Designing for Confidence

A concept for elevating the user's confidence during a statistical study with artificial intelligence

Bachelor Thesis
Stuttgart Media University
Computer Science and Media
Winter Semester 2019 / 2020

Author / Concept / Design
Stephanie Vo

Supervision
Prof. Dr. Sabiha Ghellal
Dimitri Hoffmann

Typefaces
IBM Plex Serif
IBM Plex Sans
IBM Plex Mono

Declaration of Authorship

English

I, Stephanie Vo, hereby declare on my word of honour that I have written the presented bachelor thesis with the title: „Designing for Confidence - A concept for elevating the user's confidence during a statistical study with artificial intelligence“ independently and without external help and have not used any aids other than those stated. The passages of the thesis which were taken from other works, either in wording or in meaning, are always marked with an indication of the source. The work has not yet been published or presented in any other form as an examination paper.
I have taken note of the meaning of the honorary assurance and the examination law consequences (§ 26 Paragraph 2 Bachelor-SPO (6 semesters), § 24 Paragraph 2 Bachelor-SPO (7 semesters), § 23 Paragraph 2 Master-SPO (3 semesters) or § 19 Paragraph 2 Master-SPO (4 semesters and part-time) of the HdM) of an incorrect or incomplete honorary assurance.

Stephanie Vo
Stuttgart, 28th February 2020

Deutsch

Hiermit versichere ich, Stephanie Vo, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: „Designing for Confidence - A concept for elevating the user's confidence during a statistical study with artificial intelligence“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.
Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§ 26 Abs. 2 Bachelor-SPO (6 Semester), § 24 Abs. 2 Bachelor-SPO (7 Semester), § 23 Abs. 2 Master-SPO (3 Semester) bzw. § 19 Abs. 2 Master-SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

Stephanie Vo
Stuttgart, 28. Februar 2020

Abstract

English

The capabilities of Artificial Intelligence (AI) are utilized increasingly in today‘s world. The autonomous and adaptive characteristics allow applications to be more effective and efficient. A certain subfield of Artificial Intelligence, Machine Learning, is enabling services to be tailored to a user‘s specific needs. This could prove to be useful in an information-heavy field such as Statistics. As design research from SPSS Statistics, a legacy statistical application, has indicated, statistics beginners struggle to tackle the challenge of preparing a statistical research study. They turn to several sources of information in an attempt to find help and answers but are not always successful. This leads to them being unconfident before they have even started to execute the statistical study. The adaptive features of Artificial Intelligence could help support students in this case, if designed according to established principles. This thesis investigated the question whether an AI-powered solution could elevate the users‘ confidence in statistical research studies. In order to find the answer, a prototype with exemplary User Experience was designed and implemented. Preceding research determined the domain and market offer. User research was conducted to ensure a human-centered outcome. The prototype was evaluated with real test users and the results answered the question in the affirmative.

Deutsch

Die Fähigkeiten der künstlichen Intelligenz (KI) werden in der heutigen Zeit zunehmend genutzt. Die autonomen und anpassungsfähigen Eigenschaften erlauben es, Anwendungen effektiver und effizienter zu gestalten. Ein bestimmtes Teilgebiet der Künstlichen Intelligenz, das maschinelle Lernen, ermöglicht es, Dienstleistungen auf die spezifischen Bedürfnisse eines Benutzers zuzuschneiden. Dies könnte sich in einem informationsintensiven Bereich wie der Statistik als nützlich erweisen. Wie die Designforschung von SPSS Statistics, einer statistischen Anwendung, gezeigt hat, haben Statistikanfänger Schwierigkeiten mit der Vorbereitung einer statistischen Forschungsstudie. Sie wenden sich an mehrere Informationsquellen, um Hilfe und Antworten zu finden, aber sind nicht immer erfolgreich. Dies führt dazu, dass sie unzuversichtlich sind, bevor sie überhaupt mit der Durchführung der statistischen Studie begonnen haben. Die adaptiven Eigenschaften der künstlichen Intelligenz könnten in diesem Fall helfen, die Studenten zu unterstützen, wenn sie nach etablierten Prinzipien gestaltet werden. In dieser Arbeit wurde der Frage nachgegangen, ob eine KI-gestützte Lösung das Selbstbewusstsein der Nutzer vor und während statistischen Forschungsstudien erhöhen kann. Um die Antwort zu finden, wurde ein Prototyp mit vorbildhagter User Experience entworfen und implementiert. Die vorangegangene Forschung bestimmte die Domäne und das Marktangebot. Nutzerforschung wurde durchgeführt, um ein auf den Menschen zentriertes Ergebnis sicherzustellen. Der Prototyp wurde mit echten Testnutzern evaluiert, und die Ergebnisse haben die Frage bejaht.

Table of Contents

List of Abbreviations

IBM International Business Machines (Corporation)

SPSS Statistical Package for the Social Sciences

AI Artificial Intelligence

UNeeQ User Need Questionnaire

ML Machine Learning

UX User Experience

UI User Interface

HCD Human-Centered Design

Lo-fi Low-Fidelity

Mid-fi Mid-Fidelity

Hi-fi High-Fidelity

Introduction

Artificial intelligence (AI) technologies are currently used in many products people handle every day, take for example recommendation systems in all kinds of media. Algorithms based on AI determine the content shown to consumers, enabling a personalized experience. (ELEMENTS OF AI, N.D., [W1]) Companies use AI to augment their product's efficiency and effectiveness and developers are becoming more interested in creating AI-powered applications. According to "the AI 2019 Annual Report", attendance at AI conferences continues to increase significantly. One of the largest AI and Machine Learning (ML) conferences, NeurIPS, expected 13,500 attendees in 2019, up 41% over 2018 and over 800% relative to 2012. Since 1998, the number of peer-reviewed AI papers has increased by more than 300%. (STANFORD UNIVERSITY, 2019, [W2], P.5) With the widespread attention and buzz around this technology, designers face a new challenge: Design for AI. AI should be incorporated in a meaningful way and with a user-centered approach. Design principles and best practices already exist for creating pleasant user experiences in products or services. These alone might be insufficient when it comes to AI integration for example due to ethical aspects. Research in this interdisciplinary field is ongoing and AI design guidelines have been published by corporations, emphasizing the significance to design the relationship between humans and AI diligently. (GOOGLE DESIGN, 2019, [W3])

For a subject as intricate as statistics, the use of AI is potentially beneficial. As an example, IBM SPSS Statistics is a statistical program that has been around for over 50 years, gradually being expanded with more and more features. In order to be in accordance with current users' needs and expectations, the software is now in the process of a redesign to be improved visually and user experience-wise. (STAUBER, 2017, [W4]) SPSS is a very powerful tool but with the capabilities comes complexity. A major part of the user group is comprised of students and not all of them are experts in statistics, so they are not fully certain in using it for their studies. According to the SPSS Statistics design team in Böblingen, setting up a research study to analyze in a statistical program is often intimidating for beginners. They feel insecure in the process and need several sources of help to achieve their goal. How could AI's capabilities unburden and support the user in this case?

This thesis explores ways to support the users in their research study preparation. The applicability of human-centered AI will be validated. The goal is to find out whether a solution with AI could increase the user's confidence in setting up statistical research studies. In order to research this topic, a prototype with exemplary user experience (UX) will be designed and evaluated with test users.
The paper is divided into the following chapters: theory, research, conception, prototyping, evaluation and a conclusion in the end.

• In the **Theory** section, the essence of statistics and the basic usage of the statistical package SPSS Statistics are addressed and explained. In addition to that, it is introducing the subject of artificial intelligence, its corresponding subtopics, and areas of application. The principles of several design fields such as UX design are elaborated as well.
• In the **Research** phase, the problem area of interest is defined and market analysis is done. The focus in this section is to research the user's needs and pain points intensively to lay a foundation for the subsequent steps. Personas, scenarios, use cases and goals are derived from user interviews and desk research.
• The **Conception** phase ideates and implements first concepts based on the research results. Early drafts and ideas are tested and iterated on.
• The **Prototyping** phase produces a more detailed and high fidelity prototype, incorporating the feedback collected before.
• In the **Evaluation** phase, an evaluation strategy is outlined first before executing user tests. The prototype is tested on preselected design principles and the fulfillment of user needs. Afterward, the final evaluation results are explained and discussed.
• The last chapter **Conclusion** is giving an outlook on the future of the discussed topics and the elaborated concept.

1

Theory

This thesis will try to unburden users and elevate their confidence when preparing their research study by applying artificial intelligence. The result should have a pleasant user experience. In order to utilize all aspects and components needed for this goal, this section will elaborate on the relevant topics: Statistics, AI and Design.

1.1

Project Background

The use case was provided by IBM and the SPSS Statistics design team in Böblingen. Resources and data are provided by IBM and most functionality will be based on and compared to the statistical software package IBM SPSS Statistics. The overall concept will be a general pattern for supporting the beginner users in a complex and unknown environment, in this case statistical research. What statistics are comprised of and how SPSS Statistics is mainly used will be explained.

Statistics

There are two related but separate meanings for the word "statistics". While the "field of statistics" describes the study and practice of collecting and analyzing data, "statistics" themselves are facts about or summaries of data. (CRASHCOURSE, 2018, [W5]) Essentially, the motivation behind statistics is to answer questions with the help of data. Data is the main drive force for statistics, so the reliability and accuracy of any statistics depend on the data. The importance of data quality and quantity will be presented accordingly in a later chapter. (→ DATA) There are two main types of statistics that are commonly used. (TAYLOR, 2018, [W6])

Descriptive statistics help statisticians make sense of quantitative data. They provide information like mean, median and mode which describe the central tendency. Values like the range or standard deviation demonstrate how the data is distributed. While descriptive statistics support us to extract useful information out of huge amounts of data, they only show what's already existing. (TAYLOR, 2018, [W6])

Inferential statistics enable researchers to make inferences, meaning they can make conclusions beyond the data at hand. Through samples, one can set estimations and predictions about a population without actually surveying the whole population. This way, statistical hypotheses can be tested. The results of an inferential analysis help us decide whether to reject an idea or hypothesis or not. (CRASHCOURSE, 2018, [W5])

Statistics process huge amounts of information and filter out the relevant parts. They will not deliver a definitive answer but rather facts to base a decision on during uncertain situations. (CRASHCOURSE, 2018, [W5]) Uncertainty exists in almost every scientific field so naturally, statistics are needed in multiple areas of research, for example, psychology, economics, medicine, advertising. (TAYLOR, 2018, [W6]) Acquiring data to test hypotheses, analyzing the data, planning and executing the research - there are a lot of steps to complete your research, but software applications designed for statistical analysis are of help in that process. (FARNSWORTH, 2019, [W7])

If Statistics were a superhero, it's [sic] batcall would be uncertainty, and it's [sic] tagline would be „When you don't know for sure, but doing nothing isn't an option".

Adriene Hill, Host at Crash Course Statistics
(CRASHCOURSE, 2018, [W5])

IBM SPSS Statistics

IBM SPSS Statistics, the world's leading statistical software according to IBM, was one of the first statistical applications, having been released over 50 years ago. (STAUBER, 2018, [W8]) Over the decades SPSS Statistics constantly evolved and went through feature additions, improvements, and redesigns until it reached its current state which is a modern user interface offering statistical analysis tools. (IBM, 2017, [W4]) With SPSS Statistics the user can edit and analyze their data and apply advanced statistical procedures on large data sets to get insights. The main features of SPSS Statistics will be explained.

Data View

In order to begin working with the data, one needs to open the data set first. Aside from its own, SPSS Statistics supports all common data file formats used for structured data such as MS Excel spreadsheets, plain text files, relational databases and formats from other statistical software. The data is portrayed in a spreadsheet called "data view", always showing the actual data values sorted in columns and rows. (VAN DEN BERG, N.D., [W9])

Variable View

The "variable view" is a separate but strongly related spreadsheet that displays the metadata about the data. Essentially the user can enter, find and edit information about the meaning of variables and data values in this sheet. (VAN DEN BERG, N.D., [W9]) There are different measurement levels for variables that determine what kind of values are entered and what statistical tests are applicable in their case. (PEZZULLO, N.D., [W10])

Data Analysis

Now that the user has an overview of his data, they can proceed to choose a statistical analysis (also referred to as tests or methods) to run. The analyses can be found in the menu options or in a catalog (in newer concepts) and there are around 120 analyses to select from. After choosing the appropriate test, the procedures configuration can be set such as which variables they want to inspect. (VAN DEN BERG, N.D., [W9]) Depending on number and measurement levels of variables, different statistical analyses can be performed. The data

has to match and certain assumptions need to be fulfilled for certain tests. (STATISTICS SOLUTIONS, N.D., [W11])

Output Viewer

Confirming the settings and running the analysis will produce and show the test results in the "output viewer". The results are sorted in tables with statistics on the variables chosen in the analysis configuration. (VAN DEN BERG, N.D., [W9])

