Web interface for machine learning as decision making support

Jacob Hegelund and Alexander Ljungqvist

DEPARTMENT OF DESIGN SCIENCES
FACULTY OF ENGINEERING LTH | LUND UNIVERSITY
2019

MASTER THESIS

Web interface for machine learning as decision making support

Published by
Department of Design Sciences
Faculty of Engineering LTH, Lund University
P.O. Box 118, SE-221 00 Lund, Sweden

Subject: Interaction Design (MAMM01)
Division: Department of Design Sciences
Supervisor: Joakim Eriksson
Supervisor: Björn Granvik, Softhouse Consulting AB
Examiner: Johanna Persson

Abstract

The thesis attempts to explore possibilities of implementing a user interface visualization to a machine learning algorithm. As the algorithm in the thesis does not take soft values into account, the visualization is meant to highlight its' drawbacks and aware the end user of how the results were generated. Therefore the graphical representation is aimed to fortify the algorithm as decision making support rather than decision making.

Machine learning algorithms differ thus making a general consensus towards a graphical representation non-existent. In order to emphasize the user interface aspects, the thesis writers gathered a selection of closely related phenomenon and research articles both within the fields of design process and machine learning.

The purpose of this thesis is to present the process to create a graphical implementation for a customized machine learning algorithm. The research questions consider technical and graphical aspects. How will software design and experience change based on new applicable elements of machine learning and artificial intelligence? How do one design a user friendly interface while showing and emphasizing a machine learning component?

The thesis was finalized in a Proof of Concept (PoC) version 1.0 as a web interface with root in conceptual models. The models created an entity, guiding a user towards interaction. The interaction helps the user understand the magnitude of the choices they are faced with, and furthermore help heighten the decision assurance when assigning a employee to a project.

The thesis writers seek to emit information of how a graphical representation to machine learning can be incorporated to yield an increased value of a software. The thesis itself gives an overview of machine learning basics and the design process towards applying a graphical representation on your algorithms.

Sammanfattning

Denna avhandling försöker undersöka möjligheterna att implementera ett gränssnitt för att illustrera en maskininlärningskomponent. Eftersom algoritmen i avhandlingen inte tar hänsyn till mjuka värden, är visualiseringen avsedd att lyfta fram medvetenhet för slutanvändaren av hur resultaten genererades. Därför syftar den grafiska representationen till att befästa algoritmen som beslutshjälp snarare än beslutsfattande.

Maskininlärningsalgoritmer skiljer sig ofta vilket innebär att en allmän uppfattning om en grafisk representation är idag vag. För att framhäva användargränssnittets aspekter samlade författarna ett urval av nära relaterade fenomen och forskningsartiklar inom både designprocess och maskininlärning.

Syftet med denna avhandling är att presentera processen för att skapa en grafisk implementering för en anpassad maskininlärningsalgoritm. Forskningsfrågorna behandlar både tekniska och grafiska aspekter. Hur kommer mjukvaruutveckling att förändras baserat på nya tillämpningsbara metoder av maskininlärning och artificiell intelligens? Hur designar man ett användarvänligt gränssnitt för att illustrera och betona en maskininlärningskomponent?

Avhandlingen resulterar i ett Proof of Concept (PoC) version 1.0 för webbgränssnitt baserat på konceptuella modeller. Modellerna skapar en entitet som vägleder användarens interaktion. Gränssnittet hjälper användaren att förstå valen som användaren visas för att ge användaren ett så bra bestlutsunderlag som möjligt när den ska tillsätta en konsult till ett uppdrag.

Författarna försöker ge information om hur en grafisk representation till maskininlärning kan införlivas för att ge ett ökat värde för en programvara. Avhandlingen ger en översikt för grunderna till maskininlärning och designprocessen för att tillämpa en grafisk representation av maskininlärningsalgoritmer.

Acknowledgements

We would like to thank our thesis advisor Joakim Eriksson, research engineer of Ergonomics and Aerosol technology faculty at Lund University. His door was always open whenever troublesome parts of the thesis occurred. He wanted our thesis to be the work of our own process, and only intervened guide us in a correct direction.

We would also like to thank our advisor at Softhouse Consulting, Björn Granvik, who continuously guided us towards a goal fitting for every interested party. He was an integral part in creating a thesis objective form the beginning and guided us towards what would be a final product.

We would like to thank Softhouse Consulting for the opportunity to carry out our thesis at the company.

In addition we would like to thank experts and test persons who played a very important role towards our product. Melina Katkic for superb guidance within machine learning, friends and colleagues for valuable input during design tests.

Contents

List of acronyms and abbreviations

AI = Artificial Intelligence

API = Application Programming Interface

CSS = Cascading Style Sheets

HTML = Hypertext Markup Language

KUSH = Competence system at Softhouse Consulting AB

Lo-Fi = Low-Fidelity

ML = Machine Learning

Mid-Fi = Mid-Fidelity

NLP = Natural Language Processing

PoC = Proof of Concept

SUS = System Usability Scale

TP = Test participant

UI = User interface

UX = User experience

1 Introduction

1.1 Background

On the seventh of October 2014 Zed Johnson wrote for the independent that the amount of electronic mobile devices passed the amount of living people on the earth [1]. At the time he referred to GSMA intelligence, a statistics bureau. Since then electronic devices have outnumbered the amount of human beings on the planet by roughly 1.2 billion devices. When comparing the statistics for global data on population [2] compared to electronic population [3] the electronic devices are increasing at roughly four times the human population rate. In a world where electronic devices are about to greatly outnumber the population, their data is of big importance. Almost any device work with some form of data. The data from a device alone might not contain groundbreaking information, but when the data from several thousands of devices are brought together, patterns become detectable.

Further utilization is using existing information to predict coming information, with complex mathematics. As a result of impressive development from big companies, machine learning and artificial intelligence are now major stakeholders in a data driven society.

> *"Machine learning brings the promise of deriving meaning from all that data."*[4]

Machine learning is the tool used to answer questions based on your data. As the volume of data goes beyond human comprehension and therefore human capabilities of making decision of it, machine learning has emerged as preferred method for decision making.

Machine learning is applicable all over, from agencies creating systems to detect fraud, to pattern recognition and even a small task as figuring out if a certain object is a glass of beer of not. The approach to this thesis will be the possible adaption of frontend frameworks and graphical interfaces to machine learning. The thesis is written with two aspects in consideration.

- *User Interface and Experience - How will software design and experience change based on new applicable elements of machine learning and artificial intelligence?*

- *Design process - How does one design a user friendly interface while showing and emphasizing a Machine Learning component?*

The design process is of very high importance. In order to get value from complex algorithms, a system needs concrete guidelines for users. The target group might have little to no knowledge within software and/or software design. This calls for design processes involving certain steps to validate that users are able to fully grasp and understand a system, see figure 1. The thesis will cover each element in how to create a solid software graphical system with machine learning as decision making support.

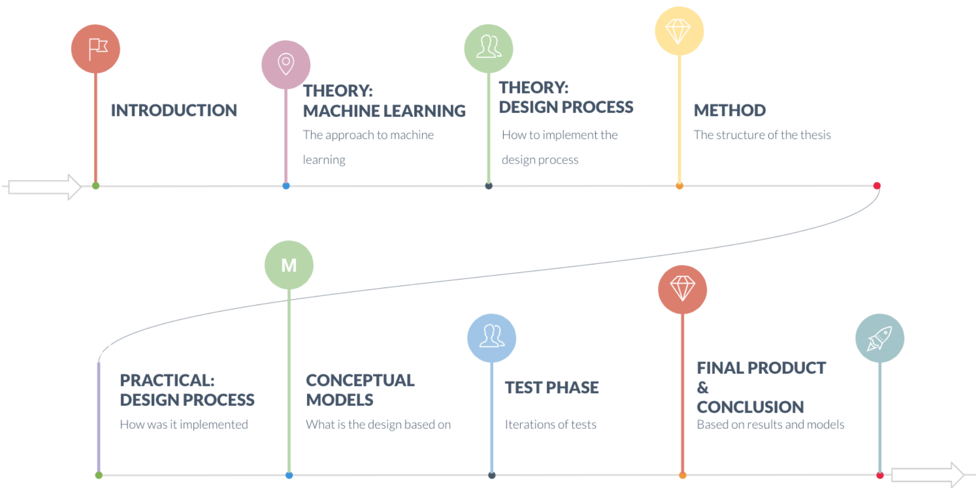

Figure 1: *Brief overview of thesis layout.*

1.2 Softhouse Consulting AB

Softhouse Consulting AB is an IT consultant company based in Malmö. Softhouse offers a broad variance of competence, mainly focused on software development and business development with an agile approach. To keep track of their different projects and the competences within the company Softhouse uses two different systems, KUSH and Stardust.

The interest for a broader AI knowledge at Softhouse combined with a great support from Softhouse to the thesis writers was the foundation for making Softhouse a great place to meet the purpose of this thesis.

1.2.1 KUSH (Competence system at Softhouse Consulting AB)

KUSH is a web based system that Softhouse Consulting uses to keep track of their employees and the employees skills and experience. The employees have their own profile where they get to set what skills they have and to what level, and what skills they want to learn and to what level they want to improve.

In this thesis we explore the opportunities to use machine learning to further develop KUSH and make sense out of all data on their employees. Can we use machine learning and the data in KUSH to match a user to a certain project? Do we need machine leaning to solve this problem? To make it reasonable using machine learning we need a problem complex enough, meaning a problem that would take to many rules and lines of code to solve ourselves.

Considering the amount of skills an employee might have and the different combinations of skills each employee can have, the problem will probably be complex enough to use machine learning. This means that to be able to illustrate which employee should be assigned to a certain project we need machine learning to solve the problem. For a user to understand what a machine learning algorithm is telling the user, design and illustration of a machine learning decision is the main focus. This leads to the human-machine interaction and designing a machine learning component. Without machine learning the illustration would not be possible and to be able to display the information and to make sense out of the data, the main focus of this thesis is the design and new possibilities of human-machine interaction. This problem shows the advantages of both developing and designing a proper machine learning component.

1.2.2 Stardust

Stardust is a system that Softhouse uses to keep track of all their projects. A project is added to Stardust with a project description and other information about the project. Stardust and KUSH are two independent systems that Softhouse uses to keep track of their employees and projects. The problem is that these systems are independent of each other and Stardust keeps track of the projects while KUSH keeps track of Softhouses' employees and the competences within Softhouse. If you want to find an employee for a certain project you need to manually do the search for the right competence in KUSH.

1.3 Problem definition

KUSH holds the data of what competences the employees at Softhouse have and Stardust holds the data about all the projects that Softhouse have. Stardust currently has no greater function for interaction with the competence system. Today there is a need to use both systems in order to find the right person for a certain project.

The goal is to create an intermediate system between these two systems. The intermediate system will be an identity and its function is to incorporate project information from Stardust with information from KUSH. The system is meant to be intelligent, meaning it will implement a few machine learning algorithms. The goal for the system is to match employees in KUSH with a project created in Stardust. It will read the project description, find keywords and match them with the most suitable employee from KUSH. To make decisions it will use machine learning and a natural language processor.

To display the result of the systems' findings, a web interface will be used. The design of the system will be the core of the design process. The design will be an implementation following Softhouse guidelines, with a distinction of what is machine learning generated information.

1.4 Final goal

The final goal of the thesis is to study the user interaction with machine learning component by creating a Proof of Concept (PoC) user interface design based on machine learning algorithms.

The PoC is a web based site to meet the needs of Softhouse consulting by create a web interface with projects overview, project page, employee profile and add employee.

When Softhouse get a new project, the project and the project description is added to Stardust. An HR manager should then add the most suitable competence for this project based on the project description and the competences for each employee in KUSH. The new system/web interface should based on machine learning and communication with both Stardust and KUSH suggest the most suitable competence. This implies that the HR manager should be able to do the following in the PoC:

1. Find a specific project.

2. Determine the best suited employee for the project.

3. Choose and add this employee to the project.

1.5 Target group and users

The purpose for the thesis is to investigate the common interaction between users and machine learning through a broader perspective. Therefore, the test iterations will use test persons with varied backgrounds and IT knowledge.

The PoC is developed for project managers at Softhouse Consulting when assigning employees to projects.

1.6 Limitations

The thesis will be limited to the extent of User Interface (UI), User Experience (UX), Machine Learning (ML) for a smaller product with a probable size of 2-4 UI pages. The limit will be within the extent of the final thesis, which is equal to one semester per thesis writer. The thesis will limit machine learning to basic features and when possible, use already created packages. Since machine learning is a major factor to this study, a lot of time will be spent on analyze machine learning, what data is available and how it can be used to then decide relevant user interface components.

2 Theory: Machine Learning

2.1 TensorFlowJS

There are a lot of different machine learning frameworks to use if you want to develop your own machine learning program. Google have released a framework called TensorFlow. This is a framework that you, as a programmer without any machine learning knowledge, can use to set up your first machine learning modules and the framework is written in Python [24]. In March of 2018 Google released a framework called TensorFlowJS which is built on TensorFlow, but is developed for the web using JavaScript [25]. This thesis mainly focuses on web development combined with machine learning, which makes TensorFlowJS the perfect framework to go with.

2.1.1 Basic knowledge to use TensorFlowJS

The TensorFlowJS framework is quick and easy to get started with, but there are some basic terms and knowledge that one preferably has to have some experience with. The list below explains the basics to some terms that the thesis writers needed to know help them to get started.

- Tensor - A tensor is an array of a data set of numerical values. The data set can have the shape of one or more dimensions, which means that the tensor array is an array containing a data set of one or more arrays. Each array in the data set should have the same length, which means that the tensor is set to a specific shape. A tensor is defined as either the input value or the output value of a model [26].

- Features - An input tensor consist of input values, where each value is labeled to a feature. This means that an input tensor is an array where each value represents a value for a feature.

- Label - An output tensor consist of output values, where the values are called labels and each label is the result of an input array.

- Model - A model is basically a function that given a specific input returns a desired output. TensorFlow provides an API that constructs models out of layers. Layers are used to build a neural network for deep learning [26].

- Layers - A layer is used to build a neural network for deep learning. There are input layers, hidden layers and output layers, where a tensor model constructs the hidden layers. A neural network can consist of one or more layers depending on the complexity [27][28]. A layer consist of a number of nodes, or neurons, and each node takes an input and produces an output through the entire network.
- Activation function - An activation function is a crucial part for a artificial neural network since an activation function lets the network handle non-linearity. The activation function defines the output, given an input, for each node [28][29]. When setting up a tensor model you have to define what activation function to use.
- Training data - To train the model and your neural network you need a larger set of training data that the model iterates through. The training data consist of a set of input data (an input tensor) and a matching set of output data (an output tensor). The model hereby learn what output a specific input corresponds to. It's important to have broad variance in the training sets.
- Test data - Test data is a smaller data set of input to verify that the model returns the correct output.

