance: Efficiency

The efficiency results for Experiment 2, compared to Experiment 1 seen on figure 5.23, have improved for the three tasks Log In, Arguments For and Hide Plot. For the five tasks Arguments Against, Training Plans, Assistive Aids, Dark Mode, and Log Out the mean task efficiency is lower than the expected efficiency. This is because at least 20% of the participants failed the tasks, which causes a lower efficiency. When this is combined with the causation effect, mentioned in section 3.7.1, from the errors, meaning the mean task efficiency falls below the expected efficiency.

A lower mean task efficiency compared to the expected efficiency can be argued to be worse than vice versa, indicating that the participants failed to complete the tasks. In contrast, a higher efficiency can indicate the participants explore the prototype or make errors, which over time can be minimized.

Performance: Learnability

The learnability score is an aggregated score based on total time on task, the total number of errors, and the mean task efficiency.

Seen on figure 5.24, is the learnability score, showing participants have gotten better at remembering how to use the prototype, even after not having used it since Experiment 1. This is also seen across all performance metrics as the average participant uses less time and makes fewer errors (see figures 5.19, 5.20 and 5.22). Since the learnability also contains efficiency, it should be noted that the learnability value for Experiment 2 therefore is lower than what it should be. This is because the participants spent less than the minimum required clicks and keyboard strokes to complete a task, which leads to failure.

Usability issues

The participants mentioned no usability issues during Experiment 2. While this can be a good thing, as it can mean that the participants are pleased with the prototype, it can also be detrimental to the usability. This is because they do not verbalize potential problems, thereby enduring, which can lead to them being angry later in time.

Semi-structured interview

The feedback from Experiment 2 as seen in table A.34 was a mixture of constructive criticism, and the participants were overall very pleased with the prototype. The type of feedback during Experiment 2 did also change compared to Experiment 1, as the feedback was more focused on optimizing the existing utility and the effect of the implemented changes.

Item 1, in table 5.10 was related to the prototype design and the reason why design 2 was bad. The participants stated that the design was more confusing, and it took longer to comprehend the information. Using this knowledge, in collaboration with the SUS and UMUX scores shown on figure 5.15, proves that design 2 was objectively worse, and will not be included in any further experiments.

Item 2 and 4 were positive feedback points related to the applied changes on the navigation drawer. The first item states that the participants did not miss the user profile, as it meant less confusion and distraction, which was emphasized as none of the participants discovered its exclusion. The second item states that the participants loved the added color to the navigation drawer, and was done due a mentioned usability issue during Experiment 1.

Item 3 and 5 were suggestions for additional functionality. Item 3 was a data filtering mechanism, enabling the case workers to sort a citizen's assistive aids as they saw fit. This suggestion was presented as the participant's definition of a useful assistive aid to base their assessment differed. Some participants wanted specific types of assistive aids to be included in the table, while others wanted the same types to be excluded. Item 5 was a list containing a citizen's registered falls, which all participants suggested.

Objectivity: Probability and Arguments

Experiment 2 had compared to Experiment 1, additional focus on increasing the objectivity of the participants. This was done by testing the two argument-designs and using arguments generated on the current understanding of the project. The feedback gathered from the semi-structured interview regarding the objectivity was related to the arguments, which is also reflected in the subset seen in table 5.11. The feedback was mostly constructive criticism, with a slightly negative tone, and questioned the usefulness of the shown arguments.

Studying the feedback shows that the first two feedback points can be seen as guidelines for constructing the arguments, as they state the arguments should be short, precise, and preferably a one-liner. This feedback is also emphasized in the study by K. Miller [32].

Items 3, 4, and 5 show the participants distrust in the probability and arguments, as they state that they would like to use it as part of their assessments, but require additional context and possibly more arguments. This possible distrust can be solved through time, as the project model gets more advanced. Adding more arguments shows the participants requirements for additional context, but also the constructive criticism. While presenting the arguments is easy from a technical standpoint, the problem lies in generating additional arguments, as more research has to be done.

5.4.6 Summary

The SEQ results show an overall improvement as the participants found the tasks easier compared to Experiment 1. This is also corroborated by the fact that the participants generally spent less time on each task and made fewer errors. The improvement is most likely due to the participants having experienced the prototype and the tasks during Experiment 1.

Despite the participants being faster and less error-prone, the efficiency shows that they performed worse as they failed to complete some tasks. While the improvements do not conclusively say that the prototype is as good as it gets, it shows that the participants can familiarize themselves with the prototype to a degree where it is usable with only a small amount of hands-on time.

The prototype's good usability is also partly reflected by the CSUQ results despite some of the results being marginally worse for Experiment 2 compared to Experiment 1. This worsening is likely to come from a statement reminding the participants what a perfect score is.

As there were multiple designs for the participants to score, the result from this is evident, with the best design being design 1. This design had the simplicity the participants wanted, without any irrelevant information.

The objectivity can be increased as the feedback can help define how and which arguments should be presented. While the feedback can help future development, the current state of the objective support shows distrust to some of the participants.

5.5 Experiment 3

5.5.1 Purpose

The third experiment was much like the second, as it used the same methods to allow for a comparison to detect a potentially optimized user experience. The experiment also conducted a user acceptance test to find any missing functionality to verify the accomplishment of the user requirements.

5.5.2 Experiment Design

Experiment 3 was conducted in person on the 29th of April 2021 with four participants, as one was on vacation. An important thing to note is that one of the remaining four participants was replaced due to sickness, and the substitute had no previous experience with the prototype. Since the experiment was conducted in person, as seen in figure 5.25, the participants were provided with a 24" screen, mouse, and keyboard to resembled their daily workstation.

The experiment was divided into two parts, which consisted of the methods

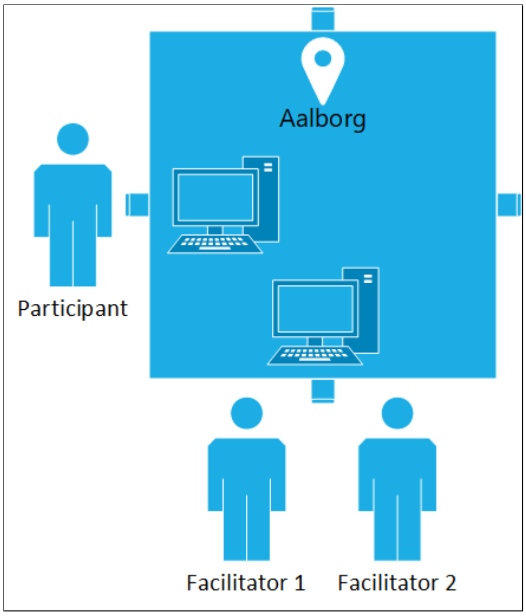

Figure 5.25: Experiment Setup for Experiment 3.

shown in table 5.12. Part one consisted of four individual usability tests to measure the user experience and interviews for more in-depth feedback with a total duration of 45 minutes for each participant. The second part consisted of a 30 minutes user acceptance test.

Table 5.12: Methods applied for Experiment 3.

Method	Types
Data Recording	Notes, Audio, Video
Interview	Semi-Structured
Usability Test	Performance, Usability Issues, SEQ, CSUQ, SUS, UMUX
Acceptance Test	User Acceptance Test

Prototype design

Shown in figures 5.26 and 5.27 is iteration 5 of the prototype design, for respectively case 1 and case 2. Only a few significant changes has been made between Experiment 2 and 3 as shown in table 5.13, with the complete changelog located in section A.4.1.

Table 5.13: Significant changes between Experiment 2 and 3.

Id	Change description	Reason
1	Added a tab for information regarding a citizens registered falls for case 2 (see figure 5.27).	The addition was requested during the semi structured interview (see number 82 in table A.16 and 58, 59, and 60 in table A.34).
2	Added SHAP values, provided through the ML API.	To place each argument correctly, according to the SHAP value as mentioned in section 3.5.
3	Removed the Color Blind Mode from the Navigation Drawer.	Research showed (see section 3.4.4) that the color scheme chosen in the prototype already supported people suffering from red-green color blindness.

Case 1 - Rehabilitation Training

The probability itself was unchanged compared to Experiment 2, but the number of arguments was reduced to four, to be able to reflect one of the four available SHAP values (see section 3.5 for more details). The navigation drawer color was changed and the icons made white for greater contrast.

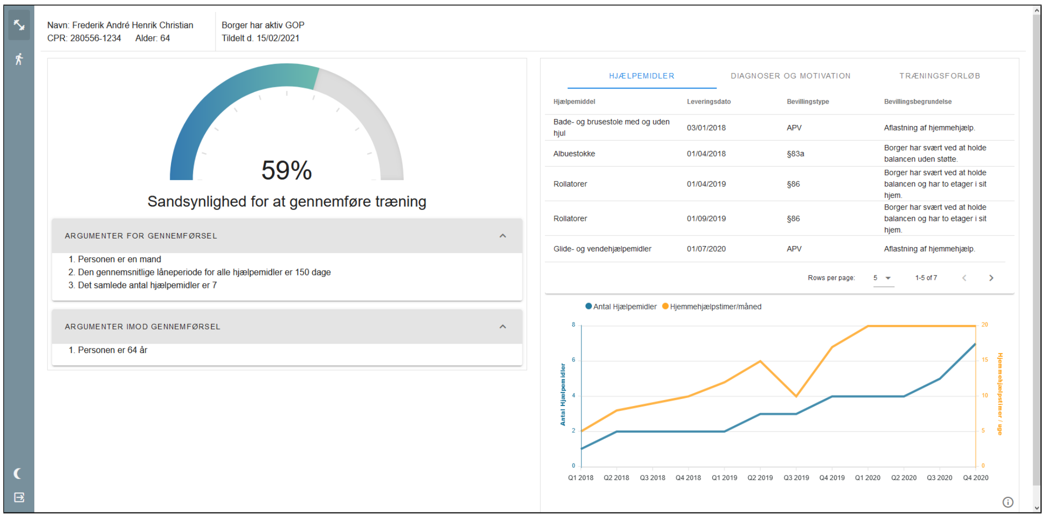

Figure 5.26: Prototype Design - Case 1 for Experiment 3.

Case 2 - Fall Preventive Training
Just like case 1, the probability was unchanged compared to Experiment 2, but the number of arguments was reduced to four, to be able to reflect one of the four available SHAP values (see section 3.5 for more details). Figure 5.27 shows the *REGISTRERET FALD* tab.

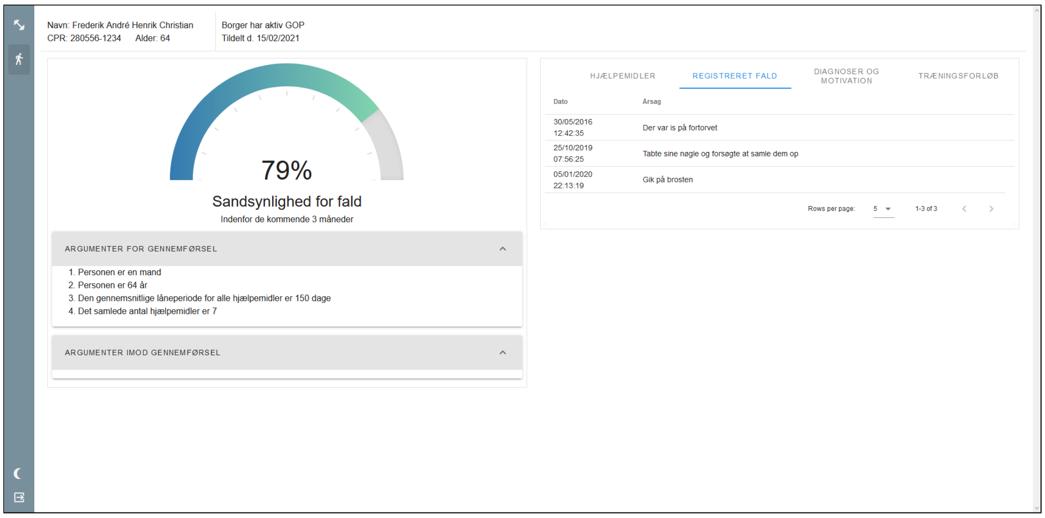

Figure 5.27: Prototype Design - Case 2 for Experiment 3.

5.5.3 Experiment Procedure

Before the experiment began, an introductory meeting was held with the participants to describe the changes and additions made since Experiment 2. The meeting also included an agenda explaining how the experiment would be conducted.

The experiment itself consisted of an in-person usability test, with 11 tasks to complete shown in table 5.14, in the order of execution. Due to some of the changes shown in table 5.13, one task was added and another removed. Furthermore, some of the task descriptions were changed to increase the task's understandability.

Table 5.14: English version of the SEQ used during Experiment 3. (Danish Version is found in table A.36).
Status: A=Added, M=Modified, R=Removed, U=Unmodified

Task #	Identifier	Task description	Status
1	Log In	Log on the website. Use the already entered username, and enter a random password.	U
-	Colorblind	Activate colorblind mode on the website.	R
2	Arguments For	Find an argument that states the citizen is able to complete training.	U
3	Arguments Against	Find an argument that states the citizen is unable to complete training.	U
4	Hide Plot	Hide one plot in the graph that shows assistive aids and home help hours over time.	M
5	Training Plans	Find the citizens training plans.	U
6	Assistive Aids	Find the citizens assistive aids.	U
7	Diagnoses	Find the citizens diagnoses and motivation.	U
8	Dark Mode	Enable dark mode on the website.	U
9	Fall Prevention	Find the probability showing a citizens chance of falling during the next three months.	M
10	Registered Falls	Find the citizens registered falls.	A
11	Log Out	Log out of the website.	U

The SEQ questionnaire followed each task to identify the perceived task difficulty, while both performance and usability issues were gathered. After all tasks were completed, the participants answered the CSUQ, SUS, and UMUX, followed by the semi-structured interview. Once all individual tests were completed, two participants stayed to complete the user acceptance test (UAT). The whole experiment session was recorded through audio and screen-capture, with note-taking during the semi-structured interview.

5.5.4 Results

SEQ

Figure 5.28 shows the mean Single Ease Question (SEQ) results for Experiment 1, 2, and 3. The results from Experiment 3 shows that the participants perceived the tasks as easier or equally as easy as during Experiment 2, with the exception being Arguments For and Fall Prevention.

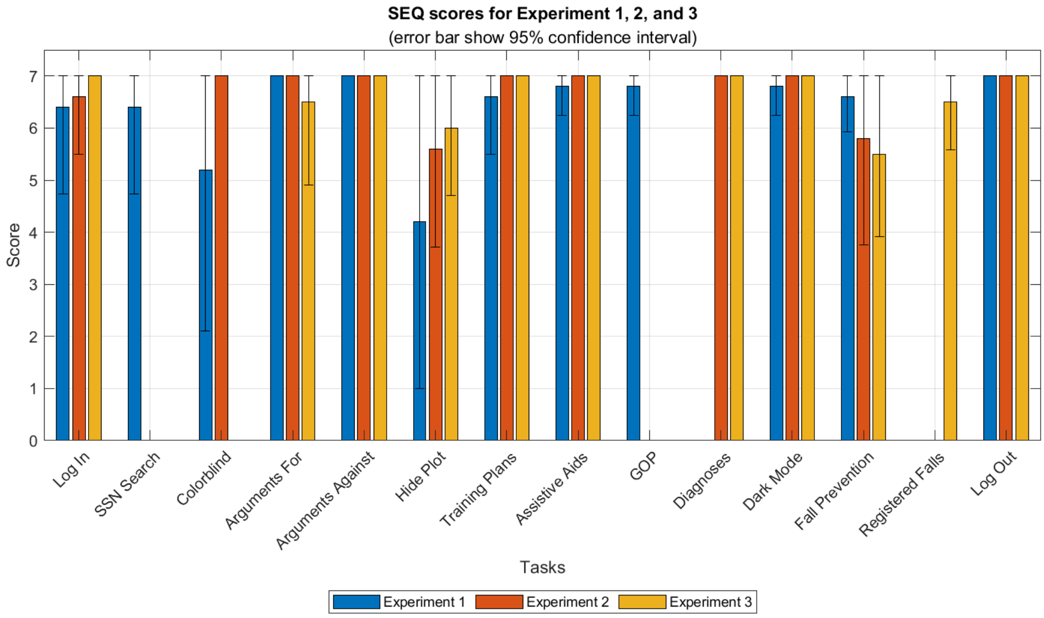

Figure 5.28: Mean SEQ scores for the tasks performed for Experiment 1, 2, and 3.

The Fall Prevention task also has the worst SEQ score of all the tasks for Experiment 3, which for Experiment 1 and 2 was the Hide Plot task. Furthermore, the SEQ score for the Fall Prevention task has decreased across all experiments, while the Hide Plot task has increased.

CSUQ

The CSUQ scores seen on Figure 5.29 shows the mean Computer System Usability Questionnaire (CSUQ) scores for Experiment 1, 2, and 3, with 1 and 7 being the best and worst scores, respectively.

Since Experiment 2, all CSUQ scores have worsened, with the largest found

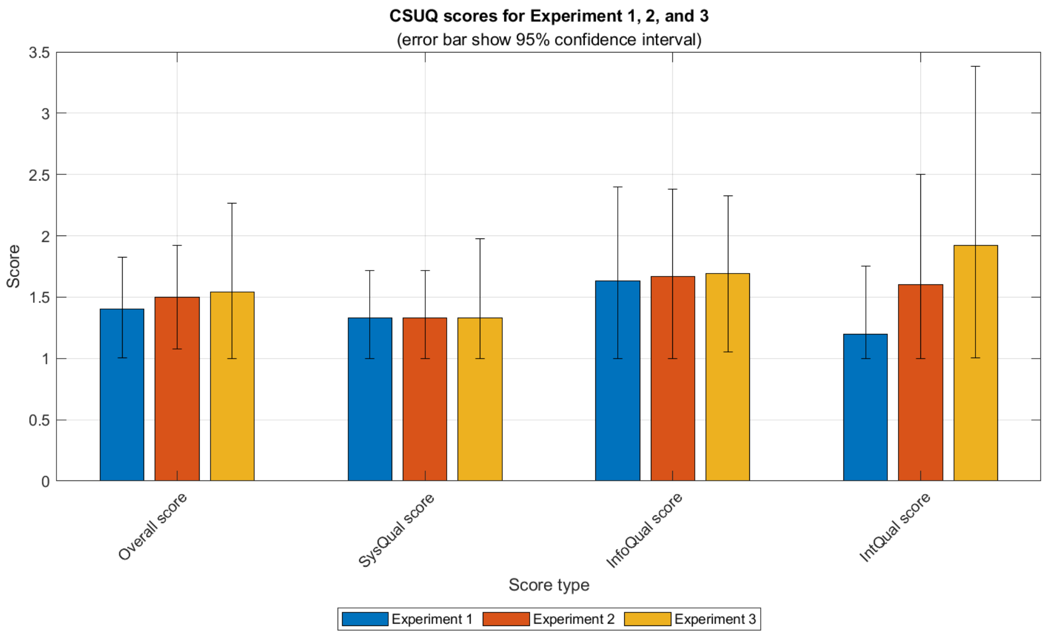

Figure 5.29: Mean CSUQ scores for Experiment 1, 2, and 3.

in the IntQual score. The three other scores have all seen a relatively minor worsening, compared to the IntQual score, except for the SysQual which is identical to Experiment 1 and 2. The confidence interval of the Overall, SysQual, and IntQual scores have also expanded, signifying less agreement between the participants.

SUS & UMUX

Figure 5.30 shows the mean System Usability Score (SUS) and the Usability Metric for User Experience (UMUX). A confidence interval is not shown for the UMUX as all participants submitted the same score. Comparing the SUS and UMUX results shows a similarity, as they received 93,1% and 95,8%, respectively.

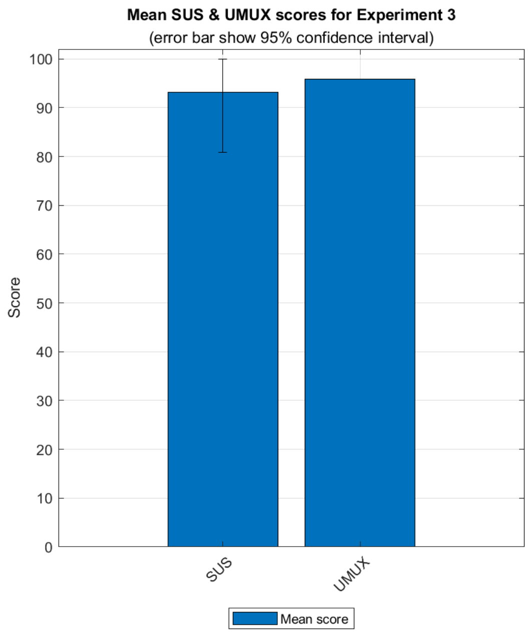

Figure 5.30: Mean SUS and UMUX score for Experiment 3.

Performance: Task success

In figure 5.31 the levels of success for Experiment 3 are shown. The graph shows that 25% failed to complete three tasks, and one needed assistance to complete the Fall Prevention task.

Figure 5.32 shows the completed number of tasks for Experiment 1, 2, and 3. Focusing on the results for Experiment 3 shows that the participants performed as well or better compared to Experiment 2, with the exception of tasks Arguments Against and Diagnoses.

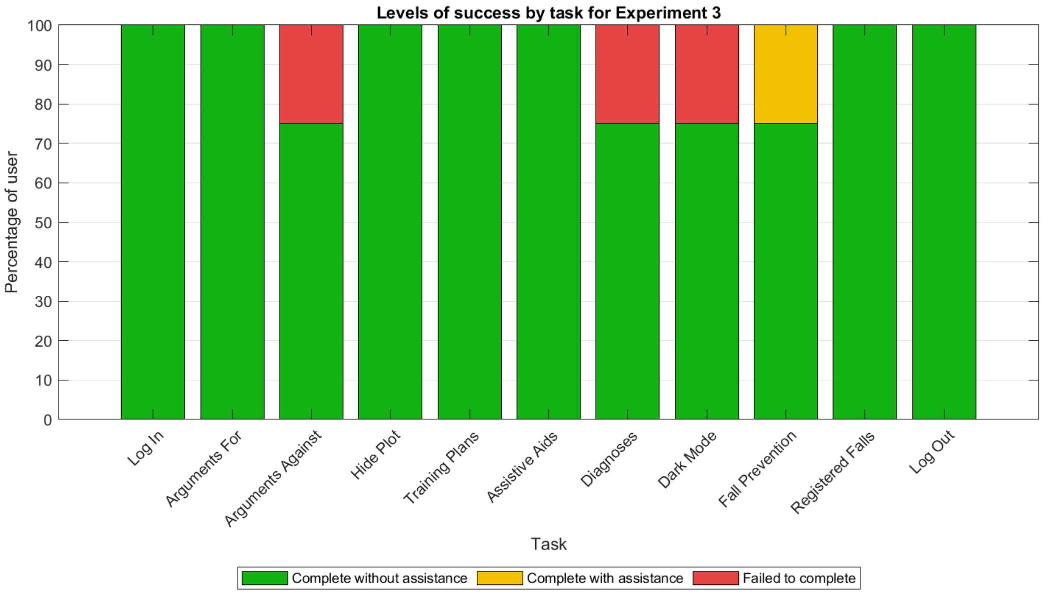

Figure 5.31: Percentage of participants for each task that could complete the tasks without assistance, with assistance, and failed to complete the task for Experiment 3.

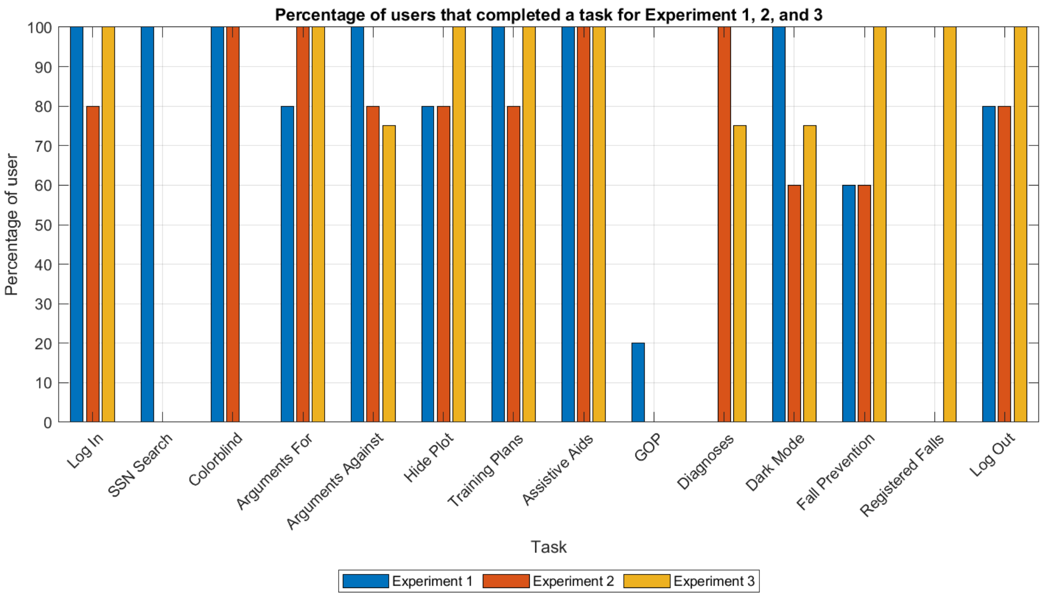

Figure 5.32: Percentage of participants that completed each task for Experiment 1, 2, and 3.

Performance: Task time

The participant's mean task completion time and mean total time on task, for each executed task for Experiment 3, is shown on figure 5.33. Looking at the graph, it shows that the two tasks Hide Plot and Fall Prevention have the largest amount of time spent before reaching the completion criteria. While the mean task completion time for these two tasks is high, the difference between it and the mean total time on task is comparable to some of the other tasks.

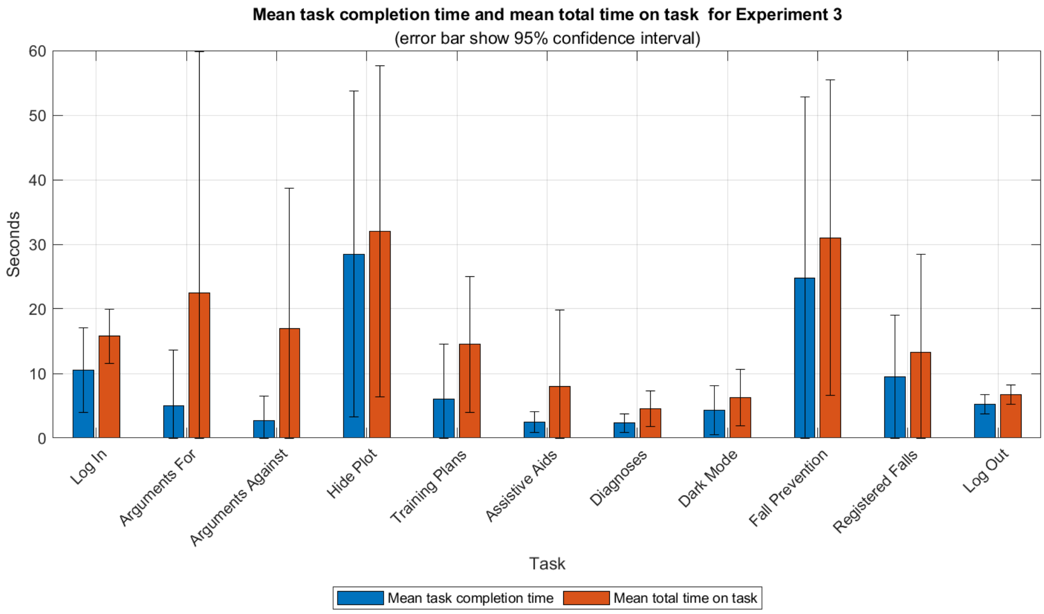

Figure 5.33: Mean task completion time and mean total task time on task for Experiment 3.

In figure 5.34 the mean task completion times for Experiment 1, 2, and 3 is seen. Comparing Experiment 3 to Experiment 2 shows that the participants performed better during five out of the ten shared tasks, and worse during the other five. A comparison between Experiment 3 and Experiment 1 shows that the participants performed better in seven out of the eight shared tasks, but worse in the Fall Prevention task.

Comparing the total time on task times, seen in figure 5.35, for Experiment 2 and 3, the participants spent for Experiment 3 more time in eight out of ten tasks.

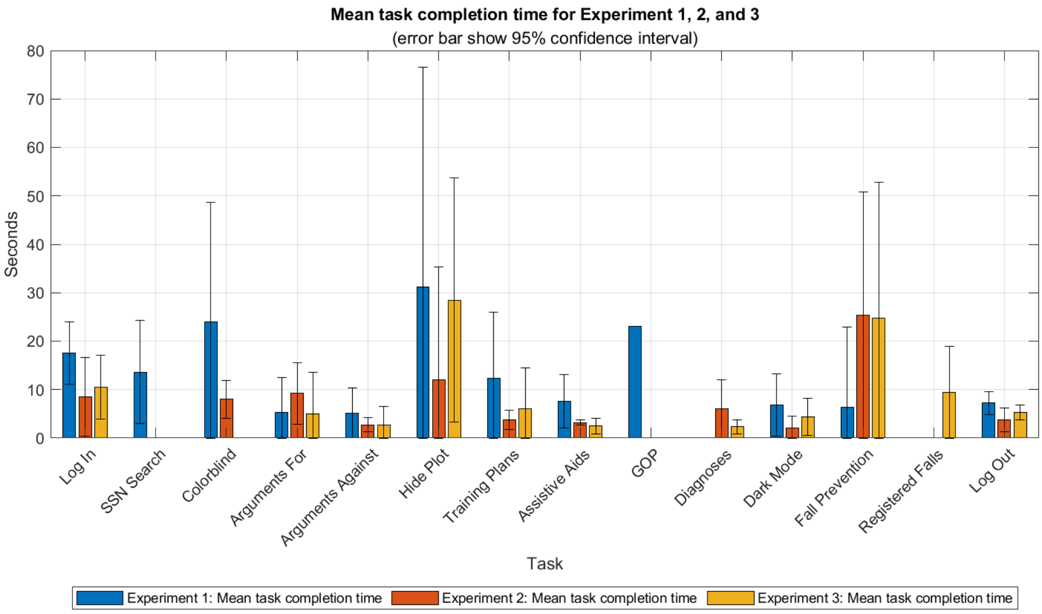

Figure 5.34: Mean task completion time for Experiment 1, 2, and 3.

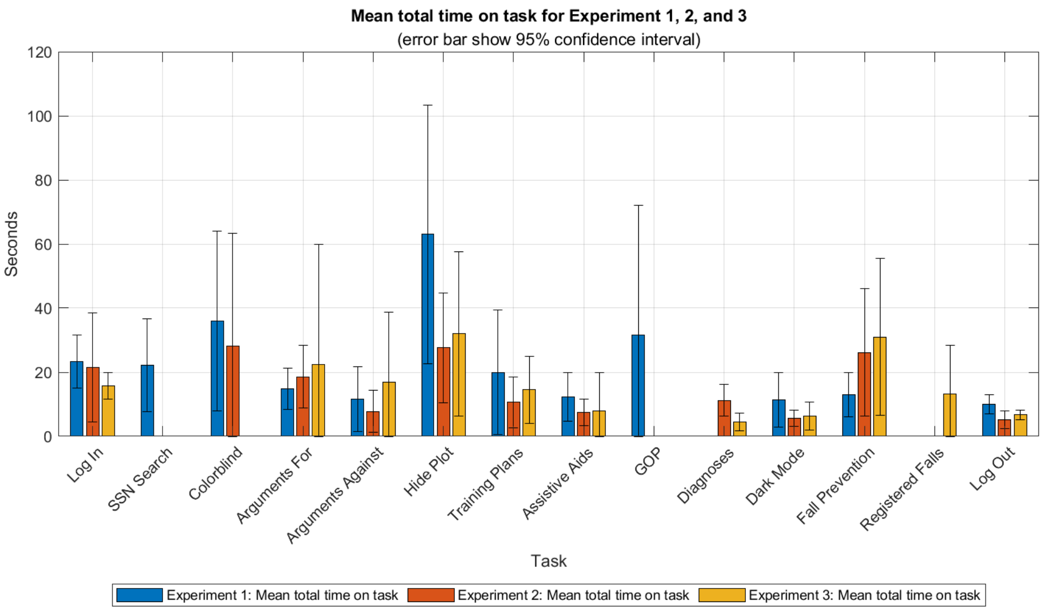

Figure 5.35: Mean total time on task for Experiment 1, 2, and 3.

Performance: Errors

Looking at figure 5.36 it is worth noting that the stacked bars each represent their part of the total number of errors. E.g., task Arguments Against shows a total of three errors, with one error being *Random click error* and two errors being *Tabs click error*. Additional information regarding the error types can be found in section A.1.

The graph shows that all tasks with errors have more than two error types, except for the tasks Training Plans and Dark Mode, having respectively three and two errors of the same type. Comparing the errors made between Experiment 2 and 3, the errors seem to have normalized. This means that a specific task does not contain a comparatively large number of errors, but these have instead been spread out among the tasks.

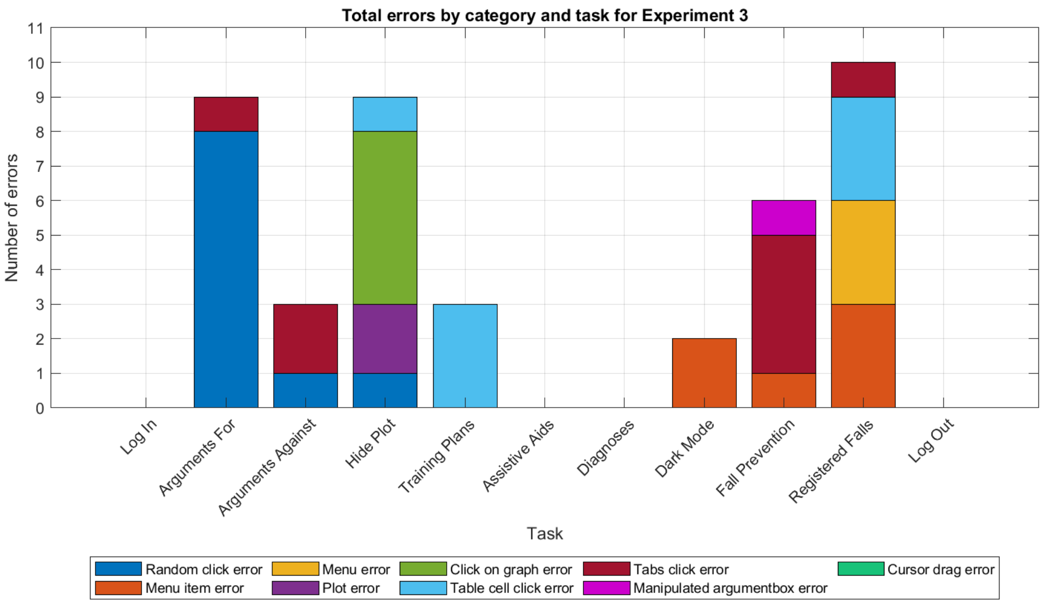

Figure 5.36: Categorised number of errors and total number of errors for each task for Experiment 3.

In figure 5.37 the total number of errors made during Experiment 1, 2, and 3 is seen, with Experiment 3 having more errors made compared to Experiment 2. The graph shows that during Experiment 3, 42 errors were made, distributed across the four participants. During Experiment 1 and 2 the five participants collectively made 58 and 31 errors.

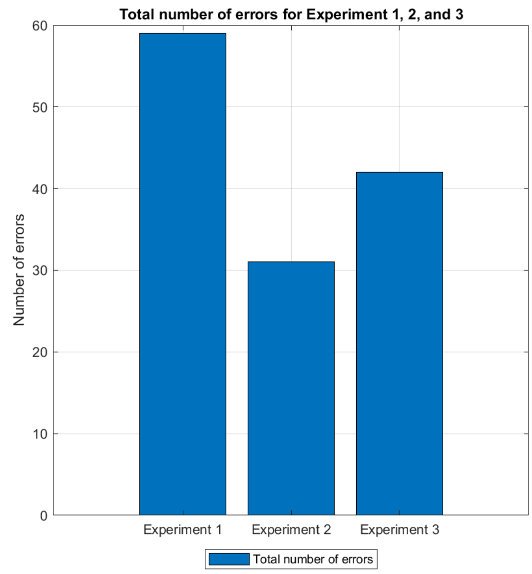

Figure 5.37: The total amount of errors for Experiment 1, 2, and 3.

Performance: Efficiency

Figure 5.38 shows the mean and expected efficiency results across Experiment 1, 2, and 3. The graph shows that for the tasks Arguments For, Hide Plot, and Registered Falls in Experiment 3, the participants has more than tripled the mean efficiency compared to the expected efficiency. Excluding these tasks, along with the Fall Prevention, the remaining seven tasks are close or identical to the expected efficiency. The Diagnoses task has a mean task efficiency below the expected efficiency, indicating that not all participants completed the task.