Reporting

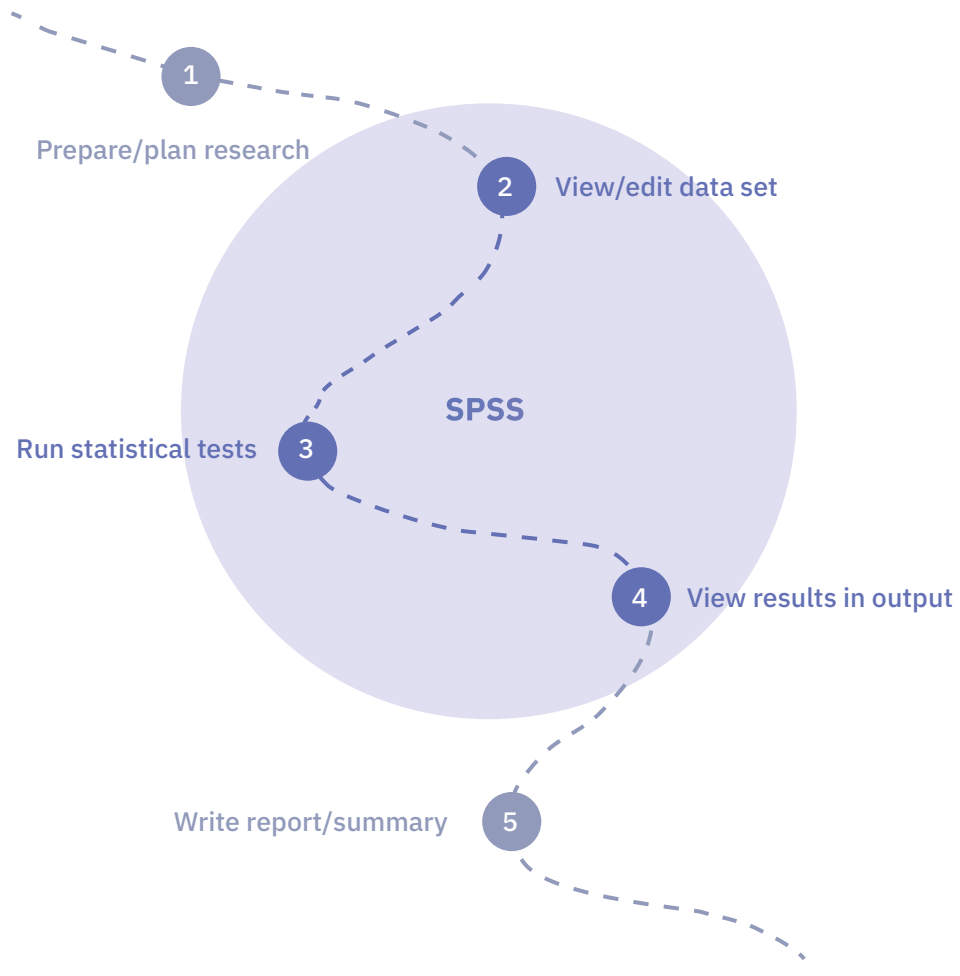
The output items which are usually tables and charts can then be simply copied and pasted into other programs like text editors to summarize the research outcomes for a report. (VAN DEN BERG, N.D., [W9])

Syntax Editor

There is another way to run an analysis in SPSS Statistics other than selecting it manually in the menu options. The syntax editor allows advanced users to type and run SPSS code called "SPSS syntax". Certain commands will produce the same results as running the analysis from the menu. SPSS syntax can be saved and shared easily between other users but the most important aspect is the replicability that is enabled through it. Reproducing results can just be done with copy and paste. (VAN DEN BERG, N.D., [W9])

The user's workflow in SPSS Statistics basically consists of the following actions. They open and view their data files and edit the data if needed to improve the data quality. Afterward, they can visualize their data in tables and charts and/or run descriptive and inferential statistics. SPSS Statistics offers a broad set of statistical tests. And lastly, the results in the output can be saved and processed. (VAN DEN BERG, N.D., [W9]) This work becomes easier with experience and practice. Thus, seasoned statisticians have grown to be skilled and accustomed to SPSS Statistics. In contrast to that, students hailing from other fields of research like economics or the social sciences might not be too familiar with it.

According to the SPSS design team in IBM Studios Böblingen, they usually encounter SPSS Statistics for the first time in their studies in an introductory statistics course. The SPSS design team is currently working on an extensive redesign to improve the user experience and visuals of the program. The new version aims to be more user-friendly and accommodating for beginners, as their research has uncovered that a lot of beginners are currently struggling with SPSS.



[FIG1] A typical workflow in statistical research and activities SPSS.

1.2

Artificial Intelligence

This chapter explains the definition and usage of artificial intelligence and its subfield machine learning. In addition, the topic of data is explained. In order to design and utilize AI, one needs to understand the basic underlying concepts.

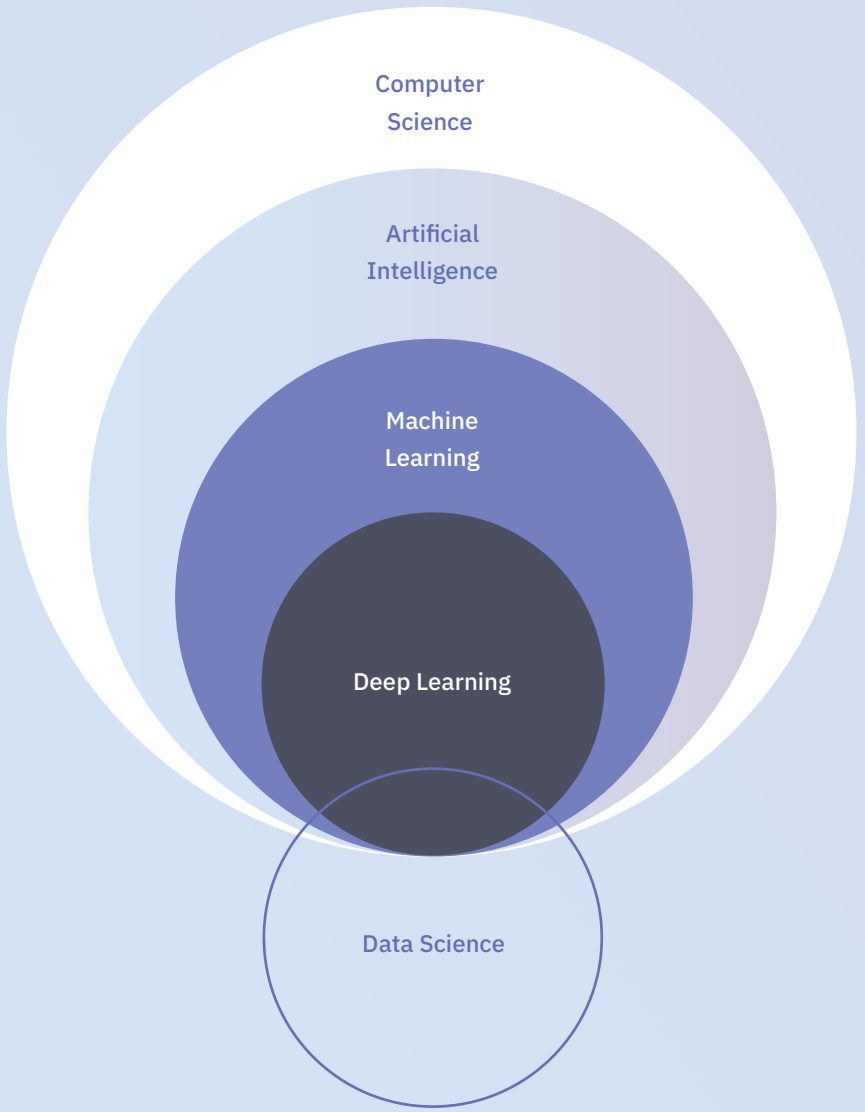
Definition of AI

The field of artificial intelligence (AI) is steadily progressing and along with it the interest in the technology. AI research and development are growing which can be seen in the rising number of papers being published across the continents. (STANFORD UNIVERSITY, 2019, [W2], P.5) In order to be able to discuss and converse about AI, one needs to know what AI technologies are and how they work. However, people's perception and understanding of AI differ greatly. (ELEMENTS OF AI, N.D., [W1])

AI is able to fulfill tasks much more efficiently and effectively than humans. For example the "AI Index 2019 Annual Report" states that "the time required to train a large image classification system on cloud infrastructure has fallen from about three hours in October 2017 to about 88 seconds in July 2019." (STANFORD UNIVERSITY, 2019, [W2], P.5) Such an action is not even possible for humans. By feeding the AI system (also referred to as a machine) data, it can recognize patterns, make predictions, and provide insights. AI systems are developed to amplify our own abilities and intelligence and allow us to be more productive and confident in our work. (IBM, N.D., [W12])

Thanks to science-fiction media, the image of omnipotent, human-surpassing artificial beings has been spread in the general public which mostly does not represent the current state of artificial intelligence. The form of AI that humans actually utilize is called narrow or weak AI. It handles one task or problem of human intelligence as opposed to the general AI that is depicted in science fiction movies for example. General AI describes a machine that is able to handle any intellectual task. While the concept of general AI is still far off from our current state of research, narrow AI is advancing steadily with results. (IBM, N.D., [W12])

Narrow AI has two subcategories: rule-based and example-based AI. The rule-based AI works with algorithms tailored for the goal in question. It follows clear instructions step by step to fulfill its purpose. The example-based AI processes data of different forms to create models through patterns the machine found in the data. This model can be applied to solve new problems. The latter approach is possible



[FIG2] Taxonomy of AI according to Elements of AI.

due to machine learning, a subfield of AI. (IBM, N.D., [W12]) AI itself is related to and includes several other disciplines. (→ [FIG2])
Due to the broad spectrum of definitions, AI counts as a "suitcase word", a term composed by cognitive scientist and AI researcher Marvin Minsky which describes a word encapsulating more meanings than the one you intended to convey. As a consequence, it is important to sort out what exact meaning is referred to when discussing a subject. (ELEMENTS OF AI, N.D., [W1])

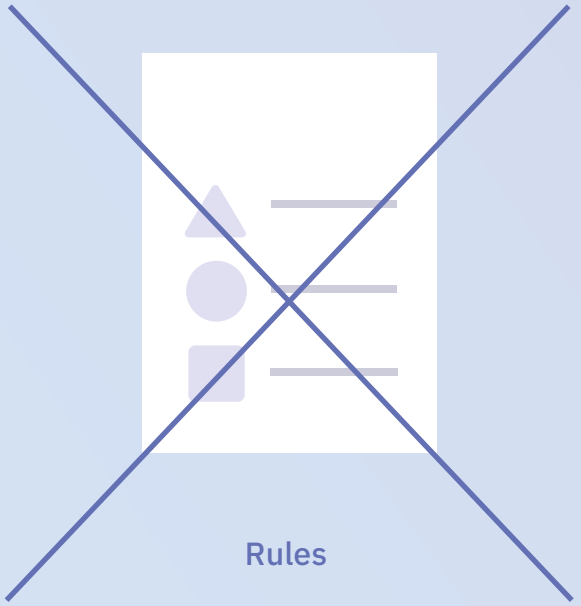
The AI that will be utilized in this thesis is of the narrow kind. It is implemented and optimized to fulfill a specific purpose. Narrow AI is applied in several products or services that people use maybe without even noticing. (ELEMENTS OF AI, N.D., [W1])
The ride-hailing service company Uber, for example, is investing extensively in their AI research and areas of application span from transportation and mobility to customer support and driver-partner navigation. Uber combines several AI techniques to drive their vision forward. Improving location accuracy with sensing and perception so that clients can find their driver faster, and leveraging computer vision for face recognition to increase safety and to validate drivers more efficiently are two of the many accomplishments they shared this year. (GHAHRAMANI, 2019, [W13])
Another use case of AI in daily life is personalized recommendations. Social media platforms, search engines, and music streaming services use AI to filter the content and try to display material relevant to the user first. (ELEMENTS OF AI, N.D., [W1]) The media-services provider Netflix, for example, does not only personalize the recommended movie or show titles, but they also show different artwork in the catalog depending on what the user has preferred in the past. When they have been watching a lot of romance titles before, it is more likely that the title artworks will be love themed and feature romantic scenes. Certain actors and actresses are depicted depending on the user's previous taste. (CHANDRASHEKAR ET AL., 2019, [W14]) While the actual algorithm behind the Netflix recommendations won't be discussed, the basic principles behind the concept will be explained in the next chapter. (→ MACHINE LEARNING)

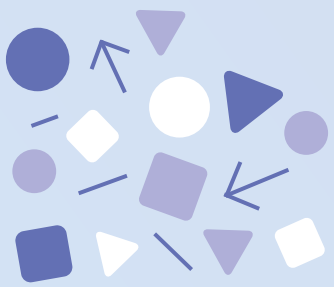
These examples give us an idea of what AI is capable of, but it is still difficult to pin down what exactly artificial intelligence is. A system simulating human intelligence could be called AI, but there are also people defining AI as statistics, business analytics and manually encoded if-then rules. To describe AI in the context of this document, the interpretation of the Elements of AI course by the University of Helsinki will be referenced. Deciding on the characteristics of AI will make it easier to describe the concept of AI. The two core properties listed in the course are autonomy and adaptability. The machine could be called artificially intelligent if it is able to perform tasks in complex environments on its own without guidance and if it can improve its performance by learning from past successes and failures. Of course, that is by far not enough to compare it to human intelligence. (ELEMENTS OF AI, N.D., [W1])

While a machine might be able to learn certain things from a set of data and apply the learned knowledge in situations it was trained for but hasn't encountered yet, a (healthy) human being can fall back on memories and experiences that might not be related to the situation at all. This is where a human is considered more intelligent. The challenges of AI differ from those of humans. The Deep Blue vs Kasparov matches in 1997 is a great example of this. Deep Blue, an IBM computer, beat the world chess champion Garri Kimowitsch Kasparov after a six-game match by utilizing its massive calculating abilities to determine the best moves possible. (IBM, N.D., [W6]) Computing many alternative move sequences at a rate of billions of computations a second is something a human could never accomplish. However, the computer Deep Blue's bigger struggle proved to be picking up the chess pieces and placing them on the board without knocking it over, a task that should be fairly easy for the average person. (ELEMENTS OF AI, N.D., [W1])
But the discussion on how to enable AI to be equal to or even surpass human intelligence is a different topic. For the scope of this paper, AI should be autonomous and adaptive.