2.1.2 Supervised machine learning

Supervised machine learning is when a function (TensorFlow model) is given data sets with input values labeled to specific output values as training data. The function should then be able to make future predictions for unlabeled input data, and produce the correct corresponding output data [30]. What makes supervised machine learning is the labeled output data, in contrast to unsupervised machine learning where the function is given input data without being labeled to specific output data and the function finds the patterns itself [31]. When using TensorFlowJS in this project, the thesis writers will create models that uses supervised machine learning.

2.2 Natural Language Processing (NLP)

Natural language processing (NLP) is a subsection of artificial intelligence used for interpreting human language. The methods for interpreting language range between statistical and machine learning methods, to a more rule-based and algorithmic approach.

The statistics analytics company SAS [32], number 131 on Forbes list over America's largest companies [33], defines natural language processing as a method to break down language into smaller elemental pieces.

The pieces are then utilized to understand the relationship and examine how they interact in order to create meaning [32]. A way to utilize NLP is to use an iterative force to gather keywords. Keywords can be set manually in order for the algorithm to locate and process, and can then have their environment analyzed to create patterns that might be of interest to the end user.

2.3 Data

"Every learner must embody some knowledge or assumptions beyond the data it's given in order to generalize beyond it. This was formalized by Wolpert in his famous "no free lunch" theorems, according to which no learner can beat random guessing over all possible functions to be learned" [34].

The most essential part of a functioning machine learning product, is the data the model is applied on. The data in machine learning constructs the base pillar of a neural network. The data is what machine learning algorithms work on, and if it's not sufficient, neither will the results be. According to Melina Katkic, co-founder of Barrel.ai[35] and Nordaxon[36], the data actually outweighs the model in terms of impact on the system.

When using machine learning, as mentioned above, the data is the most crucial part. To start designing what models to use, one has to analyze what data is available and in this case what kind of data KUSH and Startdust holds. Analyzing the data also helps deciding what features the system should provide, and thereby also what to design. Since this thesis use supervised machine learning, it gives the structure and purpose of the data an even greater importance for a well trained machine learning model.

2.3.1 KUSH data

Each employee at Softhouse add their competences in KUSH. They register what skills they have and to what extent they manage this skill. Each skill is grouped to a skillgroup, which creates a group of skills that is considered to be connected to each other. This can then be used to train a tensor model to evaluate to which group of skills an employee has strengths and to what extent.

By using the average skill level in every skillgroup for each employee, we get an overview of each employees' skillgroup levels. For each employee, the average skillgroup levels can then be used as features (input values) for a tensor model and the average skillgroup level, in relation to the overall skill level for the employee, as the output for the training data. This gives us a tensor model that, given keywords from Stardust (skillgroups), can suggest the most suitable employee based on skillgroup levels.

As mentioned in section 1.2.1 KUSH, an employee can also add skills they want to learn and to what extent. This data could also be used to match an employee to a project.

2.3.2 Stardust data

Projects are registered to Stardust with a project description. By fetching all skillgroups from KUSH, a NLP should be trained to analyze the project description and find the keywords (skillgroups) for a project.

2.3.3 False positive

Depending on the data sets and consequently how you train your machine learning model, the results' reliability can vary a lot. A poorly trained model can give you false positive results, meaning a result that seems to meet the expected value but that is actually false. The probability to get a false positive result should be as low as possible [37]. To achieve a low probability of false positive is depending on what data to use, the variety and amount of the data and what features and labels to use.

2.3.4 Overfitting

Overfitting means that a machine learning model might learn a pattern to well, in other words the training data lacks enough variety. This implies that the model might not be able to predict future input values correctly [38]. One thing that might help avoid overfitting is to make sure that the training data have enough variance which once again also implies enough amount.

2.4 Narrow AI

The term AI is nowadays a more general term of what people might think that artificial intelligence is, such as a robot acting as a human. The term can be divided into the subsections machine learning and narrow AI. Machine learning and narrow AI differs from general AI since it focuses on solving smaller but still complex problems such as image recognition. The most common applications today uses narrow AI and is used to solves smaller problems to learn and act as a human for only a certain field instead of completely act as a human being [39].

This project focuses on building smaller models to each solve a narrowed problem.

3 Theory: Design process

To give the reader content to the context: designing a product, this chapter aims to show how the thesis writers approached the design process. The thesis writers decided that the core of the process should relate to the fidelity prototype principle, in which you have Low fidelity - Mid fidelity - High fidelity [11]. To better substantiate the most suitable process, the thesis writers altered the main pillars as the prototype levels were not optimal for the design process execution. The design process is constituted by three main pillars that are supposed to be equivalent to the fidelity prototype levels:

- Flow Sketching
- Wireframes
- Proof of Concept versions

These symbolize the different stages in the design process. Between the process stages, confirmation of each phase is done. Unless a level has been confirmed, reiterations on the current level will be held. The reiteration idea comes from the agile approach the thesis writers chose for the project [5].

3.1 Brainstorming

It's very common in a design process to begin with knowledge from the projects' participants. In order to create an idea friendly environment where the whole project group is invested in ideas, a brainstorming session is suitable [6]. The word brainstorming was originally introduced by Alex F. Osborn in 1953 through his book Applied Imagination: Principles and Procedures of Creative Thinking. A brainstorming session is based upon his four concept rules of brainstorming: [7]

- No criticism based on ideas: To not prohibit participants from coming up with any type of ideas. The focus should be to generate ideas, not defend them.
- The more the merrier: The belief that quantity will generate quality is one thing Alex F. Osborn was certain of. The more ides generated by the session, the likelier it is that at least one holds value.

- Benefit from each other: Just because someone else brought the idea to life doesn't mean you can't build on it. Brainstorming is not a single player game, rather cooperation is what will generate value.
- It's not stupid if it works: Come up with any kind of idea, just bring something to the table. An idea never spoken of will never become something. As Alex himself said:

> *"It is easier to tone down a wild idea than to think up a new one."*
> [7]

3.2 Flow sketching

Visualization early in the process holds value in form of a brief overview of the system. Flow sketching is a perfect suit for the cause. A flow sketch is a plain technique to achieve a fast, cost efficient way to explore key concepts of the design. Normally, flow sketches are conducted with most basic materials such as paper and pen. Flow sketching holds the possibility to rapidly vary design ideas in order to claim a suitable design for more stable and serious development. The main reason for using Flow sketching is as the name suggests, validating the flow in the product. Flow sketching should only contain the most basic logical structure as the sole purpose is to show interested parties design paths and logical structure.

3.2.1 Paper sketching

Paper sketching is an easy technique which utilizes paper models to replicate functional applications. A paper sketch represents the desired system, created with paper and pen, and is often used as a first draft. The technique is smart and valued highly because of the rapid level of adaption. As a paper sketch is conducted early, major uncertainties may arise which can be remediated during iterations of developing.

3.3 Wireframes

Building upon the idea to step further into the design process, chosen counterpart to Mid-Fi prototyping was wireframes. Wireframes are a visual skeletal representation of a product structure.

A wireframe prototype is a prototype that contains basic functionality based on the flow sketch. A wireframe typically gives functionality for traversing between pages and contains interactive areas throughout the prototype. To create a slightly functional graphical design, the framework Sketch suits the cause. Sketch is a framework for creating fast prototypes and holds possibilities to make quick changes. A tool like Sketch can be utilized effectively as flow and logical routes are very comprehensible. Using wireframes is a cost- and time efficient way to work with functional navigation. As navigation is easier compared to earlier stages of the design process, validity aspects in terms of design choices can be established. The main purpose of a wireframe sketch is to enhance the validity of a prototype structure. Wireframes are not supposed to contain choices in terms of graphical design, rather showing only the logical layout and interactive elements.

As Garrett and James state in their book "The elements of User Experience" [15], a wireframe must incorporate all variants of navigation systems and all interface elements that require any type of functionality in a web page. What matters about wireframes are that the page layout, information-, interface- and navigation design come together to form a unification of the product to be. A wireframe can be seen as a template for further structural- and design implementations.

3.4 Proof of Concept (PoC) versions

The PoC versions is the last step in chosen design process. A PoC typically includes functionality thus requiring more time to develop. The PoC should be a reflection of the design process and the input from each previous step. The PoC can also be seen an iterative process where UI/UX flaws can be updated properly. The main reason for creating a PoC is to ensure and validate the usability for the targeted audience.

A PoC is created with similar methods as the main product, giving it similar forms of interaction. A product with a close to finished design can be interacted with while costing less time to create. The PoC give a good pillar to insinuate the amount of resources for creating the product. A PoC of this kind also gives every party involved possibilities to understand the involvement of each participant. The involvement will most likely differ from different parties, and knowledge of the work balance between parties could give the project manager a possibility to better estimate cost.

The PoC arrangement in this thesis is divided in three parts. The first PoC is a draft (0.9), meaning it still needs to be honed in order to become a final product. The middle PoC is the first updated version based on the first test iteration and conceptual models and the last PoC is version (1.0) and equal to a final product based on all previous test iterations.

3.5 Usability testing

Usability tests on a fidelity level prototype help locate shortcomings of the design and can support ideas of improvement. The goal with usability testing is to gain a greater insight in how a user experiences a product and based on the experience, improve the product [8]. A usability test is often set up for a test participant to achieve a certain goal or task.

3.5.1 Cognitive walkthrough

To get feedback early in the design process, a cognitive walkthrough was made on two different prototypes. The usability evaluation technique cognitive walkthrough is a technique that is often used early in a design process. It is a technique for evaluating whether a design meets a purpose to the user and can be done on early prototypes to give the designers guidelines on how the design is experienced by the user. A cognitive walkthrough is cheaper than complete usability tests both in terms of money and time and is a great way to find design flaws or get general input from users on how they experience the product early in the design process. As stated by John Rieman, Marita Franzke and David Redmiles in their report Usability Evaluation with the Cognitive Walkthrough, a cognitive walkthrough should consist of the following three prerequisites [43]:

- A general description of who the users will be and what relevant knowledge they possess

- A specific description of one or more representative tasks to be performed with the system
- A list of the correct actions required to complete each of these tasks with the interface being evaluated.

3.5.2 Think aloud protocol

A proposition from Cambridge claimed that verbal data accounts for data as much as any other source of data [9]. That idea was built upon and in 1982, Clayton Lewis founded the think aloud method [10]. He explains his findings in his text "Task-Centered User Interface Design: A Practical Introduction" where he explains the thinking-aloud method as:

> *"A method for studying mental processes in which participants are asked to make comments as they work on a task. The method is appropriate for studying the cognitive problems that people have in learning to use a computer system."*

Utilization of a think aloud protocol may help designers widen the aspect of their product while taking in the testers' cognitive measurements. Jakob Nielsen (author of the book usability engineering [12]), goes as far as saying:

> *"Thinking aloud may be the single most valuable usability engineering method."*

Further he claims in an article 19 years later, that he stands by his word and that the longevity as #1 method is a good indication of value [13].

3.5.3 System Usability Scale (SUS)

To measure the usability of the PoC, usability tests followed by a System Usability Scale (SUS) survey were conducted. The SUS has become a widely used method to measure the usability of a system, originally created by John Brooke in 1986 [44]. The survey consists of ten statements were the test participant decides on a scale from

1-5 if they strongly disagree och strongly agree with the statement. The statements alternate positive and negative statements about the system. The result is a score from 0-100 where 68 is the break point. A score above 68 is considered above average, and therefore good usability, and below 68 is considered below average and therefore not good usability of the system. One advantage with SUS is that it can be used on small sample size and still give reliable results [45].

4 Method

The methodology for the thesis has roots in the design process and the fidelity levels of design will act as key elements. The thesis will require an fusion of qualitative and quantitative measurements to reach a suitable result. The quantitative part of methodology will be data collection from user tests and modeling from machine learning. The qualitative part of the methodology is an evaluation regarding the understanding of interactive and intelligent user interfaces.

The quantitative approach is straight forward, consisting of test conduction and machine learning algorithms. The qualitative approach can be difficult, as the goal is to find soft values. By soft values, the thesis writers mean values that cannot be quantified and are strictly personal, reflect a certain persons' experience and mapping. These values will differ between test tasks and are not likely to be the same, making data analysis hard. To overcome and extract knowledge from soft values, a cognitive walkthrough is an alternative. The cognitive walkthrough will take in consideration a persons' certain experience and profession. A test subject will dive into their projection of the profession that will illustrate the preferred end user.

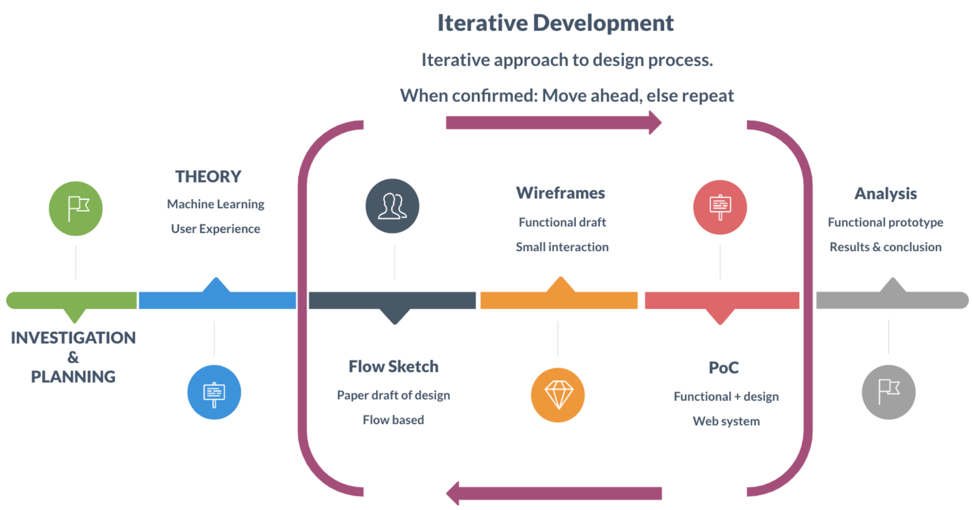

Figure 2: *Method for thesis execution, from design process to proof of concept.*

Figure 2 illustrates the execution of the design process. The entire process is agile, meaning iterations and phases are integral for the proceedings in the process. The core consist of three main pillars: Flow sketching, Wireframes and a Proof of Concept. In the figure, the process encloses these three to show that they are agile and iteration based.