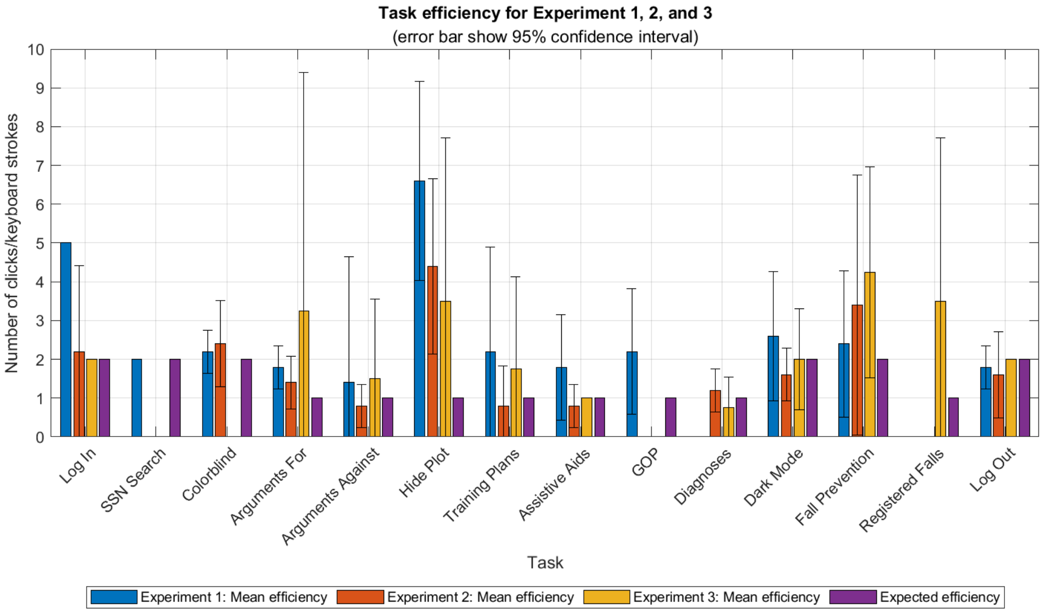

Figure 5.38: Mean efficiency scores for the different tasks for Experiment 1, 2, and 3.

Performance: Learnability

Figure 5.39 shows the aggregate learnability of the participant for Experiment 1, 2, and 3. The learnability score is an aggregated score from total time on task, the total number of errors, and the mean task efficiency score, where a lower score is better. The learnability score for Experiment 3 is worse compared to Experiment 2.

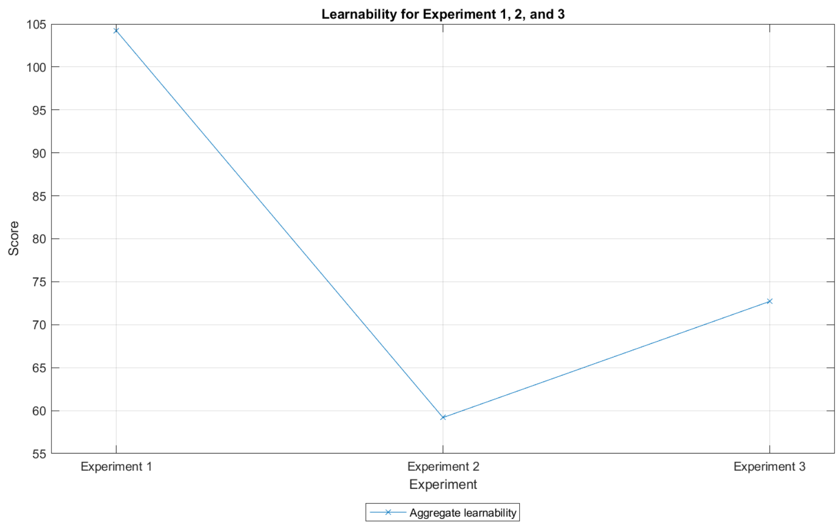

Figure 5.39: Aggregated learnability score for Experiment 1, 2, and 3.

Usability issues

During the usability test, the participants did not verbalize any usability issues. While this does not mean that no usability issues exist, it only means that the participants did not verbalize any issues during the test.

Semi-structured interview

As Experiment 3 was performed in a controlled environment, it was possible to obtain feedback, which was mostly positive. Table 5.15 shows a translated subset of all the feedback during the semi-structured interview in Experiment 3. Looking at the feedback indicates a positive experience with the prototype.

Table 5.15: Subset of translated feedback from semi-structured interview for Experiment 3.

Id	Feedback
1	The font size is perfect.
2	The prototype is very user friendly.
3	It was easy to remember how the prototype worked.
4	If the prototype was available today, i would use it frequently.
5	I thought i had to enter a lot of information, but i was a pleasantly surprised.

Objectivity: Probability and Arguments

During the semi-structured interview, some of the feedback was related to the probability and arguments, regarding the increasing of the case workers' objectivity. This is shown in table 5.16, and has been translated from the danish version found in table A.50. The feedback was both positive and negative, and expressed by the participants, stating that some of the arguments was more useful compared to others. Furthermore they wanted to use the probability and arguments as part of their work.

Table 5.16: Subset of translated feedback from the semi-structured interview related to the provision of objective support for Experiment 3.

Id	Feedback
1	The arguments regarding age and gender seems irrelevant.
2	The arguments regarding the loan period and number of assistive aids are useful.
3	I would like additional arguments.
4	Why is an argument an argument? I miss an explanation.
5	I feel supported in my decision-making process, because the probability provides super guidance.

Acceptance Test

Table 5.17 shows an overview of the tasks for the acceptance test, each marked as either completed (✓) or failed (✗). More details regarding the steps involved throughout each task is presented in chapter B.

Table 5.17: Executed task scenarios during the acceptance test.

Id	Task description	Pass/Fail
1	Log into the prototype.	✓
2	See the citizens probability for completing rehabilitation training.	✓
3	See both the positive and negative arguments for the citizens probability.	✓
4	See the citizens information.	✓
5	See the citizens diagnoses and motivations.	✓
6	See the citizens training plans.	✓
7	See the citizens assistive aids.	✓
8	See the citizens probability for falling within the following three months.	✓
9	See the citizens registered falls.	✓
10	Change the contrast polarity from positive to negative.	✓
11	Log out of the prototype.	✓

5.5.5 Discussion

SEQ

The trends of improvement and worsening seen in the SEQ scores from Experiment 2 has continued for all tasks for Experiment 3, except for the Arguments For task as seen in figure 5.28. These trend continuations do not reveal that a new participant has joined the experiment, demonstrating the prototype's usability and the tasks understandability.

Studying the decline in the Fall Prevention task, it is seen that the only factor in the task has been a rewording of the task description (see table 5.14), as the functionality has not changed. This means that the decline can be attributed to the rewording and understandability of the task. While the worsening is expected, it is also unexpected that the participants do not know how to complete the task.

An improvement trend in the perceived difficulty is seen for the Hide Plot task, over the course of Experiment 1, 2, and 3. The trend of the increasing SEQ score, combined with a new experiment participant shows that the task exhibits high understandability. Furthermore, the task also has a low number of errors compared to Experiment 1 and 2, indicating that the participants have retained knowledge of how to complete the task, or have no need to explore the prototype.

CSUQ

The results from the CSUQ show the opposite trend of the SEQ results, with the results continuing to worsen as seen in figure 5.29. An explanation for this is unclear as the expected result for the CSUQ would improve as the participants have found the tasks easier to complete as seen in figure 5.28. The most likely explanation is that the participants answer more conservatively to leave room for improvements.

While the degradation continues across all three experiments, it is most significant for the IntQual score, and is most likely explained by the new participant. Despite the worsening, it does not necessarily mean that the participants find the prototype less usable as an Overall score of $1,5$ between 1 and 7 is excellent.

SUS & UMUX

Like the CSUQ, both the SUS and UMUX seen in figure 5.30, state that the prototype has an excellent usability, with a score above 90 [65, p. 204]. Comparing the SUS to the UMUX shows that the SUS scores a little lower than the UMUX. This was also the case when these were used during Experiment 2 to find the best argument-design. It is uncertain if this trend would continue, but by comparing all the SUS and UMUX scores, it could be assumed that the UMUX would always yield a better score compared to the SUS with the pilot group.

A reason for this could be the structure of the questionnaires, as the participants tend to submit near-perfect scores, and the scales for the SUS and UMUX are between 1 and 5 and 1 and 7, respectively. This means that choosing the second-best score, for all statements, in both questionnaires would make the UMUX have a larger score compared to SUS.

Performance: Task success

As shown in figure 5.31 the three tasks Arguments Against, Diagnoses, and Dark Mode were failed and the Fall Prevention task required assistance to be completed, which was caused by a single participant. Studying the recordings of the usability test shows the reasons for the failed Arguments Against and Diagnoses tasks, being that the participant thought the tasks were completed earlier, causing them to just score the task. Even though the participant did not perform the tasks, it is seen in figure 5.28 that the SEQ scores for both tasks received a perfect score, meaning that the participant also scored it as such. The participant can only do this confidently, as they have been part of all previous experiments.

For the Dark Mode task, the participant indicated that they knew how to complete the task, but did not do it, as the results were not to their liking. This confirms the suspicion from Experiment 2 of the participant knowing how to complete the task. It is important to note that this is only true for one of the participants failing the task during Experiment 2 as seen in figure 5.32.

Comparing the task success rate between Experiment 2 and 3 as seen in figure 5.32, it is important to remember the difference in the number of participants, meaning that even though the result might seem worse, they might not necessarily be. This worsening is shown in the Arguments Against task, where only one participant fails the task for both Experiment 2 and 3. Excluding this task, overall improvement of the success rate is seen across all tasks, despite the experiment containing a new participant.

Performance: Task time

For Experiment 3 it is important to keep in mind that a new participant was part of the experiment. This is seen in the Arguments For task's confidence interval for the total time on task in figure 5.33, as this is the largest of all the tasks. Furthermore, the participant asked about the tasks' complexity and was pleasantly surprised, resulting in faster total time on task times for the following tasks.

Comparing Experiment 3 to Experiment 2 shows that the average participant completes half of the tasks faster, and the other half slower as seen in figure 5.34. The Hide Plot task has the most significant increase in the mean task completion time. Looking at the recordings shows that the new participant had a negative impact on this task's measurement. Additionally, the recordings also revealed that another participant had forgotten how to complete the task, further exacerbating the results.

For the Fall Prevention task, it is seen in figure 5.34 that the participants on average spent roughly the same amount of time on completing the task during both Experiment 2 and 3. Since the way of completing the task has not changed, this result is likely due to the new participant and the rewording of the task description as seen in table 5.14. Looking at the total time on task for the Fall Prevention task in figure 5.35, the participants spent the largest amount during Experiment 3.

Furthermore, for the total time on task, seen in Figure 5.35, the participants are slower in eight out of ten tasks when comparing Experiment 2 and 3. This is due to the new participant's need to explore the prototype.

Performance: Errors

For Experiment 3 the four participants made 11 additional errors compared to the five participants during Experiment 2 as seen in figure 5.37. This means that if the original number of five participants were attending the experiment, the total number of errors would likely have increased. Accounting for the new participant, who single-handedly made all the Random click error for the Arguments For task, would improve the overall result, although the total number of errors for Experiment 2 would still be lower.

For Experiment 1 and 2 the largest error generating task was the Hide Plot task, which for Experiment 3 has been replaced by the Registered Falls. The Registered Falls task has been added for Experiment 3, as seen in table 5.14. The errors for the Registered Falls task indicate that the participants might have been influenced by the previous task, Fall Prevention, which requires them to use the navigation drawer. Additionally, the wording of the tasks might have influenced the participants as both relates to a citizen falling. Furthermore, the number of errors produced by the Registered Falls task is reflected in the task completion time and the total time on task for the average participant seen in figure 5.33.

Looking at the errors produced by the Hide Plot task, shows the number of error types have reduced from five in Experiment 2 (see figure 5.21) to four in Experiment 3 (see figure 5.36). This reduction indicates the participants familiarity with the prototype, as there is less need to explore, further showing that the new participant can deduce how to complete the Hide Plot task, without causing additional errors.

Performance: Efficiency

Comparing the efficiency results from Experiment 3 to Experiment 2 some of the results have improved while others have worsened as seen in figure 5.38. One of these improvements is seen for the Hide Plot task, as the efficiency has been reduced thereby being closer to the expected efficiency.

Focusing on the Fall Prevention task, the increased error count seen in figure 5.36 is reflected in the mean task efficiency, which has increased for each experiment. This trend is likely due to the rewording of the task for each experiment. Comparing the results between Experiment 3 and Experiment 1 and 2 shows an overall better mean task efficiency, except for the tasks Arguments For and Fall Prevention as seen in figure 5.38. These two tasks can be explained with the new participant, and participants forgetting how to complete a task as mentioned in section 5.5.5.

Tasks such as Arguments Against, Training Plans, and Log Out seems to have a worse result for Experiment 3 due to a higher mean efficiency as seen in figure 5.38. However, the results are better as the mean efficiency is equal to or higher than the expected efficiency, indicating that the participants completed the tasks.

Performance: Learnability

The learnability score is an aggregated score based on total time on task, the total number of errors, and the mean task efficiency, which all have greater mean values for Experiment 3 than Experiment 2, resulting in a worse learnability score as seen in figure 5.39.

The expected learnability, based on the results from Experiment 1 and 2, would have been better for Experiment 3, than the actual result. It can, however, be argued that the learnability for Experiment 2 was artificially low, as the participants failed numerous tasks, thereby lowering the efficiency, which is a part of the aggregated learnability score. Therefore, a comparison between Experiment 3 and Experiment 2 the artificially low efficiency should be accounted for, along with a new participant, making the learnability score for Experiment 3 higher than it presumably otherwise would have been.

Usability issues

The participants mentioned no usability issues during Experiment 3. While this can be a good thing, as it can mean that the participants are pleased with the prototype, it can also be detrimental to the usability. This is because they do not verbalize potential problems, thereby enduring, leading to them being angry later in time.

Semi-structured interview

The five items as seen in table 5.15 represents the general feedback received during the semi-structured interview for Experiment 3.

Item 1 is related to the font size, which was mentioned as a usability issue during Experiment 1 as seen on table 5.4. During Experiment 3 the participants were in a controlled environment, meaning that they had to perform the tasks on a provided 24” screen and a supplied keyboard and mouse. This environment matches their daily work environment, meaning that the usability issue mentioned during Experiment 1, was not an issue with the prototype but rather an issue with their available equipment at the time of the experiment.

Items 2 and 3 are related to the general user experience, with participants stating that the prototype was user-friendly and easy to remember how to use. The performance metrics do not corroborate the participants perceived usability, as they have gotten worse at completing their tasks, which is reflected by the learnability score seen in figure 5.39. Despite the contradiction, the users also stated, in item 4, that they would use the prototype daily, which would completely alleviate the learnability issue.

Item 5 stems from the new participant, who feared a need to “enter a lot of information”. This fear was mitigated, as no information had to be entered, but the participant only had to consume the information presented.

Objectivity: Probability and Arguments

The feedback from Experiment 3 regarding the objectivity was, like Experiment 2, mostly constructive criticism, with a slightly negative tone. This indicates that the participants still have a problem with the arguments.

Studying items 1, 2, and 3 shows that the participants find two of the arguments irrelevant while being able to use the two other arguments as seen in table 5.16. Furthermore, the participants desired additional arguments, as they did not find the total number of four arguments adequate. This presents further challenges with the currently used way of generating the arguments, described in section 3.5, as SHAP is an interpretability method and not an explainability method. Therefore, using only SHAP makes it impossible to provide the case workers with the context for the arguments they desire, mentioned by item 4.

Despite the participants requiring additional arguments and context surrounding them, they still felt supported in their decision-making, as the probability provides "super guidance" as stated by item 5.

Acceptance Test

As shown in table 5.17, all tasks were completed successfully, without any issues or crashes occurring during the test. The participants in the test had no further comments or annoyances regarding the test itself or the prototypes functionality.

5.5.6 Summary

A new participant was introduced to the experiment while two regular participants were absent, resulting in only four participants. The experiment was conducted in person in contrast to Experiment 1 and 2, which were conducted remotely.

The participants scored the tasks as being easier compared to Experiment 1 and 2, which is likely due to them gaining experience. While the participants scored the tasks as being easier, their performance dispute this as it showed they generally performed worse compared to Experiment 2. The learnability is a good representation of this worsening.

The participants using the CSUQ scored the usability as being very high, although a slight worsening was seen compared to Experiment 1 and 2. The SUS and UMUX corroborate these findings of high usability.

During the semi-structured interview, the participants gave both positive and negative feedback, mostly centered around the general user experience. The case workers objectivity has also been increased as they found the probability very useful in their decision making process. Some of the arguments were more useful to the participants than others, but additional and more adequate arguments were required.

The executed acceptance test with the participants indicated no missing functionality in the prototype, as all tasks were passed.

Chapter 6

Discussion, Conclusion & Future Work

6.1 Discussion

Having used the user-centered approach to develop the prototype has been advantageous, as it has allowed for the inclusion of the case workers in the design process. While this could have been done regardless of the approach, the user-centered approach helped define the structure under which this should be done. Furthermore, using the user-centered approach helped concretize how the prototype should be tested. The empirical evidence gathered further adds value to the usage of the method. The user-centered approach also requires the practitioners to apply the methods two other principles in an iterative process.

By developing multiple iterations of the prototype and performed multiple experiments has helped define the prototype and increase its usefulness by optimizing the user experience. The prototype's usefulness is seen in all the methods used in the usability tests and the semi-structured interviews. Quantifying the usability has been done through the Computer System Usability Questionnaire (CSUQ), System Usability Scale (SUS), and Usability Metric for User Experience (UMUX). While the CSUQ results have worsened over time, the overall score went from $1,4$ for Experiment 1 to $1,5$ for Experiment 3. This worsening can be deemed negligible, meaning that the method still indicates good usability, as the worst possible score is $7,0$.

The CSUQ is a multi-faceted questionnaire, which is reflected in the statements and the possible scores. Because of this, part of the CSUQ is not useful in quantifying the prototype's usability, as some of the statements are not relatable, and therefore unused. Furthermore, the statements could have a detrimental effect on the final score if the participants misunderstood the statements. This potential problem was why the SUS and UMUX questionnaires were included in Experiment 3, as they are not multi-faceted and only focuses on the overall usability. This sole focus means that it could be beneficial to use the SUS and UMUX in future experiments along with the CSUQ to verify the results.

Another way to quantify usability has been done using the SEQ and performance metrics. These show slightly differing results, as the participants find the tasks easier over time, but their performance states that they have gotten worse at some of the tasks. This is valuable information as it can show discrepancies between the user's opinions and their actual performance. It is essential to note that it is possible to look at only one metric and disregard the other. However, this behavior can be detrimental to the perception of usability. While the SEQ is a beneficial addition to this thesis, it can also worsen perceived usability. The Fall Prevention task was rated increasingly difficult by the participants, despite no functional changes was added to the prototype. The increasing difficulty rating indicates that poorly written tasks can worsen the results, signifying possible non-existing problems.

Utility has also been part of the prototypes usefulness and has been evaluated along with the case workers, using User Acceptance Testing (UAT). The UAT ensured that all the user requirements were implemented and functioned according to the users expectations. Having said that, the inclusion of functional testing could have been beneficial, as it could have proved that the functional changes not reflected in the usability test, have been implemented. Such an example could be the usability issue for the GOP to only be shown in the citizen information, when it is active (see number 7 in table 5.4). As the usability test did not account for multiple different citizens, only the case with the active GOP was presented to the participants. The functional testing would here be able to prove that the functionality had in fact been implemented. Alternatively, the UAT could have been extended to account for more than one citizen, thereby providing the participant with a larger variety.

Providing decision support using the project model has made it possible to supply the case workers with a probability for case 1 and case 2 with the addition of SHAP values for interpreting the probability. However, feedback from the semi-structured interviews during Experiment 1, 2, and 3 suggests further research and advancements to the project model is required. This is due to the participants repeatedly requesting additional as well as more adequate arguments to explain the probability. As mentioned in section 2.5, the prototype in this thesis has been categorised as a non-knowledge based CDSS, as it relies on machine learning rather than a compiled list of expert knowledge. Integrating a knowledge base in the project model could provide the evidential knowledge that is required by the case workers to accommodate their needs in terms of more explanation behind the probability. If however a knowledge base is ever to be integrated, it should not be designed to make the users become less self-reliant [13].

6.2 Conclusion

Throughout this thesis, research has been done, and experiments have been conducted to develop a prototype to provide decision support for case workers in Aalborg Municipality, Denmark. To achieve this, four goals have been established:

1. Develop a useful prototype that displays information from both the machine learning model and the case workers currently used system.
2. Optimize the pilot group's user experience with the prototype.
3. Increase the objectivity of a case worker's assessment when assigning rehabilitation training.
4. Increase the objectivity of a case worker's assessment when assigning fall preventive training.

Goal #1
Developing a useful prototype has been achieved through research, user involvement, and user evaluation. The research used has been centered around recommendations for developing a user interface both in general and for a CDSS and how the users should be involved in the development process. The involvement of the users, according to the research, has proven beneficial as it showed contradictions between their requirements and the state of the art research.

Evaluating the prototype's usefulness has been done by employing the methods described in sections 3.6 to 3.8. The results indicated very high usability among the participants, which were most prominently seen in the CSUQ, SUS, and UMUX results. Since the term "useful" is a combination of usability and utility, the utility has also been validated. This was done using User Acceptance Testing (UAT), which resulted in high utility, as the prototype delivered the functionality required by the users.

Some of the required functionality was to show information from the machine learning model and the case worker's current system. The information from the machine learning model was retrieved through an established connection in the developed prototype. The information from the case worker's current system was mocked data, as a connection was not made. However, a proof was made to show that actual data would be possible to retrieve in the future.

Goal #2

Optimizing the user experience is more of a process rather than a definitive result. It is essential to look at the results over time to see if the improvements have had any positive impact on the user experience. This impact has been measured both qualitatively and quantitatively, with the employed methods described in sections 3.6 to 3.8.

These methods have looked at the users' performance during the experiments and their interpretation of the perceived usability. Using the prototype showed a performance increase over time across the participants. This is corroborated by the user's perceived usability, as they were tasked to rate their interactions with the prototype, stating how difficult the interactions were. Furthermore, the users' feedback has gotten more positive throughout the experiments.

Goal #3 & Goal #4

Increasing the objectivity of the case workers has been achieved using a machine learning (ML) model. The ML model has provided a probability, indicating if a citizen should receive rehabilitation training or is at the risk of falling within the next three months.

Both probabilities have, through the experiments, been stated by the case workers to be useful guidance when making assessments, indicating increased objectivity for goals #3 and #4. While objectivity has increased, the case workers also required further understanding of what had contributed to the probability.

These contributions have been presented as textual arguments that combine the mocked data from the case workers current system with an interpretation method called SHAP, also provided by the ML model. SHAP can show the positive and negative contributions behind a probability, enabling the case workers to interpret the arguments to see why the probability has its given value.

SHAP has a limitation, as it only shows the contributions to the probability and not the underlying explanations for the contributions. Because of this limitation, it is impossible to use SHAP as required by the case workers. A recommendation for future iterations of the ML model should be to integrate a knowledge base capable of providing more in-depth explanations behind the probability.

6.3 Future Work

Assuming the project model responsible for providing the probability is completed to a satisfactory degree, it is possible to use the prototype and further develop it into a product. From the interviews with the case workers it has been clear that the desire for an actual product is present, as it according to them has the potential to make their work easier. Before it is possible to release a product there are many things that needs to be completed, where most of these are technical related, as the general usability and utility is as desired by the case workers. Some of the technical things to implement is the connection to Cura's API, such that it is possible to retrieve the real data. While there are 98 municipalities in Denmark, not all municipalities use Cura, which means that there are additional systems that the product should be able to interface with. Before these systems are connected it is recommended that it gets tested with real data from Cura, such that it is possible to know if the product is still valid to those municipalities using Cura. Assuming that the validity is still present, additional systems should be implemented in iterations, such that errors can be carefully corrected along the way.

While the connection to Cura is of utmost importance for the product, there are also other technical things to implement, such as basic unit, integration, end-to-end and system tests. Implementing these tests ensure a higher level of product stability, as it will then be possible to discover new errors when changes to the product is made.

Even though there is much technical work to be done before a product release is possible, certain certifications also seems appropriate to acquire, as the product is intended for the Danish municipality workers. Another currently unknown factor is how the product would be received across all municipalities. Because the core premise of the product is to present the user with a score based on machine learning, there is likely going to be some who will try to reject or block the project. Should this be the case there is also a need to work on the public perception of the product, as the product is not intended to replace the case workers, but rather inform them based on statistics and thereby support their decision.

Appendix A

Experiment Data

A.1 Error opportunities

Table A.2 shows a list of the error opportunities presented in the results, as some of the categories have been combined to create a better overview.

Table A.1: List of all possible error opportunities in the in the prototype during testing

Error type	Error description
Random click error	A left or right mouse click in a location not defined by the other error types.
Menu item error	A click on a menu item, which is not in the set of actions for the given task.
Menu error	A mouse hover on the menu, which is not in the set of actions for the given task.
Hide plot error	Hiding one of the plots in the graph showing assistive aids over time and home help hours.
Show plot error	Showing one of the plots in the graph showing assistive aids over time and home help hours.
Right click on graph error	A right mouse click on the graph showing assistive aids over time and home help hours.
Left click on graph error	A left mouse click on the graph showing assistive aids over time and home help hours.
Table cell click error	A left or right mouse click on cell in a table showing either assistive aids, training plans, or registered falls. Table pagination errors are also counted.
Tabs click error	Switched to another tab.
Wrongly opened Argumentbox error	Opened argument expansion panel when it not in the set of actions for the given task.
Wrongly closed Argumentbox error	Closed argument expansion panel when it not in the set of actions for the given task.
Cursor drag error	Dragged the cursor on the graph showing assistive aids over time and home help hours.

Table A.2: List of error opportunities with categories combined for overview.

Error type	Error description
Random click error	A left or right mouse click in a location not defined by the other error types.
Menu item error	A click on a menu item, which is not in the set of actions for the given task.
Menu error	A mouse hover on the menu, which is not in the set of actions for the given task.
Plot error	Represents both *Hide plot error* and *Show plot error* from table A.1.
Click on graph error	Represents both *Right click on graph error* and *Left click on graph error* from table A.1.
Table cell click error	A left or right mouse click on cell in a table showing either assistive aids, training plans, or registered falls. Table pagination errors are also counted.
Tabs click error	Switched to another tab.
Manipulated argumentbox error	Represents both *Wrongly opened Argumentbox error* and *Wrongly closed Argumentbox error* from table A.1.
Cursor drag error	Dragged the cursor on the graph showing assistive aids over time and home help hours.

A.2 Experiment 1

Vær sikker på at alle referencer er til ting der er i appendix og ikke rapporten

A.2.1 Changelog

Table A.3: Full Changelog between the initial design and the design used throughout Experiment 1

Id	Change description	Reason for change
	Added Functionality	
1	Added a page for case 2.	At the time of developing the initial design (see figure 4.2), case 2 did not exist. When it did, it was considered relevant for the project and thus included.
2	Added a Navigation Drawer.	To be able to access both case 1 and case 2 and additionally functionality if needed.
3	Added a tab for the citizen's Diagnoses and Motivation.	To get a more clean separation between the information provided by the ML API and Cura. The pilot group approved this change (see number 38 in table A.16).
4	Added a horizontal spacer between the citizens information, GOP and the rest of the prototype.	To separate the information.
5	Added a User Profile	To see who has logged into the prototype.
6	Added a SSN Search Field.	For the case workers to search for a citizen to obtain a result from both the ML API and Cura.

Continued on next page

Id	Change description	Reason for change
	Table A.3 – continued from previous page	
7	Added a Dark mode.	A function located in the navigation drawer for the end users to change the color contrast polarity to reduce eye fatigue.
8	Added a Color blind mode.	A function located in the navigation drawer for the end users to change the color scheme to enable support for red-green color blindness.
9	Added a Login page.	A safety measure to prevent unauthorized users from gaining access to the prototype. It also acts as a verification measure for Cura as mentioned in section 4.3.2.
10	Added Log out.	Logging in should also entail the ability to log out.
11	A tooltip.	To inform the user how to hide and show the plots on the graph in the tab "Hjælpemidler".
	Modified Functionality	
12	Changed the probability description text.	To make it more clear that the text described the probability, and was positively commented during the semi-structured interview (see number 22 in table A.16).
13	Changed the font-size of the probability description.	To make sure that the text was clearly visible, and was positively commented during the semi-structured interview (see number 22 in table A.16).
14	Changed the alignment of the citizens name and SSN in the citizen information for case 1 and 2.	To achieve more cohesion.

Continued on next page

Table A.3 – continued from previous page

Id	Change description	Reason for change
15	Changed the text color on the header of the argument expansion panels.	To match the text color used throughout most of the prototype.
16	Changed the font-size on the header of the argument expansion panels.	To make the header text more readable, and to match the font-size on each tab header ("Diagnoser og Motivation", "Hjælpemidler" and "Træningsplaner").
17	Made the header of the argument expansion panels change contrast when expanded.	To show a clear visual difference between expanded versus not expanded.
18	Changed the column header text from "Dato for udlevering" to "Udleveringsdato" in the table listing the citizens assistive aids in the tab "Hjælpemidler".	Terminology correction given by the pilot group during R&D2.
19	Changed the column header text from "Dato for aflevering" to "Afleveringsdato" in the table listing the citizens assistive aids in the tab "Hjælpemidler".	Terminology correction given by the pilot group during R&D2.
20	Changed the column header text from "Begrundelse for bevilling" to "Bevillingsbegrundelse" in the table listing the citizens assistive aids in the tab "Hjælpemidler".	Terminology correction given by the pilot group during R&D2.
21	Changed probability gauge to use a new library.	Customization limitations (see - section 4.4.3).
22	Changed the background color to white.	
	Removed Functionality	
23	Removed text clipping from the probability description text.	Improve readability.

Continued on next page

Table A.3 – continued from previous page

Id	Change description	Reason for change
24	Removed the user badge on the top right corner for case 1 and 2.	The functionality was moved to the navigation drawer to make the prototype more scannable (see section 3.4.2).

A.2.2 Prototype design

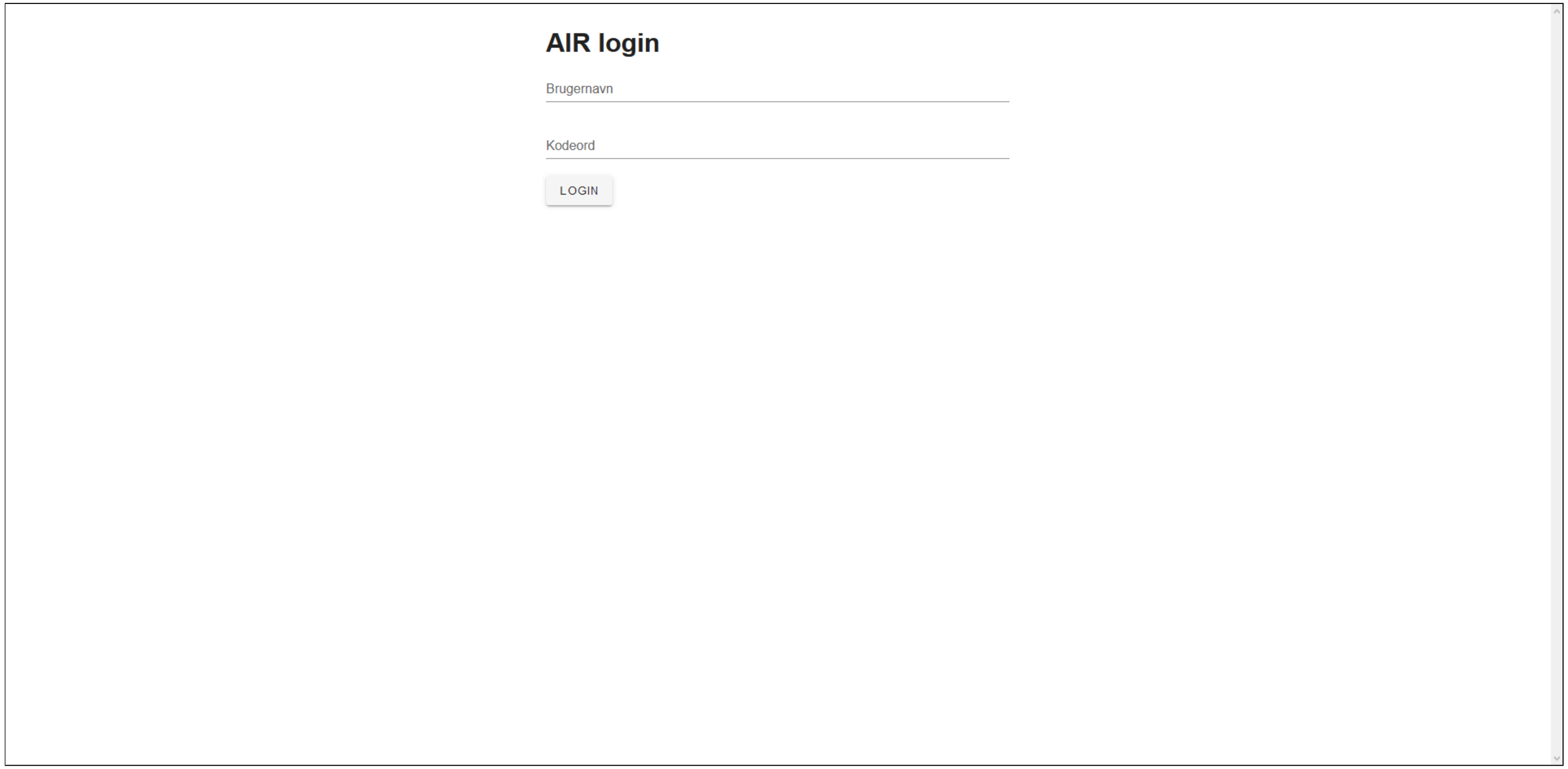

Figure A.1: Prototype login page

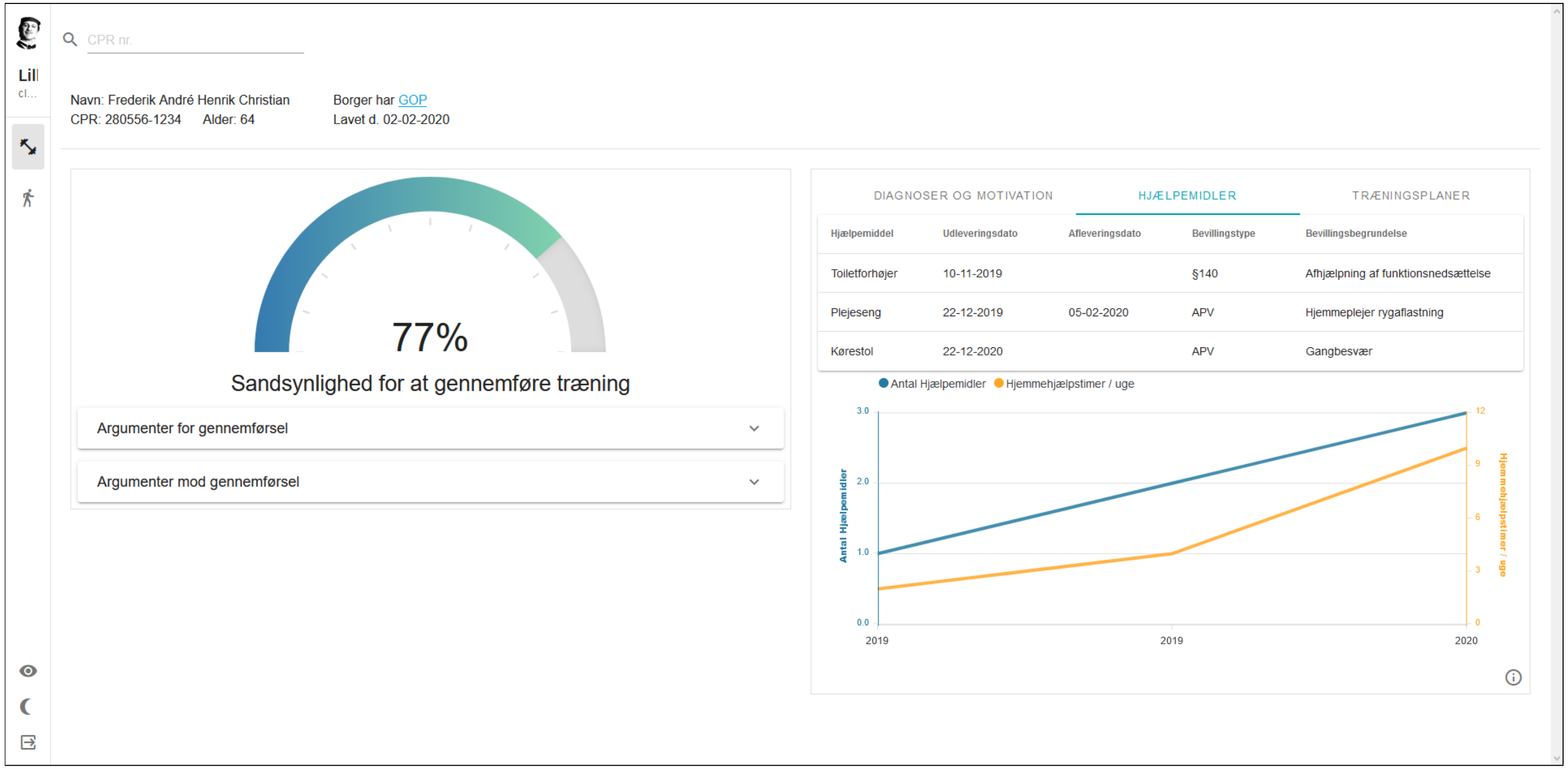

Figure A.2: Prototype Case 1 - Showing assistive aids tab

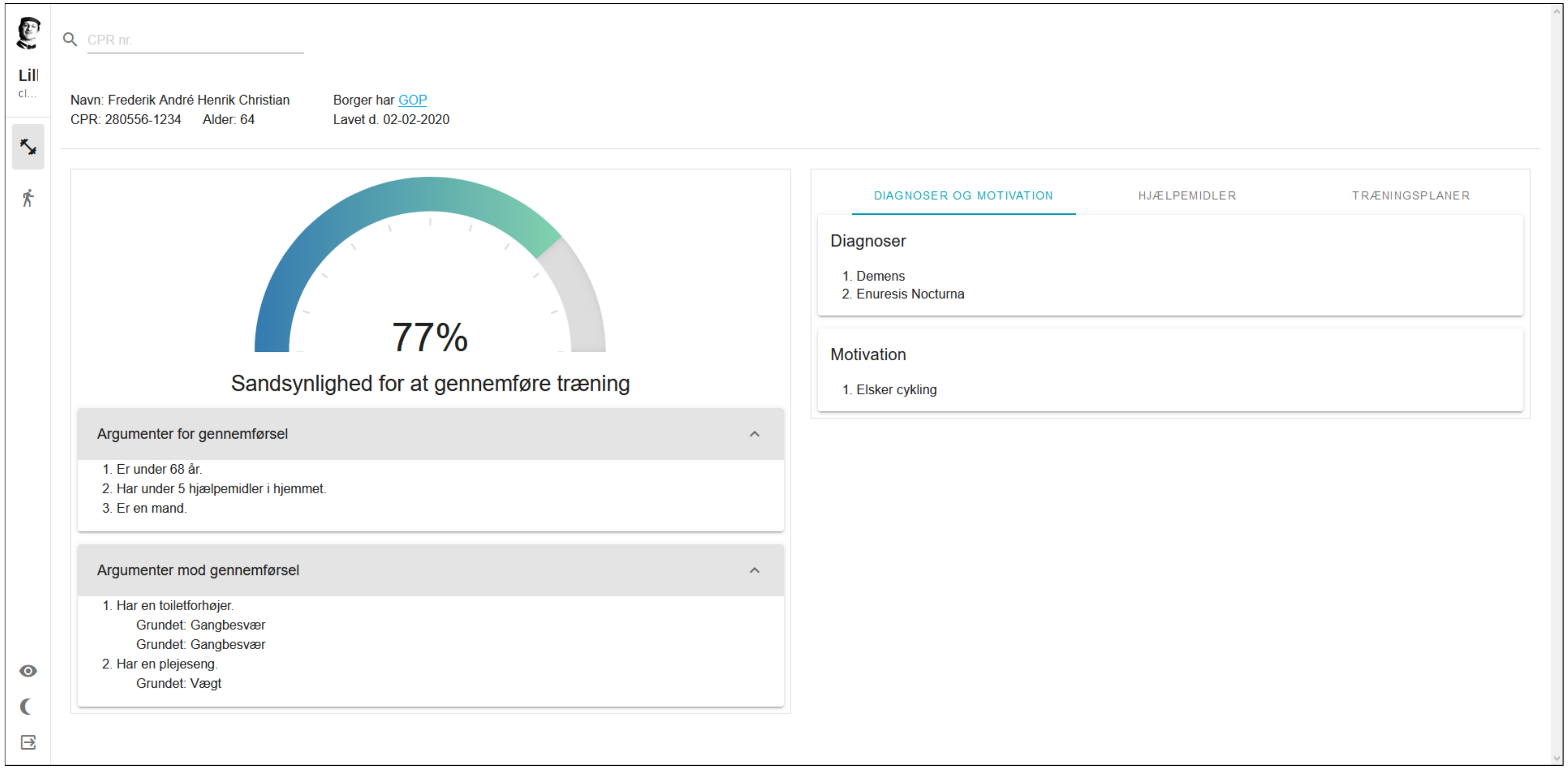

Figure A.3: Prototype Case 1 - Showing arguments for and against and the diagnosis and motivation tab