What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle.

Marvin Minsky
(MINSKY, 1988, [L1], P. 308)





[FIG3] Rules vs. examples according to IBM Design for AI.

Machine Learning

Machine learning is a subfield of AI and enables adaptivity and AI solutions that improve their performance with more experience or data. (ELEMENTS OF AI, N.D., [W15]) With machine learning, AI can solve problems on its own without being instructed specifically.

In order to reach that level, the machine has to learn first. An untrained machine is essentially like a human baby, it does not know anything yet but will learn essential skills over time. Humans take up information from their environment with their given senses. They then learn how to process that data either through experience or education from others.

So in the case of AI, the system has to be fed with data and either rules or examples on how to handle that data. As mentioned in the chapter before, it is the example-based approach that allows the AI solution to be adaptive and accomplish new tasks. (→ [FIG3]) Without explicit rules and instructions, the machine has to discover the patterns and regularities itself in order to apply them in the real world. (IBM, N.D., [W16])

The roots of machine learning are in statistics, which can also be thought of as the art of extracting knowledge from data.

Elements of AI
(ELEMENTS OF AI, N.D., [W8])

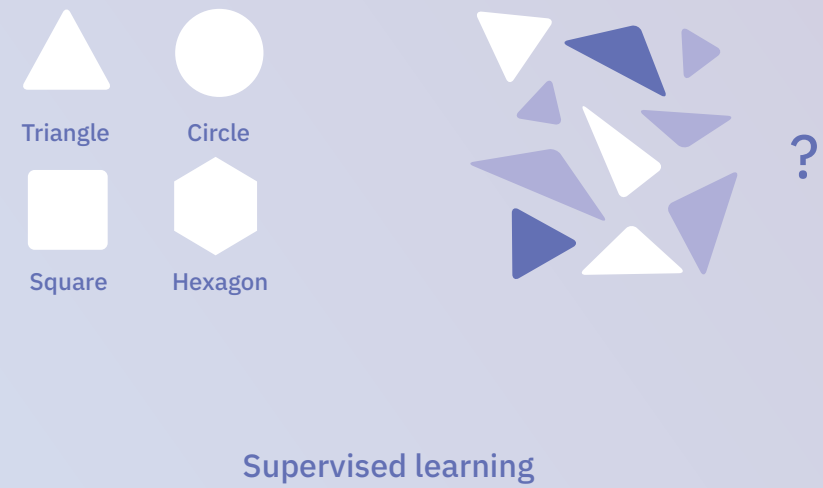
There are several ways to train a machine: supervised learning, unsupervised learning and reinforcement learning. (ELEMENTS OF AI, N.D., [W17])

Supervised learning (→ [FIG4])
The machine receives an input of labeled data as examples and has to make predictions based on the given data. If the outcome is supposed to be a class of an item, classification is used. In case of a number value, a form of regression is used. Supervised learning means providing the model with information on what kind of data it is handling and what kind of characteristics are typical. It is important to separate training data and test data in order to avoid mistakes. (ELEMENTS OF AI, N.D., [W17])

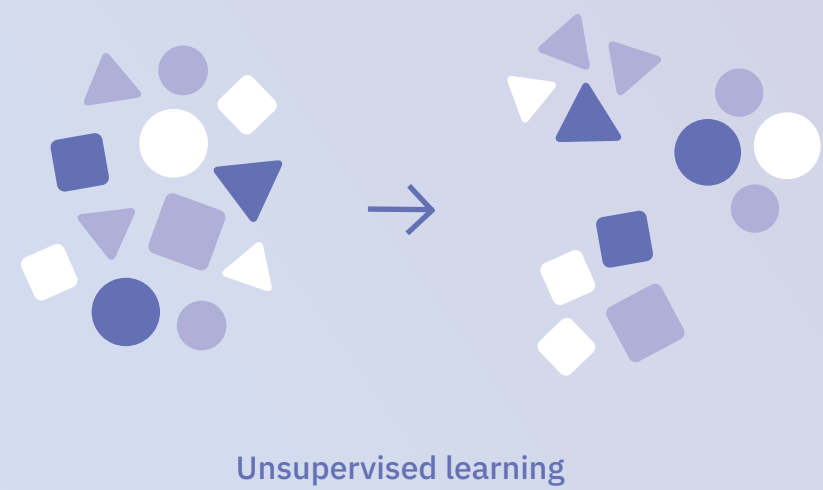
Unsupervised learning (→ [FIG5])
The machine receives an input of unlabeled data as examples and has to determine the structure of the data on its own. Detecting patterns and groupings but also anomalies and outliers are the focus in unsupervised learning. This way of learning is often used to reduce data for relevance and sort data visually to find connections. (IBM, N.D., [W16])

Reinforcement learning
An AI agent has to operate in an environment and based on its interactions with it, the AI agent gets rewarded for positive choices and punished for negative choices. After several tries, it should be able to make sense of the feedback and determine the right behavior for that environment. (IBM, N.D., [W16])

These categories are not necessarily excluding each other. Combined forms also exist, like semi-supervised learning which is partly both categories. (ELEMENTS OF AI, N.D., [W17])



[FIG4] Supervised learning according to IBM Design for AI.

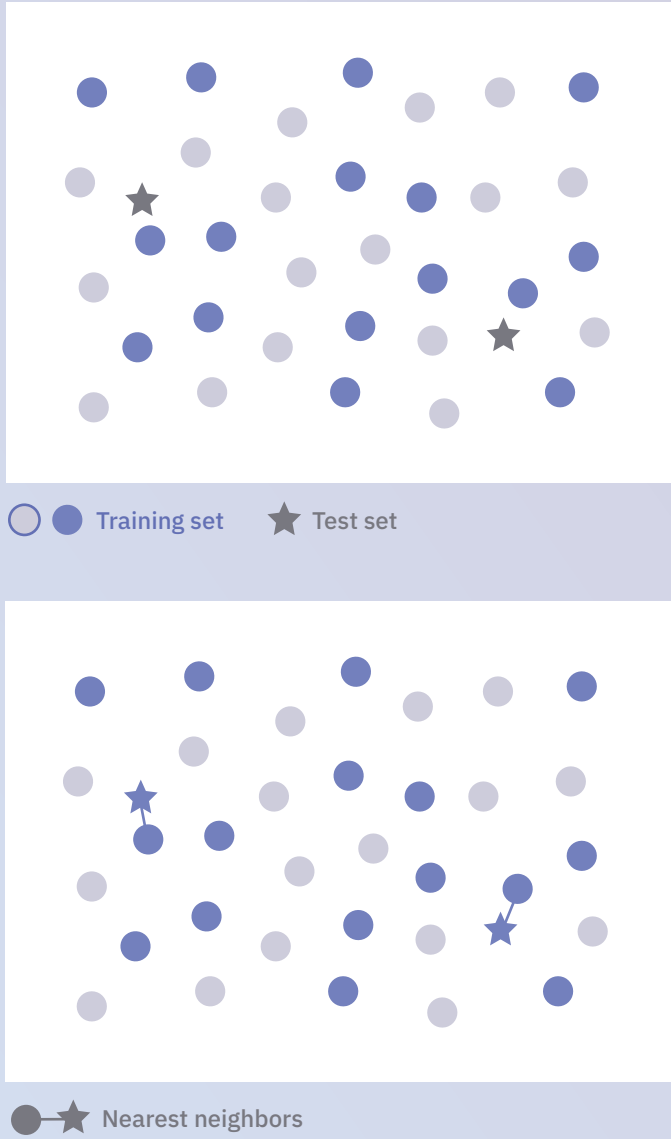


[FIG5] Unsupervised learning according to IBM Design for AI.

Nearest neighbor classifier

An example of supervised learning is the “nearest neighbor classifier”. It classifies a given item from the test data through prediction. The test item will take on the same label as the training item which is “closest” to it in nature. If the data was to be visualized in a two-dimensional plot, the proximity of the item points would indicate that their coordinate values have the smallest distance between them in comparison to other points. (→ [FIG6]) (ELEMENTS OF AI, N.D., [W18])

However, not all datasets can always be represented graphically, so the geometric distance can not be derived from that. In that case, the test item will be assigned a label according to the training item with the highest number of matching features. This concept of classification is often used to predict user behavior. While the example with the Netflix recommendation system from the last chapter might not use this exact version since it requires a lot of other factors as well, it is most likely based on or uses something similar to this algorithm. This is called collaborative filtering: using other users’ data instead of manually entering the data to determine the result. Titles popular with users, which have a similar movie preference to yours, have a high chance of being to your taste as well. Users with similar past behavior tend to also have similar future behavior. Collaborative filtering sorts out content the machine deems incompatible with your preferences, so this can create a filter bubble. The term “filter bubble” is used to describe a situation where the users only see content that is in line with their own values and views. Seeing only things you enjoy might be convenient but it is sometimes important or beneficial to be conscious of other options. An explore section could enable that. (ELEMENTS OF AI, N.D., [W18])



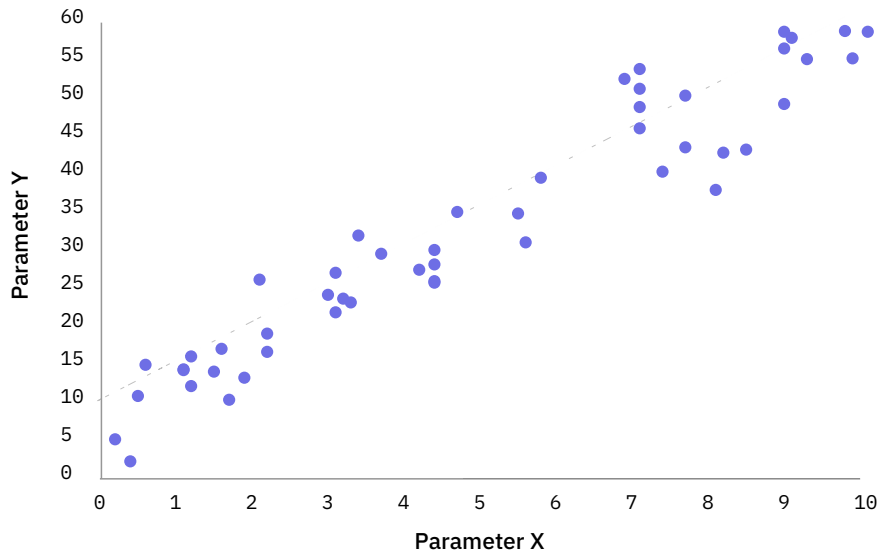
[FIG6] Nearest neighbor classifier according to Elements of AI.

Regression

Another form of supervised learning is the so-called regression. It is also used to do predictions, however, in contrast to classification, regressions have a numerical value as their output. This section will address two kinds: linear regression and logistic regression.(ELEMENTS OF AI, N.D., [W19])
The predicted value is calculated by adding up the effects of the feature variables, also called linear combination in case of linear regression. A direct example is a grocery shopping bill, where the total price is determined by adding the prices of the individual products after multiplying the quantity with the respective unit price. In this case, the quantity of the products would be the input and the total cost would be the output of the regression. The output is dependent on the quantity and unit price where the latter is generally called coefficient or weight in regressions. The starting value given before adding any effects is called intercept. In the shopping bill example, that value would be zero since we are only paying for the products.
If both the input and weights are known, then the only thing left to do is to calculate the output by multiplying and adding. But what if only the input and output are specified while the weights are not disclosed? The weights can also be produced but there are several ways to set the weights and still get the same output, so this is where machine learning comes in. Finding the weight where the predicted value matches the actual output most accurately is a task usually solved by machine learning. The accuracy increases with several output values given in the training data. (ELEMENTS OF AI, N.D., [W19])

The computed values help us make predictions about certain situations like click rates in online advertising or crime rates in a certain area. It is important to keep in mind, that the input values alone are most likely not enough to represent real-life circumstances and it is impossible to include all factors that are actually influencing the result. So linear regressions rather give us information on associations, potential causes and outcomes, which are still very valuable to make predictions. (ELEMENTS OF AI, N.D., [W19])
The nearest neighbor technique determines a class or label and the linear regression produces a numerical output. The method where you transform the linear regression's results (numbers) into predictions about classes is called logistic regression. That is possible by assigning labels to certain value levels. A simple example would be allocating the label A to any output value greater than zero and label B to values less than or equal to zero. Naturally, logistic regressions can also take on more than two labels or classes. (ELEMENTS OF AI, N.D., [W19])

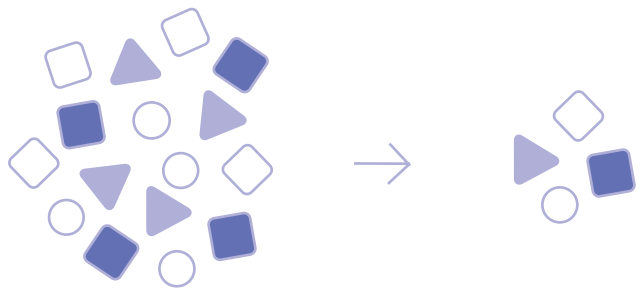
In addition to the mentioned methods, there are countless other approaches for machine learning to extract knowledge as patterns and correlations from data and with that become more "intelligent". They are not foolproof and each of them has their imperfections but producing an imperfect but good prediction is better than being completely clueless. Aspects that are affecting the accuracy and reliability of the results are the difficulty of the problem to be solved, the applicability of the machine learning method chosen and the data quantity as well as data quality. (ELEMENTS OF AI, N.D., [W19])



[FIG7] Visualisation of a linear regression according to Elements of AI.