Between the process parts there are elements for enhancing the value of the previous and coming processes. Prior to the flow sketching process there is mind mapping or brain storming. A brain storming can be seen as an informal concept recipe to what the system needs to contain. A flow sketch occurs as a reasonable next step, using the ingredients and knowledge gained from a brainstorming process. To ensure the value of the flow sketch and excel into wireframes, the thesis writers created many flow sketches and discussed which would suit best. The validity of wireframes roots in a discussion based approach between the thesis writers and the supervisor at Softhouse. When both parties considered the wireframes iterations successful, the design process advanced. The last step in the design process is the PoC. The PoCs are created in different versions, as they will be altered upon test feedback. There will be three iterations of testing the PoC, where each iteration can be seen as a complete test phase. Each test iteration will have test participants with different prerequisites in order for the thesis writers to get feedback from more than one type of user.

Based on the flow sketches, wireframes and a first iteration of cognitive walkthrough conceptual models is developed by the thesis writers. These models are implemented to a second and third iteration of cognitive walkthrough.

The process will follow the structure of an agile process and especially the Scrum approach. The thesis strongly incorporates the key elements of Scrum. Scrum is closely related to an agile approach, but goes further in terms of specifications and definitions, to become a more transparent process which is easier to handle [16][5]. The main element of Scrum is a sprint. A sprint is a shorter interval where a specific goal is set to get one step closer to the final product. As Ken Schwaber and Jeff Sutherland State in their writing "The Scrum Guide", the core of an agile software process lies within a development sprint, which is one month or less [18]. In this project, sprint intervals of one week were set. This approach creates an ever changing product that can easily be applied to demands from external parties.

4.1 Software development

In order to supply the design process with landmarks, an analysis is initiated in consideration of an initial understanding of machine learning, components and what type of data is available. Based on the analysis the design process can take shape. In order to substance choices, and give content to design process, development cycles need to proceed accordingly. The development cycle is based the Manifesto for Agile Software development and Scrum [17][18].

The development cycle follow an agile structure with five key elements:

- Analysis of requirements
- Design (architectural)
- Implementation
- Testing
- Analysis of results

In the beginning of an iteration, requirements are specified and thoroughly written. The requirements are then used to create a development design overview as foundation for the implementation step. The implementation step is solely development. During development it is important to create a product that is testable, in order to ensure the value for forthcoming iterations. The analysis of results will lay base to next iteration, completing the development cycle. A new iterations begins directly after a conclusion has been drawn form a previous.

5 Designing a product

5.1 Brainstorm

Brainstorming in a friendly environment creates possibilities for everyone to be part of the direction for the project. If participants are unavailable to attend a meeting to brainstorm there are several tools for brainstorming. As the thesis writers were not able to sit down together, a framework for creating a valid mind map as result from the brainstorming was used. The result was a mind map, see figure 3, containing the most integral parts of the project. The mind map is feasible milestone to further develop and work towards a flow sketch.

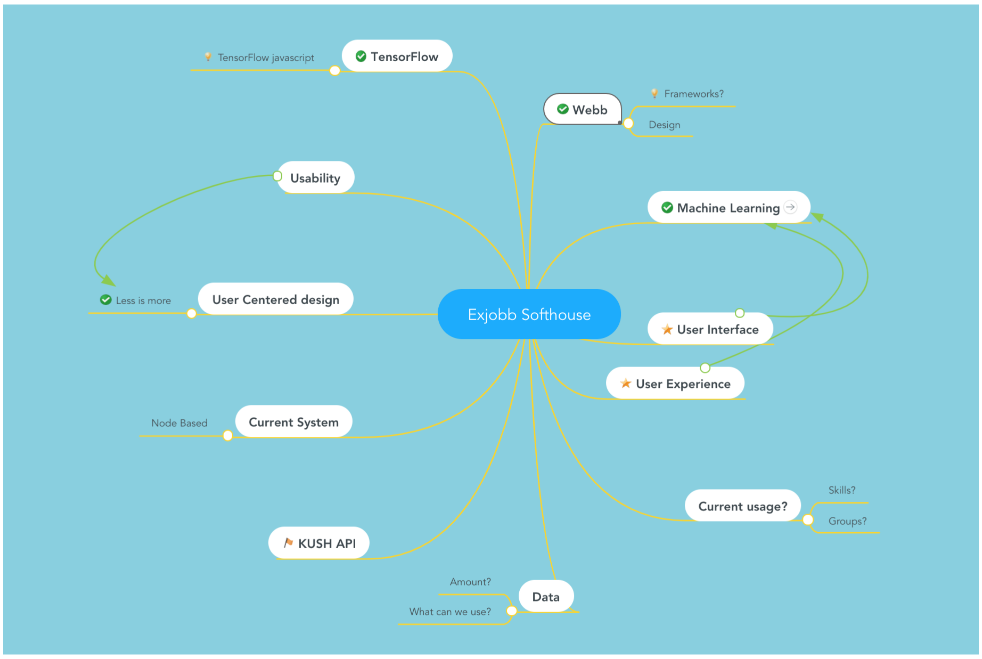

Figure 3: *Mind map from brainstorming session.*

5.2 Flow sketch

A very rapid way to generate an overview of the system is a flow sketch. The sketch is strictly supposed to give a modest picture of the flow and base pillars of the system and only contain the most basic design. Generally a sketch is done with paper and pen, removing a viewers' thought of superficial parts of the system.

Figure 4 shows an early illustration of the flow through the system. Each step in the figure is then explained.

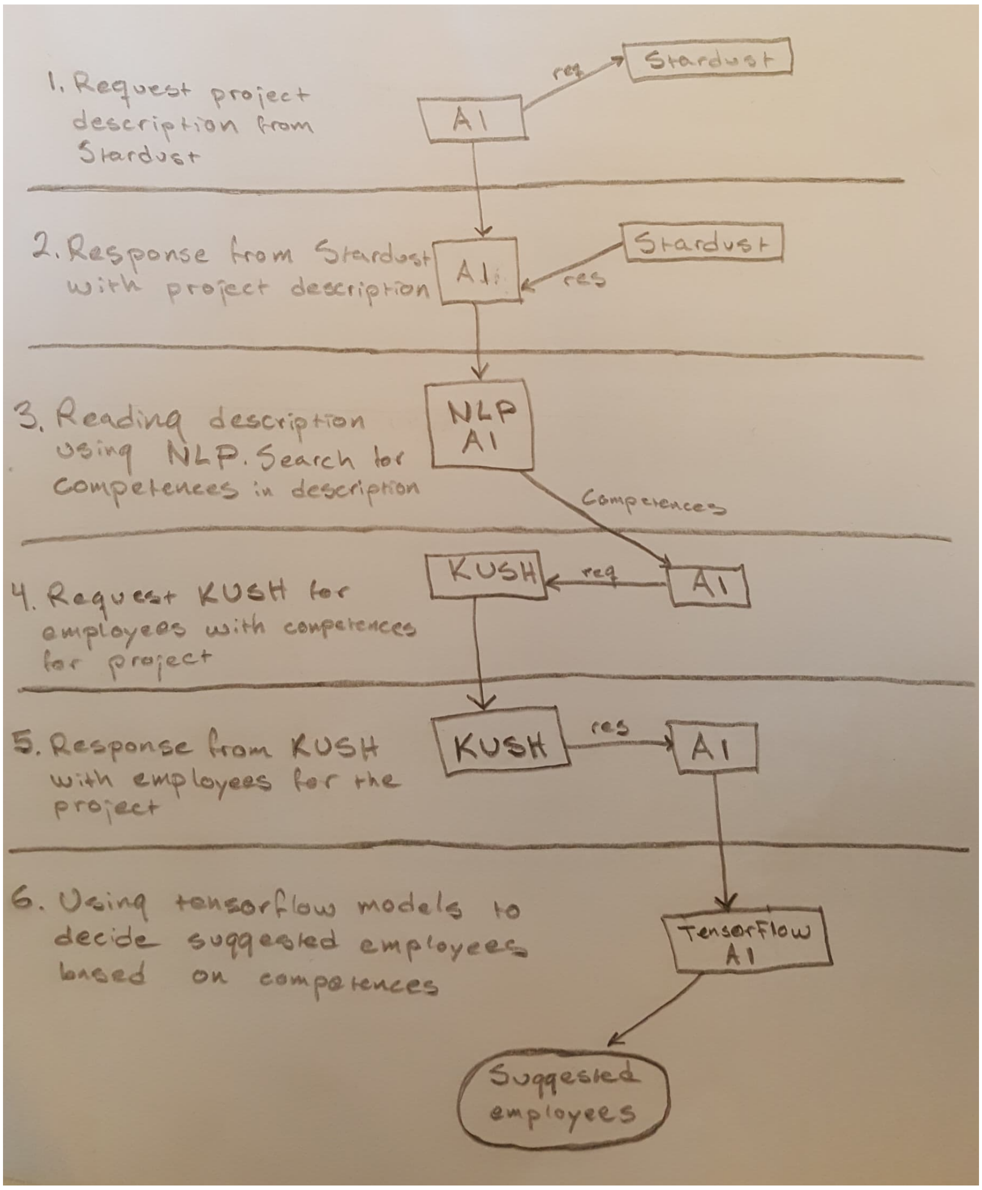

Figure 4: *Illustration of the flow for the system.*

1. Request project description from Stardust.

2. Getting response from Stardust with project description.

3. Reading project description using NLP. Finds keywords and uses the keywords to request KUSH.

4. Request KUSH for employees based on keywords in project description.

5. Getting response from KUSH with employees based on project description.

6. Using TensorFlow models to decide suggested employees based on project description.

Figure 5 shows an early drawing of the system overview. The AI component request Stardust for a project description, uses NLP to read the project description and find keywords. Then the AI component requests KUSH for employees based on the keywords and uses TensorFlow models to decide the most suitable employee for the project.

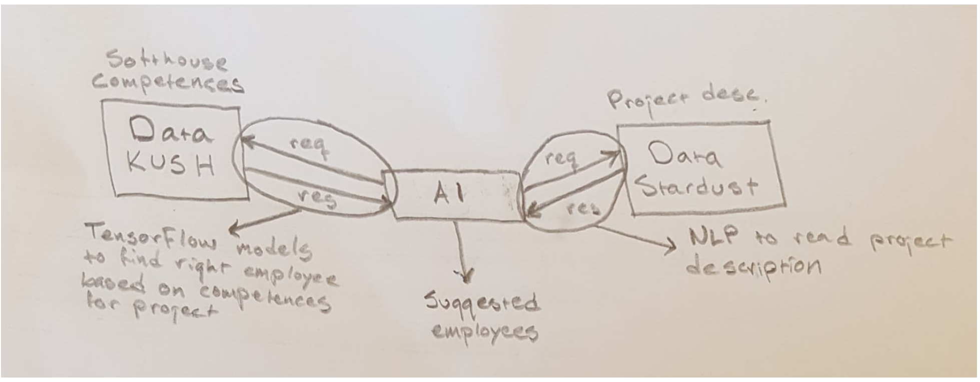

Figure 5: *Illustration of how the system works.*

The sketch created held the possibility to navigate by hand and is very applicable to feedback through various type of tests. The chosen test was navigational while thinking out loud. The tests were executed between the thesis writers to create a common idea of the product flow.

5.3 Wireframes

Following the flow sketching, a more developed sketch of the system took form. The main reason is to create a first implementation with a functional tool. The functional tool will help ensuring the flow of the system without being design specific.

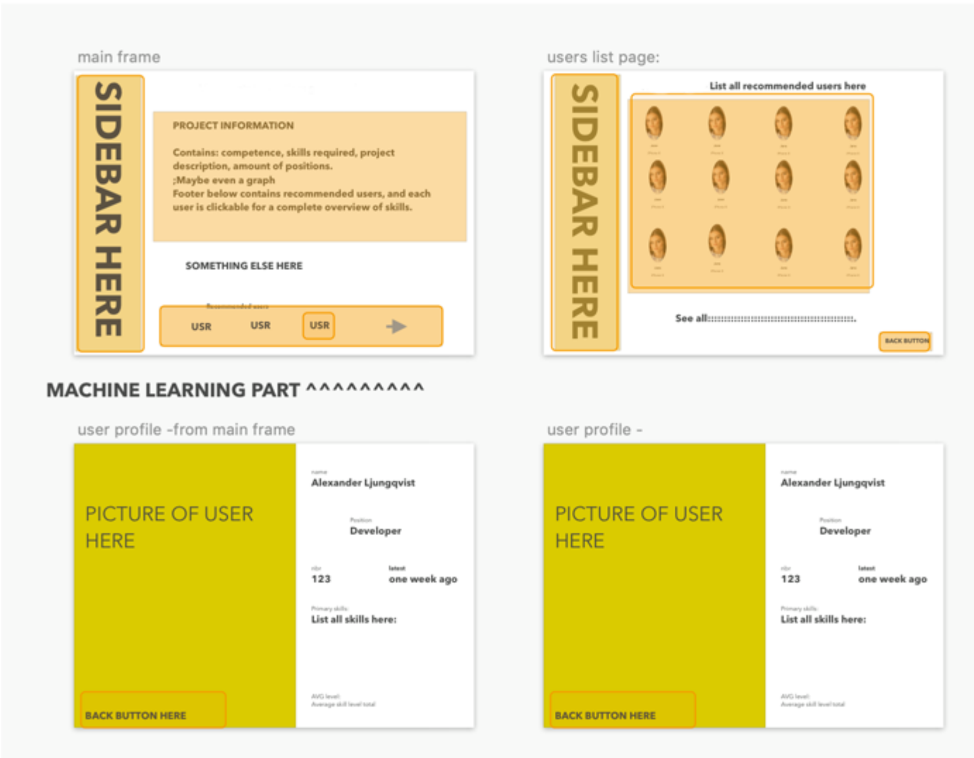

Figure 6: *Wireframes.*

Figure 6 shows a basic system easy for testing, as the sketch app contains functions for travelling between pages. The design is kept very basic to eliminate the risk of testers focusing on graphical design rather than the flow. Once the flow is decided and suit the needs from interested parties the prototype can evolve towards a proof of concept.

5.4 Proof of Concept (PoC)

M. Walker with co-writers [19] state that there is insufficient research to indicate which level of fidelity and media will produce the best feedback from users. Their conclusion was that it is not of highest importance to use a certain type of prototyping but rather knowing your medium and your end users. Due to time limit, and a significant amount of time spent on machine learning hence data analysis, a PoC of a web page was used.