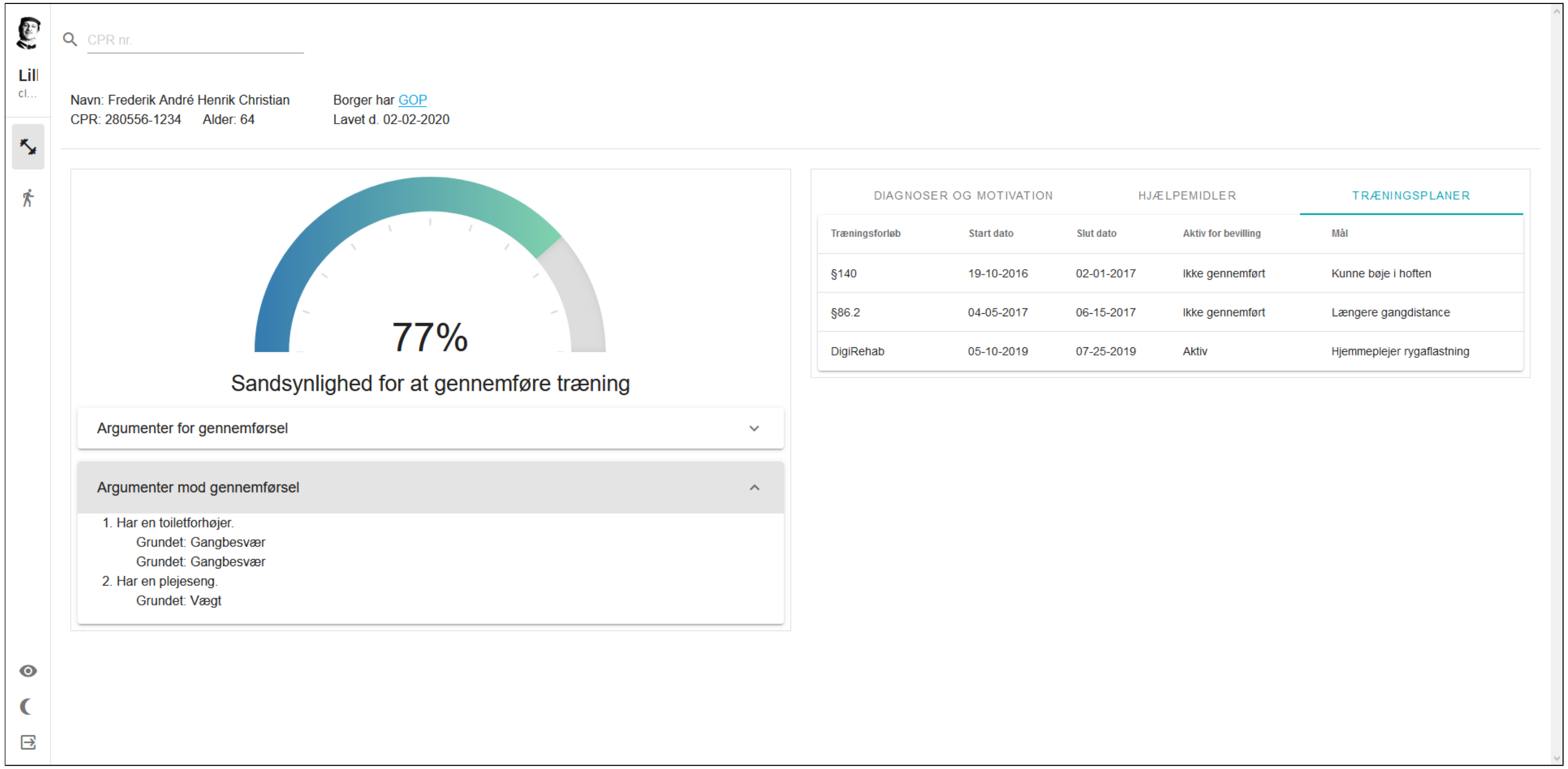

Figure A.4: Prototype Case 1 - Showing arguments against and training plan tab

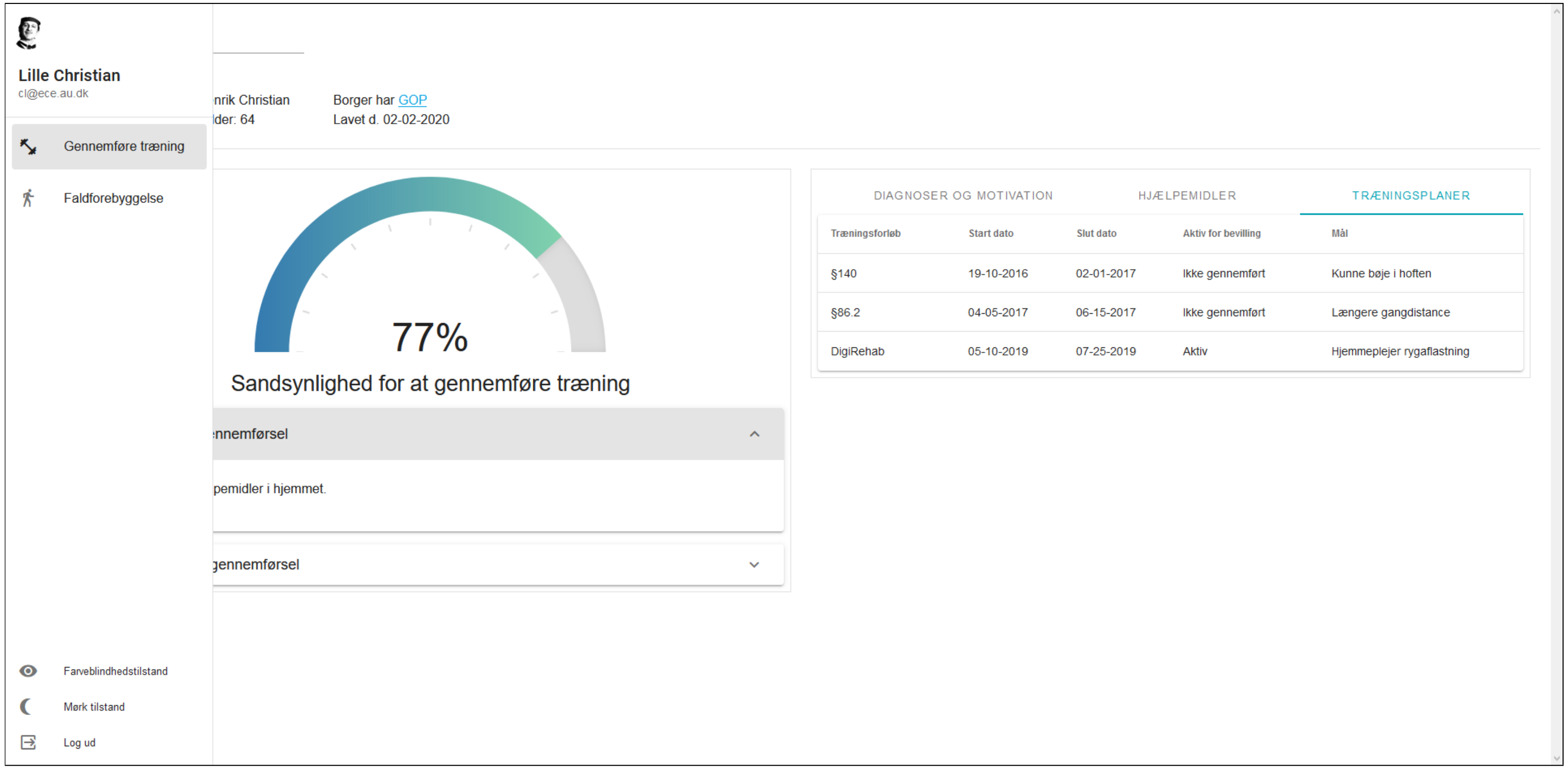

Figure A.5: Prototype Case 1 - Showing arguments for, the navigation drawer and the training plan tab

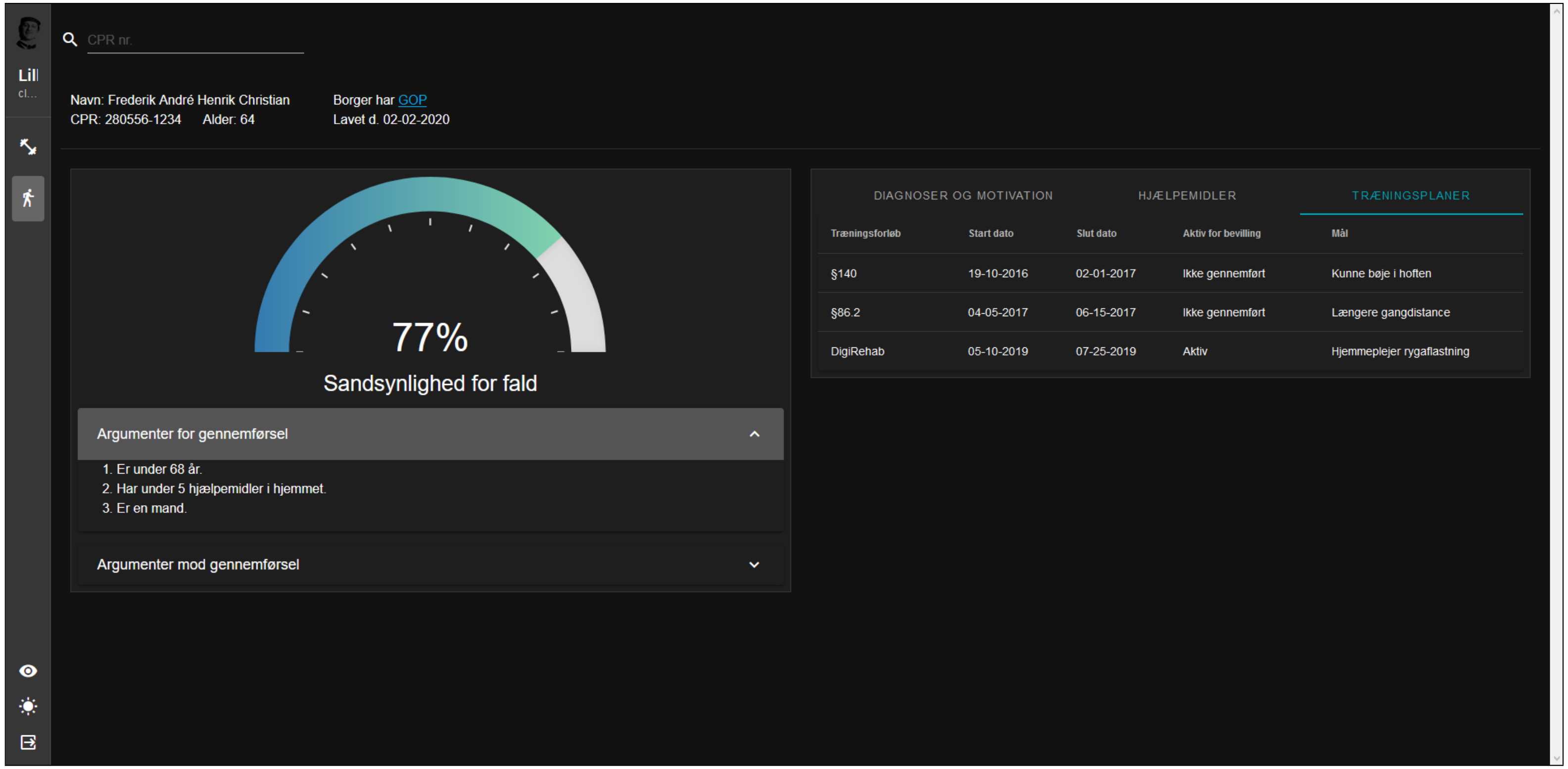

Figure A.6: Prototype Case 2 - Showing dark mode, arguments for and the training plans tab

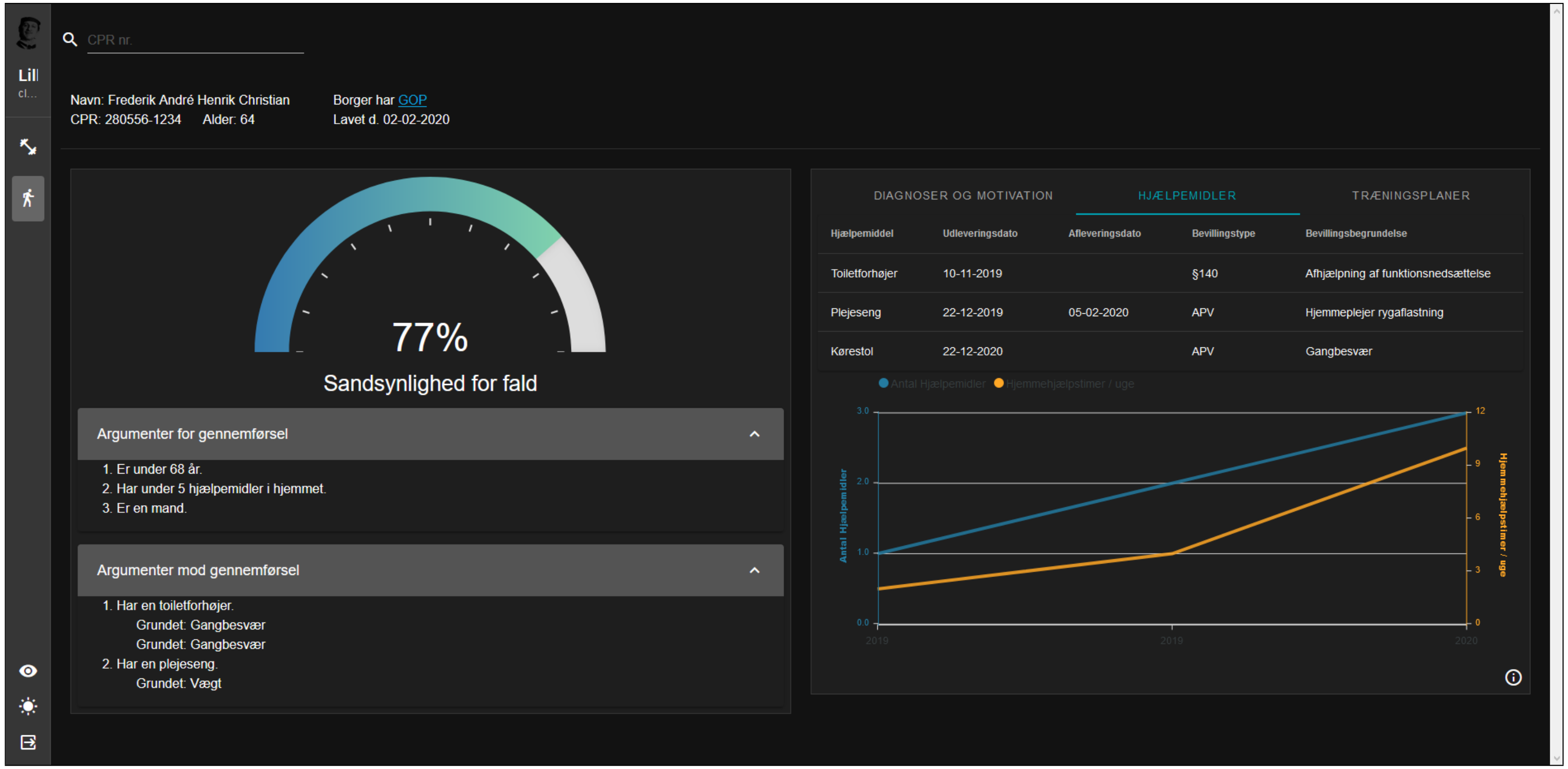

Figure A.7: Prototype Case 2 - Showing dark mode, arguments for and against, and the assistive aids tab

A.2.3 Post-task questionnaire

Table A.4: Danish version of the SEQ used during Experiment 1

Task #	Identifier	Task description
1	Log In	Log ind på hjemmesiden. Brugernavnet og koden er den samme som til Cura.
2	SSN Search	Søg efter en borger, der har CPR nummeret: 010151-0101.
3	Colorblind	Aktiver farveblindhedstilstand på hjemmesiden.
4	Arguments For	Find og vis argumenterne for at borgeren kan gennemføre træning.
5	Arguments Against	Find og vis argumenterne imod at borgeren kan gennemføre træning.
6	Hide Plot	Skjul den ene kurve i grafen.
7	Training Plans	Find og vis borgerens træningsplaner.
8	Assistive Aids	Find et af borgerens hjælpemidler der er udstedt i 2020.
9	GOP	Find og vis borgerens genoptræningsplan.
10	Dark Mode	Aktiver mørk-tilstand på hjemmesiden.
11	Fall Prevention	Find og vis siden for faldforebyggelse.
12	Log Out	Log ud af hjemmesiden.

Table A.5: SEQ answers for the individual tasks and users during Experiment 1

Task \ User	User 1 [1;7]	User 2 [1;7]	User 3 [1;7]	User 4 [1;7]	User 5 [1;7]	Mean [1;7]	UB [1;7]	LB [1;7]
Log In	7	4	7	7	7	6,4	7,0	4,7
SSN Search	4	7	7	7	7	6,4	7,0	4,7
Colorblind	1	7	5	7	6	5,2	7,0	2,1
Arguments For	7	7	7	7	7	7,0	7,0	7,0
Arguments Against	7	7	7	7	7	7,0	7,0	7,0
Hide Plot	1	3	7	7	3	4,2	7,0	1,0
Training Plans	7	7	5	7	7	6,6	7,0	5,5
Assistive Aids	7	7	6	7	7	6,8	7,0	6,2
GOP	7	7	6	7	7	6,8	7,0	6,2
Dark Mode	7	7	6	7	7	6,8	7,0	6,2
Fall Prevention	7	6	6	7	7	6,6	7,0	5,9
Log Out	7	7	7	7	7	7,0	7,0	7,0

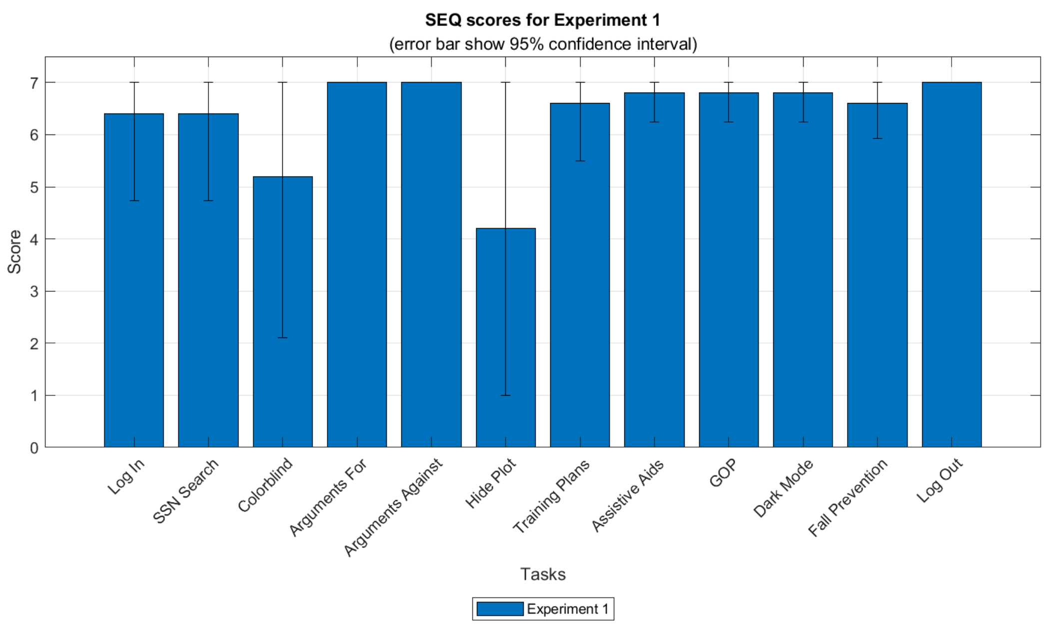

Figure A.8: SEQ answers represented on a graph for Experiment 1

A.2.4 Post-session questionnaire

Table A.6: CSUQ questions + 1 used during Experiment 1, 2, and 3

Question #	Question description
1	Samlet set er jeg tilfreds med hvor let det er at bruge hjemmesiden.
2	Det var simpelt at bruge hjemmesiden.
3	Jeg kunne udføre mit arbejde hurtigt ved hjælp af hjemmesiden.
4	Jeg føler mig komfortabel med hjemmesiden.
5	Det var nemt at lære at bruge hjemmesiden.
6	Jeg føler jeg hurtig blev produktiv ved at bruge hjemmesiden.
7	Hjemmesiden giver fejlbeskeder der tydeligt fortæller mig hvordan jeg løser problemet.
8	Hvis jeg laver en fejltagelse på hjemmesiden, kan jeg hurtig fortsætte med mit arbejde.
9	Hjemmesidens informationer (online hjælp, skærmmeddelser og anden dokumentation) var klare og tydelige.
10	Det er nemt at finde den information jeg har brug for.
11	Hjemmesidens informationer hjalp mig til nemmere at kunne udføre mit arbejde.
12	Informationens organiseringen på hjemmesidens skærmbilleder er klar og tydelig
13	Grænsefladen* på hjemmesiden er behagelig.
14	Jeg kan godt lide grænsefladen* til hjemmesiden.
15	Hjemmesiden har alt den kunnen og funktionalitet jeg forventede.
16	Jeg er generelt tilfreds med hjemmesiden.
17	Jeg kan se mig selv bruge hjemmesiden som en del af mit arbejde.

Table A.7: CSUQ metric data for Experiment 1

Statement \ User	User 1 [1;7]	User 2 [1;7]	User 3 [1;7]	User 4 [1;7]	User 5 [1;7]			
1	2	1	2	1	1			
2	2	1	1	1	1			
3	2	2	NA	1	1			
4	1	2	1	1	1			
5	1	1	2	1	1			
6	1	3	NA	NA	1			
7	NA	NA	NA	NA	3			
8	NA	NA	NA	NA	NA			
9	NA	1	NA	1	NA			
10	NA	1	2	1	1			
11	2	2	3	NA	1			
12	2	2	3	NA	1			
13	1	2	1	1	1			
14	1	NA	1	1	1			
15	NA	2	NA	1	1			
16	1	2	2	1	1			
17	1	1	2	1	1			
Score \ User	**User 1 [1;7]**	**User 2 [1;7]**	**User 3 [1;7]**	**User 4 [1;7]**	**User 5 [1;7]**	**Mean [1;7]**	**UB [1;7]**	**LB [1;7]**
Overall	1,4	1,7	1,8	1,0	1,1	1,4	1,8	1,0
SysQual	1,5	1,7	1,5	1,0	1,0	1,3	1,7	1,0
InfoQual	1,5	1,5	2,7	1,0	1,5	1,6	2,4	1,0
IntQual	1,0	2,0	1,0	1,0	1,0	1,2	1,8	1,0
Completeness	69%	81%	63%	75%	88%	75,2%		

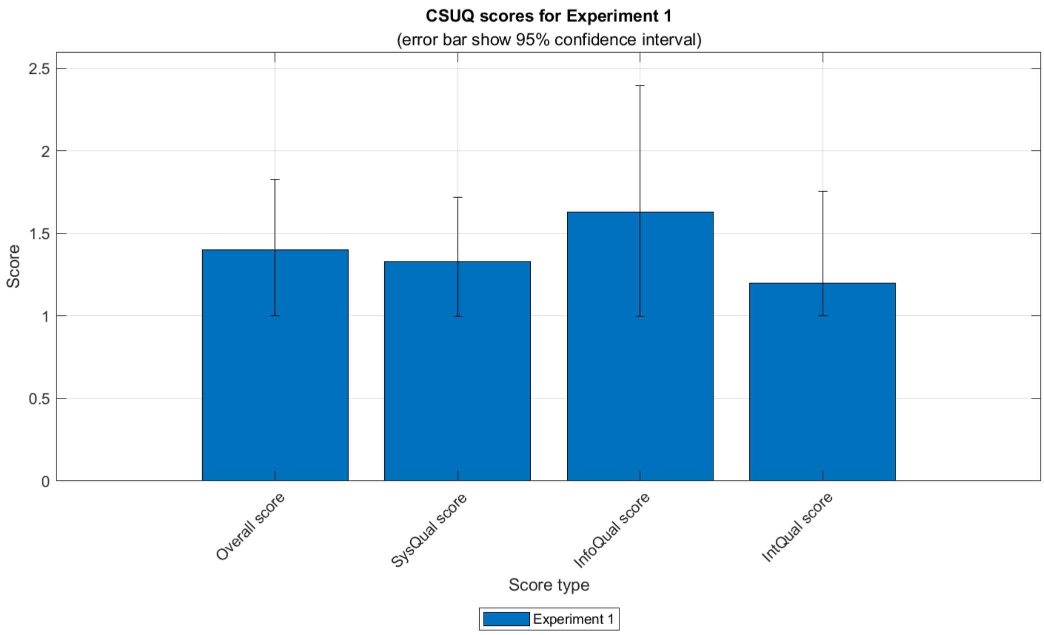

Figure A.9: Mean CSUQ scores for Experiment 1

A.2.5 Task success

Table A.8: Task success for Experiment 1 - Completed tasks are marked with (✓) and failed tasks are marked with (✗)

Task \ User	User 1	User 2	User 3	User 4	User 5	Mean
Log In	✓	✓	✓	✓	✓	100%
SSN Search	✓	✓	✓	✓	✓	100%
Colorblind	✓	✓	✓	✓	✓	100%
Arguments For	✗	✓	✓	✓	✓	80%
Arguments Against	✓	✓	✓	✓	✓	100%
Hide Plot	✗	✓	✓	✓	✓	80%
Training Plans	✓	✓	✓	✓	✓	100%
Assistive Aids	✓	✓	✓	✓	✓	100%
GOP	✗	✗	✗	✓	✗	20%
Dark Mode	✓	✓	✓	✓	✓	100%
Fall Prevention	✗	✓	✗	✓	✓	60%
Log Out	✓	✓	✗	✓	✓	80%

Table A.9: Received task assistance for Experiment 1 - Tasks with received assistance are marked with (✓) and tasks without needed assistance are marked with (✗)

Task \ User	User 1	User 2	User 3	User 4	User 5	Mean
Log In	✗	✗	✗	✗	✗	0%
SSN Search	✗	✗	✗	✗	✗	0%
Colorblind	✓	✗	✓	✗	✗	40%
Arguments For	✗	✗	✗	✗	✗	0%
Arguments Against	✗	✗	✗	✗	✗	0%
Hide Plot	✓	✓	✗	✗	✓	60%
Training Plans	✗	✗	✗	✗	✗	0%
Assistive Aids	✗	✗	✗	✗	✗	0%
GOP	✗	✓	✗	✓	✗	40%
Dark Mode	✗	✗	✗	✗	✗	0%
Fall Prevention	✗	✗	✗	✗	✗	0%
Log Out	✗	✗	✗	✗	✗	0%

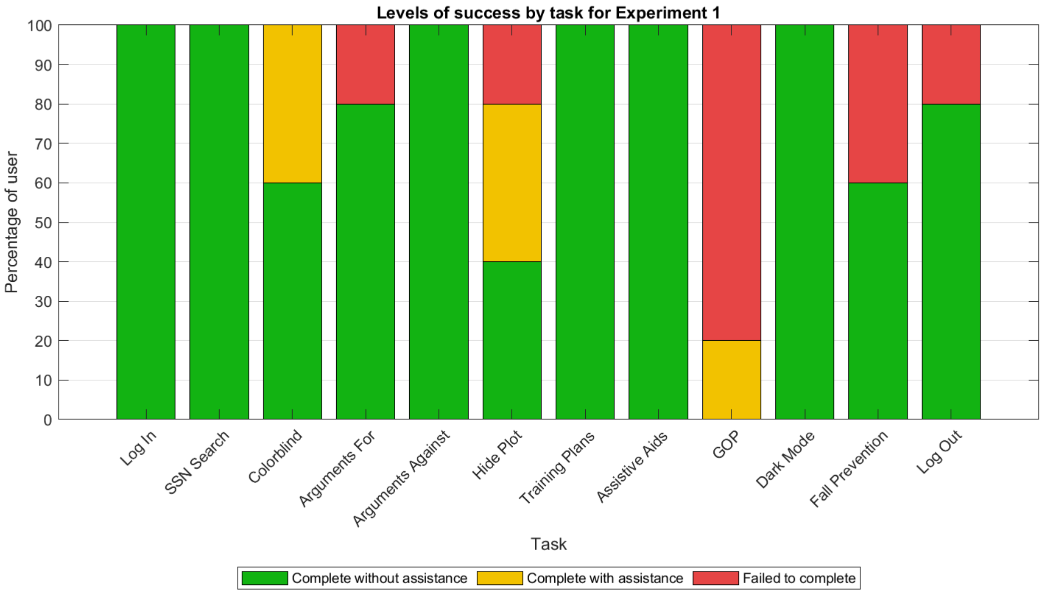

Figure A.10: Percentage of users for each task who completed a task without assistance, with assistance, or failed to complete the task for Experiment 1

A.2.6 Task time

Table A.10: Task completion time for Experiment 1

Task \ User	User 1 [s]	User 2 [s]	User 3 [s]	User 4 [s]	User 5 [s]	Mean [s]	UB [s]	LB [s]
Log In	17	22	21	19	9	17,6	24,0	11,2
SSN Search	26	19	8	6	9	13,6	24,2	3,0
Colorblind	57	14	28	8	13	24,0	48,7	0
Arguments For	-	4	12	2	3	5,3	12,5	0
Arguments Against	4	4	12	5	1	5,2	10,3	0,1
Hide Plot	-	16	17	18	74	31,3	76,6	0
Training Plans	6	5	16	30	5	12,4	25,9	0
Assistive Aids	15	8	5	6	4	7,6	13,1	2,1
GOP	-	-	-	23	-	23,0	-	-
Dark Mode	15	3	9	3	4	6,8	13,3	0,3
Fall Prevention	-	14	-	3	2	6,3	22,9	0
Log Out	8	9	-	6	6	7,3	9,6	4,9
Mean	18,5	10,7	14,2	10,8	11,8	13,4		
Mean			13,2					

Table A.11: Total time on task for Experiment 1

Task \ User	User 1 [s]	User 2 [s]	User 3 [s]	User 4 [s]	User 5 [s]	Mean [s]	UB [s]	LB [s]
Log In	24	29	27	25	12	23,4	31,7	15,1
SSN Search	38	31	16	11	15	22,2	36,7	7,7
Colorblind	73	25	40	15	27	36,0	64,0	8,0
Arguments For	14	15	23	13	9	14,8	21,2	8,4
Arguments Against	17	9	23	5	4	11,6	21,7	1,5
Hide Plot	96	83	26	30	80	63,0	103,4	22,6
Training Plans	10	10	35	39	6	20,0	39,5	0,5
Assistive Aids	20	8	7	18	9	12,4	20,0	4,8
GOP	15	86	14	37	6	31,6	72,0	0
Dark Mode	23	9	12	6	7	11,4	19,9	2,9
Fall Prevention	11	20	16	13	5	13,0	20,0	6,0
Log Out	13	12	8	9	8	10,0	12,9	7,1
Mean	29,5	28,1	20,6	18,4	15,7	22,5		
Mean			22,5					

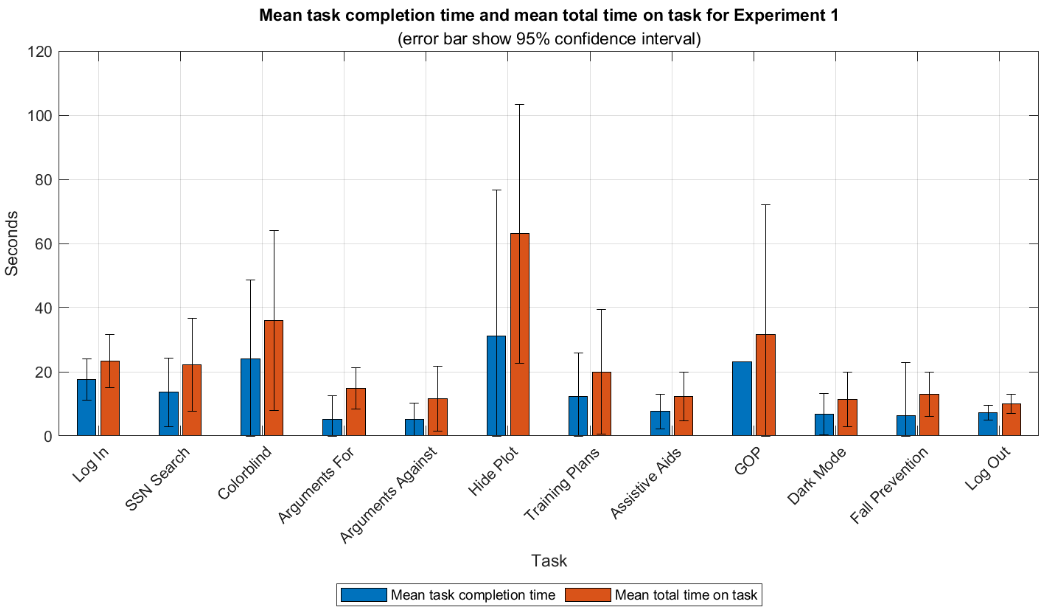

Figure A.11: Mean task completion time and mean total time on task for Experiment 1

A.2.7 Errors

Table A.12: Errors for Experiment 1

Error type \ User	User 1 [#]	User 2 [#]	User 3 [#]	User 4 [#]	User 5 [#]	Sum [#]
Random click error	6	0	0	0	0	6
Menu item error	3	1	0	3	0	7
Menu error	0	1	0	2	1	4
Hide plot error	0	1	1	0	1	3
Show plot error	0	1	0	2	1	4
Right click on graph error	0	1	0	0	0	1
Left click on graph error	3	0	0	1	2	6
Table cell click error	0	2	0	7	5	14
Tabs click error	1	1	1	1	1	5
Wrongly opened Argumentbox error	1	1	1	1	0	4
Wrongly closed Argumentbox error	2	0	2	0	0	4
Cursor drag error	1	0	0	0	0	1
Sum	17	9	5	17	11	59