Data

Data is essentially the backbone of artificial intelligence. Like in statistics, both the quantity and the quality are immensely important to achieve usable results with AI. The data quality is sometimes suffering because of neglect or minimization to save resources and time. Collecting the right data that is relevant to the task at hand and making sure it is clean, without gaps and does not have any significant outliers are very critical tasks. (IBM, N.D., [W20])



[FIG8] Population to sample according to IBM Design for AI.

One way to collect appropriate data is to take a sample that represents a population. In the field of statistics, a population is a specific set of individuals sharing features that are relevant to us. It is self-evidently more challenging to audit every person in that population, requiring more resources, time and effort, even if this would ensure exact results. Samples need to represent the population's characteristics as accurately as possible to avoid receiving misleading results about the population's entirety. (IBM, N.D., [W20])
Another aspect one needs to scrutinize is what kind of sample is needed. That depends on the question about the population we want to answer. The most common samples used are simple random, stratified and clustered. (TAYLOR, 2018, [W6]) Simple random sampling takes a subset of a population and each individual in that subset had the same probability of being selected. (HAYES, 2019, [W21]) Stratified sampling retrieves individual samples out of so-called "strata", subgroups in your population. In cluster sampling, the samples are entire clusters that appear naturally in your population. (GLEN, 2019, [W22])
After gathering enough applicable data and making sure it is complete, it is necessary to sort and organize the data. Related information should be put together in one place to enable a better overview and consistency should run through the whole dataset. (IBM, N.D., [W20])
Organizing and visualizing your data can contribute to spotting patterns and trends before actually doing any analyses. Especially visual representations like for example pie charts or scatter plots often make things very clear from the beginning. (GLEN, 2019, [W22]).
These measures help to understand your data better and recognize biases. Biased data favors or discriminates against certain things or people. Biases may not be apparent at first sight but unrepresentative samples are enough to create a shift in favor of the overrepresented. Subjective data entries, tags, and labels all contribute to biased data and while it may not be critical, it is still an inaccurate depiction of reality. (IBM, N.D., [W20])

In conclusion, it is difficult to guarantee your data to be complete, inclusive and relevant to the goal. That will remain a big challenge for machine learning. However, it is important to be conscious of the significance of data quality and quantity. (IBM, N.D., [W20])

In AI the quality of data is almost everything, as we tend to say, „garbage in, garbage out".

Teemu Roos, Leader of AI Education
at the Finnish Center for Artificial Intelligence
(AALTO UNIVERSITY, 2019, [W23])

1.3

Design

The challenge is to simplify a statistical study with the help of AI. Several factors play into the success of the concept. It will not reach its goal if it is not properly designed, following UX practices and abiding by the rules and ethics of AI. Design principles from multiple disciplines will be applied to ensure a human-centered result with a pleasant user experience.

User Experience Design

The term "user experience (UX)" was first coined by Don Norman in the early 1990s, when he was working at Apple. (NORMAN, 2013, [L2], P. XIIIF) User experience is comprised of all forms of the end-users' interaction with the company, its services, and its products. The focus and most important goal of good UX is to meet the user's needs in every aspect. That requires close cooperation between multiple disciplines like engineering, marketing, graphic design and many more. (NORMAN & NIELSEN, N.D., [W24])

The user experience of a product encompasses three major factors: its look, feel and usability. The look of a product makes it attractive and desirable to use. It pleases the user's expectations and needs visually. Content needs to be conveyed in an understandable and clear way so it is not just about making it "pretty". The product will have a pleasant feel if they are delightful to own and use. The feel of a product goes beyond functionality. However, the product should still be usable. Otherwise, the experience is deemed to be disappointing altogether. (INTERACTION DESIGN FOUNDATION, N.D., [W25])
These factors are handled by user interface design (look), interaction design (feel), and usability engineering (usability), essential subsets of user experience design that often intersect with each other.

I invented the term because I thought human interface and usability were too narrow.

Don Norman in an interview
(LYONNAIS, 2017, [W21])

User experience design often involves a solid understanding of the user's psychology and behavior. Comprehending how the users perceive products allows UX designers to iterate on their concepts and achieve the goals of the user more effectively. (RIGOPOULOS, N.D., [])
According to Marc Hassenzahl and Sarah Diefenbach in their paper "Well-being, need fulfillment, and Experience Design", any positive experience eventually stems from psychological need fulfillment. Fulfilled psychological needs like competence, relatedness, popularity, stimulation, and security will contribute to the user's well-being and result in a pleasant user experience. In experience design, these psychological needs can be taken as guidance or inspiration for the design. (HASSENZAHL & DIEFENBACH, 2012, [L3], P. 1)

Competence is the feeling that you are very capable and effective in your actions rather than incompetent or ineffective.

Relatedness is the feeling that you have regular intimate contact with people who care about you rather than feeling lonely and uncared of.

Popularity describes the feeling that you are liked, respected, and have influence over others rather than feeling like a person whose advice or opinion nobody is interested in.

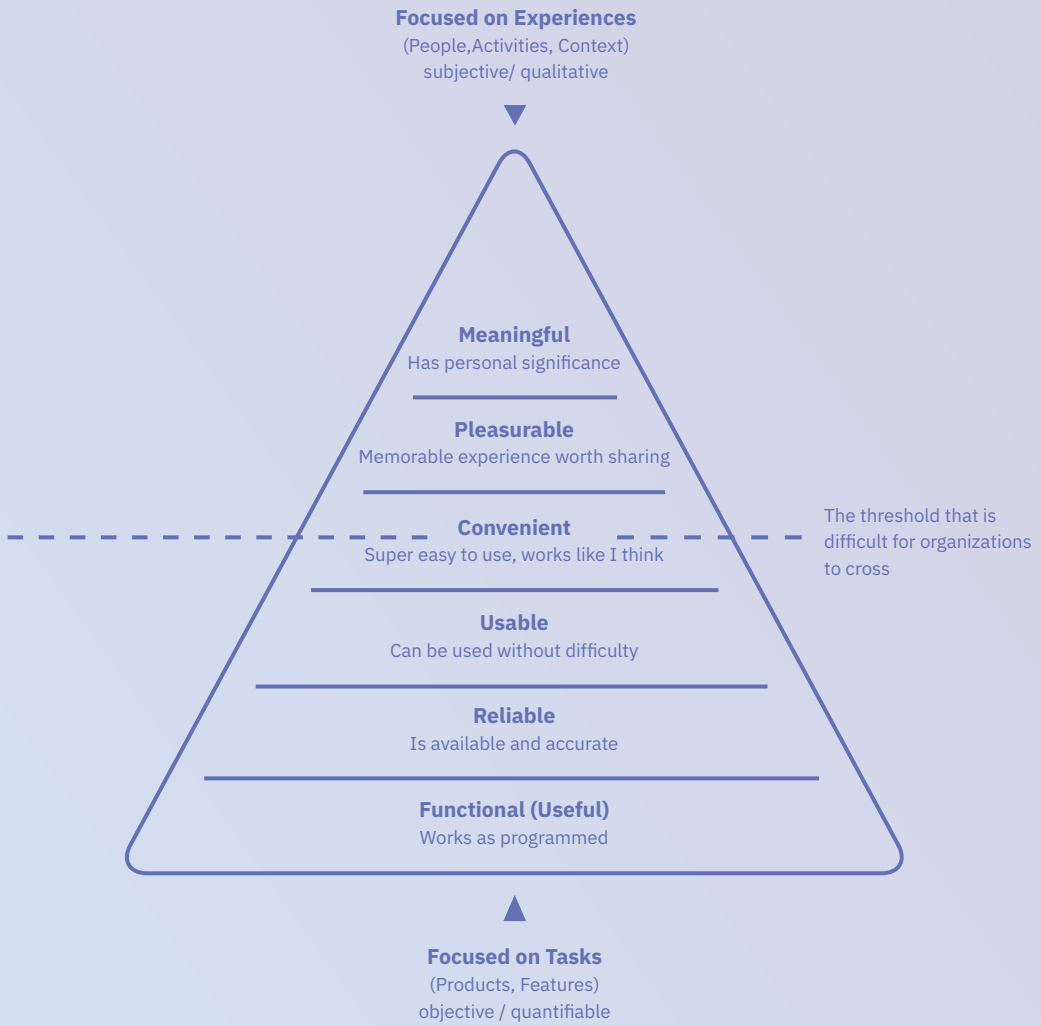
Stimulation means feeling that you get plenty of enjoyment and pleasure rather than feeling bored and understimulated by life.

Security means feeling safe and in control of your life rather than feeling uncertain and threatened by your circumstances.

(HASSENZAHL & DIEFENBACH, 2012, [L3], P. 1)

Since the goal of this thesis is to raise the user's confidence in statistical research, the focus will especially be on the competence need. It is assumed that competence is a direct contributor to confidence.

There are other approaches to defining user needs. Design leader Stephen P. Anderson, for example, developed a different user experience needs model. Rather than focusing on the psychological side, he based his model on the product's maturity process. He proposed that a product can grow from "only" being functional to become meaningful to the user, with six levels of maturity available. (ANDERSON, 2011, [L4], P. 11)



[FIG9] User experience hierarchy of needs model according to Stephen P. Anderson

Functional
The minimum form of maturity a product can have is usually functionality. Being useful and solving the problem which it was made for, counts as being functional. (ANDERSON, 2011, [L4], P. 11)

Reliable
The next step would be reliability, both of the services as well as of the data. The lack of reliability will end in the user losing trust in the product and abandon it altogether. This aspect is especially relevant when personal data is involved. (ANDERSON, 2011, [L4], P. 11)

Usable and Convenient
Products that can be handled without hurdles in a convenient way are positioned around these two levels of maturity. It is insufficient to only be able to use the product, the usage should not feel "awkward". Ease of use makes a system usable, being aligned with the user's expectations makes it convenient to use. (ANDERSON, 2011, [L4], P. 11FF)

Pleasurable
While convenience targets cognition, pleasure is derived from affect and emotions. Emotionally engaging and memorable experiences evoke pleasure in the users. (ANDERSON, 2011, [L4], P. 13)

Meaningful
This is the highest level of maturity a product can reach. Meaning is personal and subjective, so you can't create something meaningful for the user per se. However, designing for meaning is possible by focusing on the preceding levels and factoring beliefs and communities into product development. Deciding on which specific experience you want your users to have and starting from top to bottom will create innovative ideas and solutions. (ANDERSON, 2011, [L4], P. 13)

Designing an experience will only have positive results when the human needs are put in the foreground and are accustomed to. Which needs should be focused on depends on the problem that has to be solved.

Usability Engineering

The usability of a product is a major contributor to its user experience. Guaranteeing good usability through usability engineering is vital in the development of any product. Without it, users cannot achieve their goals efficiently, effectively and satisfactory so they might switch to an alternative product. Hence, lacking usability will result in a loss of customers to competitors. (KOMNINOS, 2020, [W27])
Jakob Nielsen lists the following five quality components.

Learnability
This aspect ensures that users can easily accomplish basic tasks on the first encounter with a design.

Efficiency
After learning the design, users should be able to perform tasks quickly.

Memorability
Users can easily reestablish proficiency when they return to the design after a period of not using it.

Errors
Humans are prone to errors and this focuses on the number of errors the users make, the level of severity and recovery rate.

Satisfaction
Products should call forth a sense of pleasure during usage.