The PoC contains the flow from the flow sketches, the logical structure of the wireframes and bootstrap themed design as part of the user interface. The first draft of the PoC is shown in 7, 8 and 9. The draft is used for the initial iteration of the test phase. The design choices are discussion based and represent a preferred view for the thesis writers.

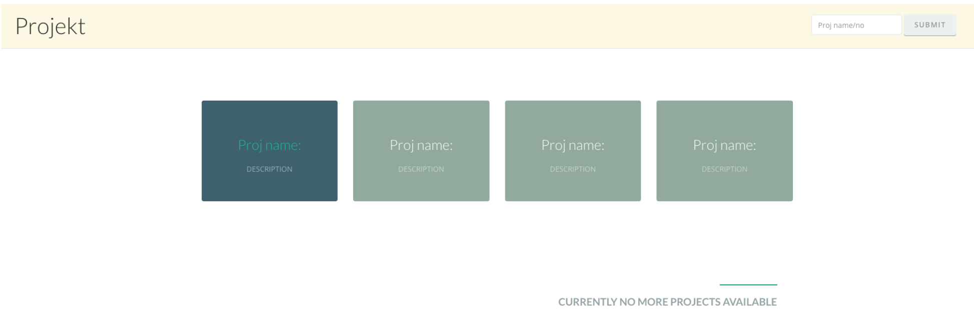

Figure 7: *Index page for PoC, showing the view of all projects.*

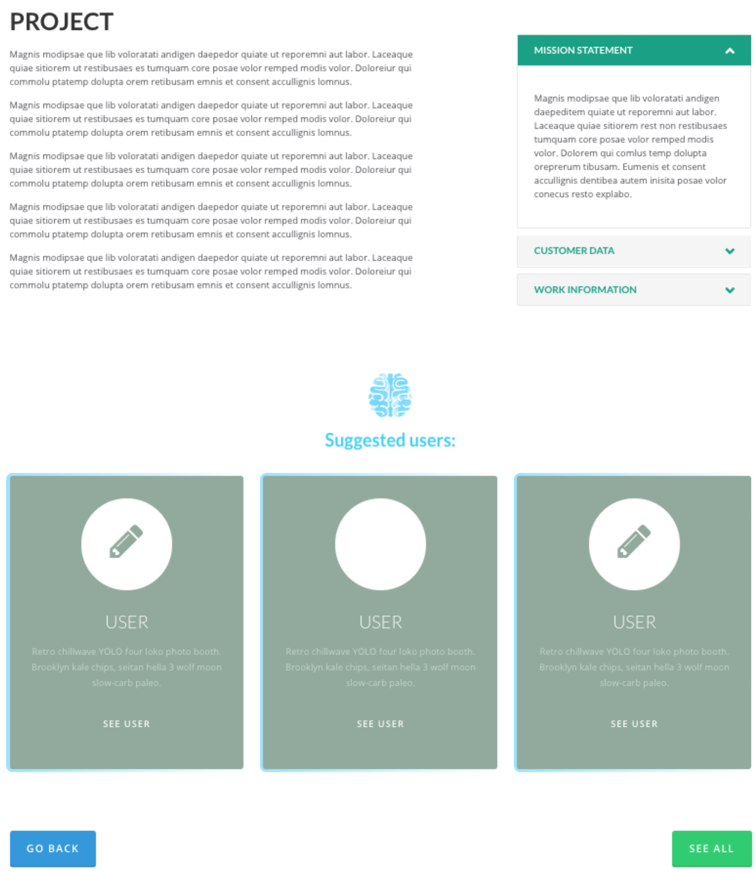

Figure 8: *Project page for first draft of PoC.*

The amount of suggested employees for the project page is limited to three, seen in figure 8, as the thesis writers considered the amount suitable. The main reason to the number of three is to create a slim design and keep the user focusing on the task. For example, if two suggested employees were to be used, a direct comparison would be more attractive, leading to possible blunders around the soft values. As well with four, the amount of time spent to find a suitable employee could, according to the thesis writers, become inefficient.

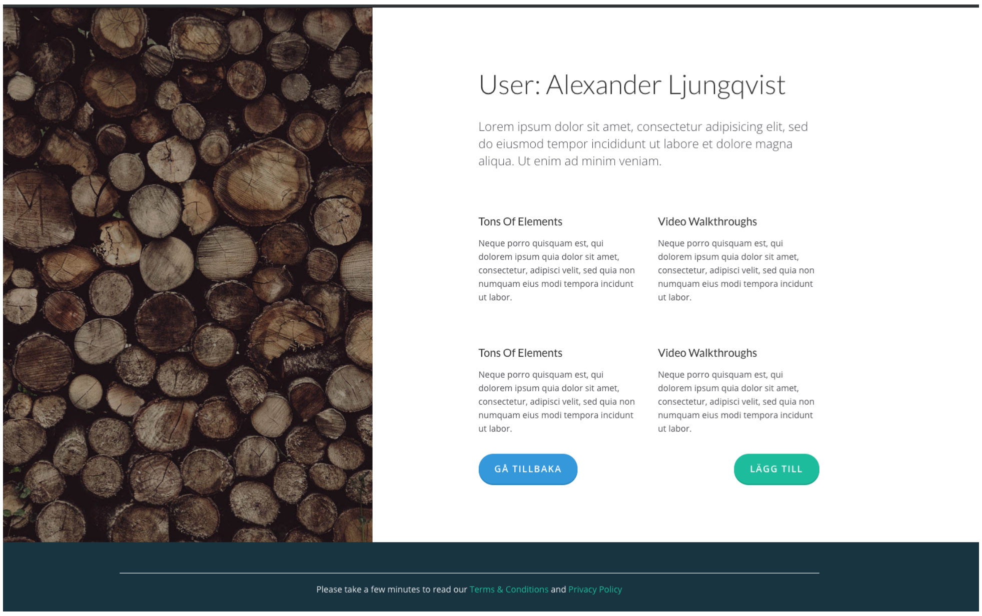

Figure 9: *Display of user profile in PoC.*

During test conduction of the PoC, main focus should be design choices. The testing of the PoC is very iteration based, small errors are gathered and fixed after each iteration. Generally a perfect design is not achieved from the beginning. An easy way to keep track of errors and tips is using a list where a test conductor takes note on each flaw they find during the tests.

6 Cognitive Walkthrough - first iteration (PoC 0.9)

To ensure the usability of the system in an early stage, test iterations with cognitive walkthrough was made. An iterative design process were used through the entire project, where continuous feedback and improvements were taken into consideration. The first cognitive walkthrough were made with two groups. One control group and one test group. The control group performed tests on more regular design, see figure 10 and the other group tested the thesis writers suggestions for AI design, see figure 11. All participants had the same prerequisites and were employees at Softhouse Consulting.

6.1 Cognitive walkthrough set up

To perform the cognitive walkthrough, six persons working on Softhouse were chosen to participate as test participants. The test participants were split into two different groups, three test participants per group, who evaluated two different prototypes, see figure 10 and figure 11.

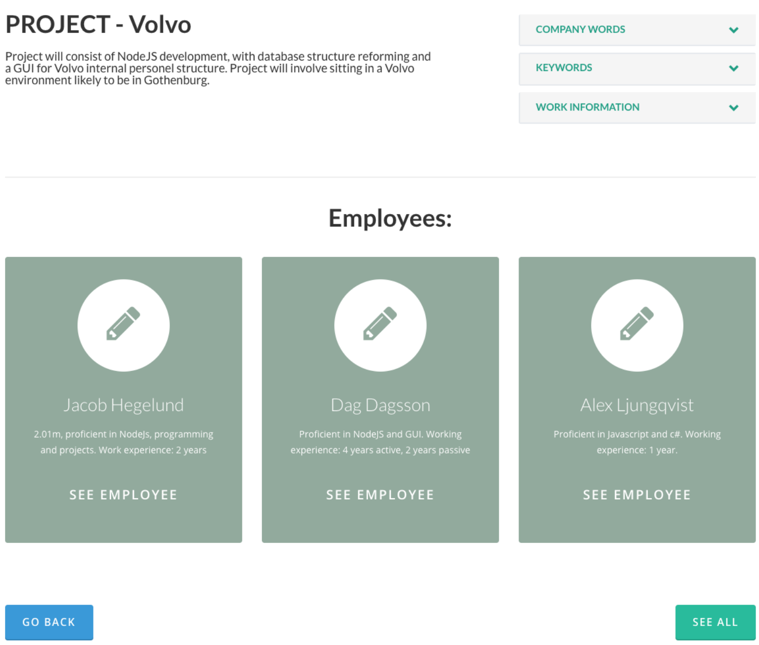

Figure 10: *No design draft of system.*

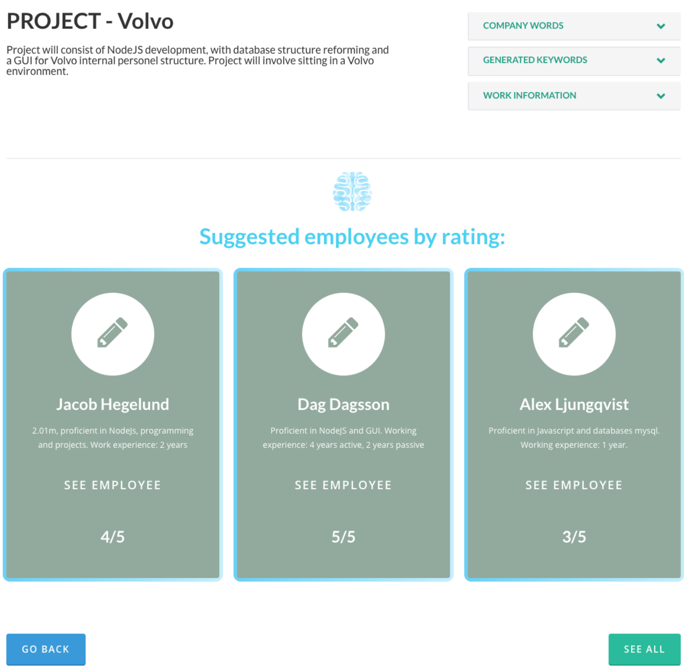

Figure 11: *Design with added brain and color features.*

Table 1 shows the process performed for each test session with each test participant.

Parts	Elements	Material	Time
Briefing	Welcome TP, TP answers Pre-Questionnaire	Pre-Questionnaire	3 min
Tasks	Inform test set up to TP, TP performs the tasks	Task descriptions, Observations protocol	3 min
Debriefing	TP answers Post-Questionnaire, Complementary questions	Post-Questionnaire, Informal debriefing	3 min
Show alternative design	Discuss alternative design	Alternative design	5 min

Table 1: *Test session overview.*

As seen in table 1, each test participant answered both a pre-questionnaire and a post-questionnaire. The pre-questionnaire focused on the test participants preknowledge regarding AI and the systems Softhouse uses, KUSH and Stardust. The post-questionnaire focused on the test participants experiences regarding the design and AI during the test. Finally a general discussion regarding the scenario, design and possible difficulties is conducted.

After the pre-questionnaire, the test participant was showed the following scenario.

> *Softhouse have got a new project from Volvo. The project and the description is added in Stardust. Your work is to add the most suitable competence for this project based on the project description and the competences for each employee in KUSH. You want to use the new system that uses AI and communicates both with Stardust and KUSH to find the most suitable competence. Think and reason through a HR managers position.*
>
> *Perform the following tasks and tell us what goes through your mind during the test, which means that you should try to describe what you see and how you reason in each step.*

The following three steps were then performed.

1. Find the Volvo project.
2. Determine the best suitable employee for the project.
3. Choose and add this employee to the project.

6.2 Cognitive walkthrough result

The pre-questionnaire showed that the test participants had different professions and thereby different experiences in personnel related questions. It also showed that the test participants had different experiences in the systems KUSH and Stardust where only one test participant had used Stardust before and knew what it was. None of the test participants have had very much experience of AI, though everyone have had some and only two of them had worked with developing or designing AI before.

As seen in figure 12, the design did affect the test participants when making their decision and only one test participant said that it just slightly affected the decision. When asking the test participants **how** the design affected them, a common answer were the three suggestions shown. The three suggestion made the decision easier, but most test participants were skeptic on how AI had made a decision and wanted to know more about how AI chose the suggestions to be confident in their decision.

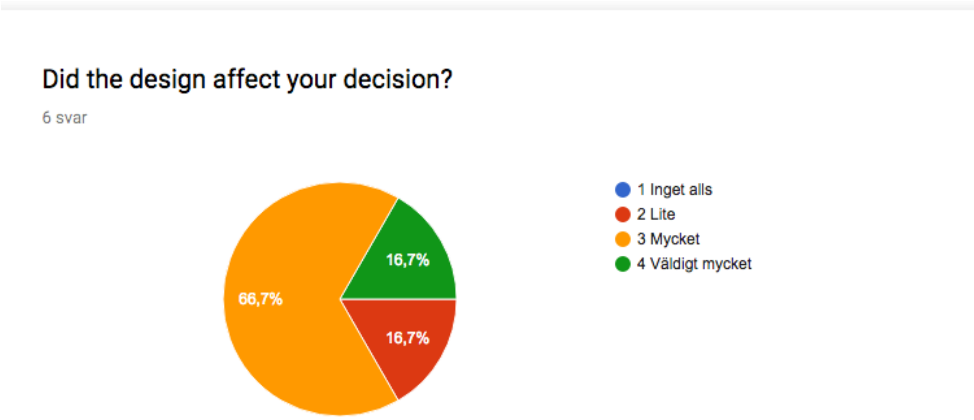

Figure 12: *Answers from post-questionnaire. Scale from 1-4 where 1 means "Not at all" and 4 means "Very much".*

Figure 13 shows that it differed through all test participants on how much the they actually reflected on that it was AI that proposed the suggestions. When looking at the answers within the two test groups, the answers are wildly spread as well and the result between the two groups did not have any significant disparity.

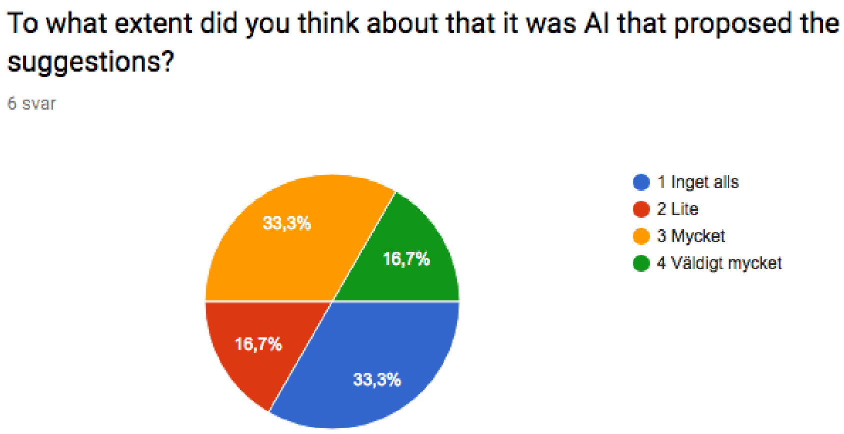

Figure 13: *Answers from post-questionnaire. Scale from 1-4 where 1 means "Not at all" and 4 means "Very much".*

Five out of six test participants said that the suggestions were helpful but because of skepticism towards AIs' decision making, at least four test participants wanted to know more about the decision making. A common opinion were that the help from AI should be provided as an option instead of the main choice. All test participants tended to want more control over the process then was available in these designs.