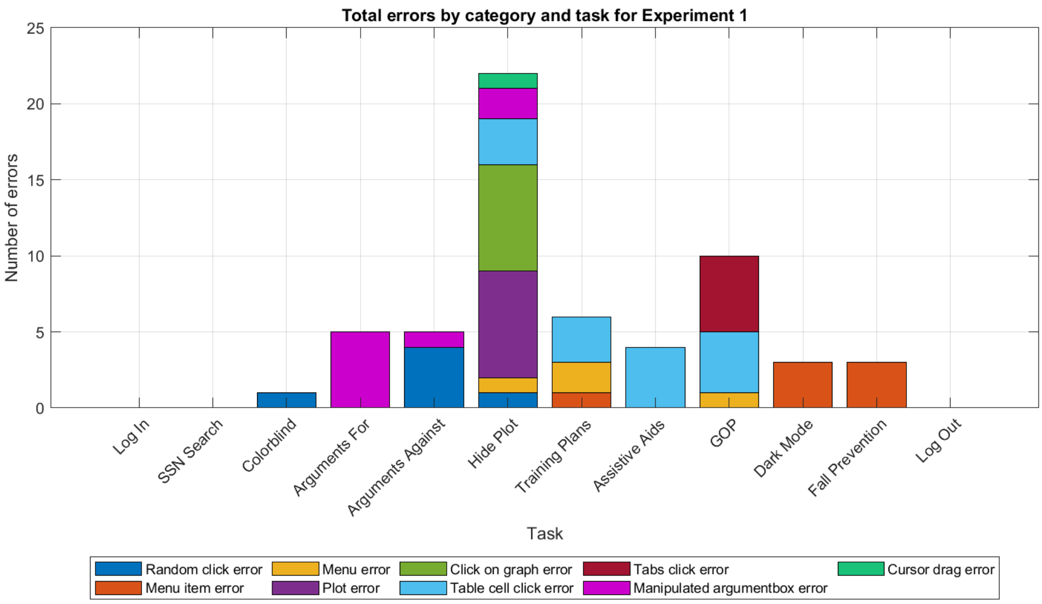

Figure A.12: Total number of errors for Experiment 1

A.2.8 Efficiency

Table A.13: Efficiency for Experiment 1

Task \ User	User 1 [#]	User 2 [#]	User 3 [#]	User 4 [#]	User 5 [#]	Mean [#]	Expected [#]	UB [#]	LB [#]
Log In	5	5	5	5	5	5,0	5	5,0	5,0
SSN Search	2	2	2	2	2	2,0	2	2,0	2,0
Colorblind	3	2	2	2	2	2,2	2	2,8	1,6
Arguments For	2	2	2	2	1	1,8	1	2,4	1,2
Arguments Against	6	0	0	0	1	1,4	1	4,6	0
Hide Plot	8	5	4	7	9	6,6	2	9,2	4,0
Training Plans	1	2	1	6	1	2,2	1	4,9	0
Assistive Aids	1	1	1	3	3	1,8	1	3,2	0,4
GOP	1	3	1	4	2	2,2	1	3,8	0,6
Dark Mode	5	2	2	2	2	2,6	2	4,3	0,9
Fall Prevention	1	4	1	4	2	2,4	2	4,3	0,5
Log Out	2	2	1	2	2	1,8	2	2,4	1,2
Sum	37	30	22	39	32	32			
Mean			32,0						

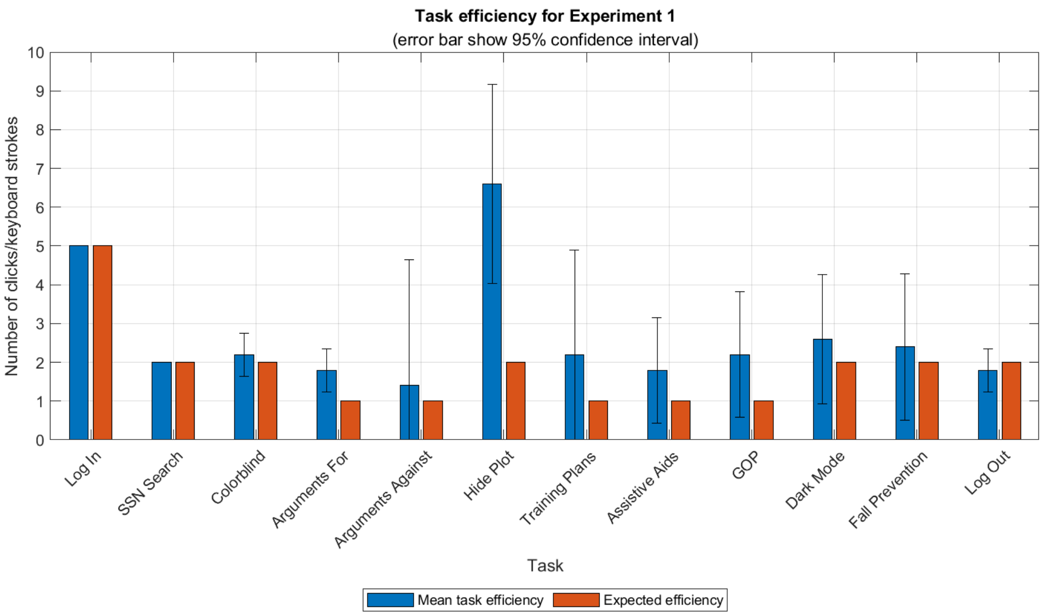

Figure A.13: Efficiency graph for Experiment 1

A.2.9 Learnability

Table A.14: Learnability for Experiment 1

Metric	Measurement
Mean task completion time	13,2
Total number of errors	59
Mean efficiency	32
Sum	104,2

A.2.10 Usability issues

Table A.15: List of usability issues and how many users drew attention to the issue for Experiment 1

Usability Issue #	Usability issue description in Danish	Presented by number of users	Severity Rating
1	Det er ikke tydeligt at argument-boksene er klikbar.	1	Minor
2	Der er for høj kontrast i teksten og tekstskriften er generelt for lille.	2	Minor
3	Det er svært at vide at der var en menu i systemet.	1	Minor
4	Man skal lede efter de tabs som ikke er aktiveret.	2	Minor
5	De inaktive tabs er for blege, og kan næsten ikke læses.	1	Minor
6	Skriftfarven i det mørke tema når graferne skal fjernes er for utydelig.	1	Moderate
7	GOP'en skal kun vises hvis der er en aktiv genoptræningsplan.	5	Moderate

A.2.11 Semi-structured Interview

Table A.16: Gathered Interview Feedback during Experiment 1

Id	Feedback
	Login
1	Jeg synes loginet var Simpelt.
2	Jeg synes loginet var Enkelt.
3	Jeg synes loginet var Logisk.
4	Loginet skal være centreret, så det er lige foran øjnene på en.
	Prototype design
5	Teksten er generelt for lille.
6	Det er rart med opdelingen af tingene fra AI og tingene fra Cura og de ting der skal bruges kan hurtig findes.
7	Opdelingen mellem toppen og det midterste er en god ide.
8	Hjemmesiden ser super godt ud.
9	Siden virker nem og overskuelig fordi man ikke får smidt for meget information i hovedet.
10	Siden er nem og overskuelig.
11	Det er godt at den fane som er markeret har en anden farve en de andre.
12	Tekststørrelsen og skriftfarven er fint.
13	Jeg vil ikke bruge det mørke tema.
14	Jeg vil gerne bruge det mørke tema.
15	Det kan misforstås at de brugte farver i højre side er den samme som i venstre side.
16	Der mangler lidt farve på siden generelt, da det er meget hvid på hvid.
17	Gerne nogle flere nuancer og måske et logo på login skærmen.
	Probability and Arguments
18	Formen og størrelsen på sandsynlighedsskalaen er meget iøjefaldende, og det er en god ting.
19	I er pisse gode. Sandsynlighedsskalaen er lige i øjet.
20	Farven på sandsynlighedsmåleren skal ikke være uniform, men om gradienten er blå-blå eller rød-gul-grøn er lige meget.
21	Farven på sandsynlighedsskalaen skal være en gradient.
22	Teksten under sandsynlighedsmåleren er selvforklarende og kan tydelig læses.

Continued on next page

Table A.16 – continued from previous page	
Id	**Feedback**
23	Indikatoren er god at have med, og den skal holdes på 10%.
24	Vil hellere have rød-gul-grøn skalaen.
25	Farven på sandsynlighedsmåleren er ligegyldig, da det er sandsynligheden der er vigtigst.
26	Jeg forholder mig kun til det tal der er blevet givet.
27	Giver et godt overblik.
28	Den blå gradient er meget behagelig for øjnene.
29	Det var ikke 100% tydeligt at man kunne trykke på argumentboksene, fordi siden var ny.
30	Det var 100% tydeligt at man kunne trykke på argumentboksene.
31	Teksten er generelt tydelig.
32	Argumenterne var lidt mangelfulde.
33	At vise og skjule argumenterne fungere godt.
34	Pilene i siden gav en, en ide om at argumentboksene kunne klikkes på.
	Diagnoses and Motivation
35	Borgerens motivation er meget vigtig at få med.
36	Borgerens diagnoser er meget vigtig at få med.
37	Det er lige meget om diagnoser står først eller motivation.
38	Diagnoser og motivation skal stå i sin egen "fane".
39	Skal kun ha præsenteret hvad borgerens motivation er og hvilket diagnoser borgeren har.
40	En begrundelse for diagnoserne vil være rart.
41	Punkterne bliver ikke forstået som en prioriteret liste, hvilket er godt.
42	Det springer i øjnene som det skal.
43	Der er de informationer der skal være.
44	"Diagnoser og Motivation" skal omdøbes til Helhedsvurdering.
	Assistive Aids
45	Det er super simpelt og nemt.
46	Der er de oplysninger som skal bruges.
47	Alle kolonner i tabellen findes i Cura.
48	Grafen til hjælpemiddelstabellen er ligegyldig.
49	Kolonnen "Udleveringsdato" skal hedde "Leveringsdato".
50	Kolonnen "Afleveringsdato" skal hedde "Hjemtagelsesdato".

Continued on next page

Table A.16 – continued from previous page

Id	Feedback
51	Kolonnetitlerne skal være tydeligere.
52	At vise inaktive hjælpemidler er ligegyldige.
53	Historik er ligegyldigt.
54	At vise både aktive og inaktive hjælpemidler er godt.
55	Det er tydeligt at grafen kan skjules.
56	Det er måske ikke brugbart at skjule graferne.
57	Farven i hjælpemiddelsgrafen kunne afspejle samme farve i hjælpemiddelstabellen.
58	Man skal kunne bladre i hjælpemiddelstabellen, da nogen borgere har mange hjælpemidler.
59	Historik over borgeren hjælpemidler vil give mening.
60	Filtrering af hjælpemidler.
61	Skjul grafen automatisk hvis en borger har ”for mange” hjælpemidler.
	Training Plans
62	Alle kolonner i tabellen findes i Cura.
63	Dataet i tabellen er ikke overflødig.
64	Kolonnen ”Mål” skal stå først.
65	Aktive GOP's skal også stå i tabellen.
66	Kolonnen ”Træningsforløb” er ligegyldig.
67	Tabellen er fin.
68	Tabellen i træningsplaner skal have en ekstra kolonne kaldet ”slutstatus” for at få en begrundelse for hvorfor en træning ikke blev afsluttet.
69	Visualisering af gennemførte og ikke gennemførte træningsplaner.
	Citizen Information
70	Det virker naturligt at borgeroplysningerne befinder sig i øverste venstre hjørne.
71	Der mangler ikke nogen borgeroplysninger.
72	Det hedder ikke ”Lavet” en GOP, men ”Udstedt” eller måske ”Tildelt”.
	Navigation Drawer (Menu)
73	Det virker ikke rart at man skal holde musen over menuen for at få den frem.
74	Det er ikke nødvendigt at se hvem der er logget ind.

Continued on next page

Id	Feedback
	Table A.16 – continued from previous page
75	Menuen fungerer fint.
76	Det er fint at menuen står i venstre side.
77	Menuen skal stå i toppen.
78	Menuen skal først foldes ud når man holder musen over den.
79	Menu i Cura så man slipper for at logge ind.
80	Menuen skal kunne ”pinnes” så menuen altid er tydelig og tilgængelig for dem der vil det.
	Case 2: Fall Preventive Training
81	Siden omkring faldforebyggelse skal indeholde det samme som for rehabilitering + en tilføjelse.
82	Registreret fald skal være tilgængelig under faldforebyggelse med tilhørende dato.
83	Noget information omkring borgerens medicinkort.
	Other
84	Det giver meget mening for mig at bruge det her system da det der skal bruges er samlet på samme side.
85	Der er potentiale.
86	Vi har gjort det mega godt!!!
87	Intet overflødigt fyld.
88	Det har meget relevans.
89	Motivation kan man først få efter man har talt med en borger, men bliver sjældent tastet ind i Cura.
90	Jeg vil gerne have en skriftstørrelsesforøgelsesfunktion.
91	Systemet skal huske hvilket ”state” man som bruger forlod system i når man logger tilbage.
92	Medtag borgerens sygdomme.
93	Skriv noget tekst der fortæller at den ene side er AI relateret og den anden er Cura relateret.
94	Vis alle borgerens aktive GOP’s i toppen (GOP, DigiRehab, eller andet).

A.3 Experiment 2

A.3.1 Changelog

Table A.17: Full Changelog between Experiment 1 and 2

Id	Change description	Reason for change
	Added Functionality	
1	Added a vertical spacer between the citizen information and the GOP.	To visually separate the citizen information and GOP.
2	Added an established connection between the prototype and the ML API.	To show a non-hardcoded probability for both case 1 and case 2.
3	Added pagination to the tab listing the citizens assistive aids.	It was mentioned during the semi structured interview that a citizen could have many assistive aids, thus making the list very long (see number 58 in table A.16).
4	Added autofill to the username field in the login page	To decrease the users effort and cognitive workload.
	Modified Functionality	
5	Changed the x-axis on the graph in the tab listing the citizens assistive aids to shown quarters instead of years.	To show a more detailed overview over time, both in terms of the assistive aids and home help hours.
6	Changed the column header text from "Udleveringsdato" to "Leveringsdato" in the table listing the citizens assistive aids in the tab "Hjælpemidler".	Terminology correction during the semi structured interview (see number 49 in table A.16).
7	Changed the column header text from "Afleveringsdato" to "Hjemtagelsesdato" in the table listing the citizens assistive aids in the tab "Hjælpemidler".	Terminology correction during the semi structured interview (see number 50 in table A.16).

Continued on next page

Table A.17 – continued from previous page		
Id	**Change description**	**Reason for change**
8	Changed the color on the navigation drawer.	Resolved usability issue found during Experiment 1 (see number 3 on table 5.4).
9	Changed the password field in the login page to be automatically focused when the user enters the login page	To decrease the users effort and cognitive workload.
	Removed Functionality	
10	Removed the User Profile.	It was mentioned during the semi structured interview as unnecessary (see number 74 in table A.16).
11	Removed the explanation content from the argument containers.	After the connection to the ML API was made, it became clear that a justification behind each argument was not possible due to the current state of the project model.
12	Removed the SSN Search Field.	The search field was no longer necessary due to the reasons explained in section 4.4.2.
13	Removed the hyperlink functionality associated with the citizens GOP in the top left corner.	It served no purpose for the participants.

A.3.2 Prototype design

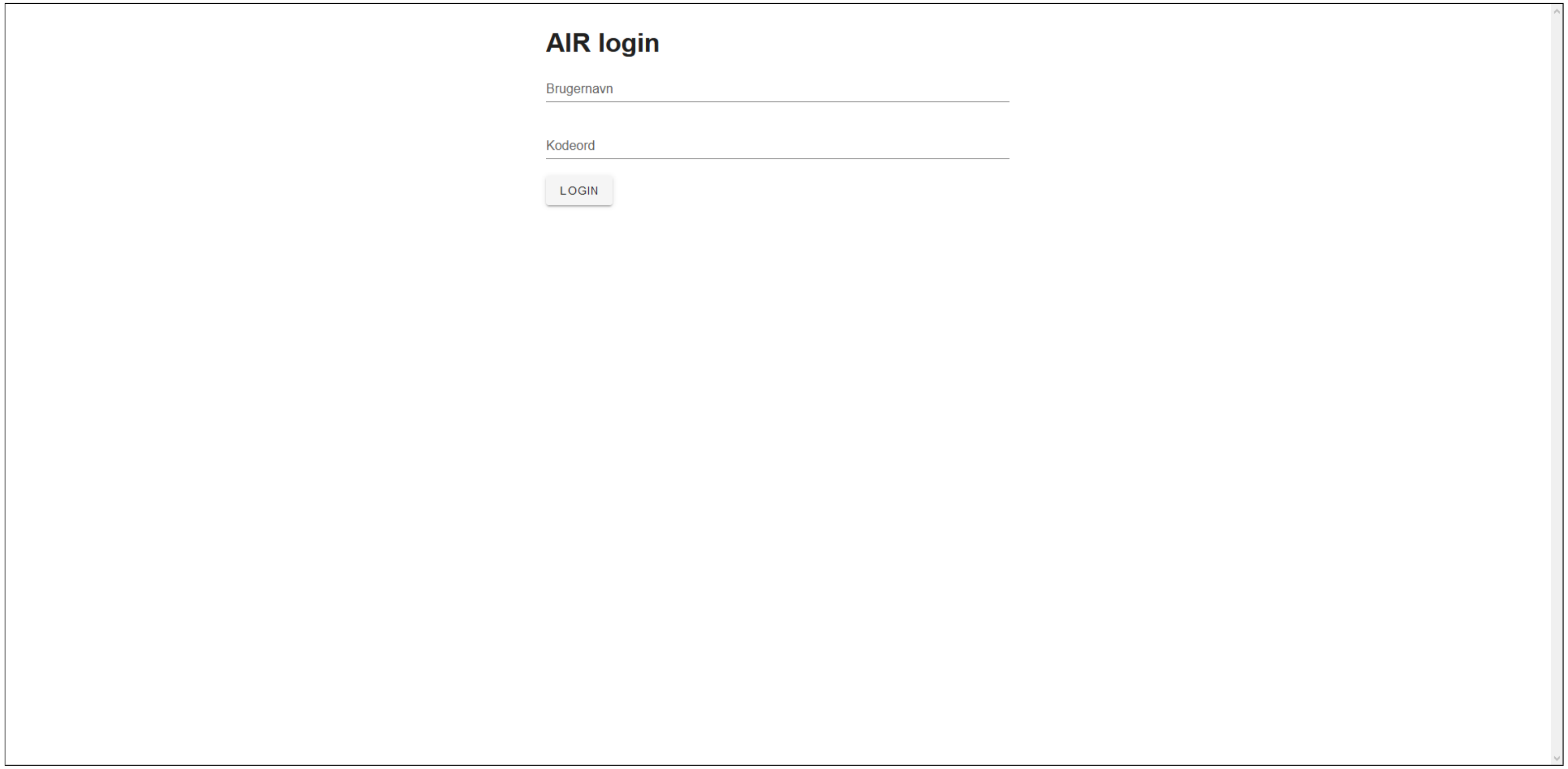

Figure A.14: Prototype login page

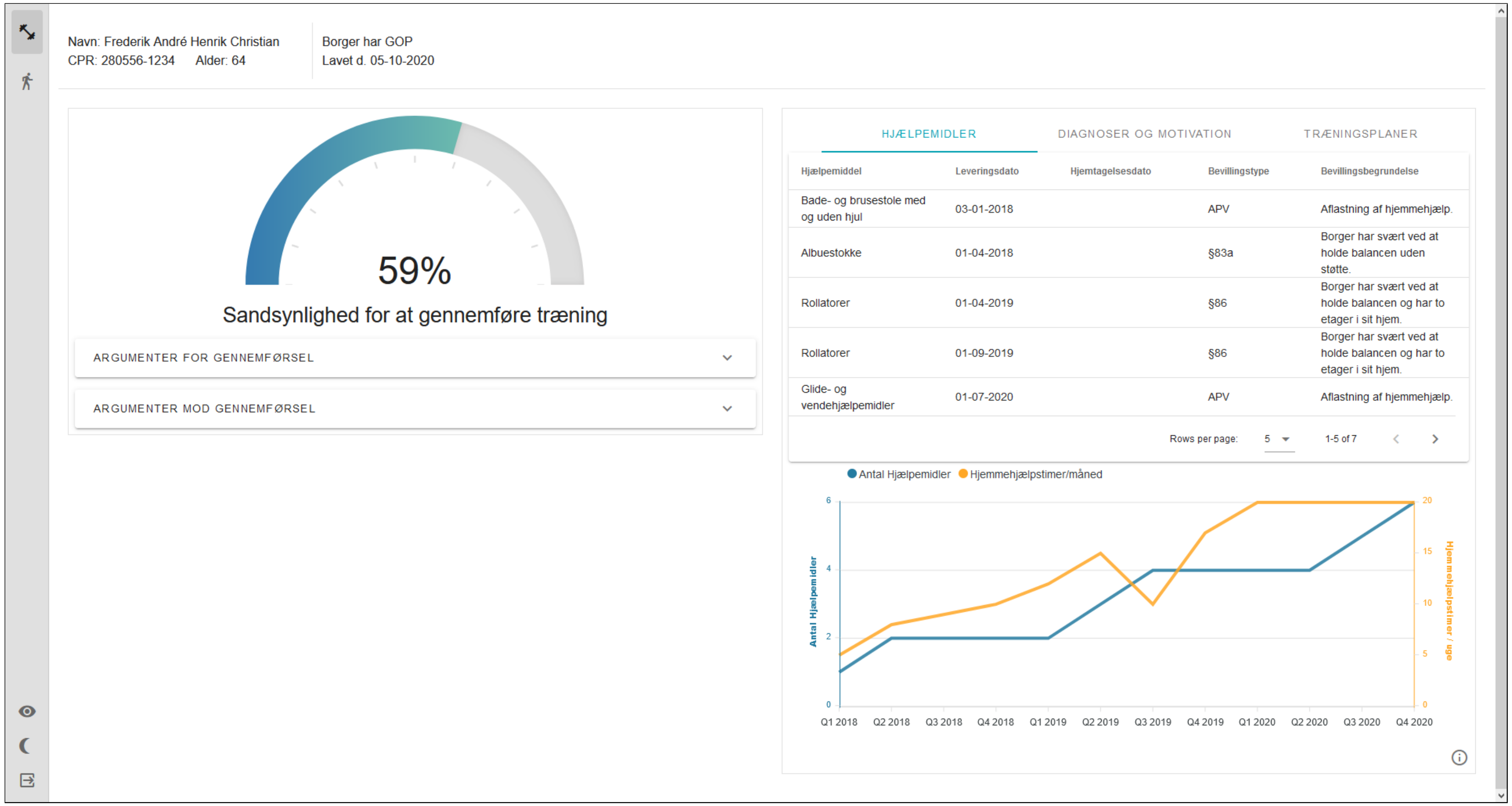

Figure A.15: Prototype Case 1 - Showing assistive aids tab

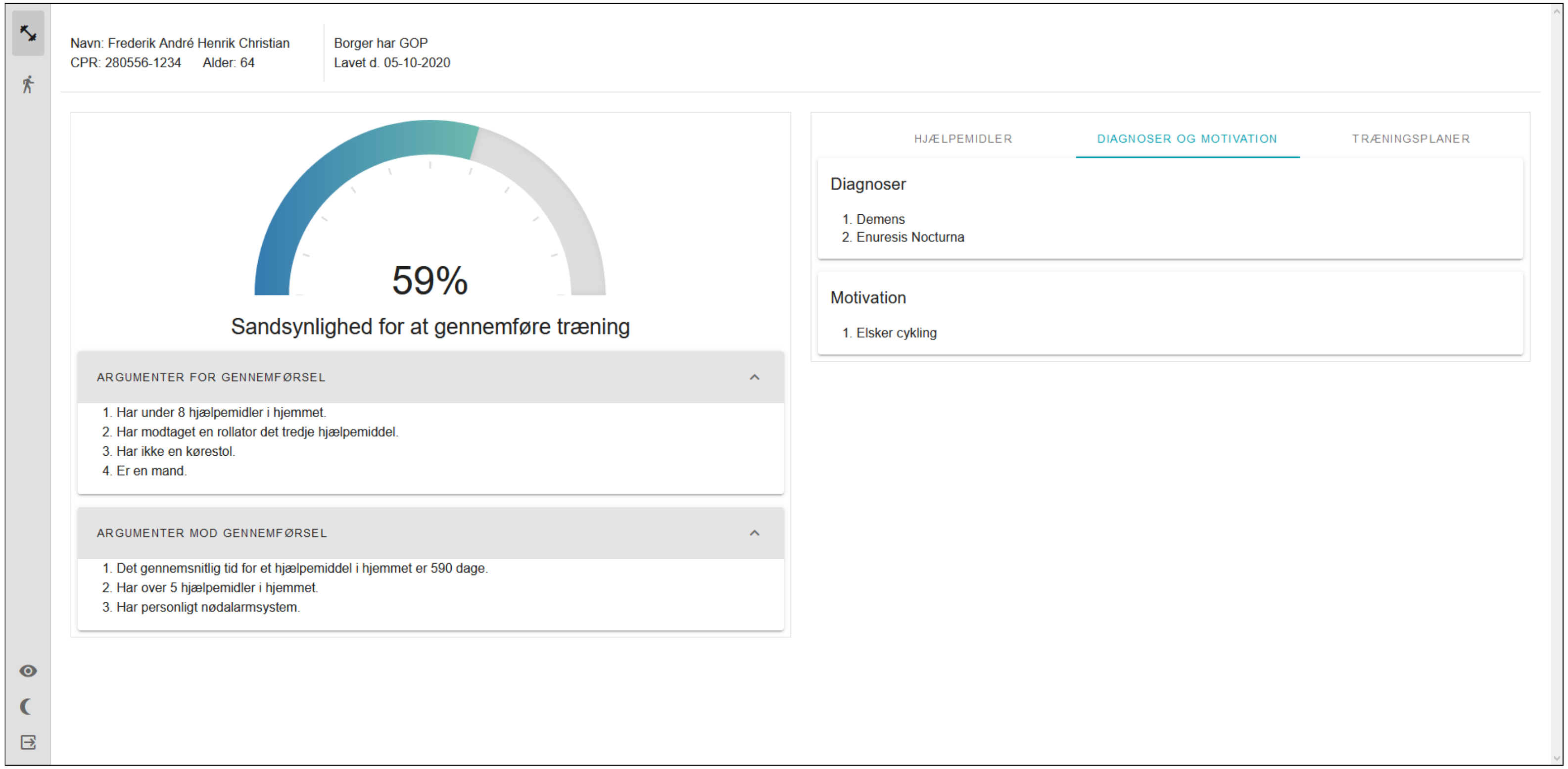

Figure A.16: Prototype Case 1 - Showing arguments for and against and the diagnosis and motivation tab

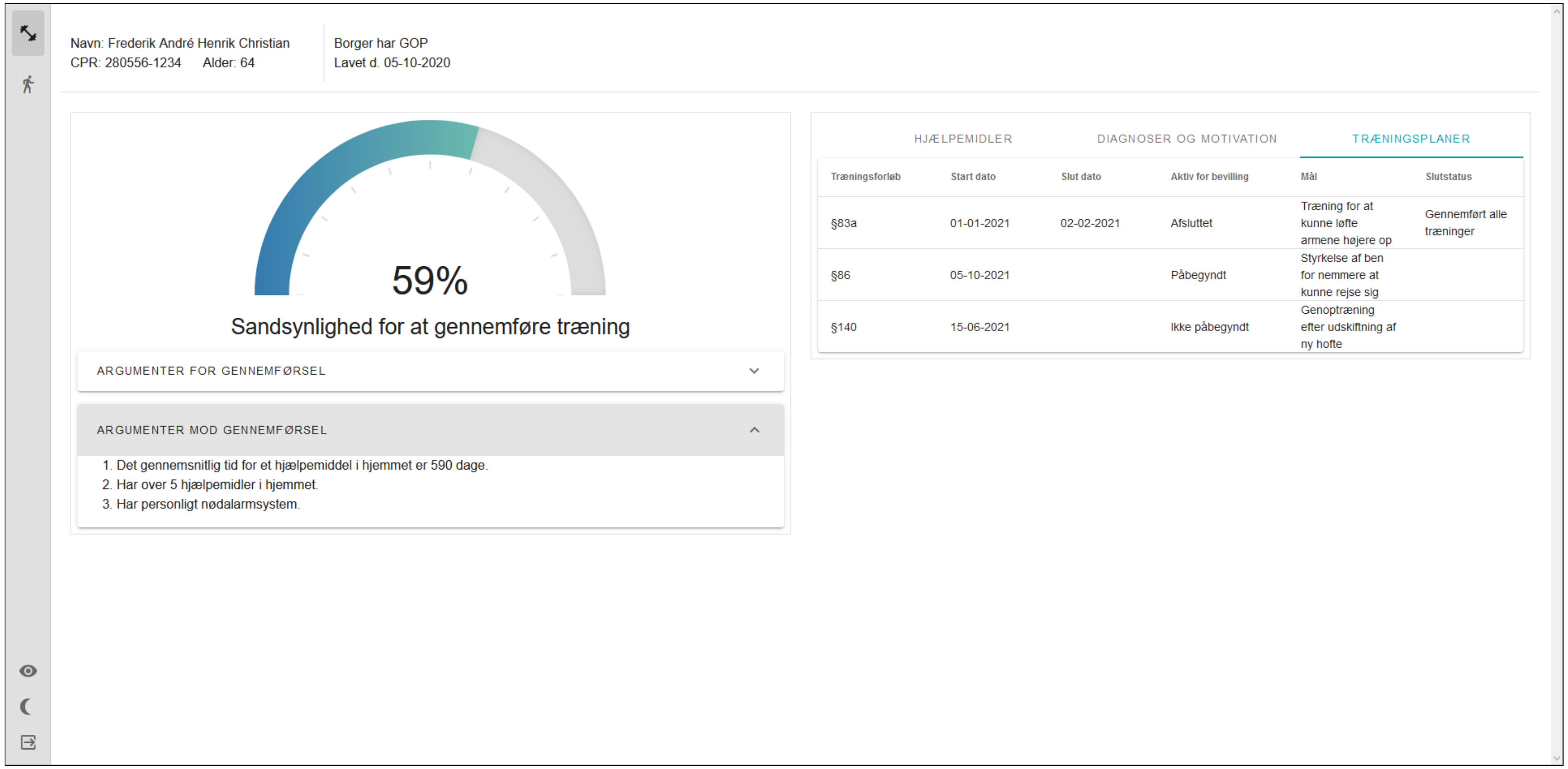

Figure A.17: Prototype Case 1 - Showing arguments against and training plan tab

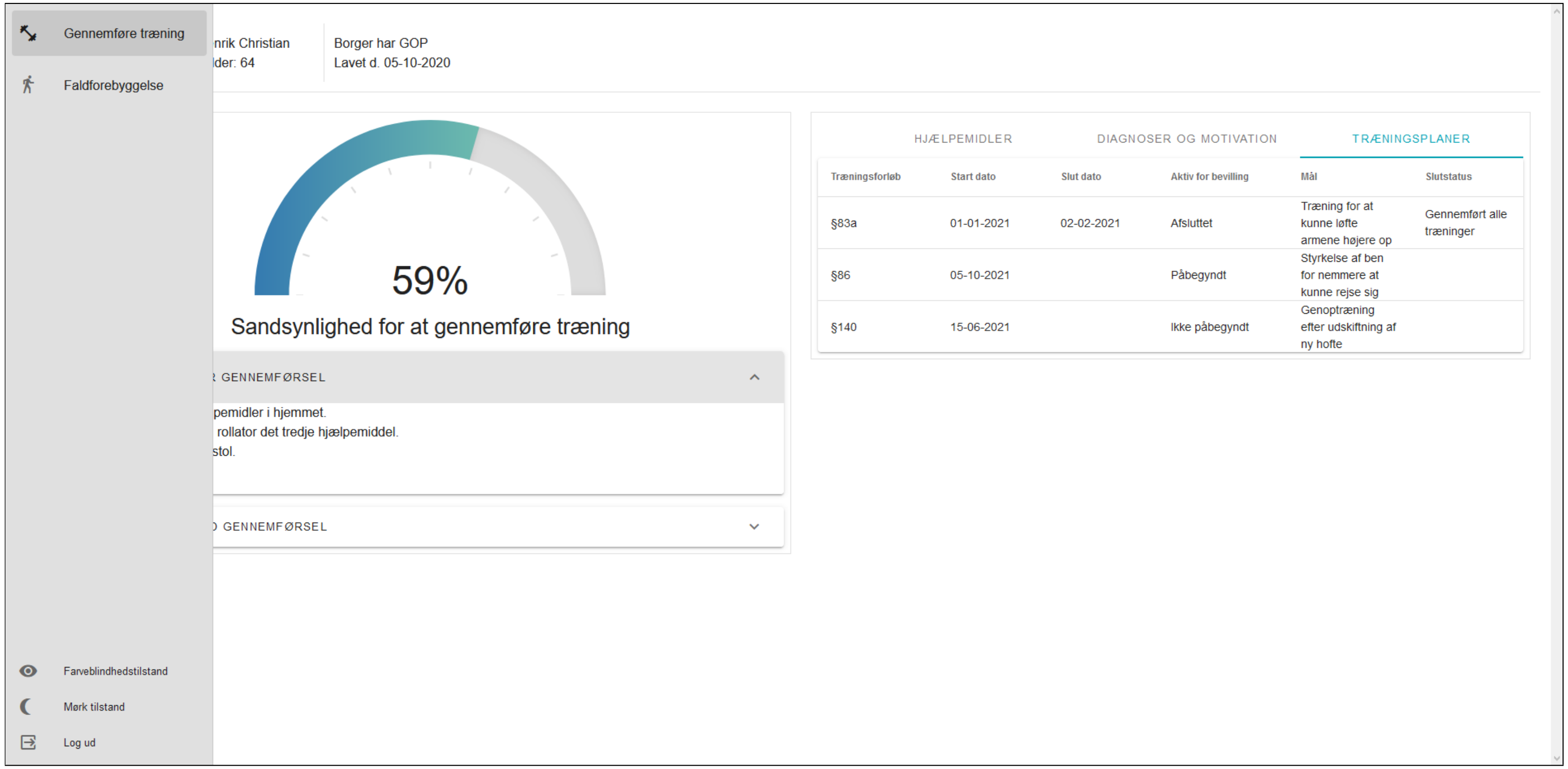

Figure A.18: Prototype Case 1 - Showing arguments for, the navigation drawer and the training plan tab

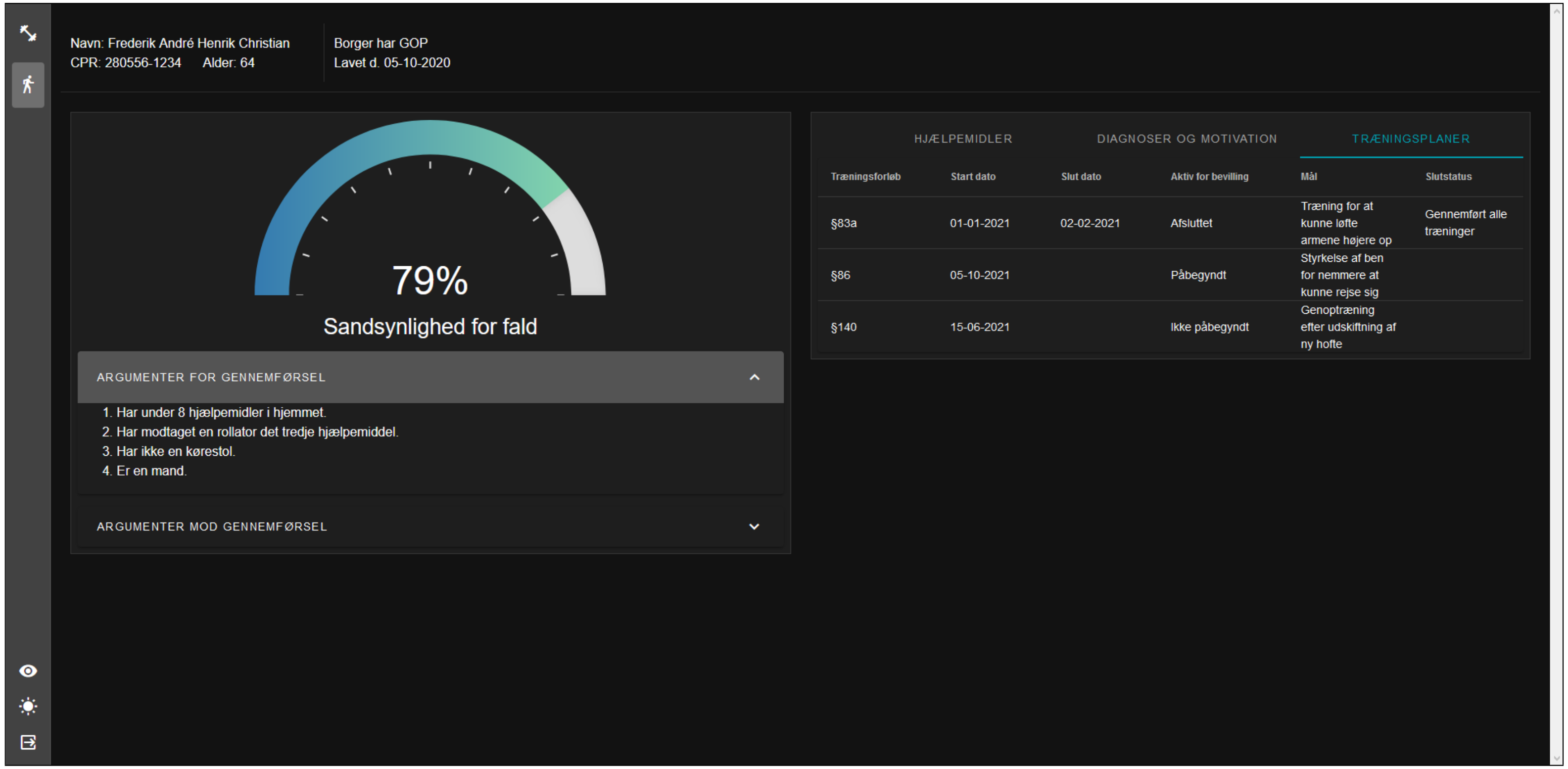

Figure A.19: Prototype Case 2 - Showing dark mode, arguments for and the training plans tab

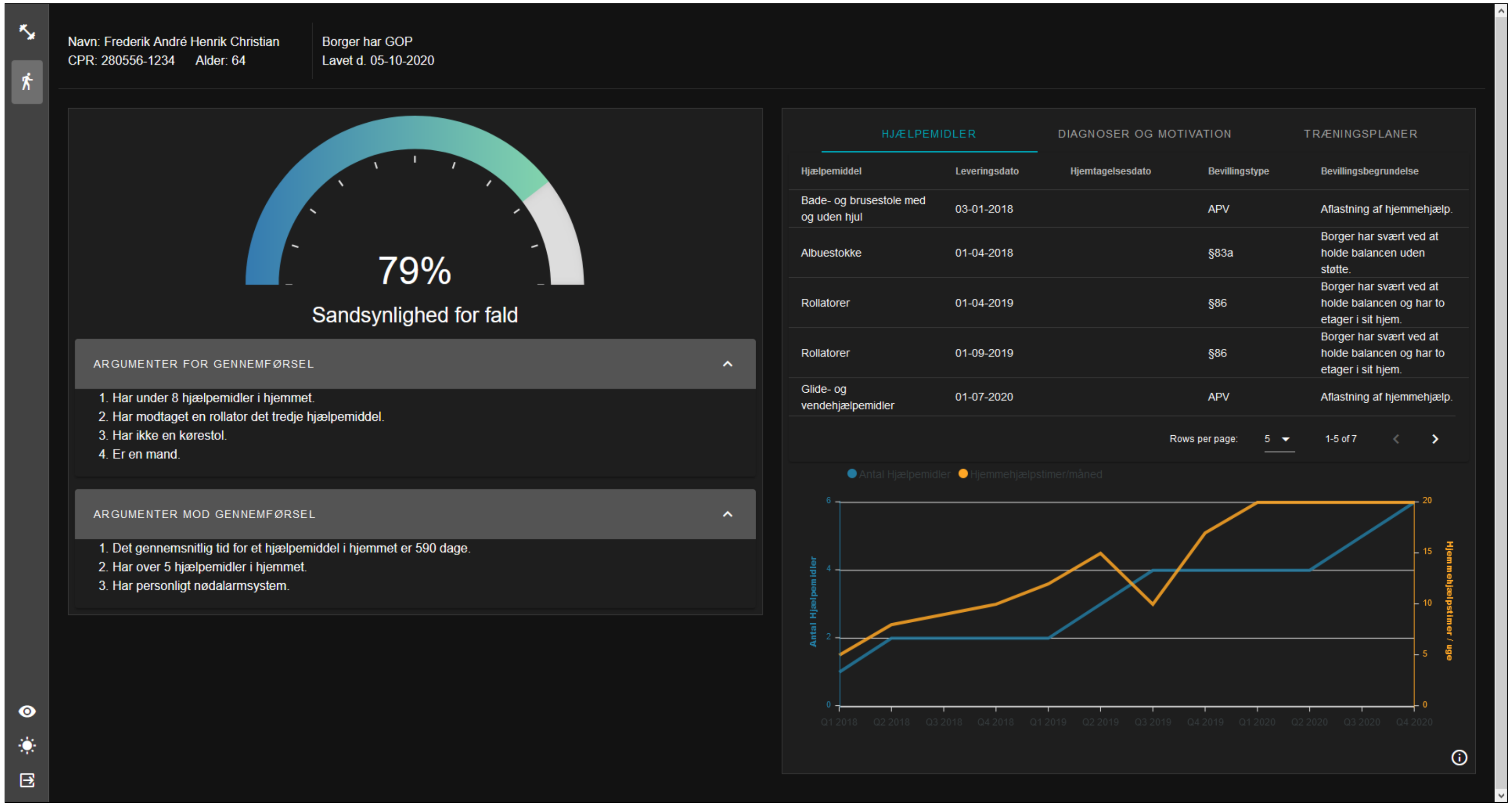

Figure A.20: Prototype Case 2 - Showing dark mode, arguments for and against, and the assistive aids tab

A.3.3 Post-task questionnaire

Table A.18: Danish version of the SEQ used during Experiment 2

Task #	Identifier	Task description
1	Log In	Log ind på hjemmesiden. Brug det udfyldte brugernavn og indtast en tilfældig kode.
2	Colorblind	Aktiver farveblindhedstilstand på hjemmesiden.
3	Arguments For	Find et argument der tæller for at en borger kan gennemføre et træningsforløb.
4	Arguments Against	Find et argument der tæller imod at en borger kan gennemføre et træningsforløb.
5	Hide Plot	Skjul en kurve i grafen der viser hjælpemidler og hjemmehjælpstimer over tid.
6	Training Plans	Find borgerens træningsplaner.
7	Assistive Aids	Find borgerens hjælpemidler.
8	Diagnoses	Find borgerens diagnoser.
9	Dark Mode	Aktiver mørk-tilstand på hjemmesiden.
10	Fall Prevention	Find sandsynligheden der viser hvorvidt en borger bør modtage faldforebyggende træning.
11	Log Out	Log ud af hjemmesiden.