(NIELSEN, 2012, [W28])

Other quality attributes exist other than these five, but an essential one is utility, which refers to the design's functionality. The product should actually do what the user needs. Usability and utility combined result in usefulness. It does not matter if something is easy if it was not the thing you were looking for. Vice versa, a system that offers exactly what you want to do but is too difficult to use is also not useful. (NIELSEN, 2012, [W28])

User Interface Design

Like mentioned before, the "look" of a product is important for the overall user experience. User interface (UI) design is the practice of creating interfaces in digital applications with a focus on looks or style. User interfaces are access points between users and systems, they have to interact with the interface in order to receive feedback or a result from the system in turn. (INTERACTION DESIGN FOUNDATION, N.D., [W29]) UI design requires consideration in both functional and graphical aspects.
In Ben Shneiderman and Catherine Plaisant's *Designing the User Interface*, principles called the "eight golden rules of interface design" were introduced. They are applicable in most interactive systems. (SHNEIDERMAN & PLAISANT, 2005, [L5], P. 74)

Strive for consistency

Similar situations should have the same sequences of actions. Terminology should be used consistently throughout the whole application, for example in prompts, menus and help screens. Visual components like color, layout, and typography should also be coherent but it is still important to highlight exceptions like warnings. (SHNEIDERMAN & PLAISANT, 2005, [L5], P. 74)

Cater to Universal Usability

Different levels of expertise and abilities require flexibility in the design. Enabling adaptive content can improve the perceived system quality, for example with shortcuts and faster pacing for expert users, and explanations for novices. (SHNEIDERMAN & PLAISANT, 2005, [L5], P. 74)

Offer informative feedback

There should be feedback from the system for every action the user takes. For small and repetitive actions, the response can be subtle, but for infrequent and major actions, it should be more significant. (SHNEIDERMAN & PLAISANT, 2005, [L5], P. 74)

Design dialogs to yield closure

Certain action sequences should be sorted into action groups with a start, middle, and end.Informing the users about their completion of a group will trigger satisfaction and a sense of accomplishment in them and they can prepare for the next set of actions. This rule can often be found in e-commerce where the checkout process is divided into smaller steps like entering the address delivery first and the payment details after. (SHNEIDERMAN & PLAISANT, 2005, [L5], P. 75)

Prevent errors

Design the system in a way so that the users cannot make critical errors that easily, for example by disabling or even hiding options that would cause errors. In case of an error, the system should detect it but be tolerant of it and offer comprehensible instructions for recovery. (SHNEIDERMAN & PLAISANT, 2005, [L5], P. 75)

Permit easy reversal of actions

Every action should be reversible. This unburdens the users since they know that mistakes can be undone and leaves room for exploration of unfamiliar options. (SHNEIDERMAN & PLAISANT, 2005, [L5], P. 75)

Support internal locus of control

Experienced users like being control of the interface and having it respond to their actions. They would rather be the initiators than the responders to the system. Unexpected actions, tedious tasks, difficulty in obtaining necessary information and not being able to reach their goal all contribute to the users' frustration. (SHNEIDERMAN & PLAISANT, 2005, [L5], P. 75)

Reduce short-term memory load

Humans can keep approximately "seven plus or minus two" chunks of information in their short-term memory, so displays should be kept simple and clean, multiple pages should be combined into a single more effective or coherent whole and the users should be given enough time to process information. (SHNEIDERMAN & PLAISANT, 2005, [L5], P. 75)

Interaction Design

Next to the usability and look of a product, the third component essential to user experience is the “feel”. That can be achieved through interaction design. The interaction takes place between users and (mostly digital) products. The term “interaction” is rather broad, a user can interact with systems through several channels like voice, motion, sound and many more. (TEO, 2020, [W30]) Gillian Crampton Smith, the director of Interaction Design Institute Ivrea, defined four dimensions of interaction design languages in the foreword of Bill Moggridge’s *Designing Interactions*. (MOGGRIDGE, 2007, [L6], P. IX) Kevin Silver, product manager at IDEXX Neo, added the fifth. (SILVER, 2007, [W31])

1-D Words

This dimension revolves around words and poetry. Words in interface components should encapsulate the action they denote and be used consistently throughout the whole system. The “tone of voice” of the dialog boxes in the system should be neither too abrupt and arrogant nor too annoyingly conversational. (MOGGRIDGE, 2007, [L6], P. XVII)

2-D Visuals

2-D languages concern visual elements like images, typography, diagrams, and icons. They can be used as supplements or replacements of words for communicating with the users. (TEO, 2020, [W30])

3-D Physical objects

Languages in this dimension are of physical form and also address the space around the user. Whether the person uses a laptop with a mouse or touchpad on his office desk, or a smartphone with his fingers on a crowded train, the interaction will differ greatly. (TEO, 2020, [W30])

4-D Time

The fourth dimension is centered around time. Languages in this dimension are for example sound, film, and animation, media forms that change as time passes. (MOGGRIDGE, 2007, [L6], P. XVIII)

5-D Behavior

This encompasses the mechanism of a product. The users’ behavior is defined through how they perform actions and how they operate the product but also how they present themselves and react to the product. (TEO, 2020, [W30])

The user has several channels or languages to interact with the product and vice versa. An interaction designer has to consider the circumstances to choose the right forms of interaction.

AI Design

Fundamentals

The practice of designing pleasant experiences for the user is thoroughly researched and guidelines have proven to be effective. With AI on the march, the question arises whether these design principles are sufficient to still guarantee a delightful UX. Until now, the focus was mostly on relationships between humans and machines. AI presents a new type of machine, one that can use human knowledge, converse with us and learn with time. (IBM, N.D. [W32]) Product designers today face the challenge of how to translate AI into a meaningful experience for everyone. (GOOGLE DESIGN, 2019, [W3])

If you aren't aligned with a human need, you're just going to build a very powerful system to address a very small—or perhaps nonexistent—problem.

Josh Lovejoy
(LOVEJOY, 2018, [W33])

Designing relationships between humans and AI requires new perspectives and considerations, as "traditional" design principles won't be enough. One principle should always prevail though and that is focusing on the users and trying to improve or enrich their life. This way of thinking has been the drive behind innovation in the last decades. The internet, smartphones, instant messaging, all these technologies came into existence in order to elevate our living standards. Our expectations of how a relationship between humans and machines should look like have been shaped by constant progression. AI raised that bar again. (IBM, N.D. [W32])
Best practices in design have to be revised or expanded, taking into account what it means to design a system that is autonomous and adaptive. When creating a meaningful, trustworthy solution for the users, designers are advised to keep in mind the following intents, especially when AI is in use. (IBM, N.D. [W32])

Purpose
The reason for the user to engage and interact with the system. This will change as the relationship between the system and the user will evolve over time. (IBM, N.D. [W32])

Value
The system enriches the users' life and augments their capabilities. (IBM, N.D. [W32])

Trust
The willingness of a user to engage emotionally with the system and investing in the relationship. This trust depends on the reliability of the system, the value it provides and the level of control the user has over the system. (IBM, N.D. [W32])

The core goal of design is to accommodate the user's needs, so when designing AI, one should be careful not to force-fit any technical capabilities enabled through AI just for the sake of it. We are designing for the user, not for the technology. (IBM, N.D. [W32])

Relationship Model

In order to utilize AI for amplifying humanity, it is necessary to establish a meaningful relationship between the user and the machine. How is a relationship designed? Forming and maintaining relationships is a natural process and usually does not require intensive planning and thinking. (IBM, N.D. [W32]) Mark Knapp, a professor at the University of Texas, developed the "relational development model" which describes the ten steps it takes for a relationship to grow, last and end. There are two stages: "coming together" and "coming apart". (COMMUNICATIONTHEORY, N.D. [W34]) Regarding AI, the focus mainly lies on the former which is comprised of the following five steps.

Initiating
First impressions are formed and judgments are made. Each party decides whether they want to pursue this relationship and move on to the phase. The AI introduces its tone, personality and presence. (IBM, N.D. [W32])

Experimenting
If there is a mutual interest, the relationship grows into an exploring phase where the parties can get to know each other and look for common interests and values. The system proves the authenticity of its actions. (IBM, N.D. [W32])

Intensifying
When the parties have enough in common they look for signs in the other that imply their interest in pursuing the relationship further. The system is able to convey multi-step and context-aware interactions. (IBM, N.D. [W32])

Integrating
Both parties are interacting naturally with each other and the relationship deepens continuously. The AI's profile is understood by both the system and the user. (IBM, N.D. [W32])

Bonding
Trust and mutual appreciation have been established between the parties, the relationship lasts until is broken off through a formal notice. This state is difficult to achieve with AI. When the other party is an intelligent machine, it is up to the designers to build a symbiotic relationship where both parties augment each other. (IBM, N.D. [W32])

The other stage „coming apart" describes the steps leading up to a relationship's end. This part is rather irrelevant for an AI-human relationship, one can usually end the connection immediately. Nevertheless, it contributes to a better overview when there is an understand of the whole model.

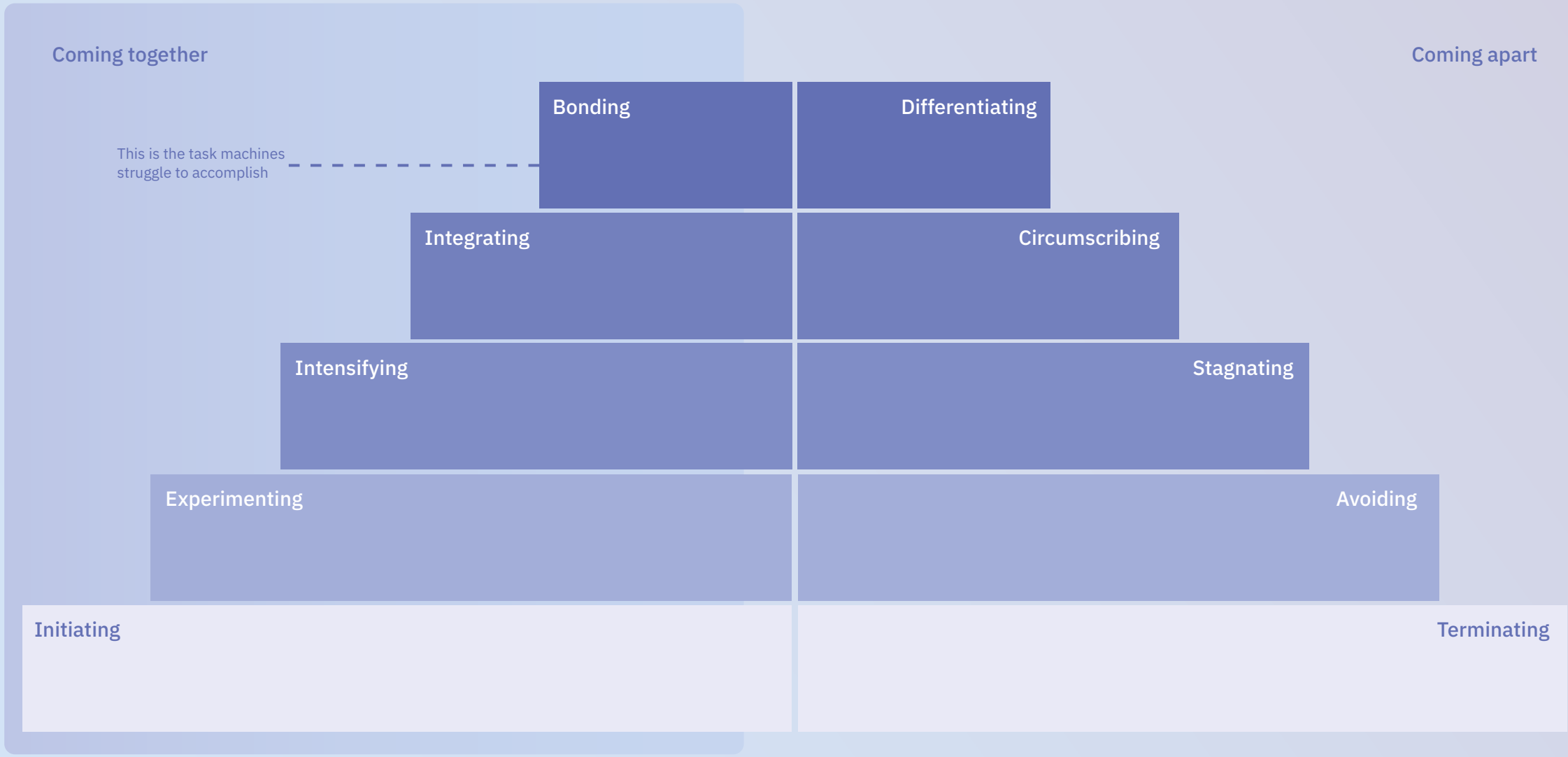
Differentiating
The individual parties start acting independently from one another and move into different directions in terms of personal development. (COMMUNICATIONTHEORY, N.D. [W34])

Circumscribing
Conversations are limited and boundaries are set up in the parties' communication. The personal space and activities of the parties continue to exlude those of the other. (COMMUNICATIONTHEORY, N.D. [W34])

Stagnating
The relationship is declining even more and the amount of communication is constantly decreasing. Relationships in this stage are not likely to continue or improve. (COMMUNICATIONTHEORY, N.D. [W34])

Avoiding
The individual parties intentionally avoid any contacts to each other and are physically detached. Any form of communication is evaded. (COMMUNICATIONTHEORY, N.D. [W34])

Terminating
The final stage of coming apart terminates the relationship. The parties take different paths and will continue their routines without the other. (COMMUNICATIONTHEORY, N.D. [W34])



[FIG10] Knapp's relational development model according to IBM Design for AI.