6.3 Discussion regarding the design propositions

The test participants did not tend to care about or be affected by the design and the different design propositions. What the test participants reacted to were the text and what it told them. The first design, see figure 10, showed no explaining text at all that AI suggested these employees or how AI suggested them. The other design, see figure 11, explained that it was suggested employees and a rating to each employee. For the first design people were guessing and wanted to know more. For the second design people reacted to the text and rating, but still wanted to know more. None of the test participants who tested the "AI-design" reacted on either the brain or the pulsating design. This suggests that people react more to something that is trying to "speak" to them. The design did not feel alive or guiding until it in some way spoke to the test participants.

Another question that almost all test participants asked were regarding the three suggestions.

Why are there three suggestions?

This can once again relate to the need of control that the test participants felt. They realized that this was the top suggestion suggested by AI, but still wanted more control over why and how these employees were suggested by the AI to make their decision. This may also relate to the fact that people do not want help if you do not ask for it. The help from AI should be the users decision to ask for, making the user feel more in control.

7 Conceptual Models

The idea of bringing task helpers to the far front of design has existed for a long time. Microsoft tried very early (dog named Bob, released 1995) and even went as far as incorporating one to Microsoft word (Clippy, released 1996). A recurrent phenomenon today is incorporating a chatbot in the UI that can be interacted with to find paths and answers, and a Chatbot Magazine article[20] state that chatbot communication will be a majority between business and customers by 2020. The goal of the task helper is interfering with the natural behaviour of exploring [21], thus creating a troublesome environment.

7.1 Machine learning as part of the user experience

Not everyone wants help, and when help is wanted you often want to ask for it yourself rather than have it forced upon you. The idea of getting help is not welcomed by all and has showed during the cognitive walkthrough.

Both of Microsofts' ideas of incorporating task helpers into their software received poor reception from its users and were abandoned by Microsoft [22]. The problem with many of the task helpers and chat bots is that they are so fond of helping that they become a liability rather than a tool for improvement. Generally, most of the tools are likely to increase your user experience if handled correctly. The threshold for interacting with these tools effectively is high and out of a normal users scope, producing a non existent relationship concluding in bad reception.

A lot of web shops today use suggested items based on bought or visited items by the user. This function is often subtle with only an explaining text followed by the suggested items. In these cases user also do not know in detail on what parameters the site has made its decisions and suggestions for the user. For a web shop this might be enough for the user to know to feel comfortable with the suggestions since it is not critical decision on what to buy or not. As seen in the cognitive walktrough, section 6.3, when it comes to a more critical decision such as assigning an employee to a certain project the user tend to be interested in which parameters the AI made its decision on. A major reason to such a reasoning is likely the seriousness of it. If a certain product chosen by a web shop seems like a good match but turn out to be bad, a possibility is to send them back and lose nothing. Whilst if wrong person is chosen for a certain project, the costs might be very severe. It could even conclude in losing a major customer, based on circumstances.

The thesis writes built upon the idea that a tool could be of help while not taking a strong presence in the interface. The goal was not to force a user to follow a certain navigational flow, rather help them finding their own flow. Simultaneously not creating a lifeless chat bot per say: "Jeff from randomCompany.io" who can only help when an exact string is used as input. Therefore the thesis writers aimed for something in between a "Jeff" and a "Clippy".

7.2 Machine learning takes shape

To substance the thought of a machine learning algorithm as part of user experience, the topic trust in artificial intelligence need to be addressed. Francesca Rossi, AI Ethics Global Leader at IBM Research, discusses how to build trust towards AI, where transparency is one of the core elements [23]. Trusting an entity based on artificial intelligence is hard, therefore making a user interaction with it equally hard. The thesis writers approached the problem in creating mapping concepts towards what is known today. The thesis writers considered user control important to the system. By user control, the thesis writer mean the ability to decide the outcome of the system, no matter the input. To further fortify the idea, thesis writers produced a couple of concepts with mapping towards daily life.

7.2.1 Brain metaphor

As the brain is the center of human intelligence, the AI in terms of pictures and scenery is generally represented by a brain. It's common for the AI to be pictured as eyes or robot looking structures.

In order to thrive in such an environment and also be able to discern the fact that it's limited to a simple UI, a concept version was a small form showing that something is alive. A beat, a pulsating force going back and forth illustrates a heartbeat and creates the concept of life. To strengthen the mapping, as seen in figure 14, the brain was chosen. The brain continuously pulsates, resembling a conscious entity. The color of the brain is designed to resemble electricity and is a futuristic approach to try and strengthen the bonds between icon and machine learning. The brain metaphor is designed to act as a core for the conceptual models, representing an entity that is aware of the goals of the user. During usage, the idea is for the user to realize that the brain is a thinking entity of itself, possibly holding knowledge that could ease the task.

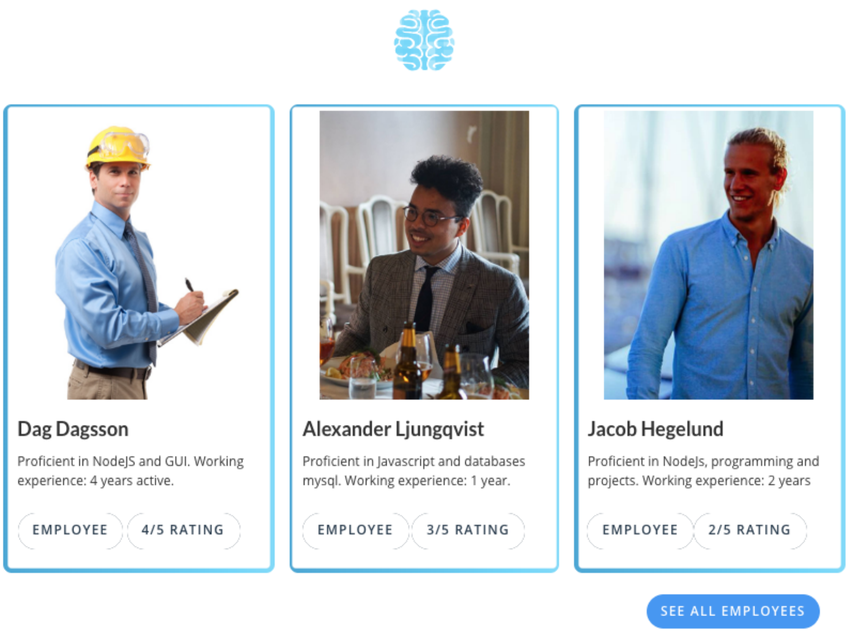

Figure 14: *Showing a pulsating brain to guide human thinking towards an organic mapping and artificial intelligence/machine learning.*

7.2.2 Revealing text

Laying all cards on the table, text revealing that the choices have been created by a machine learning algorithm, with possibility for a modal to give even more insight of what kind of machine learning algorithm has been used. This concept is straight forward and relies solely on the trust factor between user and machine. Instead of hiding the choices, this approach shows information directly. To pin this back to the UI, seen in figure 15, only a Hypertext Markup Language (HTML) tag with a modal possibility is included. HTML is the standard language for web page markup and a modal is a pop-up that is displayed on top of the page.

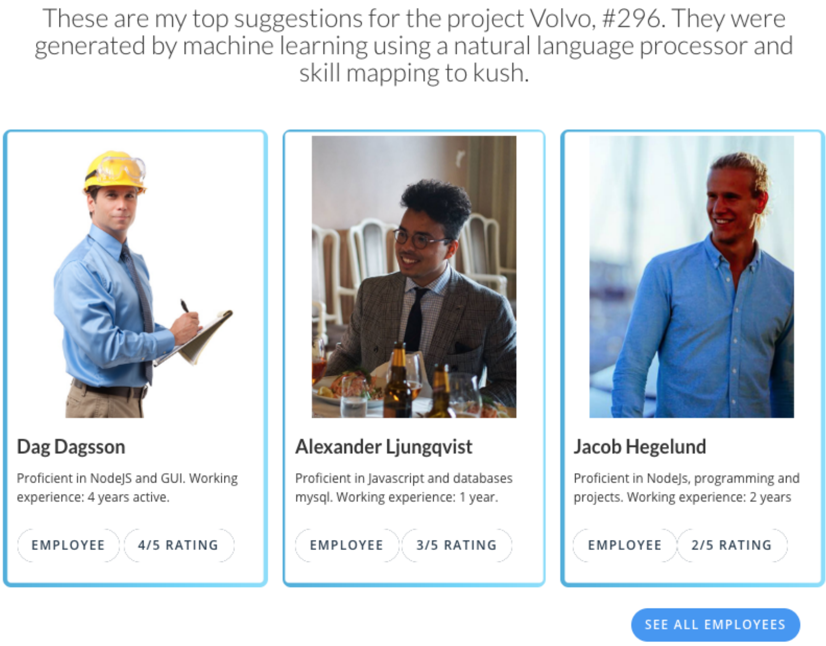

Figure 15: *Text that reveals how the choices were made.*

7.2.3 Dialogue metaphor - the interactive suggester

The concept "the suggester" is based on the functionality of Clippy. A major drawback to Clippy was its inevitable ability of being in the way. If the concept is applied without having a figure popping up every time an action is performed, the thesis writers believed the main functionality of the concept could be of help. The concept is solely text based and is programmed to appear suggestive when the interaction is bound to happen. The tool will illustrate writing when hovered over or when page reloads. Writing text give the mapping back to a chat, which is a communication mapping to a conversation.

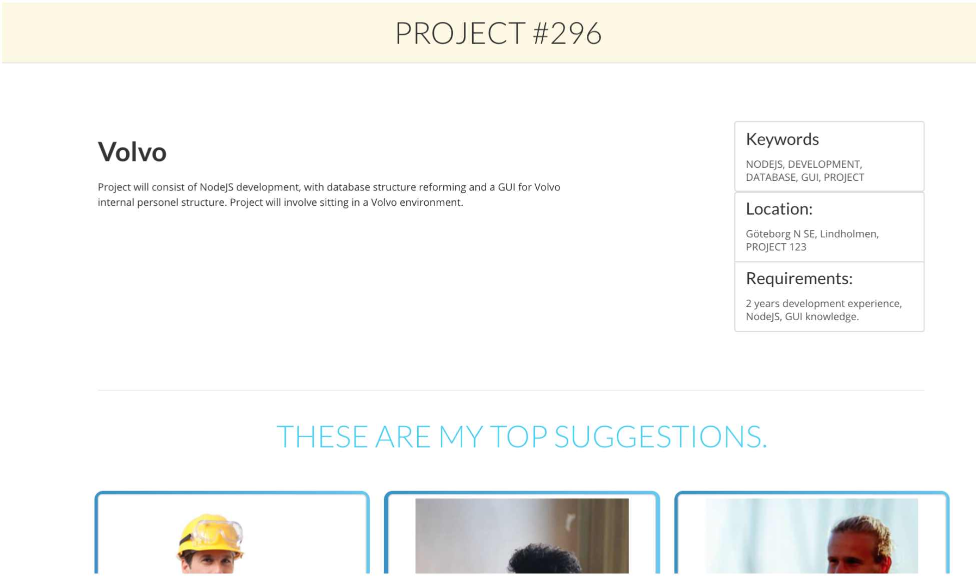

Figure 16: *Interactive text guiding the user.*

7.2.4 The combination of brain and dialogue metaphor

The proposition of the combination is to give the user a choice, if the user already knows which employee should be hired to a certain contract, the machine learning algorithm need not be consulted before decision making. The approach of combining two concepts is to benefit from the positives and negate the negatives. The thesis writers consider the combination approach the most true to both machine learning and usability. The first and major change with the combination concept is its ability to hide the machine learning solution. Instead of showing the generated top suggestions, it only shows the brain metaphor with a query saying it is ready and waiting for interaction, see figure 17.

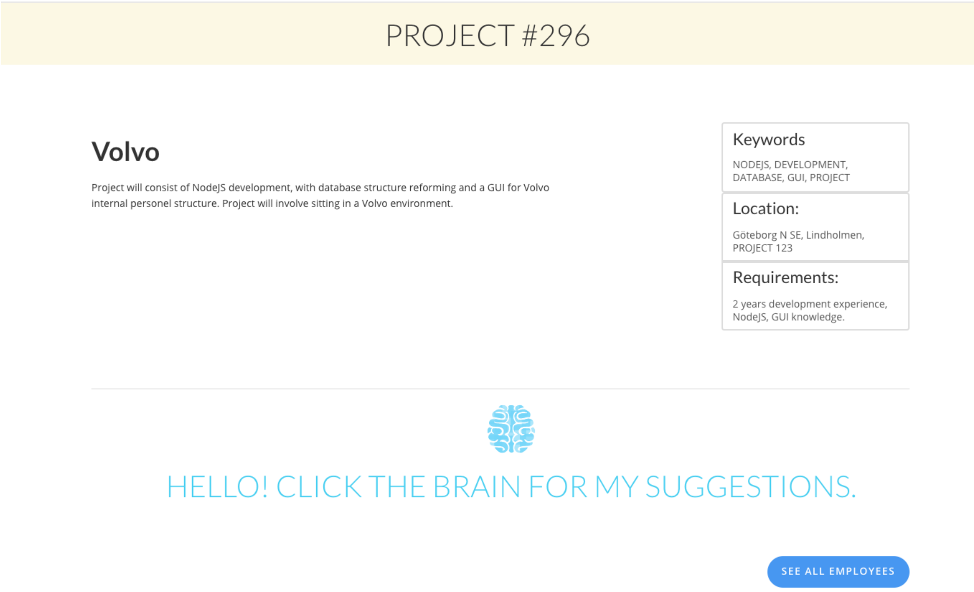

Figure 17: *Interaction in ready state.*

When the user decides to click on the pulsating brain, it responds with a query back saying that is has found its preferred candidates for a certain project. To create a further feeling of being in an interaction, text is being written in human speed for the user to perceive the interaction as a conversation. Discussion around text pace revealed a preference of conversation paced text.