Table A.19: SEQ answers for the individual tasks and users during Experiment 2.

User / Task	User 1 [1;7]	User 2 [1;7]	User 3 [1;7]	User 4 [1;7]	User 5 [1;7]	Mean [1;7]	UB [1;7]	LB [1;7]
Log In	7	7	7	5	7	6,6	7,0	5,5
Colorblind	7	7	7	7	7	7,0	7,0	7,0
Arguments For	7	7	7	7	7	7,0	7,0	7,0
Arguments Against	7	7	7	7	7	7,0	7,0	7,0
Hide Plot	7	7	4	4	6	5,6	7,0	3,7
Training Plans	7	7	7	7	7	7,0	7,0	7,0
Assistive Aids	7	7	7	7	7	7,0	7,0	7,0
Diagnoses	7	7	7	7	7	7,0	7,0	7,0
Dark Mode	7	7	7	7	7	7,0	7,0	7,0
Fall Prevention	7	7	4	7	4	5,8	7,0	3,8
Log Out	7	7	7	7	7	7,0	7,0	7,0

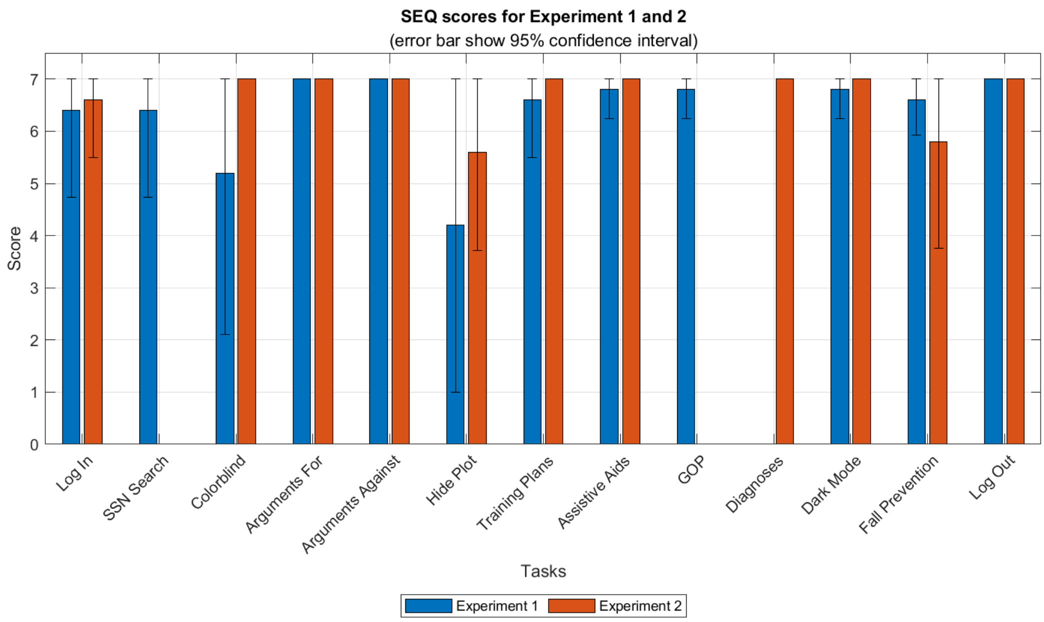

Figure A.21: SEQ answers represented on a graph for Experiment 2

A.3.4 Post-session questionnaire

Table A.20: CSUQ metric data for Experiment 2

Statement \ User	User 1 [1;7]	User 2 [1;7]	User 3 [1;7]	User 4 [1;7]	User 5 [1;7]
1	1	1	1	1	1
2	1	1	1	2	1
3	1	1	1	2	2
4	1	1	1	1	2
5	1	1	1	1	2
6	1	NA	4	2	2
7	4	NA	4	NA	NA
8	4	NA	NA	NA	3
9	1	1	NA	1	1
10	1	1	2	1	1
11	2	1	NA	1	2
12	1	1	1	2	1
13	1	1	2	1	3
14	1	1	2	1	2
15	1	1	2	2	3
16	1	1	1	2	2
17	1	1	3	1	2

Score \ User	User 1 [1;7]	User 2 [1;7]	User 3 [1;7]	User 4 [1;7]	User 5 [1;7]	Mean [1;7]	UB [1;7]	LB [1;7]
Overall	1,4	1,0	1,8	1,4	1,9	1,5	1,9	1,0
SysQual	1,0	1,0	1,5	1,5	1,7	1,3	1,7	1,0
InfoQual	2,2	1,0	2,3	1,3	1,6	1,7	2,4	1,0
IntQual	1,0	1,0	2,0	1,3	2,7	1,6	2,5	1,0
Completeness	100%	81%	81%	88%	94%	88,8%		

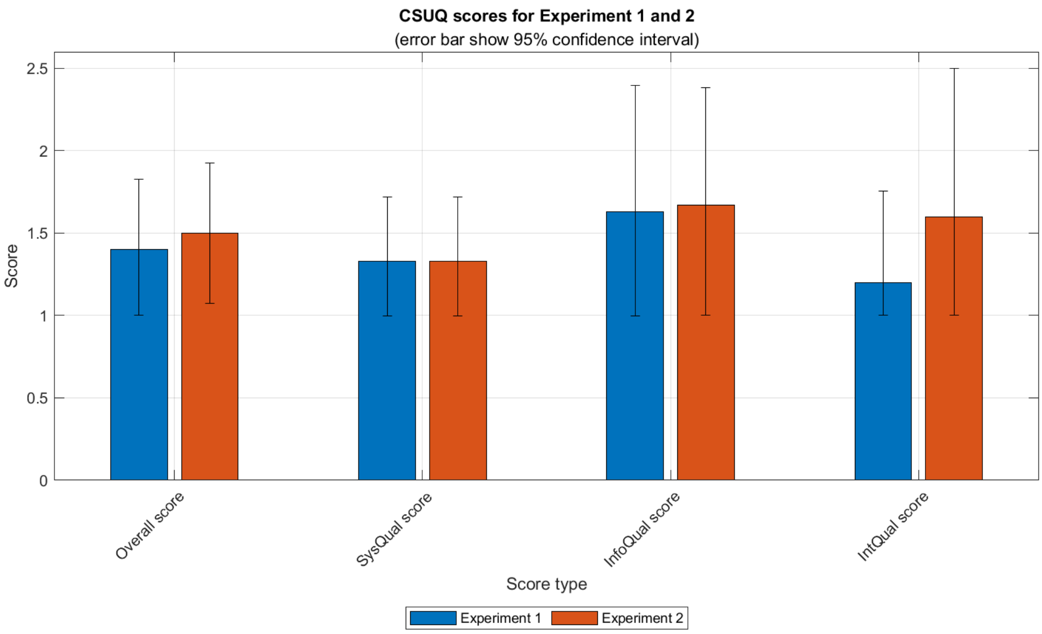

Figure A.22: Mean CSUQ scores for Experiment 2

A.3.5 SUS/UMUX

Users står som de er skrevet ind i Excel og ikke nødvendigvis som den rækkefølge de havde til experiment 1. Undersøg dette og vær sikker på rækkefølgen stemmer overens. Gør det samme for nogle af de andre tabeller

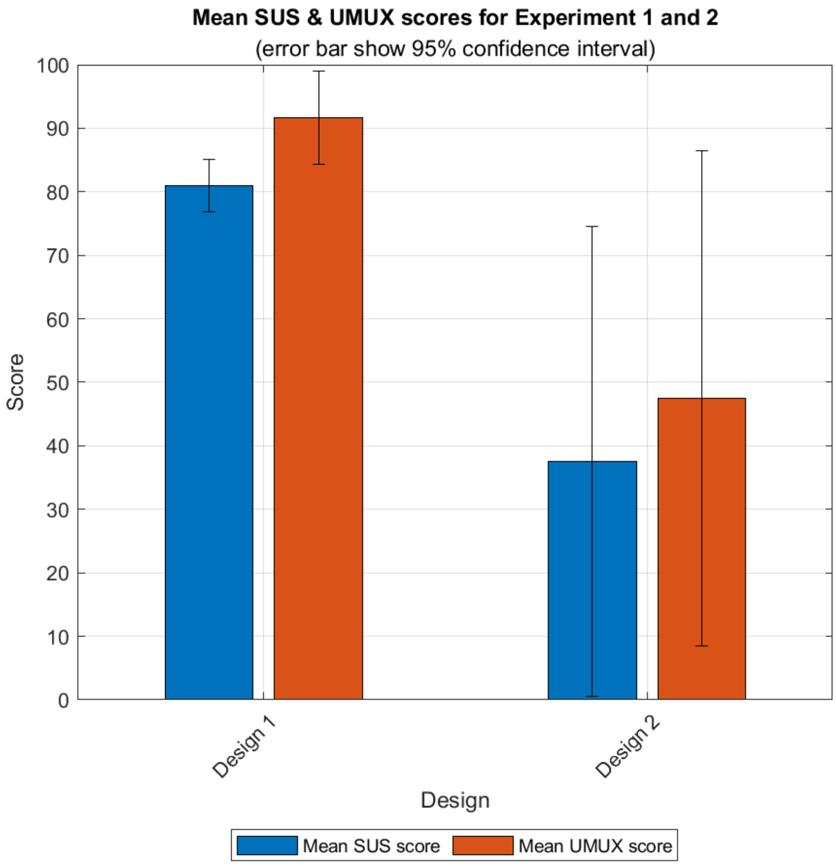

Figure A.23: The mean scores for the SUS and UMUX questionnaires and the 95% confidence interval

Table A.21: Danish SUS questionnaire given to participants during experiment 2

Statement #	Statement
1	Hvad er det seneste design har du kigget på inden dette spørgeskema?
2	Jeg tror, at jeg gerne vil kunne bruge sandynlighedsargumenterne på hjemmesiden ofte.
3	Jeg fandt sandynlighedsargumenterne på hjemmesiden unødvendigt kompliceret.
4	Jeg synes at sandynlighedsargumenterne på hjemmesiden var let at bruge.
5	Jeg tror jeg ville have brug for teknisk support for at bruge sandynlighedsargumenterne på hjemmesiden.
6	Jeg fandt funktionaliteten af sandynlighedsargumenterne på hjemmesiden var godt integreret.
7	Jeg synes at sandynlighedsargumenterne på hjemmesiden var for inkonsistent.
8	Jeg kunne forestille mig de fleste personer hurtigt vil kunne lære at anvende sandynlighedsargumenterne på hjemmesiden meget hurtig.
9	Jeg fandt sandynlighedsargumenterne på hjemmesiden meget klodset at bruge.
10	Jeg følte mig meget selvsikker da jeg brugte sandynlighedsargumenterne på hjemmesiden.
11	Jeg havde brug for at lære en masse ting før jeg kunne komme i gang med at bruge sandynlighedsargumenterne på hjemmesiden.

Table A.22: SUS metric data for Experiment 2

Statement \ User	User 1 [1;5]	User 2 [1;5]	User 3 [1;5]	User 4 [1;5]	User 5 [1;5]	Mean [1;5]	UB [1;5]	LB [1;5]
1	Design 1	Design 1	Design 1	Design 1	Design 1			
2	3	3	4	4	4			
3	2	1	2	2	2			
4	4	4	4	4	5			
5	1	1	1	1	1			
6	5	5	4	4	5			
7	3	3	2	3	3			
8	5	5	5	4	5			
9	1	1	1	3	2			
10	4	3	3	4	3			
11	1	1	1	1	1			
Mean	82,5	82,5	82,5	75,0	82,5	81,0	85,2	76,8

Table A.23: SUS metric data for Experiment 2

Statement \ User	User 1 [1;5]	User 2 [1;5]	User 3 [1;5]	User 4 [1;5]	User 5 [1;5]	Mean [1;5]	UB [1;5]	LB [1;5]
1	Design 2	Design 2	Design 2	Design 2	Design 2			
2	1	2	2	1	4			
3	5	5	4	5	2			
4	1	4	3	1	4			
5	4	4	1	3	1			
6	2	5	2	2	5			
7	3	3	5	3	3			
8	1	4	2	2	5			
9	5	4	4	5	1			
10	1	2	3	1	4			
11	5	4	4	5	1			
Mean	10,0	42,5	35,0	15,0	85,0	37,5	74,5	0,5

Table A.24: Danish UMUX questionnaire given to participants experiment 2

Statement #	Statement
1	Hvad er det seneste design har du kigget på inden dette spørgeskema?
2	Hjemmesidens sandynlighedsargumenters funktionalitet imødekommer mine krav.
3	At bruge sandynlighedsargumenterne på hjemmesiden er en frustrerende oplevelse.
4	Sandynlighedsargumenterne på hjemmesiden er let at bruge.
5	Jeg er nødt til at bruge for meget tid på at rette ting når jeg bruger sandynlighedsargumenterne på hjemmesiden.

Table A.25: UMUX metric data for Experiment 2

Statement \ User	User 1 [1;7]	User 2 [1;7]	User 3 [1;7]	User 4 [1;7]	User 5 [1;7]	Mean [1;7]	UB [1;7]	LB [1;7]
1	Design 1	Design 1	Design 1	Design 1	Design 1			
2	2	2	2	2	2			
3	7	7	7	6	7			
4	1	1	1	2	2			
5	7	7	7	6	6			
Mean	95,8	95,8	95,8	83,3	87,5	91,7	99,0	84,4

Table A.26: UMUX metric data for Experiment 2

Statement \ User	User 1 [1;7]	User 2 [1;7]	User 3 [1;7]	User 4 [1;7]	User 5 [1;7]	Mean [1;7]	UB [1;7]	LB [1;7]
1	Design 2	Design 2	Design 2	Design 2	Design 2			
2	6	3	5	4	3			
3	1	5	4	1	6			
4	7	3	5	6	1			
5	4	7	5	1	6			
Mean	16,7	75,0	45,8	16,7	83,3	47,5	86,5	8,5

A.3.6 Task success

Table A.27: Task success for Experiment 2 - Completed tasks are marked with (✓) and failed tasks are marked with (✗)

Task \ User	User 1	User 2	User 3	User 4	User 5	Mean
Log In	✓	✗	✓	✓	✓	80%
Colorblind	✓	✓	✓	✓	✓	100%
Arguments For	✓	✓	✓	✓	✓	100%
Arguments Against	✗	✓	✓	✓	✓	80%
Hide Plot	✗	✓	✓	✓	✓	80%
Training Plans	✗	✓	✓	✓	✓	80%
Assistive Aids	✓	✓	✓	✓	✓	100%
Diagnoses	✓	✓	✓	✓	✓	100%
Dark Mode	✗	✓	✗	✓	✓	60%
Fall Prevention	✗	✓	✗	✓	✓	60%
Log Out	✗	✓	✓	✓	✓	80%

Table A.28: Received task assistance for Experiment 2 - Tasks with received assistance are marked with (✓) and tasks without needed assistance are marked with (✗)

Task \ User	User 1	User 2	User 3	User 4	User 5	Mean
Log In	✗	✗	✗	✗	✗	0%
Colorblind	✗	✗	✗	✗	✗	0%
Arguments For	✗	✗	✗	✗	✗	0%
Arguments Against	✗	✗	✗	✗	✗	0%
Hide Plot	✗	✗	✗	✗	✗	0%
Training Plans	✗	✗	✗	✗	✗	0%
Assistive Aids	✗	✗	✗	✗	✗	0%
Diagnoses	✗	✗	✗	✗	✗	0%
Dark Mode	✗	✗	✗	✗	✗	0%
Fall Prevention	✗	✗	✗	✗	✗	0%
Log Out	✗	✗	✗	✗	✗	0%

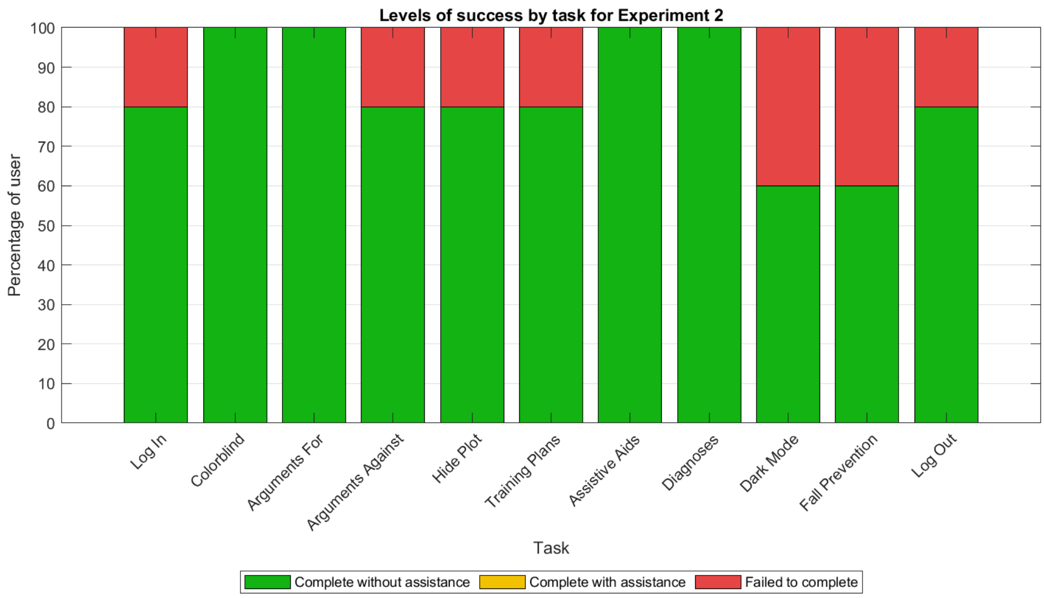

Figure A.24: Percentage of users for each task who completed a task without assistance, with assistance, or failed to complete the task for Experiment 2

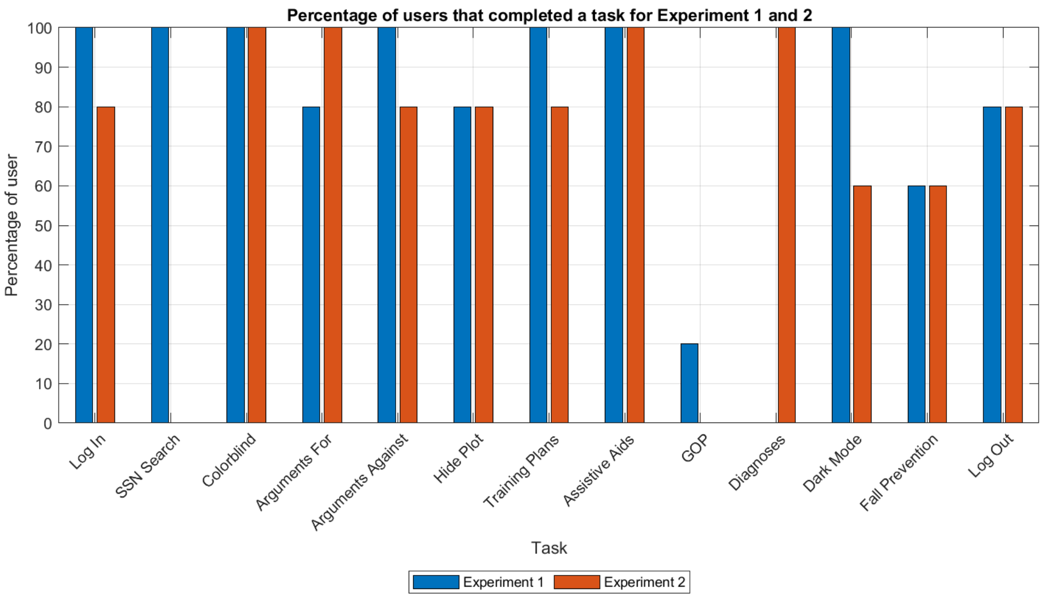

Figure A.25: Percentage of users for each task who completed a task without assistance, with assistance, or failed to complete the task between Experiment 1 and 2

A.3.7 Task time

Table A.29: Task completion time for Experiment 2

Task \ User	User 1 [s]	User 2 [s]	User 3 [s]	User 4 [s]	User 5 [s]	Mean [s]	UB [s]	LB [s]
Log In	6	-	7	16	5	8,5	16,6	0,4
Colorblind	13	9	7	5	6	8,0	11,9	4,1
Arguments For	14	7	14	2	9	9,2	15,5	2,9
Arguments Against	-	4	2	2	3	2,8	4,3	1,2
Hide Plot	-	34	5	4	5	12,0	35,4	0
Training Plans	-	4	2	4	5	3,8	5,8	1,7
Assistive Aids	3	4	3	3	3	3,2	3,8	2,6
Diagnoses	13	2	9	3	3	6,0	12,0	0,0
Dark Mode	-	2	-	1	3	2,0	4,5	0
Fall Prevention	-	28	-	34	14	25,3	50,8	0
Log Out	-	6	3	3	3	3,8	6,1	1,4
Mean	9,8	10,0	5,8	7,0	5,4	7,7		
Mean			7,6					

Table A.30: Total time on task for Experiment 2

Task \ User	User 1 [s]	User 2 [s]	User 3 [s]	User 4 [s]	User 5 [s]	Mean [s]	UB [s]	LB [s]
Log In	29	-	30	20	7	21,5	38,5	4,5
Colorblind	24	78	10	16	13	28,2	63,4	0
Arguments For	22	21	20	5	25	18,6	28,3	8,9
Arguments Against	10	16	4	4	5	7,8	14,3	1,3
Hide Plot	28	48	32	14	16	27,6	44,7	10,5
Training Plans	12	21	5	7	8	10,6	18,5	2,7
Assistive Aids	6	13	5	8	5	7,4	11,6	3,2
Diagnoses	17	12	11	10	6	11,2	16,1	6,3
Dark Mode	4	6	4	9	5	5,6	8,2	3,0
Fall Prevention	3	38	26	44	20	26,2	46,2	6,2
Log Out	2	8	6	5	5	5,2	7,9	2,5
Mean	14,3	26,1	13,9	12,9	10,5	15,4		
Mean			15,5					

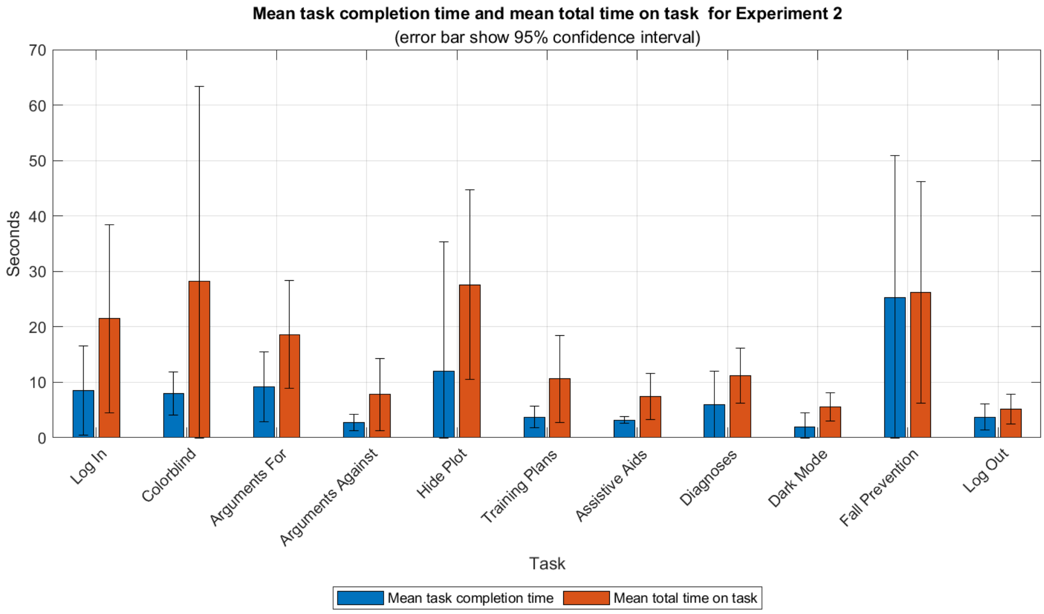

Figure A.26: Mean task completion time and mean total time on task for Experiment 2

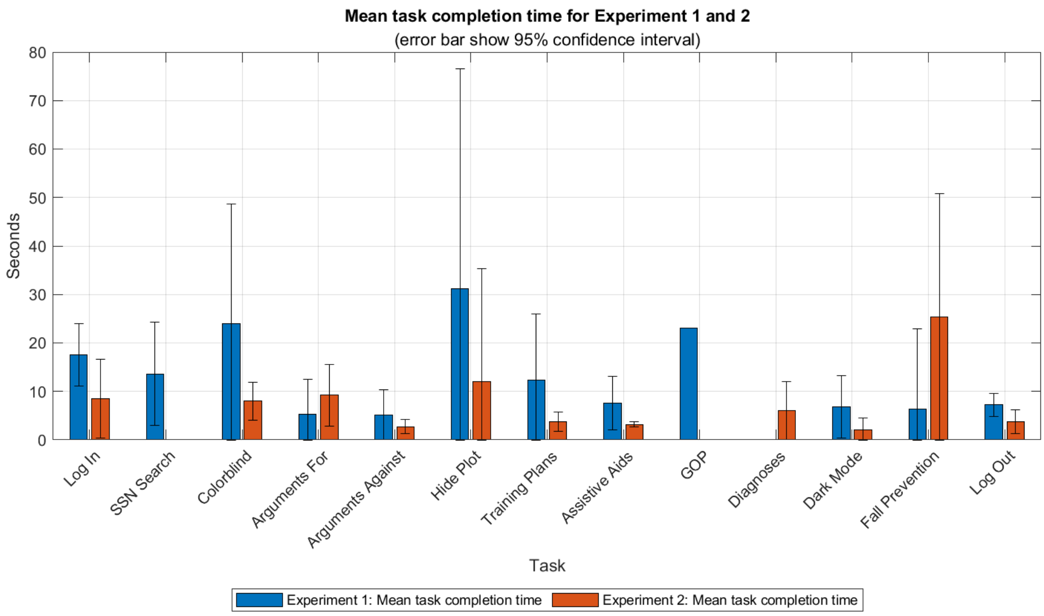

Figure A.27: Mean task completion time for Experiment 1 and 2

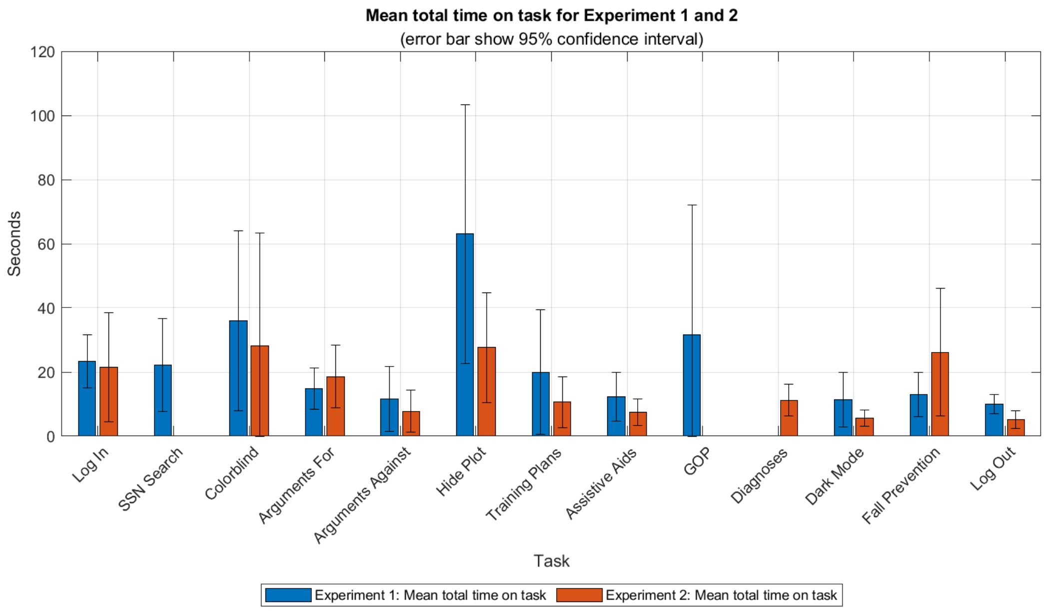

Figure A.28: Mean total time on task for Experiment 1 and 2

A.3.8 Errors

Table A.31: Errors for Experiment 2

Error type \ User	User 1 [#]	User 2 [#]	User 3 [#]	User 4 [#]	User 5 [#]	Sum [#]
Random click error	1	3	0	0	0	4
Menu item error	0	1	0	1	0	2
Menu error	1	0	1	0	1	3
Hide plot error	0	1	0	1	0	2
Show plot error	0	1	0	2	2	5
Right click on graph error	0	0	0	0	0	0
Left click on graph error	1	1	0	0	0	2
Table cell click error	0	1	0	0	0	1
Tabs click error	1	2	6	1	0	10
Wrongly opened Argumentbox error	0	1	0	0	1	2
Wrongly closed Argumentbox error	0	0	0	0	0	0
Cursor drag error	0	0	0	0	0	0
Sum	4	11	7	5	4	31

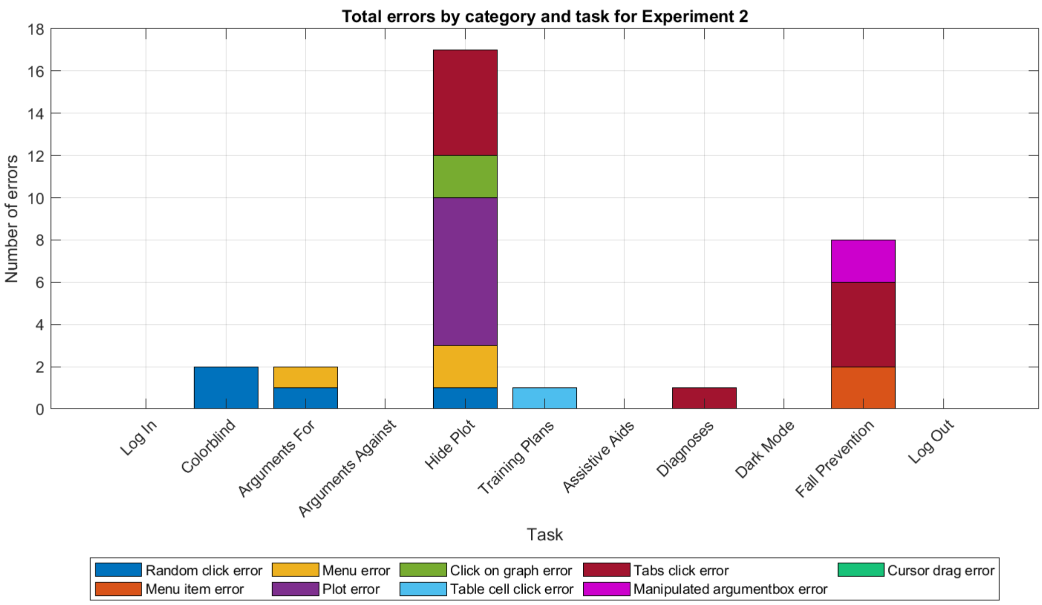

Figure A.29: Total number of errors for Experiment 2

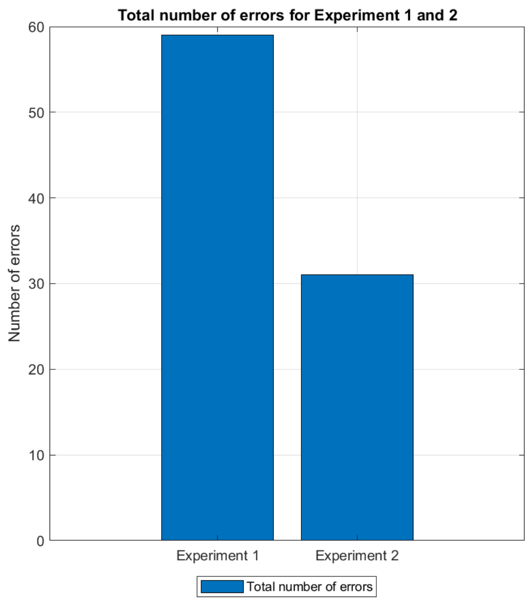

Figure A.30: Total number of errors between Experiment 1 and 2

A.3.9 Efficiency

Table A.32: Efficiency for Experiment 2

Task \ User	User 1 [#]	User 2 [#]	User 3 [#]	User 4 [#]	User 5 [#]	Mean [#]	Expected [#]	UB [#]	LB [#]
Log In	5	0	2	2	2	2,2	2	4,4	0
Colorblind	2	4	2	2	2	2,4	2	3,5	1,3
Arguments For	2	1	1	1	2	1,4	1	2,1	0,7
Arguments Against	0	1	1	1	1	0,8	1	1,4	0,2
Hide Plot	2	4	7	4	5	4,4	1	6,7	2,1
Training Plans	0	2	0	1	1	0,8	1	1,8	0
Assistive Aids	0	1	1	1	1	0,8	1	1,4	0,2
Diagnoses	2	1	1	1	1	1,2	1	1,8	0,6
Dark Mode	1	2	1	2	2	1,6	2	2,3	0,9
Fall Prevention	0	7	2	5	3	3,4	2	6,8	0,0
Log Out	0	2	2	2	2	1,6	2	2,7	0,5
Sum	14	25	20	22	22	20,6			
Mean			20,6						

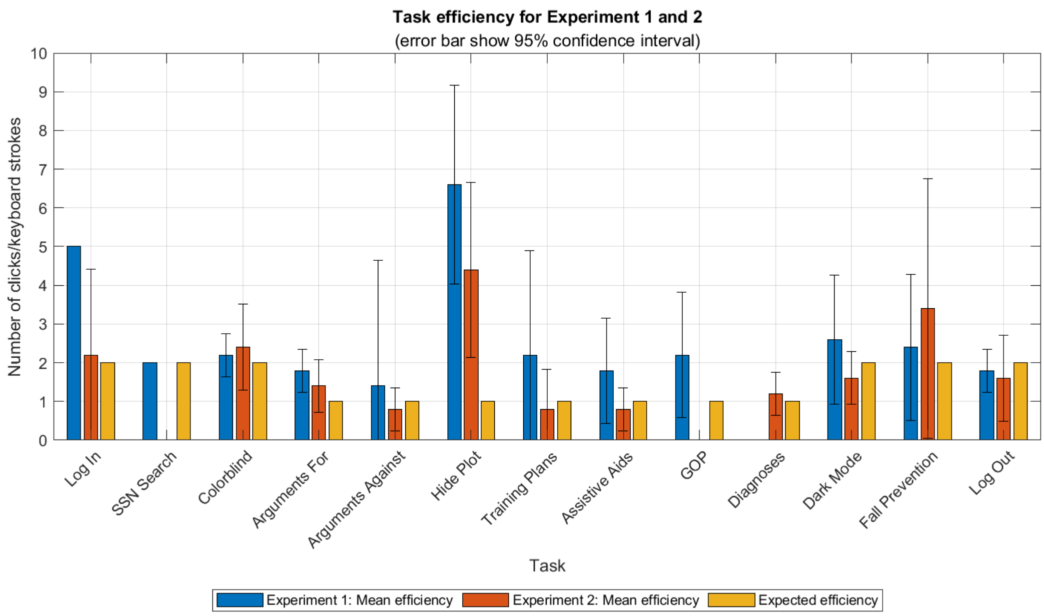

Figure A.31: Efficiency data between Experiment 1 and 2

A.3.10 Learnability

Table A.33: Learnability for Experiment 2

Metric	Measurement
Mean task completion time	7,6
Total number of errors	31
Mean efficiency	20,6
Sum	59,2

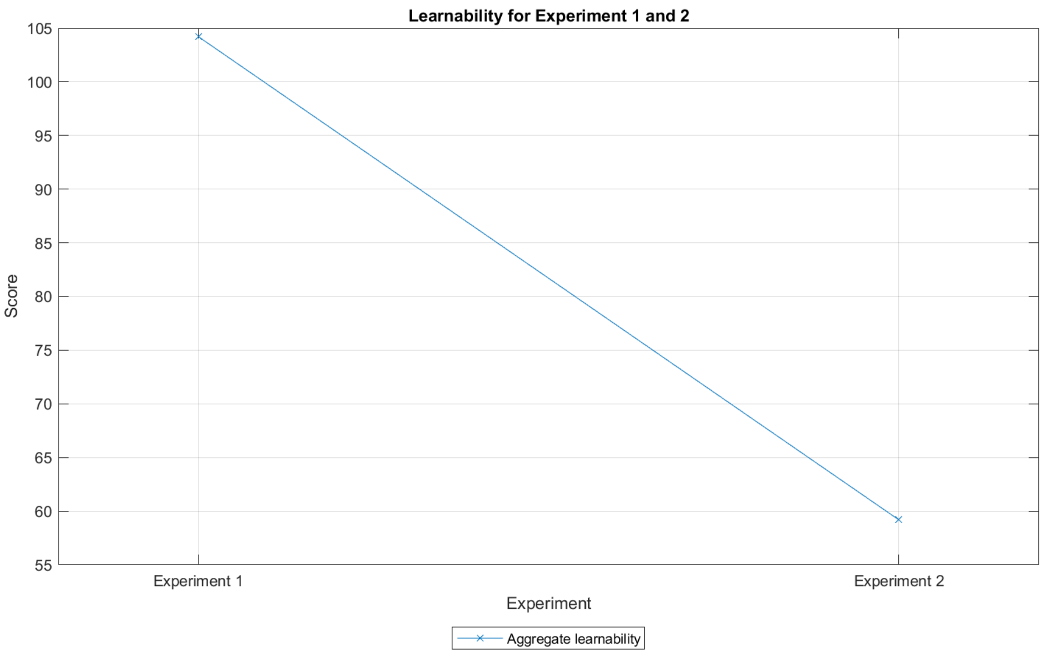

Figure A.32: Learnability between Experiment 1 and 2

A.3.11 Semi-structured Interview

Table A.34: Gathered Interview Feedback during Experiment 2

Id	Feedback
	General
1	Kønsfordelingen på arbejdspladsen består hovedsageligt af kvinder.
2	Hvis man er meget uheldig kan man godt have mere end en GOP.
	Probability and Arguments
3	De visuelle søjler tilknyttet argumenterne er ikke forstyrrende.
4	De visuelle søjler tilknyttet argumenterne er forstyrrende, der skal være et tydeligt overblik.
5	De visuelle søjler er ikke brugbare.
6	Det visuelle i argumentdelen fylder for meget.
7	Der skal forstås mere når de visuelle søjler er der.
8	Jeg ved ikke hvad de visuelle søjler betyder.
9	At få vist vigtigheden bag argumenterne er ligegyldigt.
10	Argumenterne opstillet i en prioriteret liste er så fint.