Ethics

The capabilities of AI are driving human innovation forward. However, with all those impactful AI applications appearing, the question of ethics cannot be avoided. Machines that interact with humans and build relationships with them bear responsibility for their actions but since they are artificially made that responsibility is transferred to everyone involved in the creation. Designers and developers of AI systems have to be conscious of the influence they have on the machines and society. The "Design for AI" team at IBM has defined five ethical focal areas when it comes to working with AI. They specify values AI systems should have and provide guidance for designers and developers building AI. (IBM, N.D. [W35])

Accountability

Every individual involved in the creation of an AI system and also companies invested in its development are accountable for considering the impact the system will have on the world. Through every step of the development, it is necessary to be aware of the fact that it is humans who are writing the code, defining success or failure and deciding how the system will be used. (IBM, N.D. [W36])

Value Alignment

When designing AI systems, the norms and values of the system should be aligned with those of the users. Human decision-making is a process that's driven by values derived from experiences, memories, upbringing, and cultural norms. AI machines don't have those resources to draw upon, so they have to be determined by the designers and developers. They have to collaborate to resolve whose values to focus on and why. (IBM, N.D. [W37])

Explainability

AI systems and their decision processes should be transparent and explainable. Users should be able to easily perceive, detect and understand the AI's actions. The reasoning behind the decisions and the data they are based on should be accessible. (IBM, N.D. [W38])

Fairness

AI systems require the creators to guarantee bias minimization and inclusive representation.
Humans are prone to develop biases and if they generate data or build AI systems while having a biased mindset, that will most likely show in the AI's outcomes. Representative data and diverse teams help to combat biases. (IBM, N.D. [W39])

User Data Rights

The creators and the AI system must protect user data and the users' control over their data. Unauthorized usage and sharing of personal data have to be avoided at all costs. The users should have the right to inquire about their saved data and how it is being used and have it deleted on request. (IBM, N.D. [W40])

These guidelines will evolve over time, as our values and the capabilities of AI are subject to change. Nevertheless, AI designers and developers should still always be aware of their obligation to consider ethical challenges. (IBM, N.D. [W35])

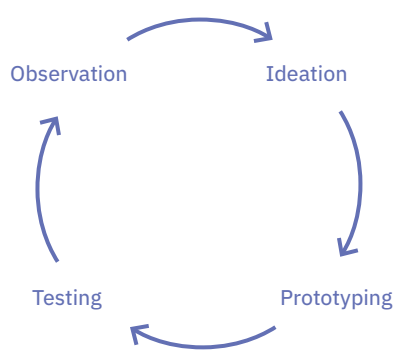
Methods

Human-Centered Design

There are different ways to approach a design process. The human-centered design (HCD) process is one of them. According to Don Norman, it „ensures that the designs match the needs and capabilities of the people for whom they are intended“. HCD puts humans and their needs, capabilities and behavior first and designs around those. (NORMAN, 2013, [L2], P. 8) Norman describes HCD as a philosophy. HCD starts with understanding the humans and their needs and difficulties. The people often don‘t know these themselves and it lies in the hand of the designers to observe and define them. (NORMAN, 2013, [L2], P. 9) There are 4 different phases in HCD.

Observation is the phase where designers research the users‘ circumstances and the underlying problem that needs to be solved. The current market offer is analyzed to find out how the problem has been approached by others. In addition to that, user research is conducted to understand the people whom the solution is for. (NORMAN, 2013, [L2], P. 222F)

Idea Generation (Ideation) produces potential solutions. Based on the findings from the research phase, numerous ideas are gathered, all catered to the defined problem. No idea should be dismissed or focused on too early, the essence of ideation lies in creative variety. (NORMAN, 2013, [L2], P. 226F)



[FIG11] The Iterative Cycle of HCD according to Don Norman.

Prototyping the ideas makes them tangible so that they can be tested later on. As it is important to get feedback early in the design process, prototypes are often not fully functional but rather demonstrate the relevant components that need validation. (NORMAN, 2013, [L2], P. 227)

Testing the prototypes with real users will uncover weaknesses and strengths of the concept. Assumptions can be confirmed and disproved. Insights gathered in this phase shape the subsequent iterations, improving the solution in the process. (NORMAN, 2013, [L2], P. 228)

These four activities are constantly repeated. HCD is an iterative design process. Each iteration produces more insights, bringing the solution closer to its goal. (NORMAN, 2013, [L2], P. 229)

The philosophy and procedures of HCD add deep consideration and study of human needs to the design process, whatever the product or service, whatever the major focus.

Don Norman
(NORMAN, 2013, [L2] P. 9F)

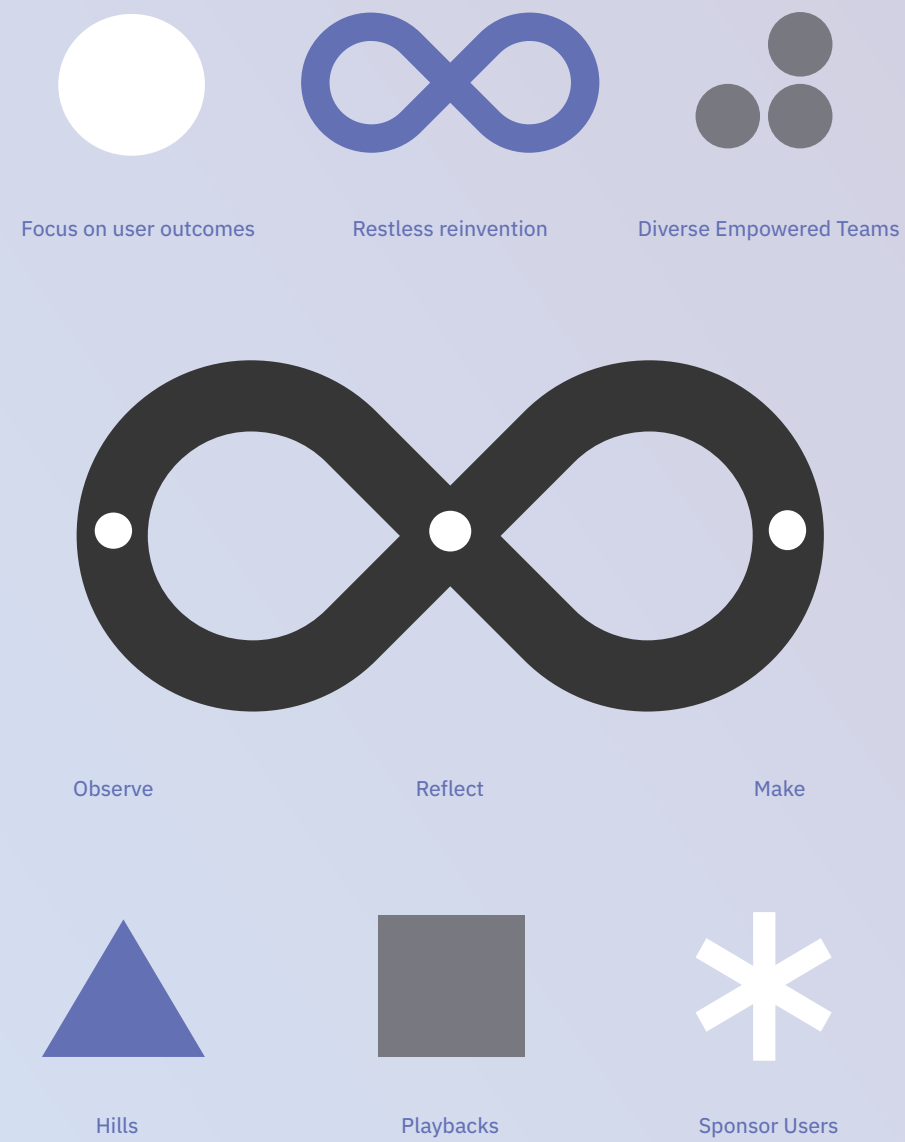
IBM Enterprise Design Thinking

Design thinking is a process that, similarly to HCD, is centered around the user. It focuses especially on creating innovative solutions. (BURMESTER, 2016, [W41]) IBM Enterprise Design Thinking is a framework developed by IBM that modified design thinking for modern enterprises. It is made up of the Principles, the Loop and the Keys. (IBM, N.D., [W42]) This thesis utilizes several aspects and methods from this framework.

The Principles guide designers to follow an approach centered around the user and his needs, accustomed to an enterprise scale.
• A focus on user outcomes by centering the design around the user's needs and goals will drive business in the right direction.
• Restless reinvention is realized by treating everything as a prototype. This should ensure constant improvement.
• Diverse Empowered Teams bring in multifaceted perspectives and opinions which contribute to a product's evolution. (IBM, N.D., [W42])

The Loop describes an iterative process, consisting of a continuous cycle of observing, reflecting and making.
• Observing means empathizing with users of a system or product and understanding their needs. It also includes testing prototypes.
• Reflecting is executed by sharing insights gained from the other steps and making sense of them is the key point in this step.
• Making is possible by transforming abstract ideas into real products that can be tested. (IBM, N.D., [W42])

The Keys help teams to stay focused and aligned in the process.
• Hills formulate a common goal or intent for the team. They define who the users are, what need is focused on and how to do that in an innovative manner.
• Playbacks take place to share the work progress and receive constructive feedback.
• Sponsor Users are invited to actively contribute to the development of a product or service. They share their views and opinions through interviews and tests. (IBM, N.D., [W42])



[FIG12] Symbols of the IBM Enterprise Design Thinking framework

2

Research

After the theoretical foundations are established, the research phase defines the domain and current market offer. The user needs and scenarios are researched through interviews to find out where the pain points lie.

2.1

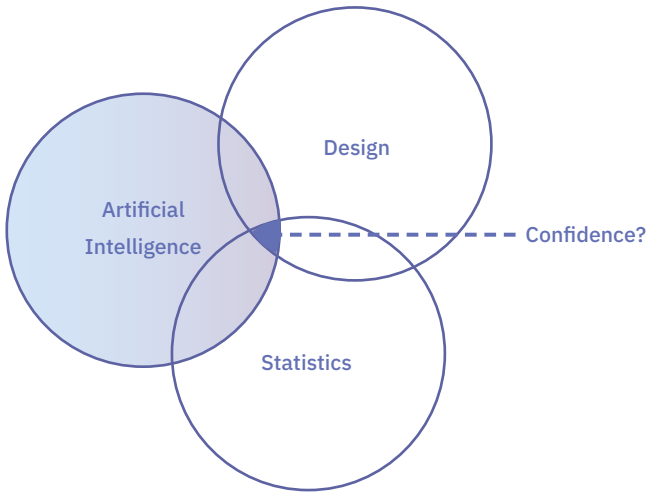
Domain Definition

The domains involved in the subsequent design process will be defined.

With all the theory laid out, the domain can be defined. The goal of this thesis is to find out whether an AI-powered could increase the user's confidence in setting up statistical research studies. This solution needs to be properly designed with the foundation of statistics and artificial intelligence in mind. Thus, the domain in question will be an intersection between statistics, AI, and design.

The use case was proved by the SPSS Statistics design team. They have provided background information and resources on the way users are currently experiencing the preparation and execution of a statistical research study. According to the findings of their design research, it is a tedious process for beginner users of SPSS Statistics. Before actually being able to perform a statistical analysis, they get stuck on preceding steps like defining the relevant variables and deciding which statistical test to use. This situation should also exist outside of SPSS Statistics, the complexity of statistical analysis does not change with the choice of the statistical tool. The beginners feel insecure and anxious if they get stuck and have to fall back on multiple sources of help.

The focus will be on the user's actions before carrying out the analysis in a statistical application which is preparing the statistical research study. This thesis will explore ways to help the beginners in the initial phase of their research.



[FIG13] Domain of the thesis

2.2

Market Analysis

To find out how other solutions are trying to support beginners in statistics, a market analysis is done. This phase will also take a look at exemplary AI-powered applications.

IBM SPSS Statistics Coach

As the software has already been introduced in a previous chapter, this part will only focus on how SPSS Statistics is handling the analysis selection and the support thereof currently.
SPSS Statistics offers guidance in the form of the “Statistics Coach” in the “IBM Knowledge Center”, the help documentation website. The Statistics Coach helps the users determine the appropriate analysis with questions concerning their research. The recommended approach is given in the end along with short instructions on how to open and set up the statistical method in SPSS Statistics.
(IBM, N.D., [W43])

Advantages
- Natural language

Drawbacks
- Static
- Separated from software
- No explanations on why recommended tests

Targeted towards
- Statistics novices

Type
- Standalone document
- Questions approach

Statkat / Statkat Module

Statkat is an online learning environment offering tools to learn and practice statistics. The first feature "Practice" is an interactive environment to exercise statistical tasks. It is a question-based tool that provides help and explanations in case of difficulties to solve the statistical problem. The second option "Structure" presents a table overview with structured information about the different statistical tests. This part also includes a side-by-side comparison to study differences between the analyses. The third and last feature "Select" helps the users select the appropriate inferential analysis according to their research question or homework assignment. These three tools aim to support students in their statistics course and contribute to their learning process. (DE VRIES, N.D., [W44])
Statkat exists as a standalone platform but is also an integrated module in the free statistical software Jamovi. The Statkat module will give the user options on how to continue upon opening a data set. Depending on the measurement levels of the variables, the module will recommend appropriate tests to proceed with and explain which test would fit better. The module also provides links to the Statkat platform for detailed explanations of the tests. (DE VRIES, 2018, [W45])

Advantages
- Clear separation between the three different features
- Explanations are given at any time
- Natural language
- Relatable, friendly tone and voice
- Freely accessible, no registration needed

Drawbacks
- Long text passages

Targeted towards
- Students who are following a course in statistics

Type
- Standalone tool or integrated module
- Table overview and questions approach

Tableau "Ask Data"

Tableau has a broad product portfolio for data analysis, organization, and management. Its strengths lie in powerful visualization tools. Tableau enables users to look at and understand their data. The data can be explored and analyzed via drag and drop. (TABLEAU, N.D., [W46])
In 2018, a feature called "Ask Data" was released in the beta version which allows users to type in questions in natural language in order to gain insights. Instead of manually inserting the variables in the dependency constellation they want, users can "converse" with their data. After selecting a data source, the tool will suggest questions to the users based on the data or they can type in their question or statement. Ask Data will instantly give a response accompanied by a visualization. Tableau will use the chart type for analyses based on best practices but it can be changed. The questions can be iterated on by adding follow-up queries or statements. It is also possible to add and use synonyms. The natural language technology behind Ask Data comprehends ambiguous or underspecified statements and offers helpful solutions. Being able to use their natural language lowers the threshold people usually have to overcome for analytics, enabling them to confidently make data-driven decisions regardless of their background. (MARKAS, 2018, [W47])

Advantages
- Plain natural language
- Quick results
- Flexibility in the wording
- Easy and intuitive