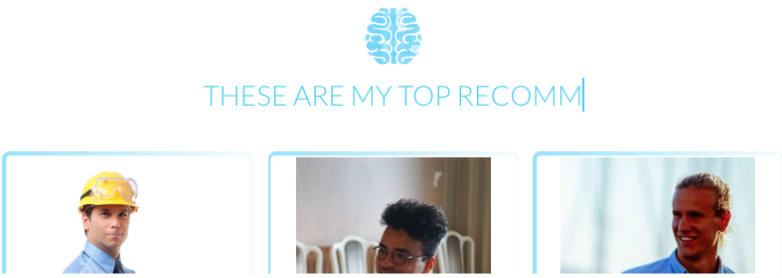

Figure 18: *Showing preferred users and typing in human speed.*

The interaction is at this time over with user being able to now decide which employee should be assigned. If none of these are suitable, user is able to navigate through all employees.

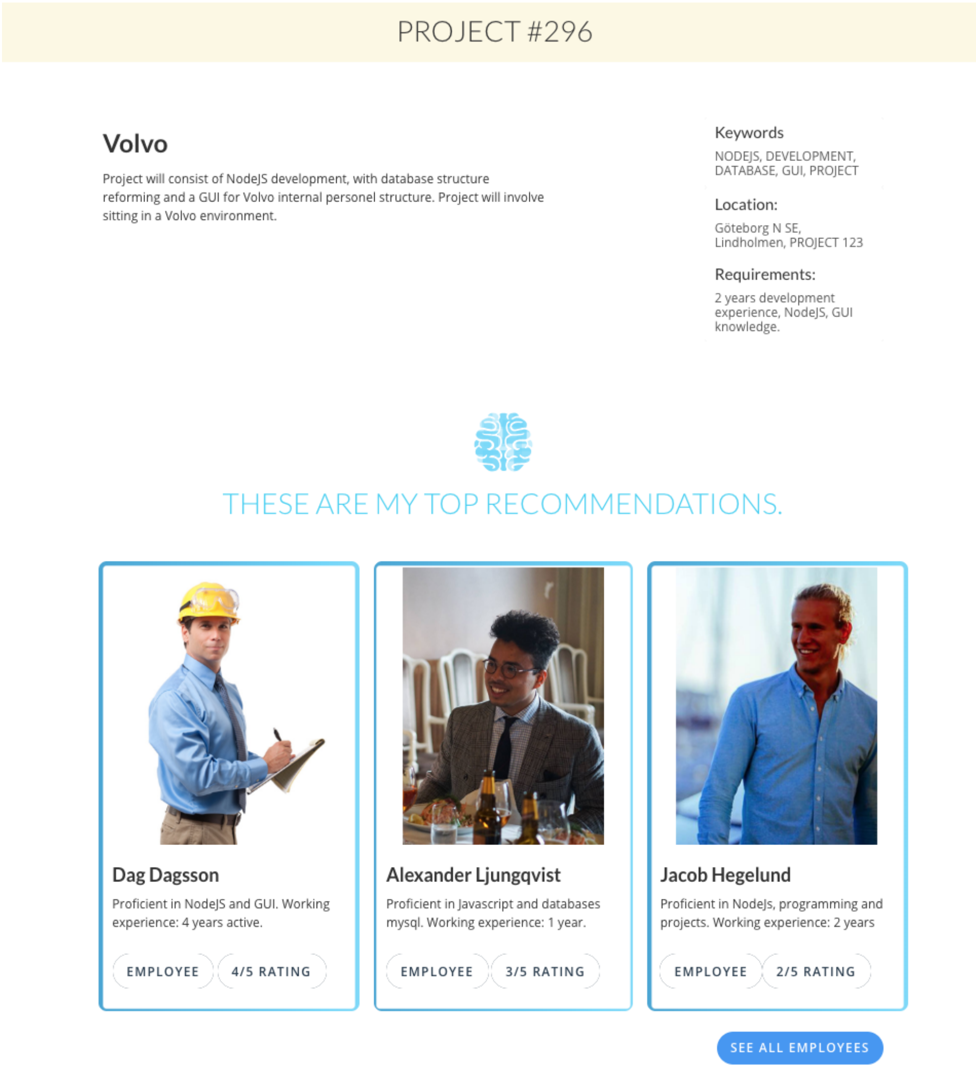

Figure 19: *View of the combination concept.*

7.2.5 Rating View

The rating view is the part of the graphical interface that exists solely to strengthen the trust between user and machine. As earlier written, the trust factor towards the unknown may vary. The main purpose of the rating view is to increase the trust between user and machine. The rating view visualizes key parameters of AI decision making to become more transparent and trustworthy. All information is not of equal importance, and to address the problem, the information was divided into three different clusters with roughly equal importance, see figure 20.

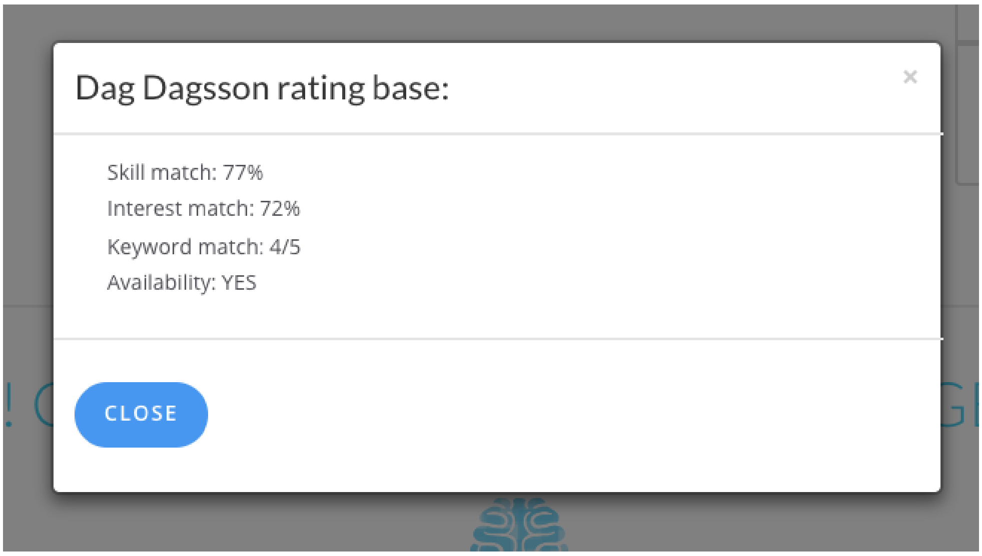

Figure 20: *Conceptual overview of rating visualization.*

The first conceptual overview of the rating overview consist of text response to the main categories the rating was based on.

8 Cognitive walkthrough - final iterations (PoC updated)

8.1 First iteration

Further usability tests were conducted to evaluate the suggested improvements. The first iteration used students from the Information and Communication Engineering Program at the Faculty of Engineering, Lund University, as test participants. To use students as test participants were based on the fact that the thesis writers wanted a broader perspective. The students might not be as business oriented in the case as employees at Softhouse.

Since the students do not have any experience with either KUSH or Stardust, nor working experience at Softhouse, the pre-questionnaire was not used. Instead a more thorough introduction about the project were made to get the students to think like an HR manager at Softhouse.

Except for the pre-questionnaire, the iteration used the same setup as the first cognitive walkthrough.

8.1.1 Result

As seen in figure 21 the design affected the test participants decision making to an average of 3, which means that it affected them much. This is the same average as in the cognitive walkthrough and the test participants once again said that the suggestions were convenient, but the test participants also said that the rating helped them to understand why these suggestions were there and therefore felt more trust in the AIs decision.

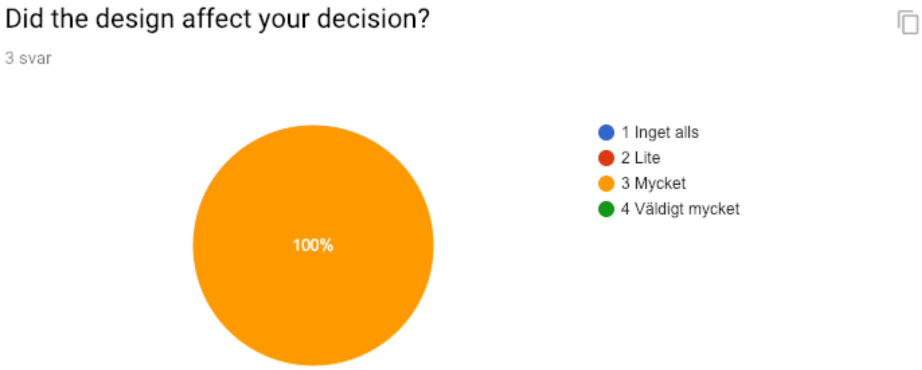

Figure 21: *Answers from post-questionnaire. Scale from 1-4 where 1 means "Not at all" and 4 means "Very much".*

As seen in figure 22 the test participants answered with an average of 3 to what extent they thought about that it was AI that proposed the suggestions. This result differs from the cognitive walkthrough, where the test participants answered with an average of 2.3. Once again the test participants mentioned the rating as a factor and also the convenience the three suggested selections gave.

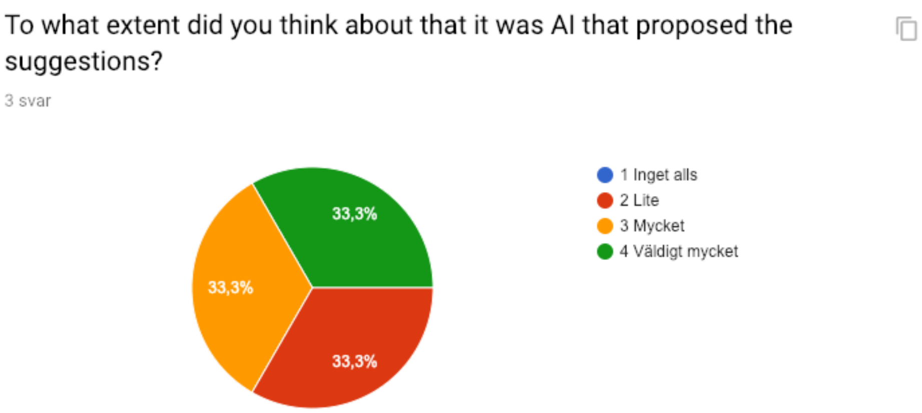

Figure 22: *Answers from post-questionnaire. Scale from 1-4 where 1 means "Not at all" and 4 means "Very much".*

8.1.2 Discussion

The test participants had to make their own decision to use the help from AI, and were also shown a dialogue that the option was possible. This might be one reason that the test participants actually thought about that AI helped them. They controlled the decision to use AI or not and had the option to ask for the help. The rating had a great impact as well. The test participants understood how the AI had made its decisions and, as mentioned above, felt more trust in the suggestions that AI proposed.

8.1.3 General design feedback

One test participant mentioned that the dialogue text should be above the brain to give a more natural flow through the scenario. Another thing mentioned was that the brain should be illustrated to be more "clickable" and also that the buttons for each user were not highlighted enough to make the user actually wanted to interact with them. The "see all" button stole the test participants attention. The test participants wanted a button for each person to directly add them to the project instead off going to the employees profile and from there add them to the project.

Even though the rating helped the test participants to make their decision, a general opinion were that they wanted more information about the rating. More explanation or illustration of what the rating actually meant. Highlighting keywords and illustrate the rating with a color schema would help the user to more efficiently find what is important.

8.1.4 Improved design proposition

Except of smaller design improvements based on the feedback from the test participants, such as more intuitive buttons, the major improvement is the rating view.

In order to give a truthful visualization of the information the thesis writers split the decision making visualization in two parts.

- **Main pillars:** Three main pillars of decision making containing the most integral parts with information about how much each employee matched given criteria. The main pillars are visualized as doughnut charts with percentage to what extent they match. Percentage and match might be considered as rough due to only showing a percentage as a result of many variables and different levels on them. The main reason to only display a small amount of information is to keep the user focused on the task itself rather than different parameters. If more information is required to complete a task, there is a button for viewing more information.

- **Further information:** Contains more direct information on what each pillar is based on. The information concealed behind the button should give a very good overview of exactly what information has been processed to create the predictions. This information should not be integral for decision making in general and is therefore concealed from the beginning. The thesis writers find consider information displayed in the portrait and rating should be enough to choose a certain employee unless a contract has specific demands.

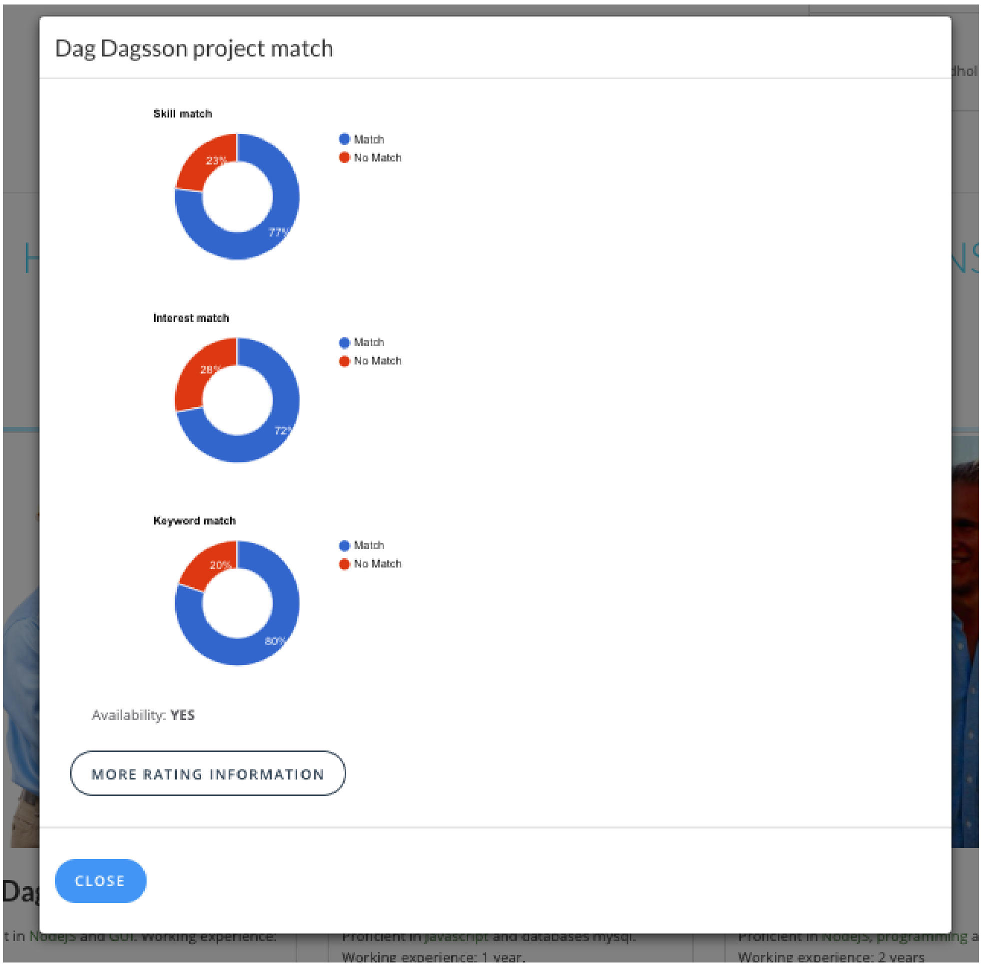

Figure 23: *Final overview of rating visualization.*

Figure 23 shows the version of the rating window after a feedback oriented test process. Following the cognitive walkthrough, minor changes got implemented in order to enhance the final visualization view.