Continued on next page

Table A.34 – continued from previous page

Id	Feedback
11	Argumenterne skal være korte og præcise.
12	Sandsynligheden og argumenterne vil kunne blive brugt i visse situationer, da en borgeres motivation kan være altafgørende i sig selv.
13	Argumenterne er fine såfremt de er en indikator for om en borger kan gennemføre eller ej.
14	Det er træls hvis argumenterne fylder mere end en linje.
15	Noget der fortæller hvordan argumenterne er prioriteret vil være smart.
16	Jeg vil kunne bruge sandsynligheden og argumenterne i min beslutningsproces, såfremt man kan have tiltro til dem.
17	At have argumenter der vægter for eller imod givet at personen er mand eller kvinde er ikke brugbart.
18	At bruge x antal hjælpemidler som et argument giver god mening.
19	At borgeren ikke har en kørestol er et fint argument.
20	Argumenter er argumenter.
21	Sandsynligheden og argumenterne vil være brugbare i de tilfælde hvor man ikke kender borgeren så godt.
22	At få vist argumenterne i rent tekst er bedst.
23	Det hjælper mig ikke at han har over 5 hjælpemidler eller har et personligt nødhjælpssystem.
24	Der mangler stadigvæk noget kontekst omkring hvorfor argumenterne er som de er.
25	Jeg kunne godt tænke mig at få noget mere at vide omkring på hvilken baggrund sandsynligheden er lavet.
26	Det kan have en positiv indvirkning i mit arbejde.
27	Jeg er i tvivl om jeg kan bruge sandsynligheden og argumenterne for sig selv, men de kan helt bestemt give mig en pejling.
28	Der må være nogle flere argumenter der spiller ind, end dem der står her nu.
29	Jeg vil være glad hvis sandsynlighedsmåleren gik fra rød til grøn.
30	Farven på sandsynlighedsmåleren er lidt blåt i blåt.
	Assistive Aids
31	Der skal kun vises et hjælpemiddel for hver type.
32	Hvis borgeren har to eller flere af samme type hjælpemiddel, så skal det vises i tabellen.

Continued on next page

Table A.34 – continued from previous page	
Id	**Feedback**
33	Der plejer at gå langt tid imellem man får det første hjælpemiddel af en slags til det næste af samme slags.
34	Hjælpemidler såsom puder, bestik, sejl, osv. Har ingen relevans.
35	Jeg vil se alle borgerens hjælpemidler, så jeg kan få et fuldt overblik.
36	Hvert hjælpemiddel i tabellen skal stemme overens med det som bliver vist på grafen.
37	At inkluder små hjælpemidler er ligegyldigt.
38	Kun de tunge hjælpemidler er vigtige at inddrage.
39	Et filter hvor man selv kan vælge hvilke hjælpemiddelstyper man er interesseret vil være en god ide.
40	Det er bedst at vise kvartaler hen af grafens x-akse.
41	At vist år på grafens x-aksen er for voldsomt og uger giver ingen mening.
42	At vise måneder på grafens x-aksen er for meget.
43	At vise måneder på grafens x-aksen kunne måske godt bruges.
44	Måske kunne hvert halve år på grafens x-akse være fint.
45	Vis kun aktive hjælpemidler, da listen ellers vil blive alt for lang.
46	Man skal ikke vise de hjælpemidler borgeren har haft fra starten af på grafen, kun udviklingen her og nu er vigtig.
47	Man skal kunne se hvilke hjælpemidler borgeren har haft fra starten af på grafen.
48	Hjælpemidlerne i tabellen skal være tydeligere.
49	Jeg vil gerne kunne se hjælpemidlerne over et bestemt tidsrum.
	Training Plans
50	Det vil være en god ide at vise slutstatus på tabellen.
51	Slutstatus er hvor langt borgeren er kommet i forhold til de mål der er sat.
52	Slutstatus fortæller om borgeren har nået sine mål, og hvad der har været årsagen til at målet ikke er blevet nået.
53	Slutstatus bliver altid lavet.
54	Slutstatus er prosa tekst og svært at få med i en tabel.
55	Der skal findes en løsning på at vise slutstatus i prosa tekst uden den bliver direkte vist i tabellen.
56	At kalde fanen for træningsplaner er nok ikke det rette ord at bruge, men måske snarere træningsforløb.

Continued on next page

Id	Feedback
	Table A.34 – continued from previous page
57	Et træningsforløb kan være aktiv, påbegyndt eller ikke afsluttet.
	Case 2: Fall Preventive Training
58	Jeg vil gerne have vist et overblik over borgerens registeret fald ligesom i Cura.
59	Måske kunne fald informationen blive placeret som en del af træningsplaner eller hjælpemidler, men jeg er ikke helt sikker.
60	Årsag til fald og dato for hvornår faldet fandt sted virker brugbart at tage med, og kunne placeres i en tabel.
	Navigation Drawer (Menu)
61	Farven er fin (I don't care).
62	Den er nem at få øje på.
63	Menuen tog lidt tid at finde.
64	Ikonerne træder ikke nok frem i menuen.
65	Jeg har ikke tænkt over hvorfor menuen er placeret i venstre side.
66	Menuen er meget mørk i mørk, der mangler blikfang.
67	Det er ligesom i Cura.
68	Jeg har slet ikke savnet brugerprofilen, og jo mindre til at forvirre jo bedre.
69	Jeg elsker den blå-grå farve.
70	Menuen er til at få øje på uden at være påtrængende.
	Other
71	Det hedder (beviliget en GOP) og ikke (lavet en GOP).
72	Jeg ved ikke om det er fordi jeg har brugt hjemmesiden før, men det føles mere naturligt end sidst.
73	Det var nemt at huske hvordan hjemmesiden virkede.
74	Hjemmesiden er simpel.
75	Hjemmesiden er overskuelig.
76	Det er rart, at selvom jeg ser på træningsplaner eller hjælpemidler forbliver resten af siden den samme.
77	Det er nemt at se hvor tingene er henne.
78	Hjemmesiden er ganske brugbar, også fordi man kan tænke sig om en ekstra gang inden man visiterer træning.
79	Hjemmesiden er super at bruge som støtte til en bevilling.

Continued on next page

Table A.34 – continued from previous page

Id	Feedback
80	Jeg vil kunne bruge hjemmesiden til at træffe en beslutning, så længe der er tillid til resultatet.
81	Hjemmesiden mangler ikke umiddelbart noget.
82	Hjemmesiden er brugervenlig.
83	Man ved hvor man skal trykke henne.
84	Det er rart at hjemmesiden er dæmpet i farverne.
85	Man skal prøve hjemmesiden af som en del af sit arbejde, før man kan sige om det vil være noget man kan bruge.
86	Jeg kunne ikke helt huske hvordan hjemmesiden fungerede, men den giver noget værdi.
87	Når jeg får en ansøgning, så er hjemmesiden mit ”go to”.
88	Jeg synes der mangler noget på hjemmesiden, men jeg kan ikke sætte ord på hvad det er.
89	Informationen fra Cura sammen med sandsynligheden og argumenterne understøtter hinanden rigtig godt.

A.4 Experiment 3

A.4.1 Changelog

Table A.35: Full Changelog between Experiment 2 and 3

Id	Change description	Reason for change
	Added Functionality	
1	Added a tab for information regarding a citizens registered falls for case 2 (see figure 5.27).	The addition was requested during the semi structured interview (see comment 82 in table A.16 and 58,59 and 60 in table A.34).
2	Added SHAP values, provided through the ML API.	To place each argument correctly, according to the SHAP value as mentioned in section 3.5.
	Modified Functionality	

Continued on next page

Id	Change description	Reason for change
	Table A.35 – continued from previous page	
3	Changed the icon color on the navigation drawer.	During the semi structured interview it was mentioned that the navigation drawer was hard to see, and that the icons was to dark (see comment 63, 64 and 66 in table A.34).
	Removed Functionality	
4	Removed the column “Hjemtagelsesdato” in the table listing the citizens assistive aids in the tab ”Hjælpemidler”.	Many explained in the semi structured interview, that showing both active and inactive assistive aids would make the list in the table too long (see comment 45 in table A.34).
5	Removed the Color Blind Mode from the Navigation Drawer.	Research showed (see section 3.4.4) that the color scheme chosen in the prototype already supported people suffering from red-green color blindness.

A.4.2 Prototype design

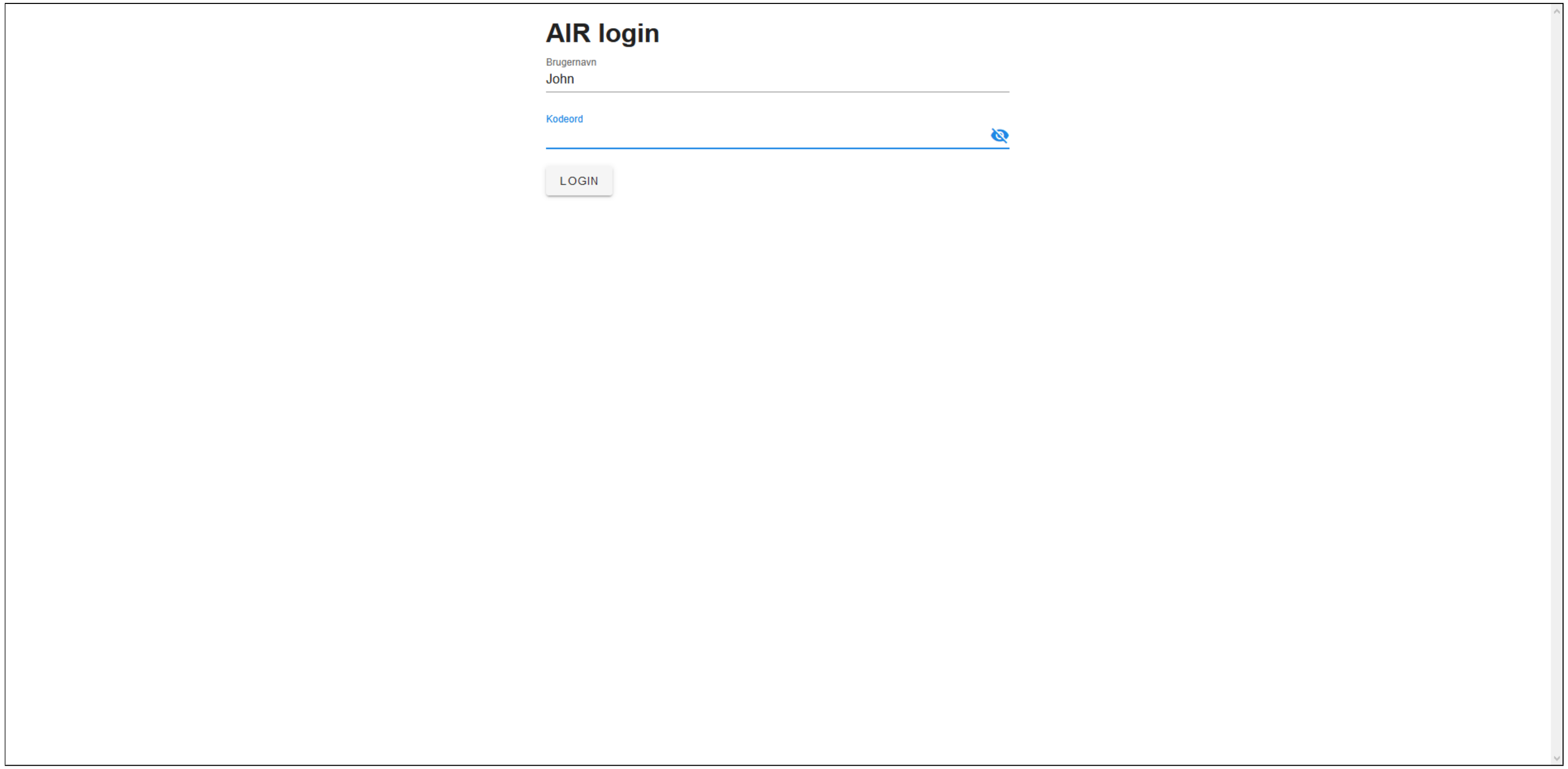

Figure A.33: Prototype login page

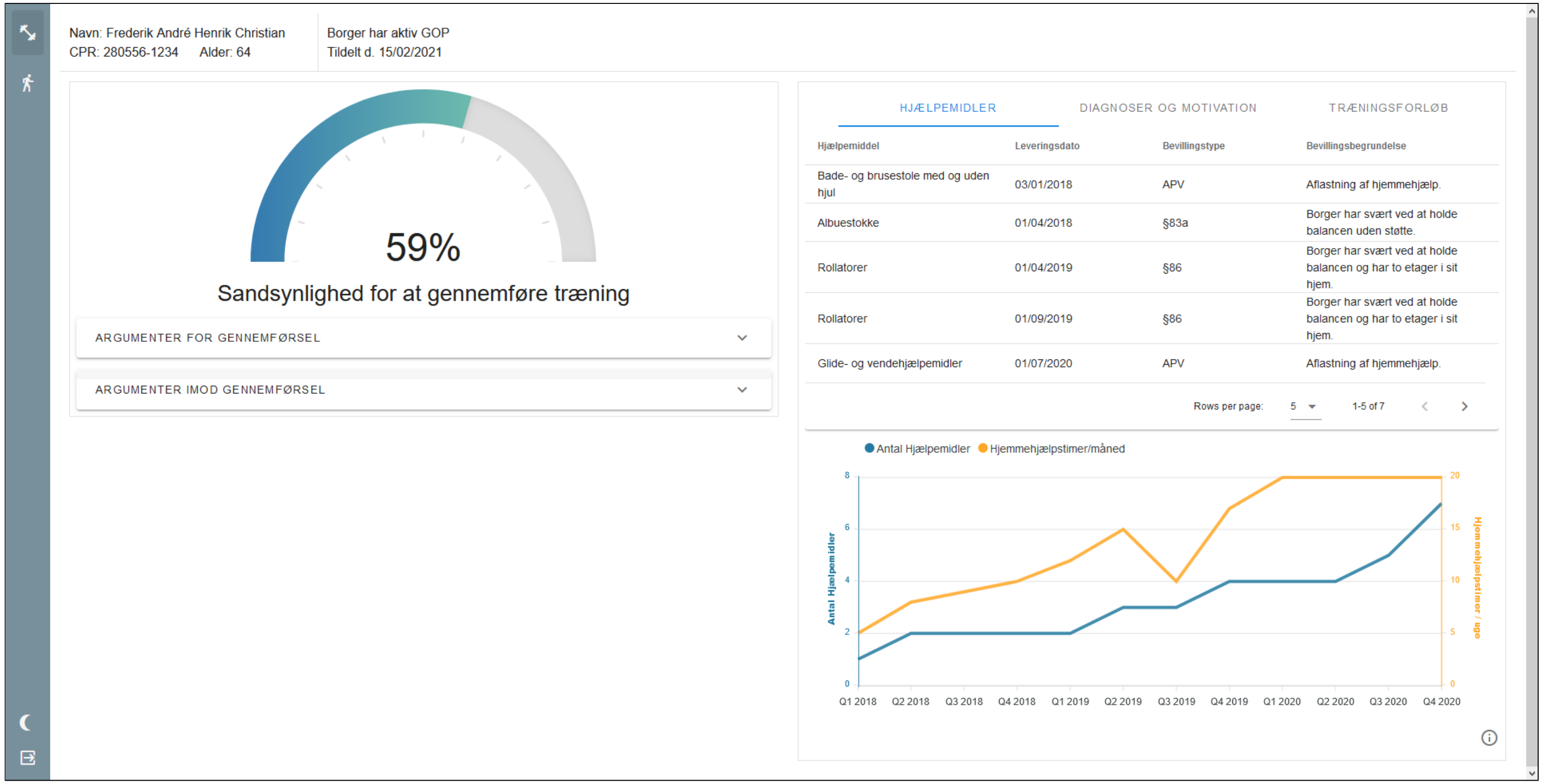

Figure A.34: Prototype Case 1 - Showing assistive aids tab

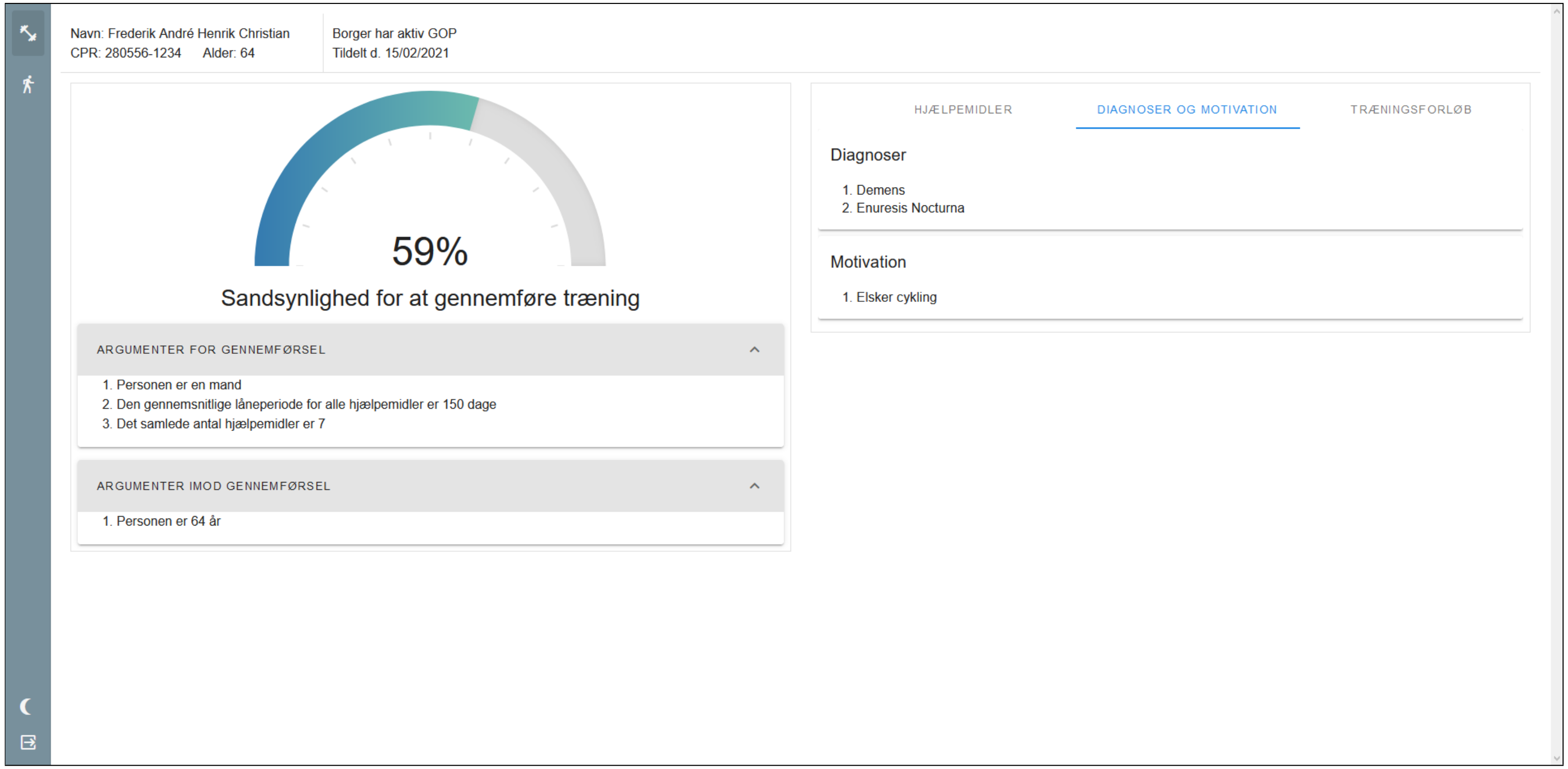

Figure A.35: Prototype Case 1 - Showing arguments for and against and the diagnosis and motivation tab

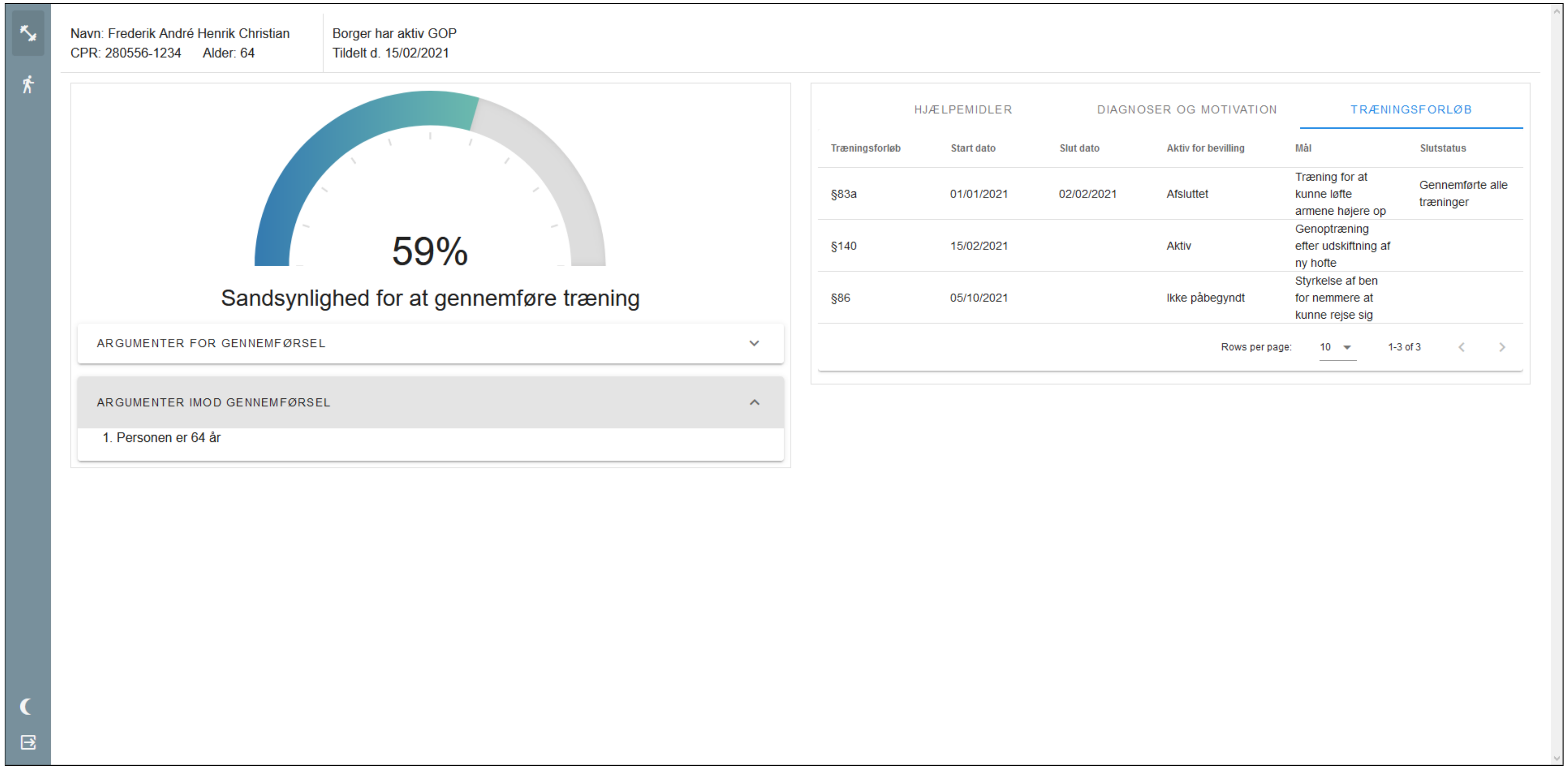

Figure A.36: Prototype Case 1 - Showing arguments against and training plan tab

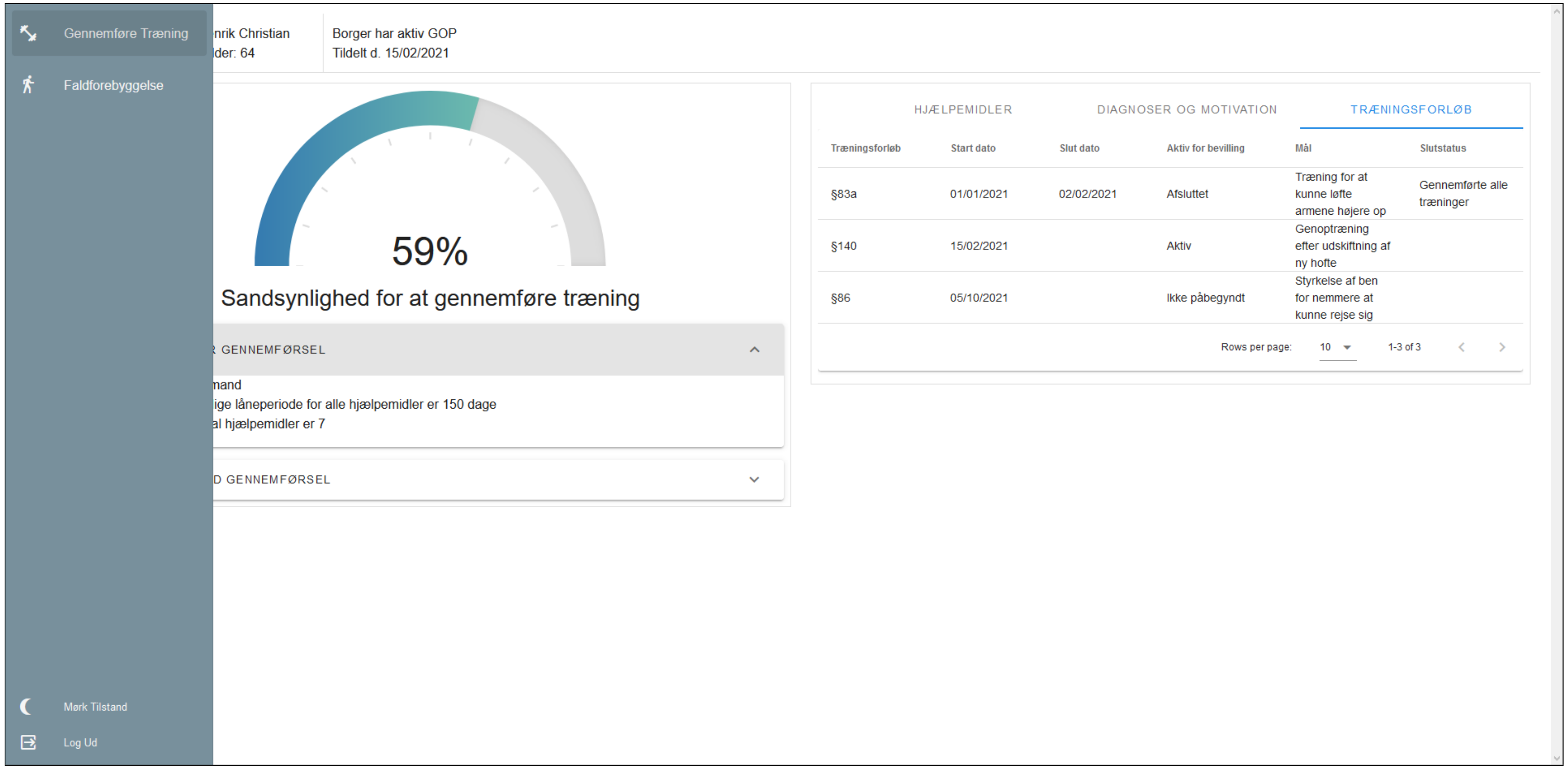

Figure A.37: Prototype Case 1 - Showing arguments for, the navigation drawer and the training plan tab

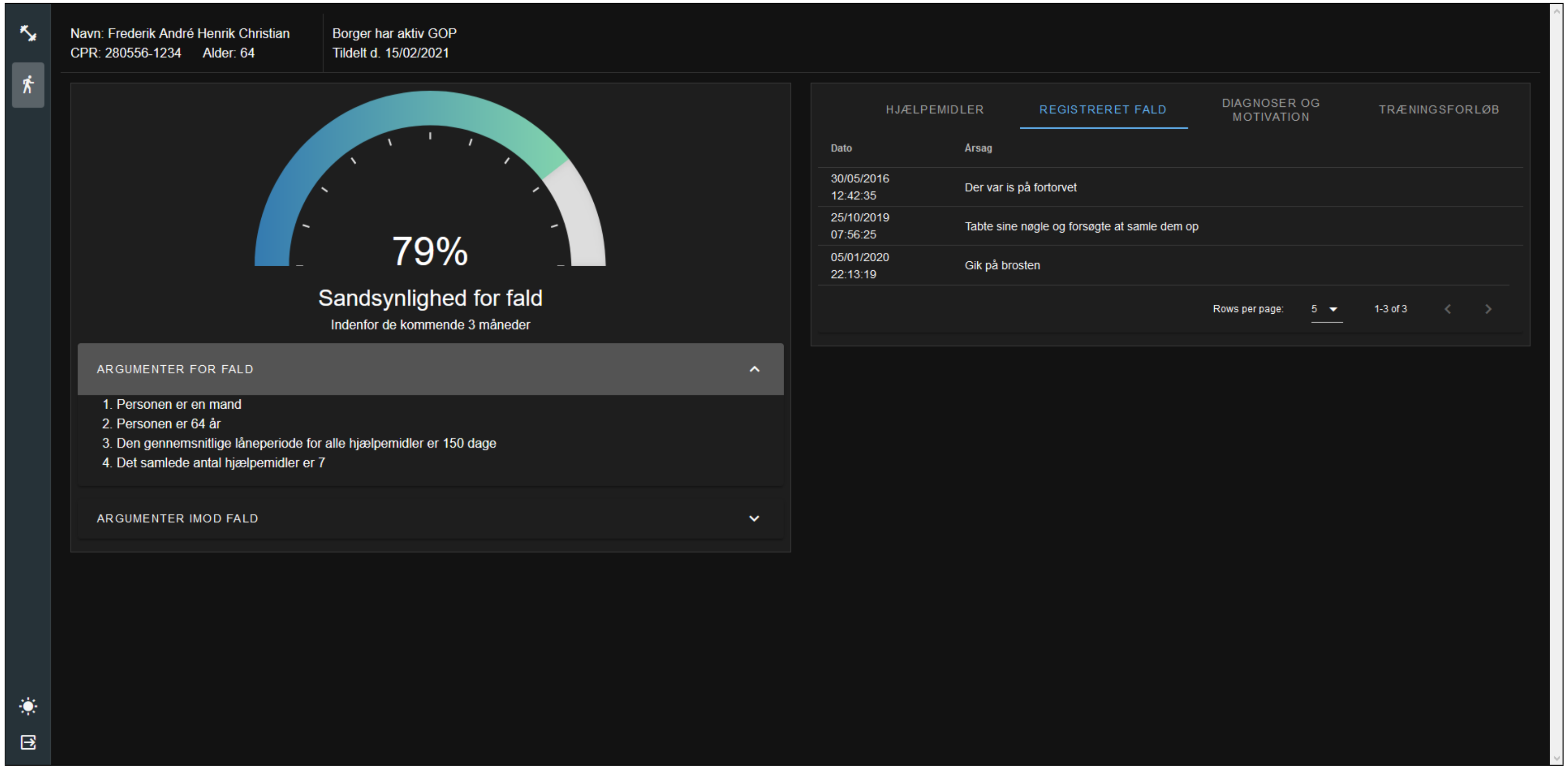

Figure A.38: Prototype Case 2 - Showing dark mode, arguments for and the registered falls tab

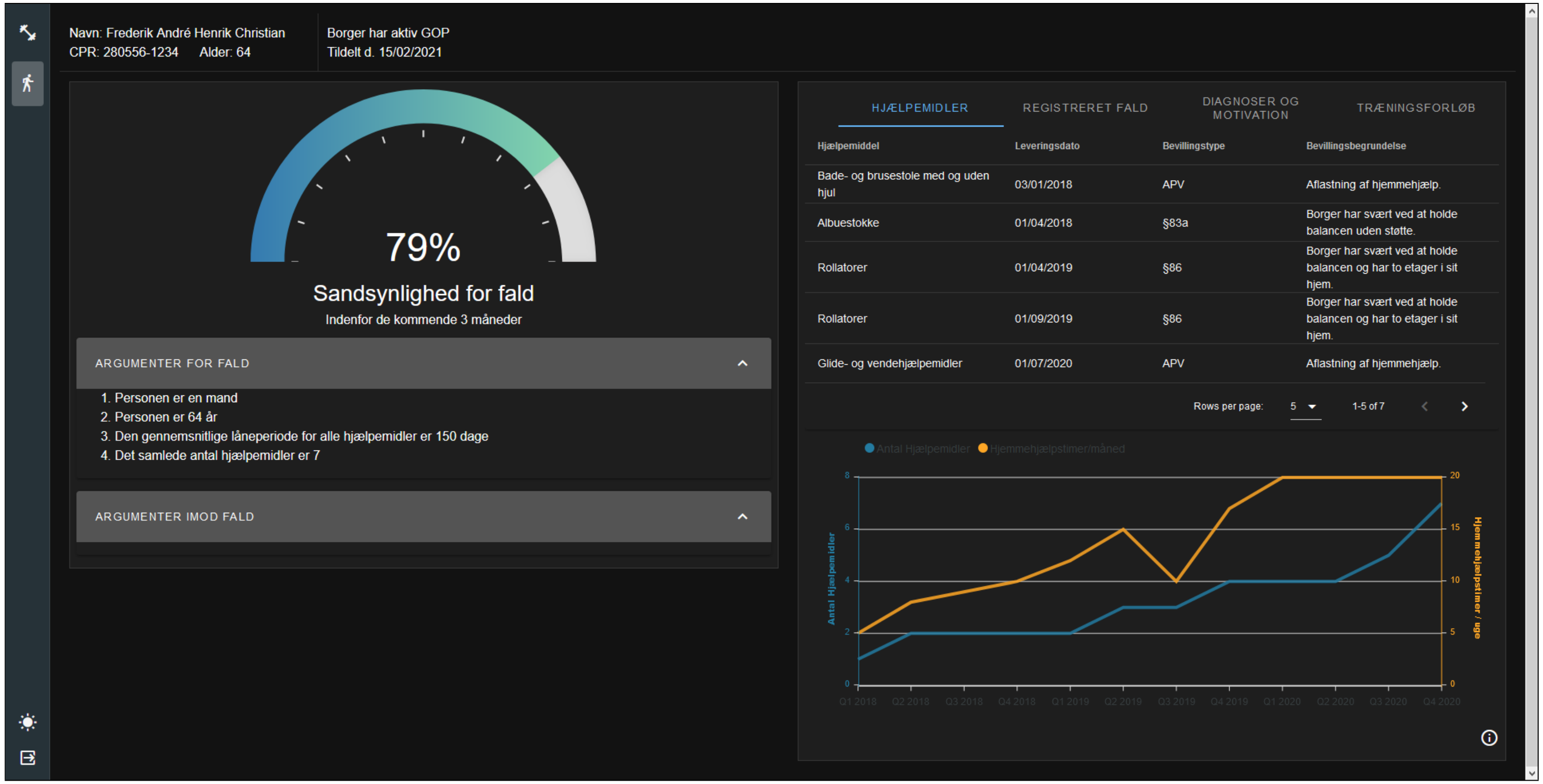

Figure A.39: Prototype Case 2 - Showing dark mode, arguments for and against, and the assistive aids tab

A.4.3 Post-task questionnaire

Table A.36: Danish version of the SEQ used during Experiment 3

Task #	Identifier	Task description
1	Log In	Log ind på hjemmesiden. Brug det allerede udfyldte brugernavn og indtast en tilfældig kode.
2	Arguments For	Find et argument der tæller for at en borger kan gennemføre et træningsforløb.
3	Arguments Against	Find et argument der tæller imod at en borger kan gennemføre et træningsforløb.
4	Hide Plot	Skjul én kurve i grafen der viser hjælpemidler og hjemmehjælpstimer over tid.
5	Training Plans	Find borgerens træningsplaner.
6	Assistive Aids	Find borgerens hjælpemidler.
7	Diagnoses	Find borgerens diagnoser og motivation
8	Dark Mode	Aktiver mørk-tilstand på hjemmesiden.
9	Fall Prevention	Find sandsynligheden der viser borgerens fald risiko indenfor de kommende 3 måneder.
10	Registered Falls	Find borgerens registreret fald.
11	Log Out	Log ud af hjemmesiden.