Drawbacks
- Limited to the dataset inserted

Targeted towards
- Individuals and businesses

Type
- Integrated feature
- Questions approach

JASP Student Guide

JASP is an open-source software program for statistical data analysis. Focusing on the simpler analyses before displaying more complicated procedures, JASP wants to accommodate the statistical practitioners in a minimalistic way. It offers common inferential statistical tests. Information on the software and teaching and learning materials are available on the blog website. (JASP, N.D., [W48]) For student support, JASP provides a regularly updated student guidebook which introduces the software, its functions, and its statistical capabilities. It has a chapter called "Which test should I use?" that contains several vertical decision trees partitioned into the different research goals. (GOSS-SAMPSON, 2019, [W49])

Advantages
• Natural language

Drawbacks
• Static
• Very long document, a lot to read

Targeted towards
• Students

Type
• Standalone document
• Vertical decision tree

Intellectus Decision Tree

Intellectus is a statistical software aiming to disrupt the status quo of being "overly technical". Made for non-statisticians from the fields of education, health care, and many other industries, Intellectus provides a platform for efficient and effective data analysis and reporting. (INTELLECTUS STATISTICS, N.D., [W50]) Their priority is to cater to statistics beginners throughout the whole work process. For example, statistical test results are interpreted and displayed in a narrative in natural language, making the reporting easier for students. (INTELLECTUS STATISTICS, 2018, [W51])
The program has an integrated decision tree that helps the user find the appropriate analysis for his planned research. The decision tree indicates visually which analysis would make sense for the objective in question. The users can go back and forth between steps and restart the question tool if they want to. Otherwise, they can opt to conduct the recommended analysis. (INTELLECTUS STATISTICS, 2019, [W52])
The Intellectus decision tree or any kind of statistical test decision tree for that matter covers the more common and basic statistical tests and may not be applicable for more complicated and advanced analyses. (STATISTICS SOLUTIONS, N.D., [W53])

Advantages
• Plain English, natural language
• Interactive tool
• Status indicator in the decision tree
• Explorative approach

Drawbacks
• Cluttered content
• Dry pathfinding process, very short explanations

Targeted towards
• Non-statisticians
• Students, universities, and businesses

Type
• Integrated feature
• Horizontal decision tree

Ada Health Companion

Ada Health is a personal health guide app that intelligently supports the user's healthcare journey. It utilizes AI algorithms and self-learning AI technology to assess the user's symptoms, to suggest a diagnosis and give advice on how to proceed next. Ada's quality of medical and scientific data is higher than that of other medical chatbots, making it stand out on the market. It can process more cases and insights and learn from interactions with the user to build a personalized profile. Ada starts by asking the user a set of questions driven by medical reasoning and machine learning to check symptoms based on the information given. It supports the user early in the healthcare journey. Ada offers a source of reliability to patients who tend to google their health symptoms before getting a consultation with a doctor. (NIKOLOVA, N.D., [W54])

The assessment takes place in the form of a conversation. Ada compares the user's symptoms with that of thousands of other users to calculate the most probable causes. In addition to symptom assessment, the users also has access to a condition library to educate themselves about possible causes and risks. (ADA, N.D., [W55])

Advantages

- Plain language, understandable
- Personalized treatment through adaptivity
- Empathic tone
- Transparency in data usage

Drawbacks

- none

Targeted towards

- Regular people without in-depth medicinal knowledge

Type

- Questions approach

Google Flights

Google Flights is an online flight booking search service through which people can purchase airline tickets on third-party websites. A distinct innovative feature is the possibility to get results even when not specifying certain values; for example, the user can search for flights by only entering the dates and budget, Google Flights will offer various destinations to choose from. (WIKIPEDIA, N.D., [W56])

Vice versa, the user can also choose a destination and receive information on every price for each day of the next year calculated by Google Flights. The cheapest flights can be detected quickly this way. When the search has been performed and the results are on display, Google Flights also uses AI to indicate whether the prices are high, typical or low compared to what you would usually find, in order to enable people to book flights more confidently. When possible, the service also informs them how the price has changed in the past months and notifies them if a rise or drop is expected or not. (GOOGLE DESIGN, N.D., [W57])

If the price does drop after the person has booked their flight already, Google Flights will refund them the difference. However, this feature is only available for select itineraries. In addition to the flight insights, there are also recommendations for the next steps like booking hotels and finding activities to do. (HOLDEN, 2019, [W58])

Advantages

- Plain language, understandable
- Consideration of AI confidence level
- Trust-building through refund in case of inaccurate prediction

Drawbacks

- none

Targeted towards

- People all over the world

Type

- Insights tool

2.3

User Research

In order to delevop the concept for the users first, user research is essential. Finding out what they think of their current situations and what their pain points are, will set the path for the rest of the design process.

Initial Interviews

In order to get insights from people who are actually affected by the challenge of setting up a statistical research, interviews were held with 4 students. They have had a statistics course in their studies before or are currently enrolled in one, and have used SPSS Statistics in that course. While a few of them were a bit more advanced, a major part of them were beginners in the field. Most of the courses were introductory to the basics of statistical analysis and the usage of statistical software. Their purpose is to prepare the students for future research studies that will be carried out by them independently.

The majority of the interviewees stated that they were currently not confident enough in their statistical skills to begin a research study on their own, let alone use SPSS Statistics with it. Even if they had a research goal in mind and the relevant data collected already, they were unsure on how to proceed. What to do with the variables and which statistical test they should use were decisions where they got stuck and required external help. They would have to depend on the support of their professors or tutors to find out what exactly to do. Google and lecture scripts were also listed as the main sources of guidance in challenging situations. As stated several of the interviewees, “constantly needing to look up information makes them feel incompetent and stupid”.

The interviews helped to understand their point of view when it came to preparing their study before actually using statistical software. It was ensured that the interviewees understood the purpose of the research and the approximate scope of the planned deliverable. They should comprehend how the collected information will be utilized to guarantee transparency and trust. Additional information inquired was their educational background and their use of technology in general.

Personas

Personas are created based on the user research outcomes. They embody the behaviors, attitudes, education, use of technology, workflow and the like of a particular user group. A thought through persona can help understanding the users better. (FAULKNER, 2018, [W65])

The following personas were constructed with the help of the preceding interviews.

Emma

Emma is a 24-year-old student, currently in her last semester in a Social Work undergraduate program. She enrolled in a mandatory statistics course in her sophomore year which was supposed to teach her the common statistical tests and the basic usage of a statistical package, in this case, that was SPSS Statistics. The usual course lesson consisted of assignments done in the classroom where the students could get support from tutors and the professor. Emma mainly did statistical analyses in groups with her classmates and the assignments had step by step instructions on which analysis to use. She just had to follow the exact steps that were listed.

Scenario

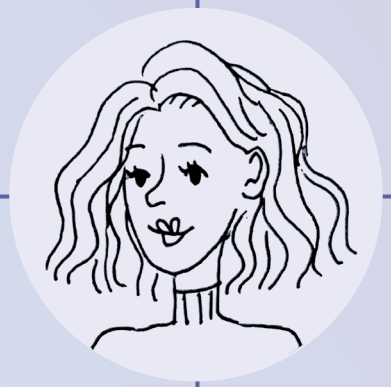
Emma is researching the mental health of youths with troubled family backgrounds. She wants to want confirm whether children of divorcees are more likely to get a mental disorder. Now that she has to do independent studies for her thesis she is at a loss on how to proceed in her specific case. She wants to use SPSS Statistics for her Bachelor's thesis and already collected quantitative data through questionnaires. However, she is unsure about which variables to select in SPSS and which statistical test is the right one to use. She looked into her statistics course notes but couldn't find a similar case.

Says

- „I don't know anything"
- „How often will I need this? Once."
- „I just enrolled in that statistics course because I had to"
- „What are all these foreign sounding words?"

Thinks

- Doing research studies is tedious
- Research studies require so many things I have no idea about
- Statistical tests?
- SPSS is hard to use



Does

- Works on her bachelor thesis now
- Looks into her university scripts to find answers
- Googles a lot
- Needs help from her statistics professor and tutor

Feels

- Insecure
- Confused
- Irritated
- Desperate
- Incompetent
- Stupid

[FIG14] Empathy map for Emma

Howard

Howard is a 22-year-old student, currently doing an Economics undergraduate program, with a Finance minor. He wants to become an analyst. Currently, he is in his second year and enrolled in mandatory courses in statistics and the principles of empirical analysis. The courses introduce the students to the statistical package SPSS Statistics. Due to his personal interest in data science, he is eager to learn how to apply statistics. He is not fully confident in conducting statistical research studies on his own yet due to the vast procedures possible.

Scenario

For a school assignment, Howard has to research the correlation between the brand of a cell phone and the buying decision of cell phone consumers. He was provided with a dataset and has a hunch about which variables are the relevant ones. However, he is unsure about the statistical test he needs to use. He resorts to Google for more information. There are a lot of platforms with a lot of explanations but he can't find tips on his specific case.

Says

- „If only I knew how to use the statistical tests properly"
- „Why is this the right answer?

Thinks

- Data science is interesting
- Being skilled in statistics will give me good job chances
- Statistics websites are good for help, it's just a lot too read

Does

- Works on his assignments diligently
- Looks up terms he does not under stand
- Googles a lot
- Explores possibilities in statistical research

Feels

- Curious
- Interested
- Like a beginner
- Happy when he finds the right answer

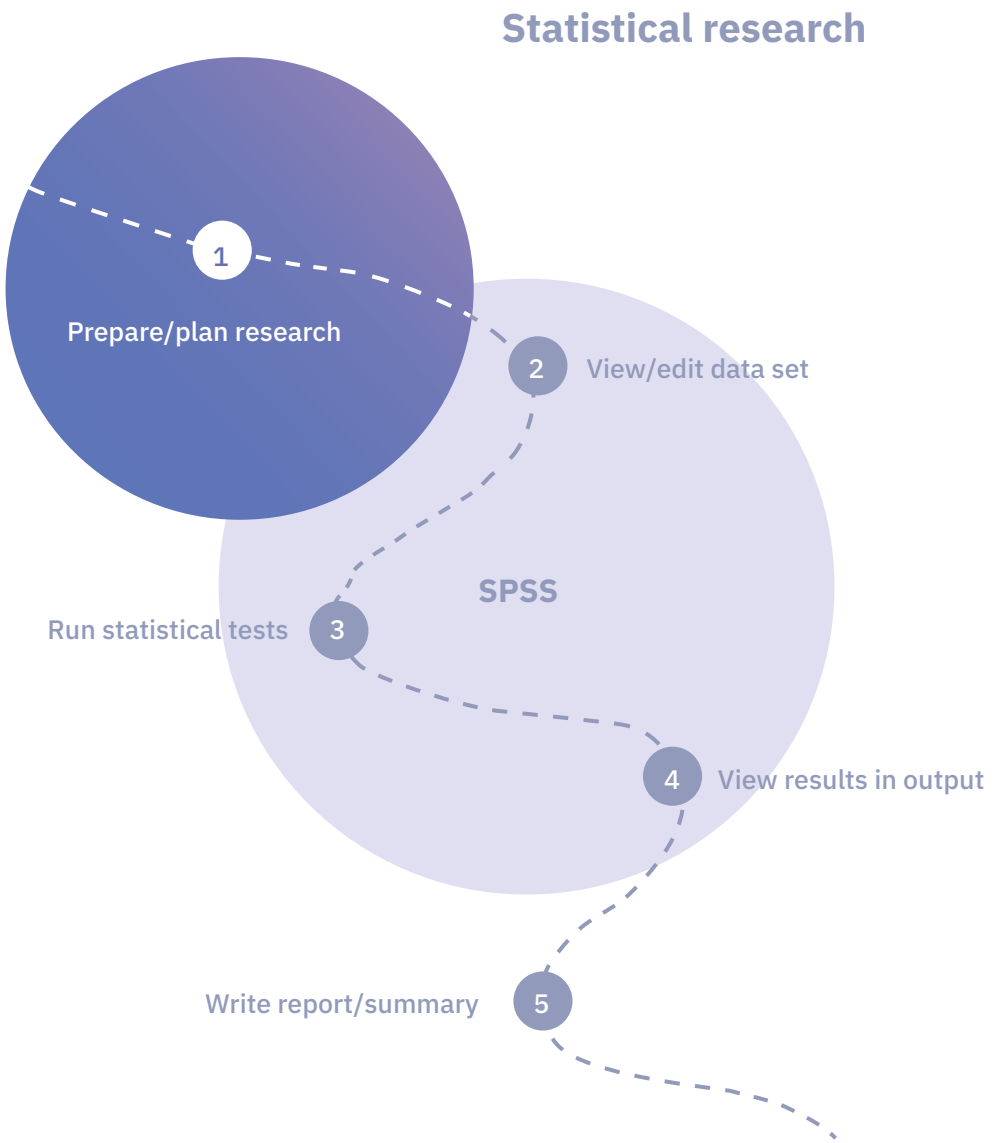
[FIG15] Empathy map for Howard

Goal Definition

The user research has contributed to refining the problem area of interest. Through the interviews it could be gathered that beginners in statistics are indeed struggling to set up their research studies due to the vast amount of information needed. Their sources of help are divided between the internet, books and professors. The students are unsure about how to handle their data and how to proceed with the statistical research. SPSS is not very beginner friendly. These circumstances combined make them feel unconfident in their skills and generally incompetent in statistical research.