8.2 Second iteration

A second round of cognitive walkthrough were conducted based on the result from the first iteration. Improvements to the design, see figure 23 section 8.1.4 Improved design proposition, were implemented based on the previous feedback. The second iteration of cognitive walkthrough used the same setup as the first iteration, but with the addition of a System Usability Scale (SUS) survey after the test. To use of a SUS survey is based on that the fact that the thesis writers want to get feedback on the usability in general and how the system is perceived by the users.

For the second iteration, test participants with less computer experience than persons studying computer science or working with at Softhouse were used to get an even broader perspective for the test results.

8.2.1 Result

Once again the test participants had problems to understand why and how to view the rating view, and the test participants also wanted a button for adding a persons in the suggested user view. A common problem was to understand how to add a person to the project and to understand the purpose of the profile view and thereby find the "Add" button.

The test participants focused a lot on comparing the keywords with each user and see if the suggested users matched their view of the most suitable person.

Figure 24 and figure 25 shows that the design did affect the test participants in their decision making but that they did not actively reflected on that AI proposed the suggestions.

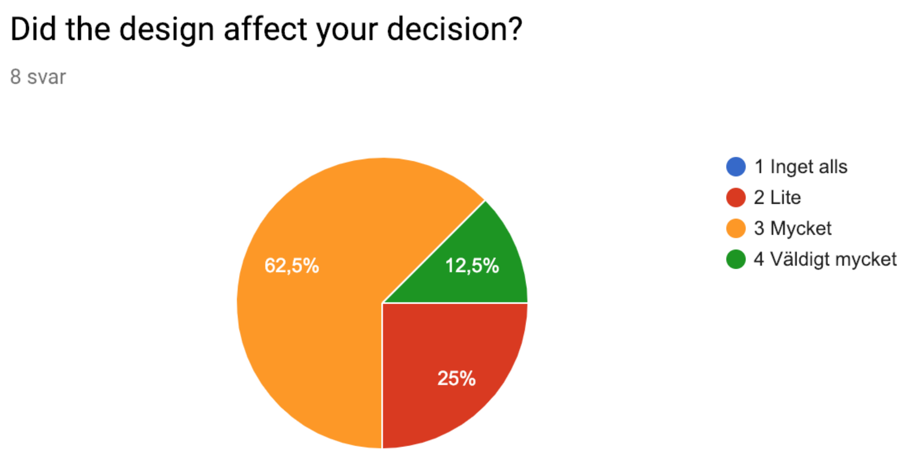

Figure 24: *Answers from post-questionnaire. Scale from 1-4 where 1 means "Not at all" and 4 means "Very much".*

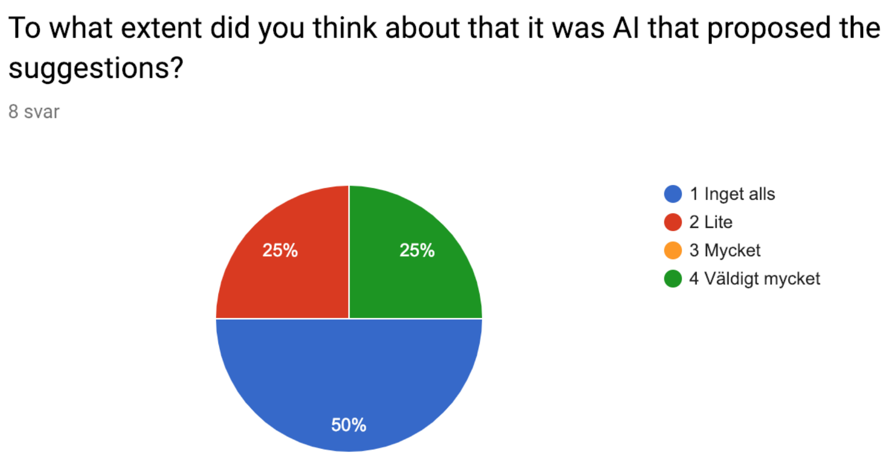

Figure 25: *Answers from post-questionnaire. Scale from 1-4 where 1 means "Not at all" and 4 means "Very much".*

8.2.2 SUS result

Test Participant	Score
1	95
2	92.5
3	90
4	92.5
5	100
6	100
7	95
8	95
Average	95

Table 2: *SUS score.*

8.2.3 Discussion

The buttons for the suggested user view were still not intuitive enough. One main reason seemed to be the label "4/5 rating" for the rating view which did not lead the test participants to click on the rating view. A new rating illustration instead of "4/5" is needed.

Seeing that almost every test participant spend a lot of time comparing keywords, this step has to be more intuitive like highlighting keywords making the keywords easier to find.

The problem finding the add button seemed to relate to the extra step the users have to do to find it. The test participants tried to find the add button when the suggested persons were displayed. Since the thesis writers want the extra step for the user to avoid the scenario where the user adds a person without thinking about the suggestion, a more consistent way would be to show the profile view in the project view in the same way as the rating is shown. This would give even more consistency to the system and probably make the user less confused than being sent to a new page for the profile. It would also make it easier to compare each profile of the suggested users.

The SUS score shows that the system is very easy to use and the test participants did not feel uncomfortable or strained using it. A score of 95, see table 2 section 8.2.2 SUS result, is considered very good compared to the average of 68, see section 3.9 System Usability Scale. This might be due to the fact that the system do not have too many interaction options and that the user quickly get an overview of what is supposed to be done. The system is still very self explanatory using, except of some minor details, and the users did not felt strained using it.

All test participants thought that the design affected their decision, see figure 24 section 8.2.1 Result, and said that the design and the flow of the system helped them to make the decision and as also seen in the SUS score thought that the design was easy to use and made the decision easier.

This iteration used as mentioned test participants with less computer experience and this might be the reason for the result seen in figure 25. In this iteration, the test participants thought less of that it was AI that had suggested the persons than both the earlier test sessions. The test participants tended to trust the AIs' suggestion without any thought on how the AI had made its suggestions. This could correlate to the fact that the test participants might not have their own idea of how AI works and therefore trusts that AI will make the correct decision.

8.2.4 General design feedback

Since a significant part of all test participants had problems understanding the rating label, one consideration was to illustrate the rating as stars or a similar illustration. This will probably make more sense to the users than just a number since most people can relate the illustration to other situations.

An add button in the suggested user view was preferred by almost all test participants instead of the option to see the profile view for each user to add a person.

9 Final PoC

The chapter final PoC aims to create a greater understanding in what exactly was created by the thesis writers, both in terms of graphical design and development wise in terms of machine learning.

9.1 Graphical Interface

The graphical interface aims to represent a possible solution to how an integration between web design and machine learning can be visualized. The main piece of the graphical representation has root in the understanding of the conceptual models. The models try to give the user a familiar feeling and invoke knowledge that the interaction is logical. The PoC is the last step in the design process where most of thesis resources were spent. It finalized in a concept based on the idea that less is more and takes advantage of every day conceptual models towards a living entity.

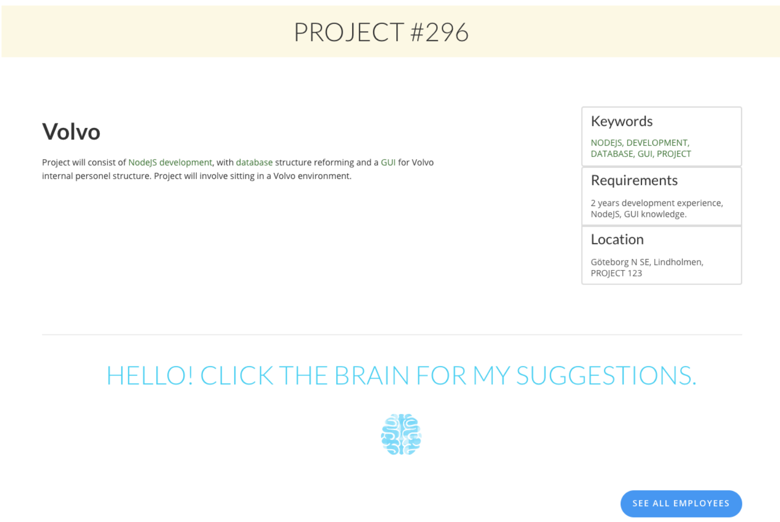

Figure 26: *Project start page.*

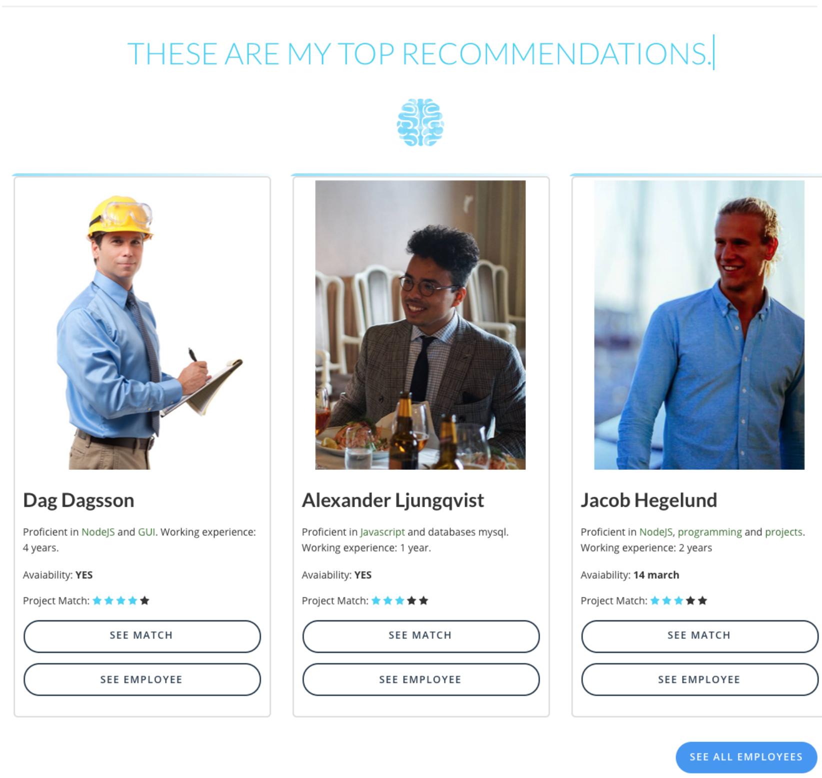

Figure 27: *Project expanded page.*

The main part of the PoC resides within one page only. Figure 26 and 27 shows the project description view. Instead of traversing between many pages, the thesis writers figured keeping the user focused on one page would ease the interaction with the conceptual models. HTML modals fit the description perfectly, which hold the possibility to traverse to a new page while still staying on the same page. Figure 28 and 29 displays the finalized rating and profile view using HTML modals.