Table A.37: SEQ answers for the individual tasks and users during Experiment 3

Task \ User	User 1 [1;7]	User 2 [1;7]	User 3 [1;7]	User 4 [1;7]	Mean [1;7]	UB [1;7]	LB [1;7]
Log In	7	7	7	7	7,0	7,0	7,0
Arguments For	5	7	7	7	6,5	7,0	5,4
Arguments Against	7	7	7	7	7,0	7,0	7,0
Hide Plot	6	6	5	7	6,0	7,0	5,7
Training Plans	7	7	7	7	7,0	7,0	7,0
Assistive Aids	7	7	7	7	7,0	7,0	7,0
Diagnoses	7	7	7	7	7,0	7,0	7,0
Dark Mode	7	7	7	7	7,0	7,0	7,0
Fall Prevention	5	5	7	5	5,5	7,0	5,4
Registered Falls	6	7	7	6	6,5	7,0	6,1
Log Out	7	7	7	7	7,0	7,0	7,0

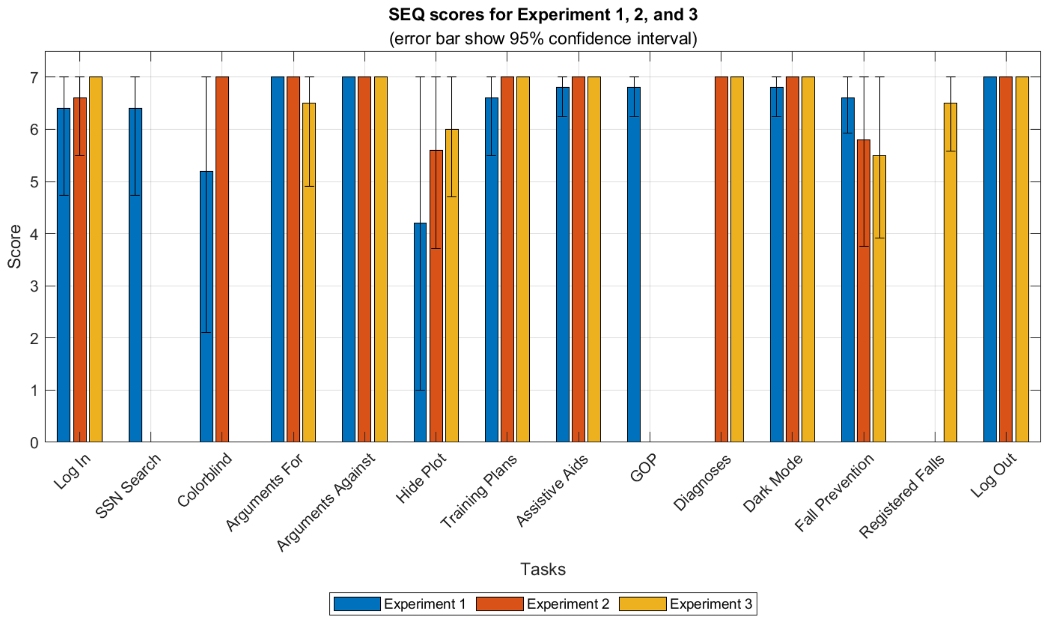

Figure A.40: SEQ answers represented on a graph for Experiment 3

A.4.4 Post-session questionnaire

Table A.38: CSUQ metric data for Experiment 3

Statement \ User	User 1 [1;7]	User 2 [1;7]	User 3 [1;7]	User 4 [1;7]			
1	1	1	1	1			
2	1	1	1	1			
3	2	1	2	NA			
4	3	1	1	1			
5	1	1	1	1			
6	3	1	3	NA			
7	NA	2	2	NA			
8	NA	1	2	NA			
9	3	2	1	NA			
10	2	1	1	2			
11	2	1	2	NA			
12	2	1	2	1			
13	NA	1	2	1			
14	NA	1	2	1			
15	3	1	3	2			
16	2	1	2	1			
17	3	1	1	3			
Score \ User	User 1 [1;7]	User 2 [1;7]	User 3 [1;7]	User 4 [1;7]	Mean [1;7]	UB [1;7]	LB [1;7]
Overall	2,1	1,1	1,8	1,2	1,5	2,3	1,0
SysQual	1,8	1,0	1,5	1,0	1,3	2,0	1,0
InfoQual	2,3	1,3	1,7	1,5	1,7	2,3	1,1
IntQual	3,0	1,0	2,3	1,3	1,9	3,4	1,0
Completeness	75%	100%	100%	63%	84,5%		

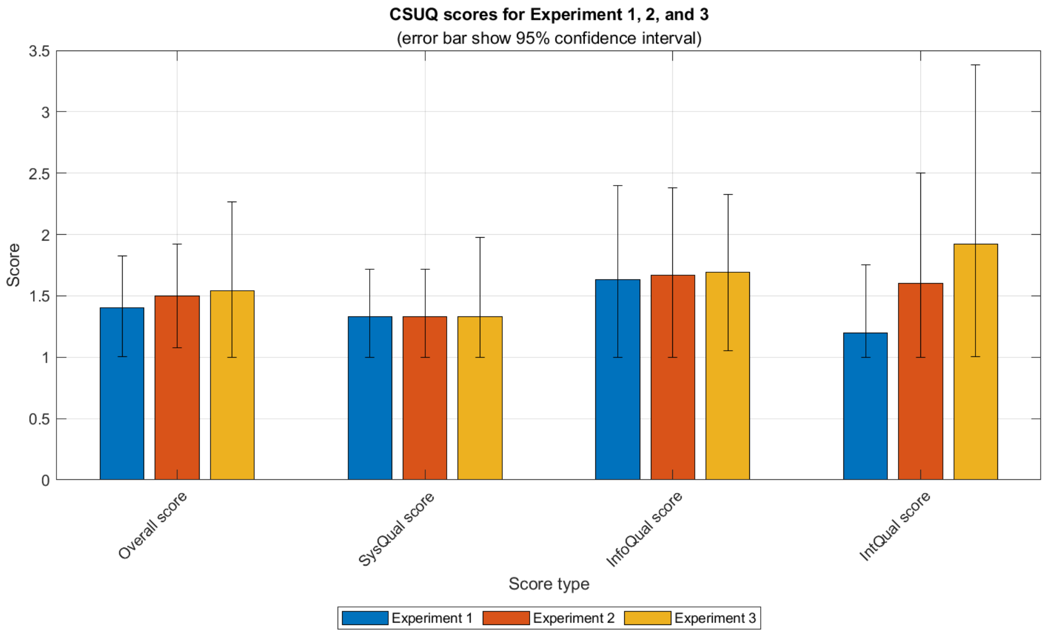

Figure A.41: Mean CSUQ scores for Experiment 3

Table A.39: Danish SUS questionnaire given to participants during Experiment 3

Statement #	Statement
1	Jeg tror at jeg gerne vil bruge hjemmesiden ofte.
2	Jeg fandt hjemmesiden unødvendigt kompliceret.
3	Jeg synes hjemmesiden var let at bruge.
4	Jeg tror jeg vil have brug for teknisk support for at være i stand til at bruge hjemmesiden.
5	Jeg fandt hjemmesidens funktionalitet godt integreret.
6	Jeg synes hjemmesiden var for inkonsistent.
7	Jeg kunne forestille mig, at de fleste personer hurtigt vil kunne lære at bruge hjemmesiden.
8	Jeg fandt hjemmesiden meget klodset at bruge.
9	Jeg følte mig meget selvsikker, da jeg brugte hjemmesiden.
10	Jeg havde brug for at lære en masse ting før jeg kunne komme i gang med at bruge hjemmesiden.

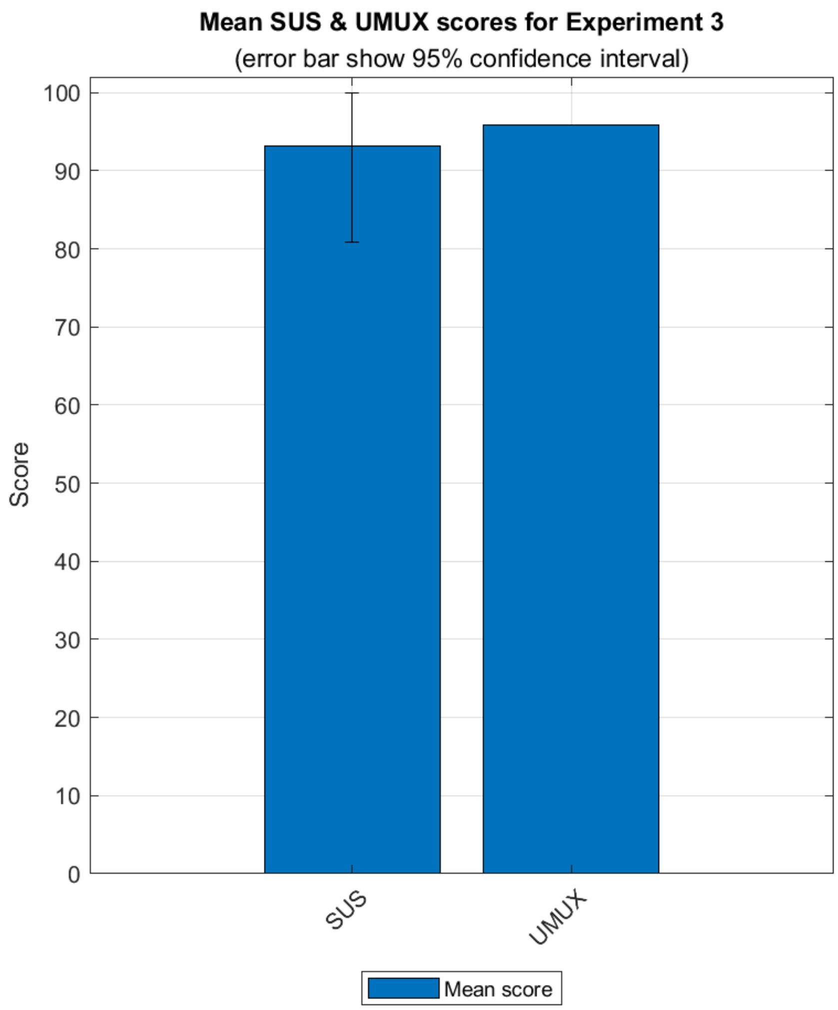

Figure A.42: The mean scores for the SUS and UMUX questionnaires and the 95% confidence interval

Table A.40: SUS metric data for Experiment 3

Statement \ User	User 1 [1;5]	User 2 [1;5]	User 3 [1;5]	User 4 [1;5]	Mean [1;5]	UB [1;5]	LB [1;5]
1	4	4	5	4			
2	2	1	1	1			
3	5	5	5	5			
4	1	1	1	1			
5	3	5	5	5			
6	3	3	1	1			
7	5	5	5	5			
8	1	1	1	1			
9	4	5	5	5			
10	1	1	1	1			
Mean	82,5	92,5	100,0	97,5	93,1	100,0	80,8

Table A.41: Danish UMUX questionnaire given to participants Experiment 3

Statement #	Statement
1	Hjemmesidens funktionalitet imødekommer mine krav.
2	At bruge hjemmesiden er en frustrerende oplevelse.
3	Hjemmesiden er let at bruge.
4	Jeg bruger for meget tid på at rette ting når jeg bruger hjemmesiden.

Table A.42: UMUX metric results

Statement \ User	User 1 [1;7]	User 2 [1;7]	User 3 [1;7]	User 4 [1;7]	Mean [1;7]	UB [1;7]	LB [1;7]
1	2	2	2	2			
2	7	7	7	7			
3	1	1	1	1			
4	7	7	7	7			
Mean	95,8	95,8	95,8	95,8	95,8	95,8	95,8

A.4.5 Task success

Table A.43: Task success for Experiment 3 - Completed tasks are marked with (✓) and failed tasks are marked with (✗)

Task \ User	User 1	User 2	User 3	User 4	Mean
Log In	✓	✓	✓	✓	100%
Arguments For	✓	✓	✓	✓	100%
Arguments Against	✓	✗	✓	✓	75%
Hide Plot	✓	✓	✓	✓	100%
Training Plans	✓	✓	✓	✓	100%
Assistive Aids	✓	✓	✓	✓	100%
Diagnoses	✓	✗	✓	✓	75%
Dark Mode	✓	✗	✓	✓	75%
Fall Prevention	✓	✓	✓	✓	100%
Registered Falls	✓	✓	✓	✓	100%
Log Out	✓	✓	✓	✓	100%

Table A.44: Received task assistance for Experiment 3 - Tasks with received assistance are marked with (✓) and tasks without needed assistance are marked with (✗)

Task \ User	User 1	User 2	User 3	User 4	Mean
Log In	✗	✗	✗	✗	0%
Arguments For	✗	✗	✗	✗	0%
Arguments Against	✗	✗	✗	✗	0%
Hide Plot	✗	✗	✗	✗	0%
Training Plans	✗	✗	✗	✗	0%
Assistive Aids	✗	✗	✗	✗	0%
Diagnoses	✗	✗	✗	✗	0%
Dark Mode	✗	✗	✗	✗	0%
Fall Prevention	✗	✓	✗	✗	25%
Registered Falls	✗	✗	✗	✗	0%
Log Out	✗	✗	✗	✗	0%

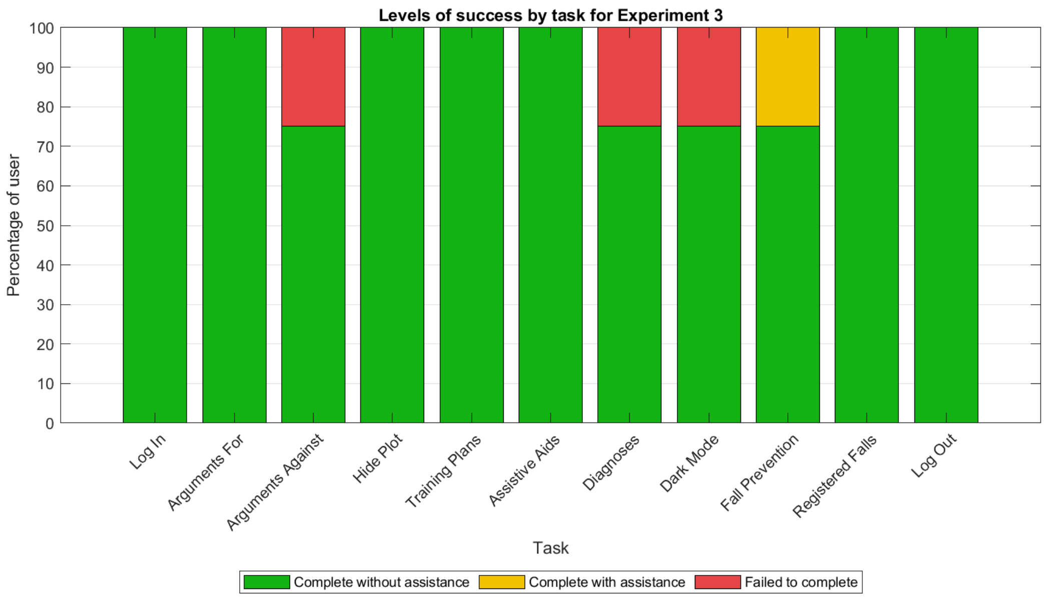

Figure A.43: Percentage of users for each task who completed a task without assistance, with assistance, or failed to complete the task for Experiment 3

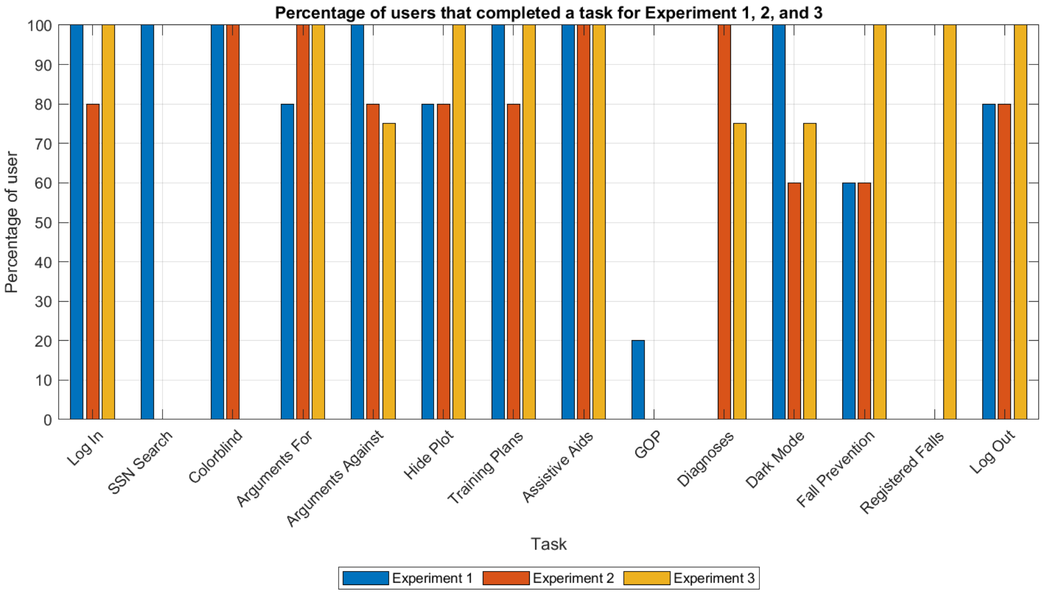

Figure A.44: Percentage of users for each task who completed a task without assistance, with assistance, or failed to complete the task between Experiment 1, 2, and 3

A.4.6 Task time

Table A.45: Task completion time for Experiment 3

Task \ User	User 1 [s]	User 2 [s]	User 3 [s]	User 4 [s]	Mean [s]	UB [s]	LB [s]
Log In	8	14	14	6	10,5	17,1	3,9
Arguments For	3	-	9	3	5,0	13,6	0
Arguments Against	3	-	1	4	2,7	6,5	0
Hide Plot	48	12	34	20	28,5	53,7	3,3
Training Plans	3	14	3	4	6,0	14,5	0
Assistive Aids	2	2	2	4	2,5	4,1	0,9
Diagnoses	2	-	2	3	2,3	3,8	0,9
Dark Mode	4	-	6	3	4,3	8,1	0,5
Fall Prevention	19	51	13	16	24,8	52,9	0
Registered Falls	8	12	16	2	9,5	19,0	0
Log Out	5	4	6	6	5,3	6,8	3,7
Mean	9,5	15,6	9,6	6,5	9,2		
Mean	10,3						

Table A.46: Total time on task for Experiment 3

Task \ User	User 1 [s]	User 2 [s]	User 3 [s]	User 4 [s]	Mean [s]	UB [s]	LB [s]
Log In	18	16	17	12	15,8	19,9	11,6
Arguments For	17	10	57	6	22,5	59,8	0
Arguments Against	21	-	23	7	17,0	38,7	0
Hide Plot	51	15	39	23	32,0	57,7	6,3
Training Plans	22	18	8	10	14,5	25,0	4,0
Assistive Aids	6	3	4	19	8,0	19,8	0
Diagnoses	5	2	5	6	4,5	7,3	1,7
Dark Mode	8	3	9	5	6,3	10,6	1,9
Fall Prevention	23	54	24	23	31,0	55,4	6,6
Registered Falls	11	15	25	2	13,3	28,4	0
Log Out	6	6	7	08	6,8	8,3	5,2
Mean	17,1	14,2	19,8	11,0	15,6		
Mean	15,5						

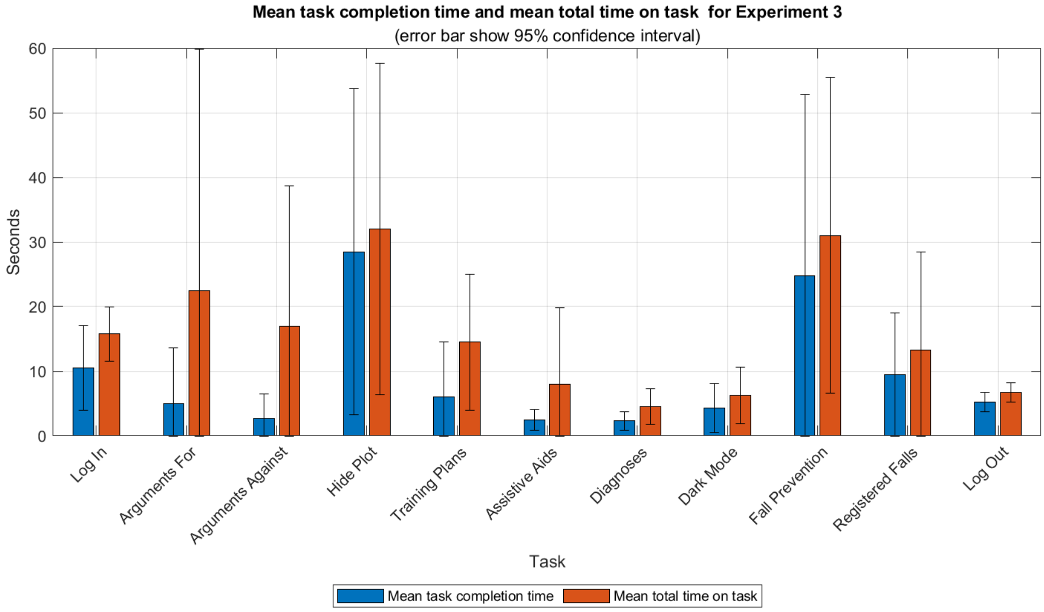

Figure A.45: Mean task completion time and mean total time on task for Experiment 3

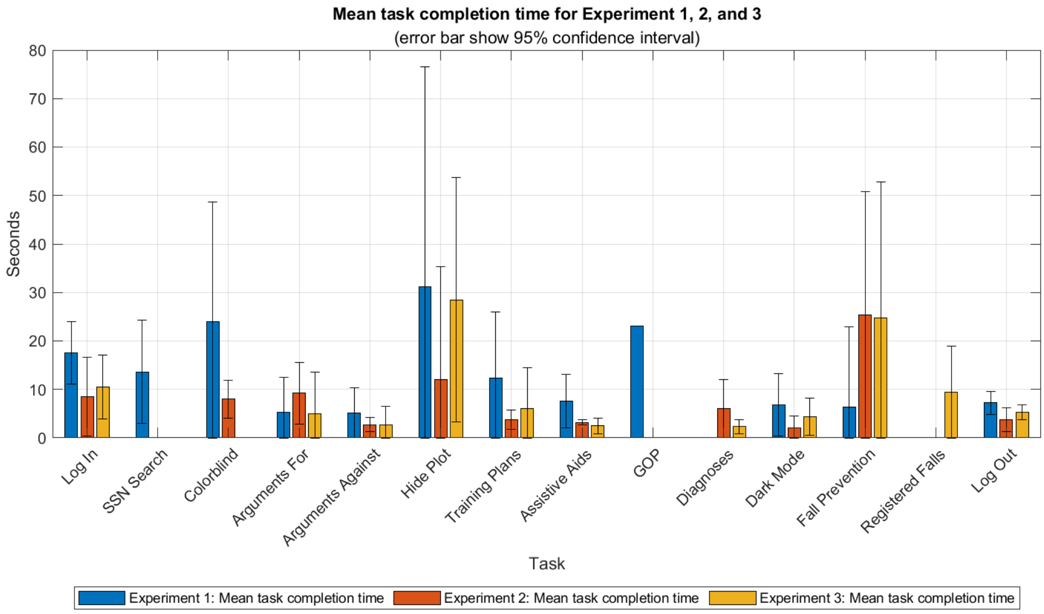

Figure A.46: Mean task completion time for Experiment 1, 2, and 3

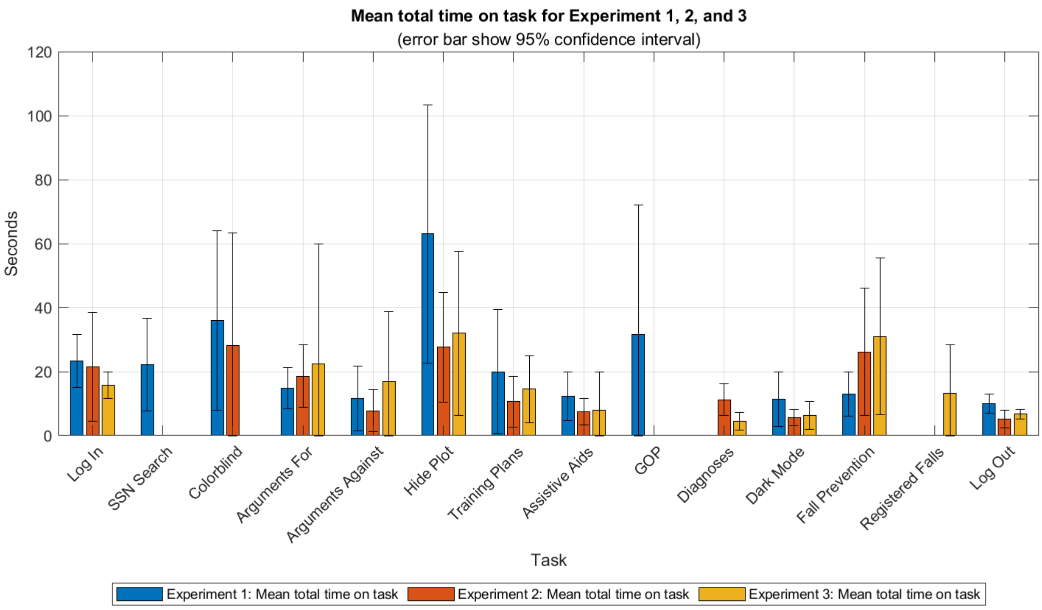

Figure A.47: Mean total time on task for Experiment 1, 2, and 3

A.4.7 Errors

Table A.47: Errors for Experiment 3

Error type \ User	User 1 [#]	User 2 [#]	User 3 [#]	User 4 [#]	Sum [#]
Random click error	0	0	9	1	10
Menu item error	2	2	1	1	6
Menu error	1	2	0	0	3
Hide plot error	0	0	0	0	0
Show plot error	0	0	2	0	2
Right click on graph error	5	0	0	0	5
Left click on graph error	0	0	0	0	0
Table cell click error	4	0	3	0	7
Tabs click error	3	3	1	1	8
Wrongly opened Argumentbox error	0	1	0	0	1
Wrongly closed Argumentbox error	0	0	0	0	0
Cursor drag error	0	0	0	0	0
Sum	15	8	16	3	42

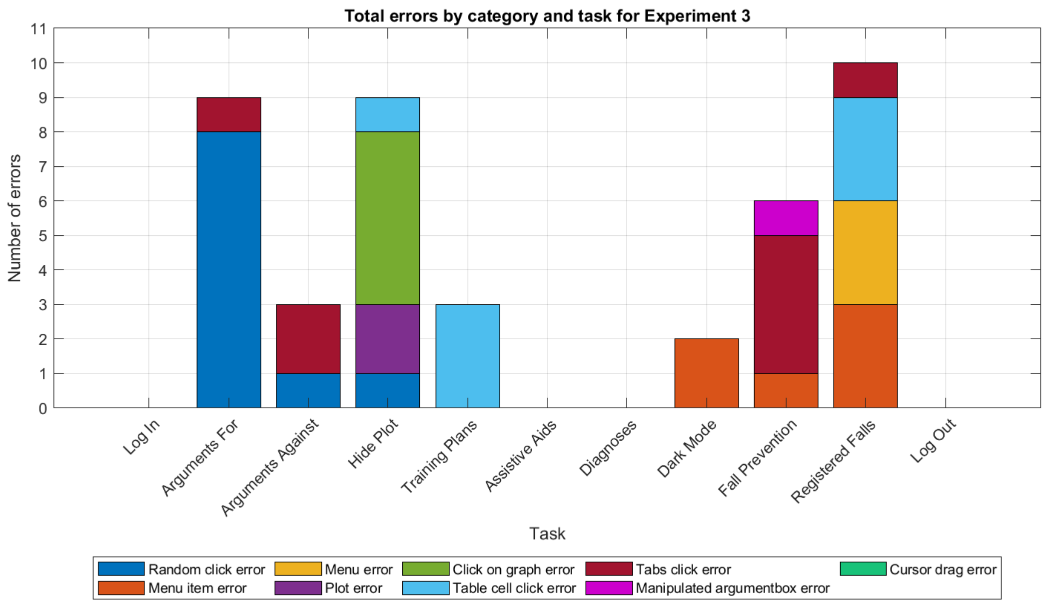

Figure A.48: Total number of errors for Experiment 3

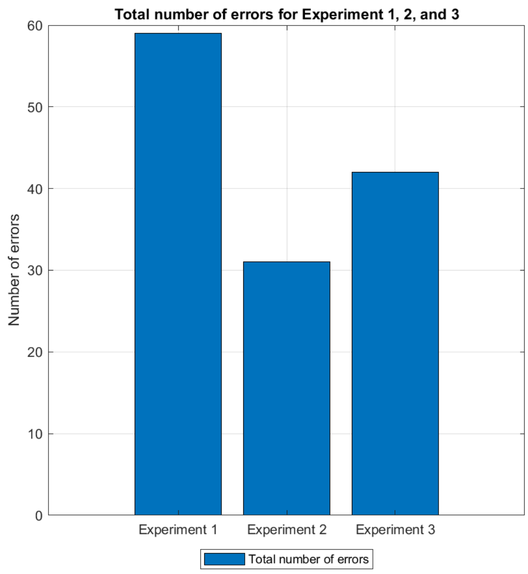

Figure A.49: Total number of errors between Experiment 1, 2, and 3

A.4.8 Efficiency

Table A.48: Efficiency for Experiment 3

User / Task	User 1 [#]	User 2 [#]	User 3 [#]	User 4 [#]	Mean [#]	Expected [#]	UB [#]	LB [#]
Log In	2	2	2	2	2,0	2	2,0	2,0
Arguments For	2	1	9	1	3,3	1	9,4	0
Arguments Against	3	0	2	1	1,5	1	3,6	0
Hide Plot	7	1	4	2	3,5	1	7,7	0
Training Plans	4	1	1	1	1,8	1	4,1	0
Assistive Aids	1	1	1	1	1,0	1	1,0	1,0
Diagnoses	1	0	1	1	0,8	1	1,5	0
Dark Mode	2	1	3	2	2,0	2	3,3	0,7
Fall Prevention	2	6	4	5	4,3	2	6,7	1,5
Registered Falls	3	5	6	0	3,5	1	7,7	0
Log Out	2	2	2	2	2,0	2	2,0	2,0
Sum	29	20	35	18	25,5			
Mean	25,5							

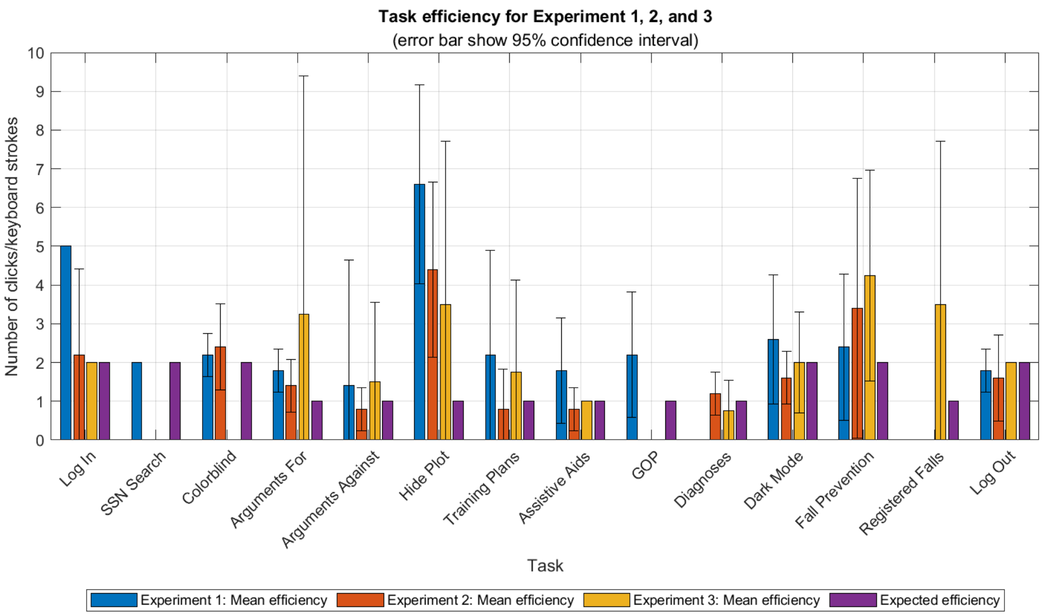

Figure A.50: Efficiency between Experiment 1, 2, and 3

A.4.9 Learnability

Table A.49: Learnability for Experiment 3

Metric	Measurement
Mean task completion time	10,3
Total number of errors	42
Mean efficiency	25,5
Sum	89,7

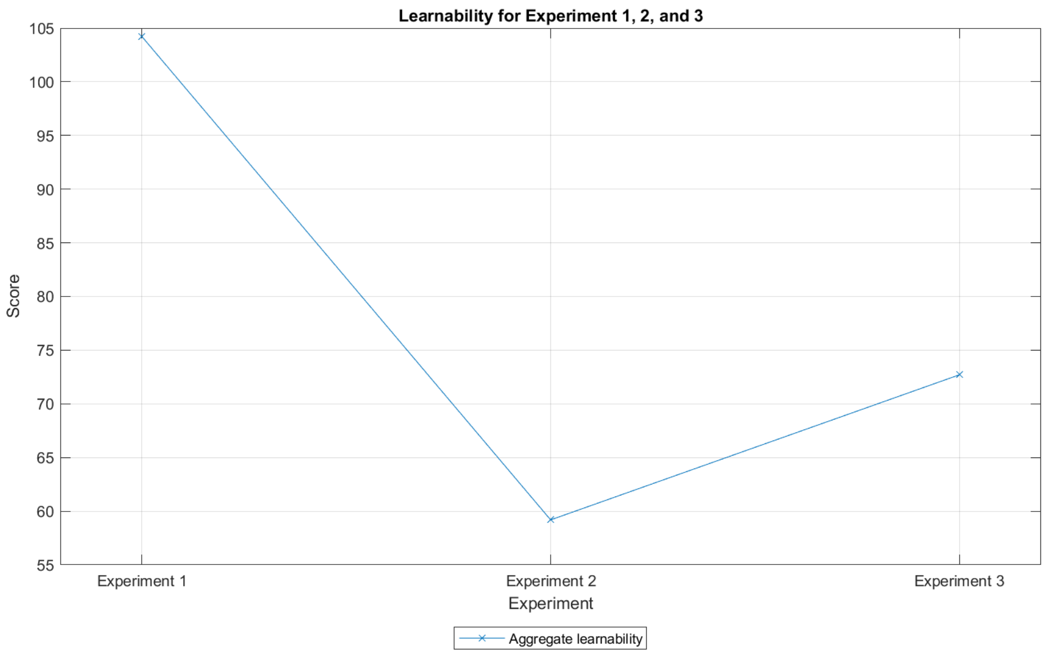

Figure A.51: Learnability between Experiment 1, 2, and 3

A.4.10 Semi-structured Interview

Table A.50: Gathered Interview Feedback during Experiment 3

Id	Feedback
	General
1	Det hele er let at læse.
2	Der vil måske være nogle læseproblemer på en bærbar, men vi sidder alle ved en stationær med to store skærme.
3	Det har alt sammen relevans.
4	Jeg savner ikke umiddelbart noget.
5	Jeg vil godt kunne bruge hjemmesiden som en del af mit arbejde, såfremt den er integreret i Cura.
6	Hjemmesiden er meget brugervenlig.
7	Jeg troede jeg var nødt til at taste en masse ting ind, men jeg blev glædelig overrasket.
8	Det er godt at det er blandet information fra Cura i tilfælde af man er uenig med sandsynligheden.