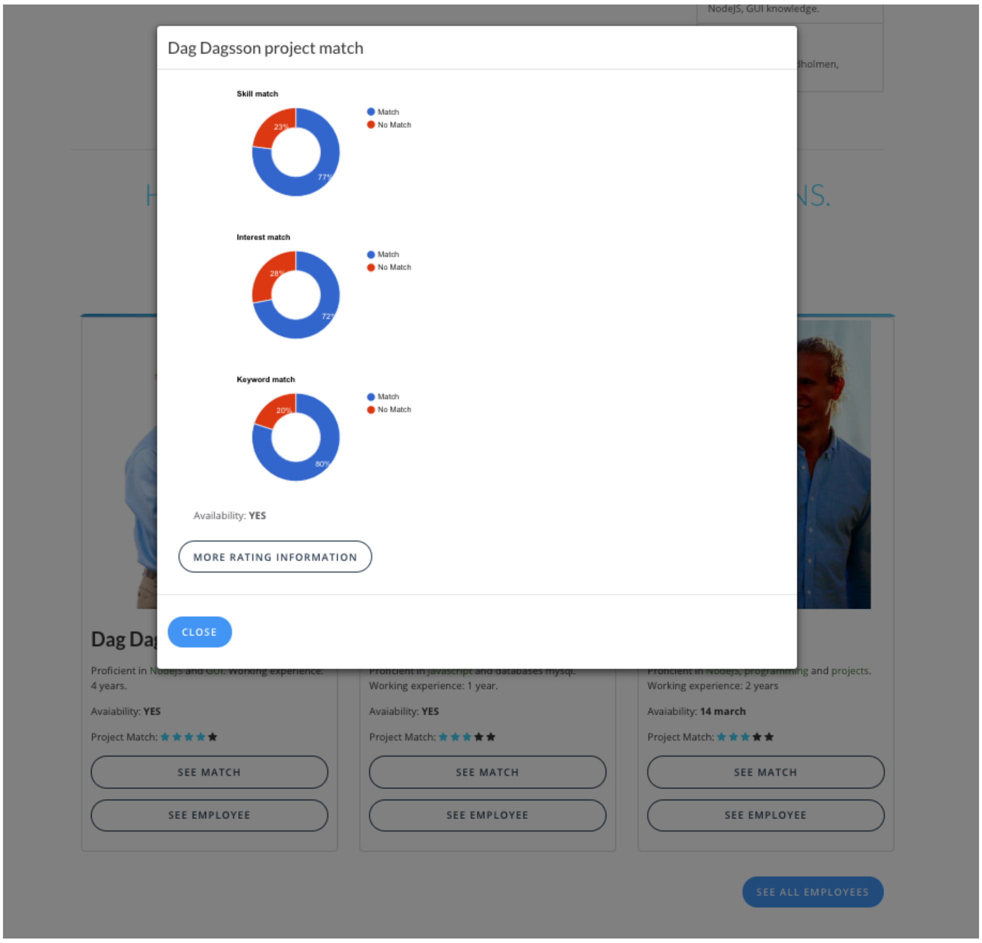

Figure 28: *Modals of rating view.*

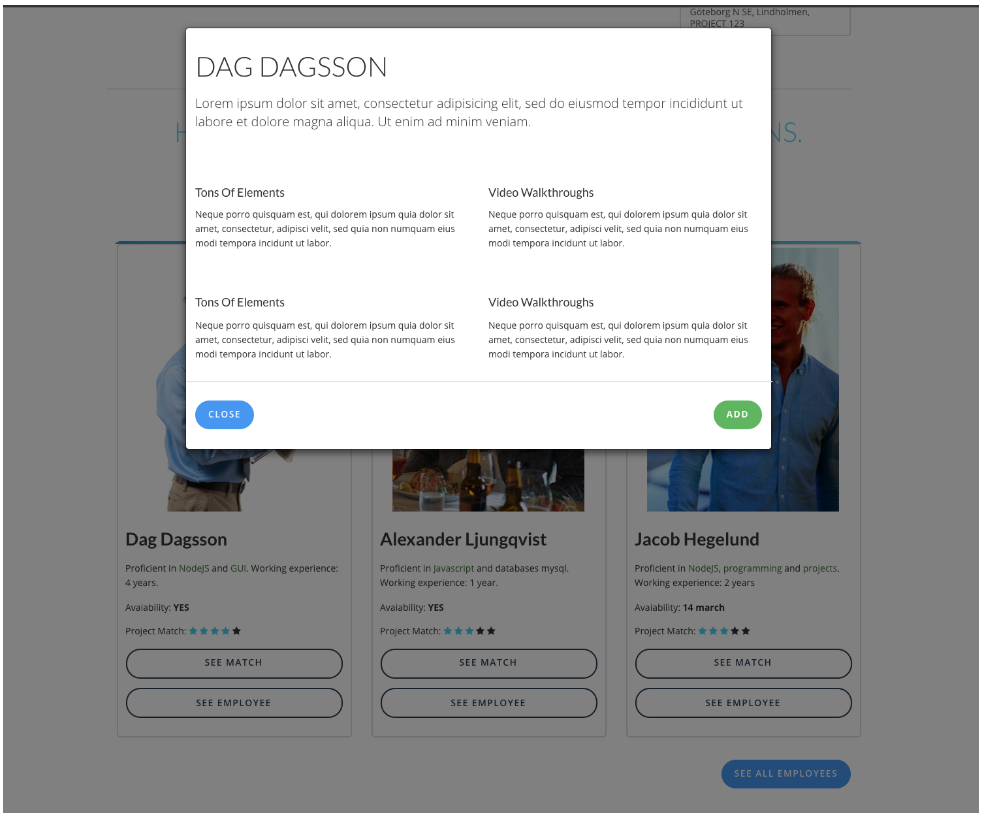

Figure 29: *Modals of profile view.*

The design was made light weight on purpose in order to allow a user to retain focus on the interactive parts of the interface. The brain is in center and has pulsating colors to resemble an organic/living piece, see figure 30.

Figure 30: *Showing three different stages of the brain to symbolize a pulsation.*

The major point of the design is to illustrate how machine learning can work as a proposition rather than decision. The main goal with the UI is to give the user a foundation before the user chooses which type of employee would be optimal to involve in a certain project. The entire PoC is created with normal JavaScript, HTML and CSS.

9.2 Machine Learning

The machine learning approach of the thesis was to create an algorithm that could be visualized via the user interface. The major design choices are products of machine learning development, as the output of the algorithms is the core foundation to the design. Meaning the thesis is based on the progression of machine learning development, and the understanding of how machine learning models work.

9.2.1 Natural Language Processor and TensorFlowJS

The natural language processing part of the product resides within the project description. The description is fetched from Stardust, the earlier mentioned system for keeping track of all projects, and is then processed by the library NLP.js to fetch keywords and stash them. The keywords play an important role as they are the foundation to the modelling part. The modelling part is created with TensorFlow.js and utilize methods for handling data and making predictions.

The outcome of the predictions would be a list of employees, where as the most fitting (three first), will be shown in the main part of the interface. They are shown in order of rating and employees beneath the top three can be found in the page "all employees".

10 Conclusion

We believe that there at this moment no set standard for how to develop a conceptual model towards Machine Learning, it will rather have to relate to the task. As Arthur Clarke said, "Any sufficiently advanced technology is indistinguishable from magic.", and that sums up the task itself. How do you work to create a trust towards what can be seen as magic?

The best possible answer we do not have, however the approach is what we will take on going forward. The idea of incorporating a transparent decision helper eases the process in trust building and shortens the usage threshold. Our final PoC achieved above mentioned points, and when the interface was explained, our test participants understood the design choices.

10.1 Improvements and Further development

In this thesis the Machine learning parts are very small in comparison to bigger projects, giving the idea that it might not optimal to use machine learning for the cause. This is partly correct, however, the thesis aimed to benefit from the knowledge of how Machine Learning can be incorporated rather than the efficiency of the models. The PoC is small and could have been bigger but as written in the final PoC chapter, it's designed to be a bridge between two systems. A way to build upon the PoC in this thesis is to lift back the information from the software Stardust and only use the PoC as a main system where you can add project information. It would be a smarter and easier solution for the end users as the ones who add the information into the system could possibly utilize the same system for finding employees. Today, instead you have to use two different types of software today for that task.

As the PoC does not contain very much functionality, there are development opportunities in which one could extend the functionality to make it more intuitive. One small thing that could have been implemented is similar to the button "don't show this again". If you know what the system does and how it works, it would be of good use to be able to choose if you want to click the brain every time or if you want to get the help every time/never. This could possible become irritating after many usage iterations when a trust factor with the system is established. The thesis is looking into decision support rather than decision making, and that could possibly create a gap between product and thesis.

10.2 Reflection

The iterative approach was very successful, as we are used to the approach when creating a product. Conducting tests and creating equivalent environments for each test participant can be hard due to personalities. However, the best part of the test phase was the discussions. The environment was relaxed and a dialogue made it easy for each party to communicate.

The biggest hindrance we faced during product preparation was the data. As machine learning is heavily data dependent, we needed a good source of data to create and finalize our product. As the data was not optimal, we had to adapt by customizing the data towards the task.

Initially we were not certain to whether the research questions would receive proper answers, but as the method proceeded, information and conclusions appeared giving content to the cause. Both approaches to gather information played a very big role in the entire project. The soft values were interesting as they became heavily subjective from person to person, and it's hard to extract a figurative value that can be explained to the reader. On the other hand, quantitative values are very direct and comprehensible but will not return a good view unless many values are put together.

10.3 Final Words

The machine learning approach to the thesis was to create an algorithm that would help Softhouse find fitting employees to a certain project. The algorithms needed to return values that could be visualized via the user interface. To create a higher value for both the authors and Softhouse, the authors aimed to find applicable machine learning libraries and packages as one of Softhouses' values is to increase competence. One of the goals from Softhouse was to examine the ability to create value from machine learning without extensive knowledge beforehand. Softhouse has a wish to progress into a firm that has a greater knowledge within artificial intelligence and machine learning. For this specific task, our knowledge within machine learning suited the requirements very well.

To get an understanding of how the end product could incorporate machine learning, a vast amount of resources and time were spent towards creating a fitting algorithm that could enhance the proposition of the thesis.

The reason the aspect of machine learning was involved is because we think it's of importance to discover new possibilities regarding how to incorporate machine learning in everyday life. We believe that it will come a time when machine learning overall becomes more accessible and the demand for it will raise. Similar to the way as web bureaus popped up back when internet was new, we vision machine learning in the future. You don't need to be an expert to make it work, you rather just have to understand how it functions.

References

[1] Z. Boren, (October 2014), *There are officially more mobile devices than people in the world.*, Independent https://www.independent.co.uk/life-style/gadgets-and-tech/news/there-are-officially-more-mobile-devices-than-people-in-the-world-9780518.html

[2] Current world population, (February 2019), http://www.worldometers.info/world-population/

[3] (Accessed: March 2019), https://www.gsmaintelligence.com/

[4] Y. Guo, (Accessed: 2019-02-20), Google AI, https://www.youtube.com/watch?v=F7f20VB-A50

[5] SoftHouse Consulting AB, (February 2014), *Scrum på fem minuter*, https://issuu.com/softhouse/docs/scrum_broschyr_sve_ny_121029

[6] R. B. Gallupe, A. R. Dennis, W. H. Cooper, J. S. Valacich, L. M. Bastianutti and J. F. Nunamake, (November 2017), *Electronic Brainstorming And Group Size*, Academy of Management Journal, https://journals.aom.org/doi/abs/10.5465/256377

[7] A. F. Osborn, (1957), *Applied Imagination: Principles and procedures of creative thinking*, New York : Charles Scribner's Sons

[8] TechSmith, *Usability testing basics, an overview*, (Accessed Feb 2019), http://webservices.itcs.umich.edu/drupal/wwwsig/sites/webservices.itcs.umich.edu.drupal.wwwsig/files/Usability-Testing-Basics.pdf, TechSmith

[9] K. A. Ericsson & H. A. Simon, (1984), *Protocol analysis: Verbal reports as data.*, Cambridge, MA, US: The MIT Press.

[10] C. Lewis, (1982), *Research Report RC9265*, IBM TJ Watson Research

[11] Prototyping, (Accessed March 2019), http://www.eat.lth.se/fileadmin/eat/MAMA15/Foe_3_-_Prototyping.pdf

[12] J Nielsen, (1993), *Usability Engineering*, Morgan Kaufmann Publishers Inc

[13] J Nielsen, (January 2012), *Thinking Aloud: The #1 Usability Tool*, https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/

[14] D-labs, (Accessed February 2019), https://www.d-labs.com/en/services-and-methods/medium-fidelity-prototyping.html

[15] J.J. Garrett, (2010), *The Elements of User Experience: User-Centered Design for the Web and Beyond*, New Riders Publishing

[16] K. Schwaber, (2004), *Agile Project Management With Scrum*, Microsoft Press

[17] K. Schwaber, M. Beedle, A. Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, J. Kern, B. Marick, R. Martin, S. Melloer, K. Schwaber, J. Sutherland, D.Thomas (February 2001), *Manifesto for Agile SoftWare Developemnt*, https://moodle2016-17.ua.es/moodle/pluginfile.php/80324/mod_resource/content/2/agile-manifesto.pdf

[18] J. Sutherland, K. Schwaber, *The Scrum Guide, The Definitive Guide to Scrum: The Rules of the Game*, (November 2017), https://scrumguides.org/docs/scrumguide/v2017/2017-Scrum-Guide-US.pdf#zoom=100

[19] M. Walker, L. Takayama, J. Landay, (September 2002), *High-fidelity or low-fidelity, paper or computer? choosing attributes when testing web prototypes*, Computer Science Division University of California, Berkeley

[20] brn.ai, (March 2019), *Chatbot Report 2018: Global Trends and Analysis*, https://chatbotsmagazine.com/chatbot-report-2018-global-trends-and-analysis-4d8bbe4d924b

[21] S. Hawking, (October 2018), *Brief answers to big questions*, Hodder & Stoughton

[22] L. Swartz, (June 12 2003), *Why people hate the paperclip:labels, appearance, behavior and social responses to user interface agents*, Symbolic Systems Program Stanford University, http://xenon.stanford.edu/~lswartz/paperclip/paperclip.pdf

[23] F. Rossi, 2019, *Building trust in artificial intelligence*, Journal of international affairs

[24] (Accessed: 2018-10-11), https://www.tensorflow.org/

[25] (Accessed: 2018-10-11), https://js.tensorflow.org/

[26] (Accessed: 2018-10-11), https://js.tensorflow.org/tutorials/core-concepts.html

[27] G. E. Hinton, (October 20, 2007), *Learning multiple layers of representation*, Department of Computer Science, University of Toronto

[28] R. P. Lippmann, (April 1987), *An Introduction' to Computing with Neural Nets*, IEEE ASSP MAGAZINE

[29] A. K. Jain, J Mao, K. M. Mohiuddin, (March 1996), *Artificial Neural Networks: A Tutorial*, IEEE Computer Society

[30] S. B. Kotsiantis, (July 16, 2007), *Supervised Machine Learning: A Review of Classification Techniques*, Department of Computer Science and Technology, University of Peloponnese, Greece

[31] S. Lloyd, M. Mohseni, P. Rebentrost, (November 4 2013), *Quantum algorithms for supervised and unsupervised machine learning* arXiv preprint arXiv:1307.0411

[32] Natural Language Processing, (February 2019),

https://www.sas.com/sv_se/insights/analytics/what-is-natural-language-processing-nlp.html#nlphowitworks, SAS

[33] (Accessed: March 2019), https://www.forbes.com/companies/sas/#4273c0bf6fa4

[34] P. Domingos, (2012), *A Few Useful Things to Know about Machine Learning*, Department of Computer Science and Engineering, University of Washington, https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf

[35] (Accessed: 2019-01-10), http://www.barrel.ai/

[36] (Accessed: 2019-01-10), http://www.nordaxon.com/

[37] S. Abu-Nimeh1, D. Nappa, X. Wang, and S. Nair1, (2007), *A Comparison of Machine Learning Techniques for Phishing Detection*, SMU HACNet Lab, Southern Methodist University, Dallas

[38] P. Domingos, (2012), *A Few Useful Things to Know about Machine Learning*, Department of Computer Science and Engineering, University of Washington

[39] B. Goertzel and C. Pennachin, (2007), *Artificial General Intelligence*, Springer

[40] A.F. Osborn, (1963), *Applied imagination; principles and procedures of creative problem-solving.*, New York: Scribner

[41] H. Al-Samarraie & H. Shuhaila, (December 2017), *A Review of Brainstorming Techniques in Higher Education*, VOL 27 Journal: Thinking skills and Creativity

[42] M. Rettig, (April 1994), vol 37 no 4, *Prototyping for Tiny Fingers*, ACM: New York, NY, USA,

[43] J. Rieman, M. Franzke, D. Redmiles, (May 7 1995), *Usability Evaluation with the Cognitive Walkthrough*, CHI '95 Conference Companion on Human Factors in Computing Systems, Pages 387-388

[44] (Accessed: 2019-01-30), https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

[45] A. Bangor, P. Kortum and J. Miller, (May 2009). Vol 4, pp. 114-123, *Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale*, Jounal of usability studies, http://uxpajournal.org/wp-content/uploads/pdf/JUS_Bangor_May2009.pdf