Continued on next page

Table A.50 – continued from previous page	
Id	**Feedback**
9	Hjemmesiden giver et hurtigt overblik omkring bevilling af træning.
10	Hjemmesiden er simpel på en positiv og overskuelig måde.
11	Hvis hjemmesiden var tilgængelig nu, så vil jeg bruge den jævnlig.
12	Hjemmesiden er nemt.
13	Hjemmesiden er lige til at gå til.
14	Savner lidt farve på hjemmesiden.
15	Det er meget hvidt i hvidt.
16	Hjemmesiden er let at huske.
17	Overskrifterne i fanerne på højre side kunne godt blive markeret i fed.
18	Det hele er flot stillet på.
19	Størrelsesforholdene mellem de forskellige ting på hjemmesiden sidder lige i skabet.
20	Hjemmesiden er måske lidt kedelig at se på.
21	Det skal afprøves som en del af mit arbejde, før jeg kan sige om det er noget jeg kan bruge.
22	Det er rart at tingene bliver stående hvor man forventer det.
23	Det er umiddelbart meget nemt at hoppe ind i igen.
24	Der var nogle enkelte ting der glippede, men overordnet set var det let at gå til.
Probability and Arguments	
25	Argumenterne er ikke bløde nok.
26	Hvorfor er argumentet et argument? Jeg savner forklaring.
27	Sandsynligheden er super til at give et koldt tal, som kan være med til at underbygge ens egen holdning og mening.
28	Jeg føler mig støttet i min beslutningsproces, da sandsynligheden er en super vejleder.
29	Det skal være tydeligt hvis sandsynligheden er beregnet ud fra kun træningsdata eller bliver der også brugt hjælpemiddelsdata.
30	Der må være en effekt på sandsynligheden hvis man får bevilliget træning og afhængig af hvor lang tid man har haft et hjælpemiddel.
31	Et argument imod kunne være "Sengeliggende".
32	Jeg vil godt kunne bruge lidt flere argumenter.
33	Sandsynligheden vil helt sikkert spille med i min overordnet vurdering.
34	Sandsynligheden giver en rigtig god indikator på om man bare skal afvise eller om man skal fordybe sig.

Continued on next page

Table A.50 – continued from previous page

Id	Feedback
35	Hvor kommer sandsynligheden fra?
36	Jeg er ligeglad med argumentet omkring personens alder og køn.
37	Argumenterne der omhandler låne periode og antal af hjælpemidler er brugbart.
	Registered Falls
38	Det kunne være rart at se hvad sandsynligheden har været for fald for et år siden.
39	Måden det står præsenteret på er næsten ligesom i Cura.
40	Det er godt med noget faldregistrering.
41	Datoen for fald er i Cura kun datoen for hvornår faldet er registreret, men ikke hvornår faldet er sket.

Appendix B

User Acceptance Test

This chapter contains an overview of all the user acceptance tests performed during this thesis. The test description for some tests are empty (-), meaning that it is possible to complete the test without the need to perform any actions, as the previous tests have lead the user into the correct state.

Test scenario #1 (UR1): Log into the prototype.

Test step	Test description	Expected results	Pass(✓)/Fail(✗)
1	Write a password	The password entered is shown as asterisks.	✓
2	Click on the "Login" button	The user gets logged into the application and redirected to the "Gennemføre Træning" screen.	✓

Table B.1: Test scenario #1 - Log into the prototype.

Test scenario #2 (UR2): See the citizens probability for completing rehabilitation training.

Test step	Test description	Expected results	Pass(✓)/Fail(✗)
1	-	A probability for completing a rehabilitation training course is shown as a number.	✓

Table B.2: Test scenario #2 - See the citizens probability for completing rehabilitation training.

Test scenario #3 (UR4): See both the positive and negative arguments for the citizens probability.

Test step	Test description	Expected results	Pass(✓)/Fail(✗)
1	From the page "Gennemføre Træning" click on the panel "ARGUMENTER FOR GENNEMFØRSEL"	The panel opens and shows the arguments in plain text.	✓
2	From the page "Gennemføre Træning" click on the panel "ARGUMENTER IMOD GENNEMFØRSEL"	The panel opens and shows the arguments in plain text.	✓

Table B.3: Test scenario #3 - See both the positive and negative arguments for the citizens probability.

Test scenario #4 (UR5): See the citizens information.

Test step	Test description	Expected results	Pass(✓)/Fail(✗)
1	-	The citizens information is shown in plain text.	✓

Table B.4: Test scenario #4 - See the citizens information.

Test scenario #5 (UR6): See the citizens diagnoses and motivations.

Test step	Test description	Expected results	Pass(✓)/Fail(✗)
1	Click the tab ”DIAGNOSER OG MOTIVATION”	The tab should show the citizens diagnoses and motivation in plain text.	✓

Table B.5: Test scenario #5 - See the citizens diagnoses and motivations.

Test scenario #6 (UR7): See the citizens training plans.

Test step	Test description	Expected results	Pass(✓)/Fail(✗)
1	Click the tab ”TRÆNINGSFORLØB”	The tab should show the citizens training plans in a table.	✓

Table B.6: Test scenario #6 - See the citizens training plans.

Test scenario #7 (UR8, UR9, & UR10): See the citizens assistive aids.

Test step	Test description	Expected results	Pass(✓)/Fail(✗)
1	Click the tab ”HJÆLPEMIDLER”	The tab should show the citizens assistive aids. The assistive aids should be represented in a table and a graph that also includes home help hours.	✓

Table B.7: Test scenario #7 - See the citizens assistive aids.

Test scenario #8 (UR3): See the citizens probability for falling within the following three months.

Test step	Test description	Expected results	Pass(✓)/Fail(✗)
1	Hover the menu	The menu should show the menu items "Gennemføre Træning", "Faldforebyggelse", "Mørk Tilstand", and "Log Ud".	✓
2	Click on the menu item "Faldforebyggelse"	The screen should change and show the probability as a number.	✓

Table B.8: Test scenario #8 - See the citizens probability for falling within the following three months.

Test scenario #9 (UR11): See the citizens registered falls.

Test step	Test description	Expected results	Pass(✓)/Fail(✗)
1	Click the tab "REGISTRERET FALD"	The tab should show the citizens registered falls in a table.	✓

Table B.9: Test scenario #9 - See the citizens registered falls.

Test scenario #10 (UR12): Change the contrast polarity from positive to negative.

Test step	Test description	Expected results	Pass(✓)/Fail(✗)
1	Hover the menu	The menu should show the menu items "Gennemføre Træning", "Faldforebyggelse", "Mørk Tilstand", and "Log Ud".	✓
2	Click on the menu item "Mørk Tilstand"	The contrast polarity of the screen should change from positive to negative.	✓

Table B.10: Test scenario #10 - Change the contrast polarity from positive to negative.

Test scenario #11 (UR13): Log out of the prototype.

Test step	Test description	Expected results	Pass(✓)/Fail(✗)
1	Hover the menu	The menu should show the menu items "Gennemføre Træning", "Faldforebyggelse", "Mørk Tilstand", and "Log Ud".	✓
2	Click on the menu item "Log Ud"	The user should be signed out and redirected to the login screen.	✓

Table B.11: Test scenario #11 - Log out of the prototype.

Appendix C

Systematic

C.1 Cura API Authentication

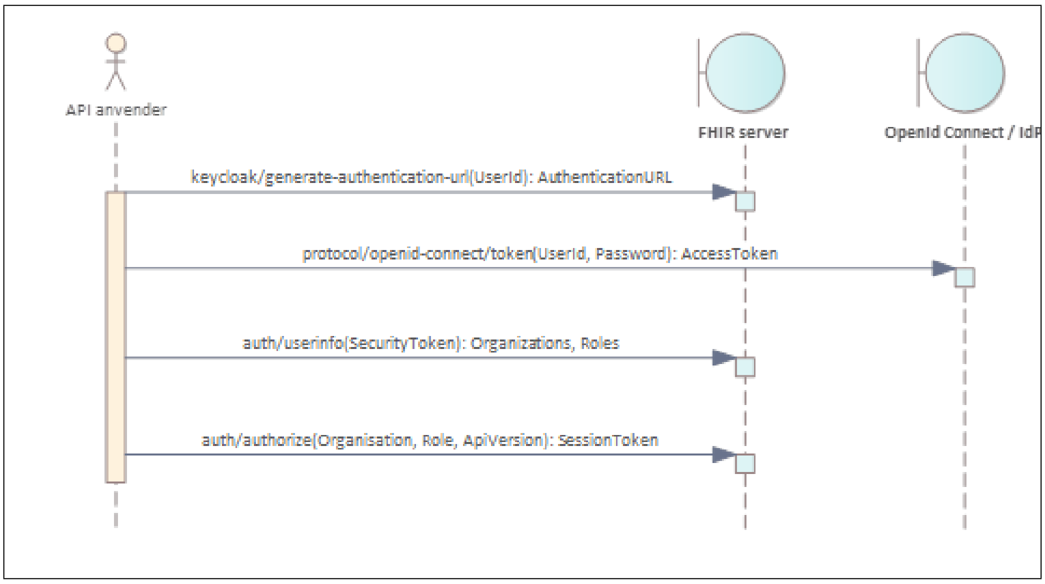

Figure C.1: Cura authentication of third party users [89]

List of Figures

List of Tables

List of Listings

Nomenclature

AI Artificial Intelligence.

Assistive aid An item to help a citizen in their everyday life e.g. wheelchair or walker.

Case 1 The assessment performed by a case worker before assigning rehabilitation training.

Case 2 The assessment performed by a case worker before assigning fall preventive training.

Case worker A person employed by a municipality to prescribe training courses or assistive aids to a citizen.

CDSS Clinical Decision Support System

Cura Refers to the Columna Cura product suite by Systematic [89]

GOP A citizen's rehabilitation plan made by a hospital.

LB Lower bound for the confidence interval

ML Machine Learning.

Pilot group A group of potential end users, consisting of five case workers, and was used to evaluate the progression of the project.

R&D1 A previous R&D project made by the group [41], which this report is a continuation of.

R&D2 A previous R&D project made by the group [42], which this report is a continuation of.

Registered fall A registration by the home helper that states when a citizen has fallen and why.

SSN Social Security Number or CPR in Danish.

Subsystem 1 Is a REST API connecting the prototype to the AI for providing the necessary information, regarding a citizens probability for either completing a rehabilitation course, or the risk of falling.

Subsystem 2 Is a REST API connecting the prototype to Cura for providing additional necessary citizen information.

UB Upper bound for the confidence interval

Bibliography

[1] *Lov om ændring af lov om social service\n(Rehabiliteringsforløb og hjemmehjælp m.v.)* Retsinformation. Dec. 27, 2014. URL: https://www.retsinformation.dk/eli/lta/2014/1524 (visited on 05/30/2021).

[2] Danmarks Statistik. *Folketal den 1. i kvartalet efter område, køn, alder og civilstand - Statistikbanken - data og tal.* URL: https://www.statistikbanken.dk/statbank5a/selectvarval/define.asp?PLanguage=0&subword=tabsel&MainTable=FOLK1A&PXSId=199113&tablestyle=&ST=SD&buttons=0 (visited on 05/15/2021).

[3] Denise Taylor. "Physical activity is medicine for older adults: Table 1". In: *Postgraduate Medical Journal* 90.1059 (Jan. 2014), pp. 26–32. ISSN: 0032-5473, 1469-0756. DOI: 10.1136/postgradmedj-2012-131366. URL: https://pmj.bmj.com/lookup/doi/10.1136/postgradmedj-2012-131366 (visited on 05/15/2021).

[4] Healthcare Denmark. "A dignified elderly care in Denmark". In: (Sept. 1, 2019), p. 36. URL: https://www.healthcaredenmark.dk/media/plvbj4yz/elderly-care-v10919.pdf.

[5] Sandra Zampieri et al. "Lifelong Physical Exercise Delays Age-Associated Skeletal Muscle Decline". In: *The journals of gerontology. Series A, Biological sciences and medical sciences* 70 (2014). DOI: 10.1093/gerona/glu006.

[6] Greta Häggblom-Kronlöf and Ulla Sonn. "Use of assistive devices – a reality full of contradictions in elderly persons' everyday life". In: *Disability and Rehabilitation: Assistive Technology* 2.6 (Jan. 1, 2007). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17483100701701672, pp. 335–345. ISSN: 1748-3107. DOI: 10.1080/17483100701701672. URL: https://doi.org/10.1080/17483100701701672 (visited on 05/15/2021).

[7] Digitaliseringsstyrelsen et al. "Digitaliseringspagt – En ny retning for det fællesoffentlige samarbejde". In: (Mar. 1, 2019), p. 6. URL: https://fm.dk/media/17965/digitaliseringspagt__en_ny_retning_for_det_faellesoffentlige_samarbejde_a.pdf (visited on 05/15/2021).

[8] AIR. *AIR API - Swagger UI*. URL: https://air-fastapi.azurewebsites.net/docs (visited on 04/05/2021).

[9] Dan Power. *What is a DSS?* Ask Dan! about DSS. Mar. 20, 2009. URL: https://dssresources.com/faq/index.php?action=artikel&id=184 (visited on 05/30/2021).

[10] Daniel J. Power. *A Brief History of Decision Support Systems*. A Brief History of Decision Support Systems. Mar. 10, 2007. URL: http://dssresources.com/history/dsshistory.html (visited on 05/30/2021).

[11] Yuri Boreisha and Oksana Myronovych. "WEB-BASED DECISION SUPPORT SYSTEMS AS KNOWLEDGE REPOSITORIES FOR KNOWLEDGE MANAGEMENT SYSTEMS". In: *Ubiquitous Computing and Communication Journal* 3.2 (Jan. 1, 2008), p. 8. URL: https://www.researchgate.net/publication/238783065_WEB-BASED_DECISION_SUPPORT_SYSTEMS_AS_KNOWLEDGE_REPOSITORIES_FOR_KNOWLEDGE_MANAGEMENT_SYSTEMS (visited on 05/30/2021).

[12] Daniel J. Power. *What is the expanded DSS framework?* Ask Dan! about DSS. June 21, 2015. URL: https://dssresources.com/faq/index.php?action=artikel&id=2 (visited on 04/15/2021).

[13] Reed T. Sutton et al. "An overview of clinical decision support systems: benefits, risks, and strategies for success". In: *npj Digital Medicine* 3.1 (Dec. 2020), p. 17. ISSN: 2398-6352. DOI: 10.1038/s41746-020-0221-y. URL: http://www.nature.com/articles/s41746-020-0221-y (visited on 03/26/2021).

[14] Eta Berner and Yang Gong. *Clinical Decision Support Systems: Theory and Practice*. 3rd ed. 2016. 324 pp. ISBN: 978-3-319-31911-7. DOI: 10.1007/978-3-319-31913-1.

[15] Jan Horsky et al. "Interface design principles for usable decision support: A targeted review of best practices for clinical prescribing interventions". In: *Journal of Biomedical Informatics* 45.6 (Dec. 2012), pp. 1202–1216. ISSN: 15320464. DOI: 10.1016/j.jbi.2012.09.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S1532046412001499 (visited on 04/07/2021).

[16] Ranielle Castillo and Arpad Kelemen. “Considerations for a Successful Clinical Decision Support System”. In: *Computers, informatics, nursing : CIN* 31 (2013). DOI: 10.1097/NXN.0b013e3182997a9c.

[17] Markets And Markets. *clinical decision support systems market Insights, Trends — Industry Report, 2025 — MarketsandMarkets™*. July 12, 2020. URL: https://www.marketsandmarkets.com/ResearchInsight/clinical-decision-support-systems-market.asp (visited on 04/15/2021).

[18] Wolters Kluwer. *Wolters Kluwer’s expert solutions combine expertise with advanced technology*. 2021. URL: https://www.wolterskluwer.com/en (visited on 04/15/2021).

[19] Zynx Health Incorporated. *Zynx Health — Clinical Decision Support Solutions for Healthcare*. Zynx Health. 2021. URL: https://www.zynxhealth.com/ (visited on 04/15/2021).

[20] Markets And Markets. *Clinical Decision Support Systems Market - Global Forecast to 2025 — MarketsandMarkets*. Markets And Markets. Dec. 2020. URL: https://www.marketsandmarkets.com/Market-Reports/clinical-decision-support-systems-market-18085342.html (visited on 04/15/2021).

[21] Wolters Kluwer. *Lippincott Advisor*. Wolters Kluwer. 2021. URL: https://www.wolterskluwer.com/en/solutions/lippincott-solutions/lippincott-advisor (visited on 04/15/2021).

[22] Zynx Health Incorporated. *Post-Acute Care*. Zynx Health. 2021. URL: https://www.zynxhealth.com/what-we-do/post-acute-care/ (visited on 04/15/2021).

[23] Nan Hou. *Helping Rehabilitation Facilities Succeed in 2021*. Zynx Health. Apr. 9, 2021. URL: https://www.zynxhealth.com/blog-zc-rehab/ (visited on 04/15/2021).

[24] Mah Laka, Adriana Milazzo, and Tracy Merlin. “Factors That Impact the Adoption of Clinical Decision Support Systems (CDSS) for Antibiotic Management”. In: *International Journal of Environmental Research and Public Health* 18 (2021), p. 1901. DOI: 10.3390/ijerph18041901.

[25] R. Bhardwaj, A. R. Nambiar, and D. Dutta. *A Study of Machine Learning in Healthcare*. ISSN: 0730-3157. July 2017. DOI: 10.1109/COMPSAC.2017.164.

[26] Devam Dave et al. "Explainable AI meets Healthcare: A Study on Heart Disease Dataset". In: *arXiv:2011.03195 [cs]* (Nov. 6, 2020). arXiv: 2011.03195. URL: http://arxiv.org/abs/2011.03195 (visited on 04/17/2021).

[27] A. Adadi and M. Berrada. "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)". In: *IEEE Access* 6 (2018). Conference Name: IEEE Access, pp. 52138–52160. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2870052.

[28] Alejandro Barredo Arrieta et al. *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI.* Dec. 26, 2019. DOI: 10.1016/j.inffus.2019.12.012. URL: https://www.sciencedirect.com/science/article/pii/S1566253519308103 (visited on 04/17/2021).

[29] Scott Weber et al. "Practitioner Approaches to the Integration of Clinical Decision Support System Technology in Critical Care". In: *JONA: The Journal of Nursing Administration* 39.11 (Nov. 2009), pp. 465–469. ISSN: 0002-0443. DOI: 10.1097/NNA.0b013e3181bd5fc2. URL: https://journals.lww.com/jonajournal/Fulltext/2009/11000/Practitioner_Approaches_to_the_Integration_of.7.aspx (visited on 04/21/2021).

[30] Anne Marie Kanstrup, Marion Berg Christiansen, and Christian Nøhr. *Patient safety informatics: adverse drug events, human factors, and IT tools for patient medication safety.* Ed. by Vassilis Koutkias et al. Vol. 166. Studies in health technology and informatics. OCLC: ocn715922210. Amsterdam ; Washington, DC: IOS Press, 2011. 274 pp. ISBN: 978-1-60750-739-0 978-1-60750-740-6.

[31] E. Kilsdonk et al. "Uncovering healthcare practitioners' information processing using the think-aloud method: From paper-based guideline to clinical decision support system". In: *International Journal of Medical Informatics* 86 (Feb. 2016), pp. 10–19. ISSN: 13865056. DOI: 10.1016/j.ijmedinf.2015.11.011. URL: https://linkinghub.elsevier.com/retrieve/pii/S1386505615300629 (visited on 04/07/2021).

[32] Kristen Miller et al. "Interface, information, interaction: a narrative review of design and functional requirements for clinical decision support". In: *Journal of the American Medical Informatics Association* 25.5 (May 1, 2018), pp. 585–592. ISSN: 1527-974X. DOI: 10.1093/jamia/ocx118. URL: https://doi.org/10.1093/jamia/ocx118 (visited on 03/26/2021).

[33] All About UX. *User experience definitions.* URL: https://www.allaboutux.org/ux-definitions.

[34] Don Norman and Jakob Nielsen. *The Definition of User Experience (UX).* Nielsen Norman Group. URL: https://www.nngroup.com/articles/definition-user-experience/ (visited on 02/24/2021).

[35] Tom Tullis and Bill Albert. *Measuring the User Experience: Collecting, Analysing, and Presenting Usability Metrics.* 2nd. Morgan Kaufmann, July 3, 2013. 320 pp. ISBN: 978-0-12-415781-1. URL: https://www.elsevier.com/books/measuring-the-user-experience/albert/978-0-12-415781-1.

[36] Interaction Design Foundation. *Key Question in User Experience Design – Usability vs Desirability.* The Interaction Design Foundation. Aug. 1, 2020. URL: https://www.interaction-design.org/literature/article/key-question-in-user-experience-design-usability-vs-desirability (visited on 02/24/2021).

[37] Jakob Nielsen. *Usability 101: Introduction to Usability.* Nielsen Norman Group. Jan. 3, 2012. URL: https://www.nngroup.com/articles/usability-101-introduction-to-usability/ (visited on 02/23/2021).

[38] Christian Fischer Pedersen. *AIR.* Apr. 14, 2021. URL: https://projekter.au.dk/air/ (visited on 04/22/2021).

[39] Interaction Design Foundation. *What is User Centered Design?* The Interaction Design Foundation. May 15, 2021. URL: https://www.interaction-design.org/literature/topics/user-centered-design (visited on 05/15/2021).

[40] Yvonne Rogers, Helen Sharp, and Jennifer Preece. *Interaction Design: beyond human-computer interaction.* 5th. John Wiley & Sons, Inc, 2019. 657 pp. ISBN: 978-1-119-54725-9.

[41] Frederik Andersen and Theis Egsgaard Hansen. "User interface for AI-Rehabilitation. Research and Development Project 1". Research and Development Project. Aarhus University, Oct. 20, 2020. 13 pp.

[42] Frederik Andersen and Theis Egsgaard Hansen. "User interface for AI-Rehabilitation. Research and Development Project 2". Research and Development Project. Aarhus University, Jan. 4, 2021. 20 pp.

[43] Katy Le. *User-centered Design Method.* Medium. Nov. 7, 2017. URL: https://medium.com/redcatstudio/user-centered-design-method-28e3aafc8c8a (visited on 03/17/2021).

[44] John M Bryson. "What to do when Stakeholders matter: **Stakeholder Identification and Analysis Techniques**". In: *Public Management Review* 6.1 (Mar. 2004), pp. 21–53. ISSN: 1471-9037, 1471-9045. DOI: 10.1080/14719030410001675722. URL: http://www.tandfonline.com/doi/abs/10.1080/14719030410001675722 (visited on 04/16/2021).

[45] Colin Eden and Fran Ackermann. *Making Strategy The Juorney: The Journey of Strategic Management*. 1st edition. London ; Thousand Oaks, Calif: Sage, Sept. 4, 2009. 520 pp. ISBN: 978-0-7619-5225-1.

[46] NN Group. *Jakob's Law of Internet User Experience (2 min. video) (Video)*. Aug. 18, 2017. URL: https://www.nngroup.com/videos/jakobs-law-internet-ux/ (visited on 03/17/2021).

[47] Sara Wilken and Stine Kinsbøl. *CURA ARBEJDSGANG VEJLEDNING TIL FUNKTIONSEVNETILSTANDE OG GENERELLE OPLYSNINGER - PDF Free Download*. DOCPLAYER. Jan. 23, 2019. URL: https://docplayer.dk/128218456-Cura-arbejdsgang-vejledning-til-funktionsevnetilstande-og-generelle-oplysninger.html (visited on 05/16/2021).

[48] Everett N. McKay. *UI is Communication: How to Design Intuitive, User Centered Interfaces by Focusing on Effective Communication*. 1st. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2013. ISBN: 978-0-12-396980-4.

[49] Steve Krug. *Don't make me think, revisited : a common sense approach to Web usability*. 3rd. New Riders, 2014. 214 pp. ISBN: 978-0-321-96551-6. URL: https://search.library.wisc.edu/catalog/9910206558602121 (visited on 04/23/2021).

[50] UXPin. *User Interface Elements Every Designer Should Know*. Studio by UXPin. Oct. 30, 2020. URL: https://www.uxpin.com/studio/blog/user-interface-elements-every-designer-should-know/ (visited on 04/13/2021).

[51] Maria De La Riva. *32 User Interface Elements For UI Designers: Your Ultimate Glossary*. June 13, 2019. URL: https://careerfoundry.com/en/blog/ui-design/ui-element-glossary/ (visited on 04/13/2021).

[52] Usability.gov. *User Interface Elements*. Usability.gov. Publisher: Department of Health and Human Services. Oct. 9, 2013. URL: https://www.usability.gov/how-to-and-tools/methods/user-interface-elements.html (visited on 04/13/2021).

[53] Luz Rello, Martin Pielot, and Mari-Carmen Marcos. "Make It Big!: The Effect of Font Size and Line Spacing on Online Readability". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI'16: CHI Conference on Human Factors in Computing Systems. San Jose California USA: ACM, May 7, 2016, pp. 3637–3648. ISBN: 978-1-4503-3362-7. DOI: `10.1145/2858036.2858204`. URL: `https://dl.acm.org/doi/10.1145/2858036.2858204` (visited on 04/12/2021).

[54] James Kalbach. *Designing Web navigation*. 1st. OCLC: ocm85829310. Beijing ; Sebastopol: O'Reilly, 2007. 394 pp. ISBN: 978-0-596-52810-2.

[55] Theresa Marie Sparks. "The Effects of Color Choice in Web Design on the Usability for Individuals with Color-Blindness". In: (May 2019), p. 61. (Visited on 04/27/2021).

[56] admin. *Color Blindness: The Most Common, Uncommon Eye Condition*. Downtown Vision. Last Modified: 2020-08-09 11:00:00 +0000 UTC. 0. URL: `https://dtvisionreno.com/blog/color-blindness-the-most-common-uncommon-eye-condition/` (visited on 04/26/2021).

[57] Colblinder. *Protanopia – Red-Green Color Blindness – Colblindor*. URL: `https://www.color-blindness.com/protanopia-red-green-color-blindness/` (visited on 04/27/2021).

[58] Colblinder. *Deuteranopia – Red-Green Color Blindness – Colblindor*. URL: `https://www.color-blindness.com/deuteranopia-red-green-color-blindness/` (visited on 03/03/2021).

[59] Henriette Eisfeld, Felix Kristallovich, and Ida Serneberg. "A qualitative study of an emerging user interface design trend". In: (May 17, 2020), p. 30. (Visited on 04/27/2021).

[60] scikit-learn. *An introduction to machine learning with scikit-learn — scikit-learn 0.24.2 documentation*. scikit learn. URL: `https://scikit-learn.org/stable/tutorial/basic/tutorial.html` (visited on 05/18/2021).

[61] Scott Lundberg. *Welcome to the SHAP documentation — SHAP latest documentation*. 2018. URL: `https://shap.readthedocs.io/en/latest/index.html` (visited on 04/19/2021).

[62] Jeff Sauro. *MeasuringU: 4 Steps to Translating a Questionnaire*. MeasuringU. Feb. 4, 2014. URL: `https://measuringu.com/translation/` (visited on 04/28/2021).

[63] Kate Moran. *Usability Testing 101*. Nielsen Norman Group. Dec. 1, 2019. URL: https://www.nngroup.com/articles/usability-testing-101/ (visited on 02/26/2021).

[64] Marieke McCloskey. *Turn User Goals into Task Scenarios for Usability Testing*. Nielsen Norman Group. Jan. 12, 2014. URL: https://www.nngroup.com/articles/task-scenarios-usability-testing/ (visited on 02/08/2021).

[65] Jeff Sauro and James R Lewis. *Quantifying the User Experience. Practical Statistics for User Research*. 2nd. Morgan Kaufmann, 2016. ISBN: 978-0-12-802308-2.

[66] Jeff Sauro. *Rating the Severity of Usability Problems*. MeasuringU. July 30, 2013. URL: https://measuringu.com/rating-severity/ (visited on 02/25/2021).

[67] Don A. Dillman et al. "Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet". In: *Social Science Research* 38.1 (Mar. 1, 2009), pp. 1–18. ISSN: 0049-089X. DOI: 10.1016/j.ssresearch.2008.03.007. URL: https://www.sciencedirect.com/science/article/pii/S0049089X08000306 (visited on 05/10/2021).

[68] James R. Lewis. "Measuring Perceived Usability: The CSUQ, SUS, and UMUX". In: *International Journal of Human-Computer Interaction* 34.12 (Dec. 2018). Publisher: Taylor & Francis Ltd, pp. 1148–1156. ISSN: 10447318. DOI: 10.1080/10447318.2017.1418805. URL: http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=131906010&site=ehost-live (visited on 05/22/2021).

[69] Jeff Sauro PhD. *10 Things to Know About the Post Study System Usability Questionnaire – MeasuringU*. Dec. 18, 2019. URL: https://measuringu.com/pssuq/ (visited on 05/21/2021).

[70] Jeff Sauro and James R. Lewis. "When designing usability questionnaires, does it hurt to be positive?" In: *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. the 2011 annual conference. Vancouver, BC, Canada: ACM Press, 2011, p. 2215. ISBN: 978-1-4503-0228-9. DOI: 10.1145/1978942.1979266. URL: http://dl.acm.org/citation.cfm?doid=1978942.1979266 (visited on 05/22/2021).

[71] Kraig Finstad. "The Usability Metric for User Experience". In: *Interacting with Computers*. Modelling user experience - An agenda for research and practice 22.5 (Sept. 1, 2010), pp. 323–327. ISSN: 0953-5438. DOI: `10.1016/j.intcom.2010.04.004`. URL: `https://www.sciencedirect.com/science/article/pii/S095354381000038X` (visited on 04/15/2021).

[72] Mehmet Ilker Berkman and Dilek Karahoca. "Re-Assessing the Usability Metric for User Experience (UMUX) Scale". In: 11.3 (2016), p. 21.

[73] James Lewis, Brian Utesch, and Deborah Maher. *UMUX-LITE: when there's no time for the SUS*. Journal Abbreviation: Conference on Human Factors in Computing Systems - Proceedings Pages: 2102 Publication Title: Conference on Human Factors in Computing Systems - Proceedings. Apr. 2013. DOI: `10.1145/2470654.2481287`.

[74] Brian Hambling and Pauline Van Goethem. *User Acceptance Testing : A step-by-step guide*. 1st. BCS, May 26, 2013. 226 pp. ISBN: 978-1-78017-167-8. (Visited on 04/26/2021).

[75] Arbejdstilsynet. *APV i døgninstitutioner og hjemmepleje - Arbejdstilsynet*. Arbejdstilsynet. May 7, 2021. URL: `http://at.dk/brancher/doegninstitutioner-og-hjemmepleje/apv/` (visited on 05/07/2021).

[76] Danske Love. *Serviceloven § 83 a*. Danske Love. May 7, 2021. URL: `https://danskelove.dk/serviceloven/83a` (visited on 05/07/2021).

[77] Danske Love. *Serviceloven § 86*. Danske Love. May 7, 2021. URL: `https://danskelove.dk/serviceloven/86` (visited on 05/07/2021).

[78] Stephen J. Bigelow. *What are the types of requirements in software engineering?* SearchSoftwareQuality. Dec. 9, 2020. URL: `https://searchsoftwarequality.techtarget.com/answer/What-are-requirements-types` (visited on 04/26/2021).

[79] World Leaders in Research-Based User Experience. *User Need Statements*. Nielsen Norman Group. Mar. 24, 2021. URL: `https://www.nngroup.com/articles/user-need-statements/` (visited on 04/26/2021).

[80] Ivy Hooks. *Writing Good Requirements*. Writing Good Requirements. 1993. URL: `https://reqexperts.com/wp-content/uploads/2015/07/writing_good_requirements.htm` (visited on 03/02/2021).

[81] David D. Walden et al. *Systems Engineering Handbook*. Fourth. Wiley, 2015. 305 pp. ISBN: 978-1-118-99940-0.

[82] Mark Coppock. *The Best Web Browsers for 2020*. Digital Trends. Section: Computing. Mar. 1, 2021. URL: https://www.digitaltrends.com/computing/best-browser-chrome-vs-firefox-vs-safari-vs-edge/ (visited on 12/29/2020).

[83] Samantha Silver. *Three-Click Rule: UX Best Practices Revisited*. Apr. 18, 2018. URL: https://info.keylimeinteractive.com/threeclicksrule (visited on 05/23/2021).

[84] Shaumik Daityari. *Angular vs React vs Vue: Which Framework to Choose in 2021*. CodeinWP. Feb. 22, 2021. URL: https://www.codeinwp.com/blog/angular-vs-vue-vs-react/ (visited on 03/08/2021).

[85] Aris Pattakos. *Angular vs React vs Vue: Which Framework Is Better? 2021*. aThemes. Jan. 25, 2021. URL: https://athemes.com/guides/angular-vs-react-vs-vue/ (visited on 03/08/2021).

[86] Andrew Hurskiy and Sofiya Merenych. *Vue vs React vs Angular - Best Front-end Javascript Frameworks - Clockwise Software*. Nov. 26, 2020. URL: https://clockwise.software/blog/angular-vs-react-vs-vue// (visited on 03/08/2021).

[87] krausest. *Results for js web frameworks benchmark - official run*. 2020. URL: https://krausest.github.io/js-framework-benchmark/2020/table_chrome_87.0.4280.66.html (visited on 03/08/2021).

[88] VueJs. *Introduction*. Vue.js. July 3, 2021. URL: https://v3.vuejs.org/guide/migration/introduction.html#overview (visited on 03/08/2021).

[89] Systematic. *Columna Cura by Systematic*. Systematic - Columna Cura. May 7, 2021. URL: //systematic.com/healthcare/solutions/care/columna-cura/ (visited on 05/07/2021).

[90] HL7. *Index - FHIR v1.0.2*. URL: https://hl7.org/fhir/DSTU2/index.html (visited on 03/05/2021).

[91] REST API Tutorial. *What is REST*. REST API Tutorial. URL: https://restfulapi.net/ (visited on 05/23/2021).

[92] Systematic. *Cura snitflade beskrivelse for 3. part integration*. Mar. 20, 2018. URL: https://www.ethics.dk/ethics/publicTenderDoc/38ddbc08-0b64-4a82-a364-7dc7ef017270/15400063-766a-438f-909d-8cf28bda23d3/download (visited on 05/11/2021).

[93] Vuetify. *Get started with Vuetify*. Vuetify. May 6, 2021. URL: https://vuetifyjs.com/en/getting-started/installation/ (visited on 05/06/2021).

[94] Openbase. *25 Best Vue Component Library Libraries in 2021*. Openbase. May 6, 2021. URL: https://openbase.com/categories/js/best-vue-component-library-libraries (visited on 05/06/2021).

[95] cometchteam. *vue-svg-gauge*. npmjs. May 6, 2021. URL: https://www.npmjs.com/package/vue-svg-gauge (visited on 05/06/2021).

[96] OpenID. *OpenID Connect — OpenID*. Aug. 1, 2011. URL: https://openid.net/connect/ (visited on 05/12/2021).

[97] Systematic. *Columna Cura*. Sept. 17, 2019. (Visited on 05/11/2021).

[98] Jakob Nielsen. *Why You Only Need to Test with 5 Users*. Nielsen Norman Group. Mar. 18, 2000. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ (visited on 05/25/2021).

[99] Wayne W. LaMorte. *Confidence Intervals for Sample Size Less Than 30*. Apr. 21, 2021. URL: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module6-RandomError/PH717-Module6-RandomError11.html (visited on 05/13/2021).

[100] Philip B. Stark. *SticiGui Confidence Intervals*. SticiGui. May 13, 2021. URL: https://www.stat.berkeley.edu/~stark/SticiGui/Text/confidenceIntervals.htm (visited on 05/13/2021).

[101] Microsoft. *Microsoft Teams*. May 26, 2021. URL: https://www.microsoft.com/en-ww/microsoft-teams/group-chat-software (visited on 05/26/2021).