The part of the students‘ workflow that will be focused on is the preparation phase. The students are planning their research beforehand and have collected/are collecting their data. They have a research goal in mind but does not exactly know how to achieve it. The goal is to support the students in their research so that they can gain the confidence to tackle the challenge and feel prepared to do so.

The next step will be to find solutions to the proposed problem independently from an AI-powered possibility. Gathering as many ideas and concepts as possible is easier without being fixated on the technology and features.



[FIG16] Focal area of this thesis

3

Conception

With the status quo defined through personas and scenarios, the process can move on to the conception phase. The researched information will be used to generate ideas and concepts centered around the user's needs.

3.1

Ideation

During the ideation part rough high-level concepts for a solution are collected. The main focus is to diverge creatively and gather ideas directed towards the defined pain points.

Big Idea Vignettes

The solution is kind of like...

Big Idea Vignettes are an IBM Design Thinking method to rapidly generate ideas and solutions. The concepts are visualized with sketches and explanations start with „it's kind of like...". A big idea describes the experience a user might have with the solution, not the features. (IBM, N.D., [W59])

... a parent. It teaches users what they need to know for their statistical research. While they are not confident on their own yet, it will carry them.

... a conversational assistant. It understands the user's needs and tries its best to help. They can communicate with it comfortably in their natural language.

... an oracle. It predicts the paths suited for the users after a consultation. They receive suggestions but it is up to them to decide what to do with the „prophecies".

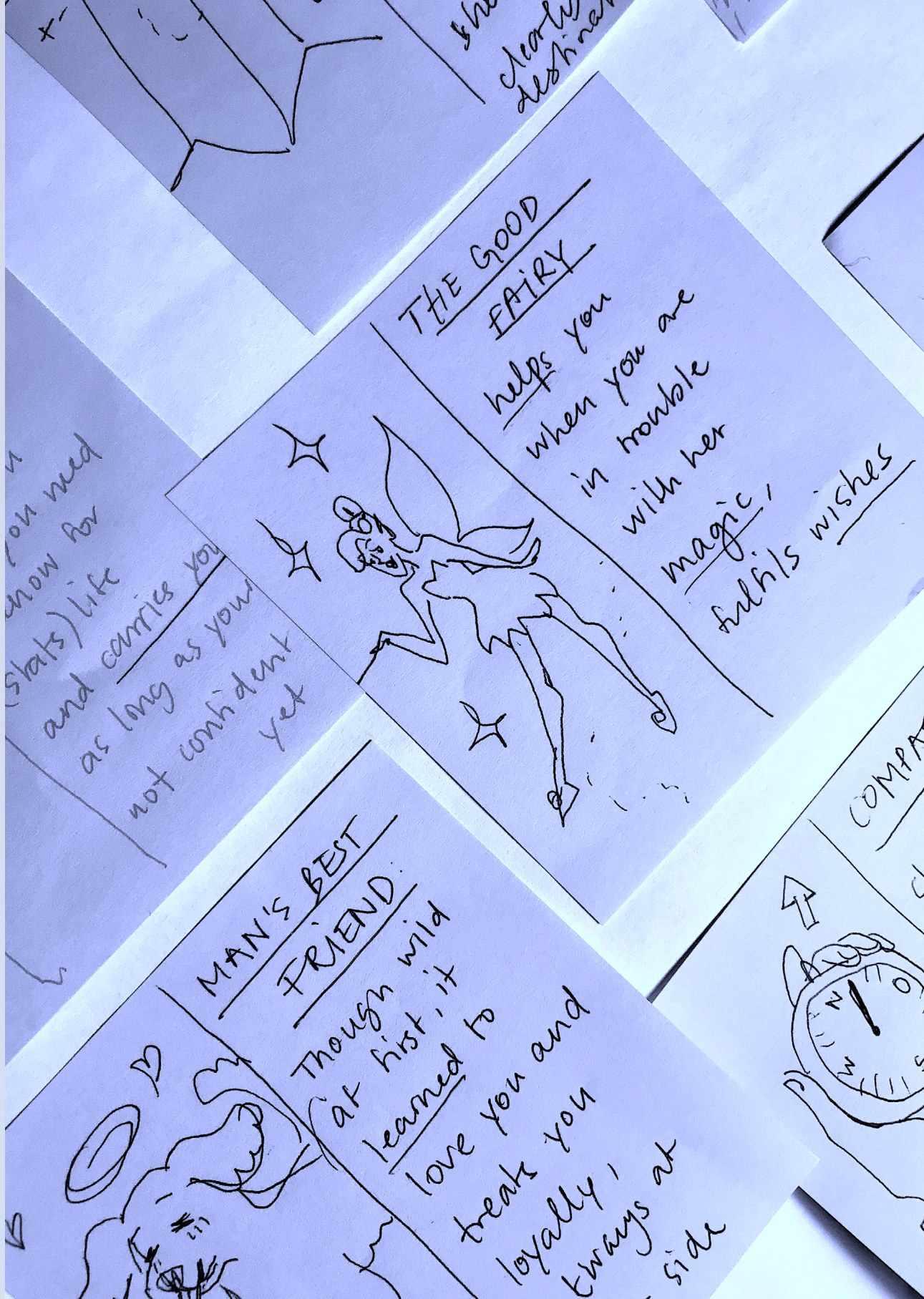
... ultra vision goggles. It filters out unnecessary information and highlights the relevant parts. The user can find what he needs more effectively.

... a fairy. It magically helps the user when they are stuck and fulfills their wishes as best as it can. Having this support by their side makes it easier for the user to tackle challenges.

... a puppy. Though untamed at first, it will grow with the user and learn to love the user. The friendly companion is loyal to the user and knows how to treat them.

... a compass. It shows the user where to go when they are lost on their adventures and offers reliability in uncertain situations.

... a map. It helps the users locate where they currently stand so they can move towards the destination.



[FIG17] Photo of the Big Idea Vignettes created during the ideation

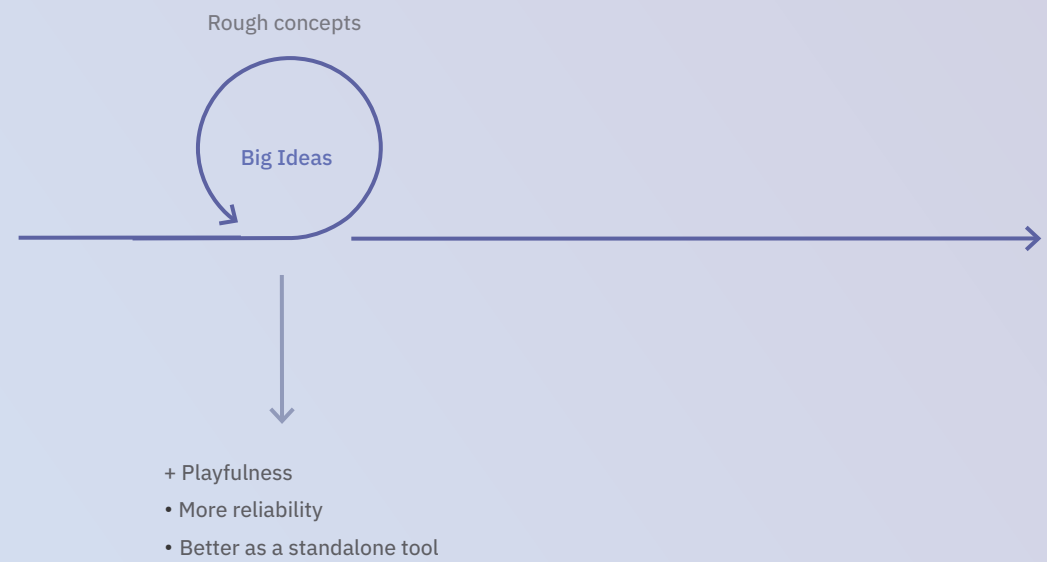
Feedback

User interviews

The big idea vignettes were presented with sketches to a couple of test users who fit the personas created in the earlier phase. They shared their impressions on the concepts and personal suggestions were added.
In general, the ideas were perceived positively. The users explained that they felt like they would receive help with these solutions. The concepts would make it less „intimidating“ to go into statistical research on their own.
A few comments regarding reliability were added. For example, every presented idea should provide explanations for their help. Recommendations, predictions, and mistakes should be clear to the user. Having that available would ease their anxiety about uncertainty and make them feel more confident in the process.
Lastly, they preferred playfulness in the ideas. Making the metaphorical approach fun would lower the threshold of entering statistical research.

Expert feedback

The concepts were also shown to other designers from IBM Studios Böblingen. It was explained to them that the implementation of an integrated tool was considered. The ideas were going in a good direction. However, one addressed the fact that the scalability of an integrated tool could be hard to manage. The person suggested that a different approach should be taken. A standalone tool such as an app fits the purpose of this thesis better which is to create a general concept for supporting the user in their statistical research studies. Moving in with a standalone tool makes it possible to iterate more quickly on drafts and there are no restrictions from any existing statistical software. They also mentioned that if the concept was to be tailored to the SPSS Statistics environment there could be issues with confidentiality.



[FIG18] Feedback from the ideation phase

3.2

Initial Concepts

Certain ideas are filtered out and combined to create first low-fidelity prototypes that can be tested already.

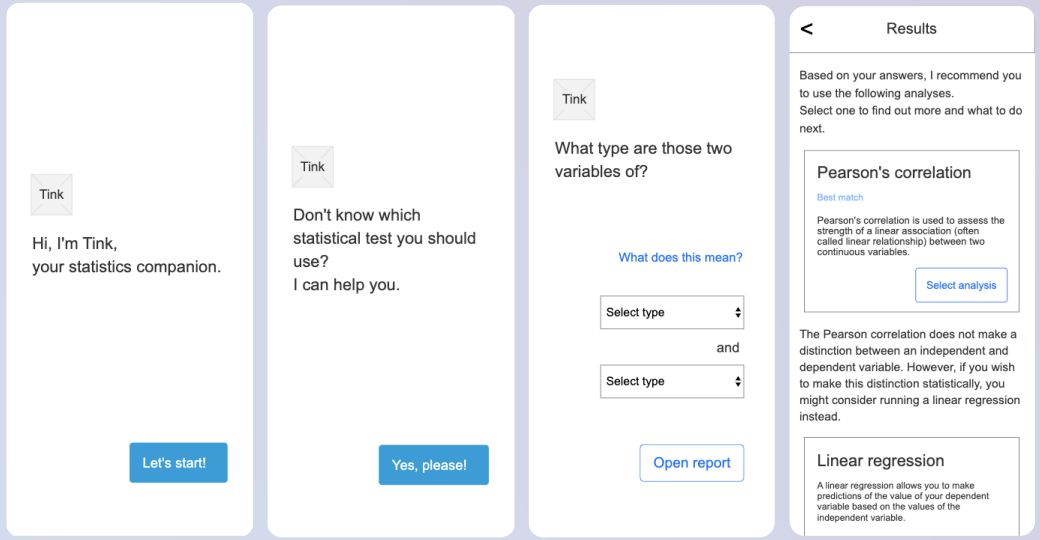
Low-Fidelity Prototypes

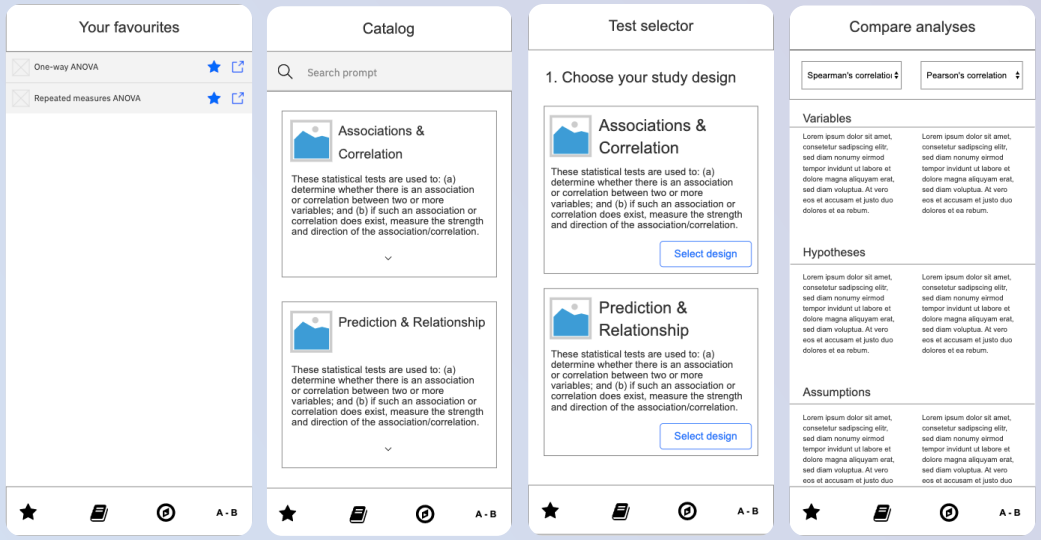
Low-fidelity (lo-fi) prototypes focus on quick testing of the concepts and less on the visuals. (BHOWMICK, 2018, [W60])

The low-fidelity prototypes were created in Axure RP, a wireframing and prototyping tool. They do not have detailed functionality but rather demonstrate an image of the whole concept. Features are portrayed superficially.

The magical fairy approach tries to accommodate Emma's preferences. It offers guidance and helps in the form of a companion app. The user is guided through the process of setting up their statistical research study in a conversational style. The main feature is the statistical test selection.

The pocket toolkit approach is directed towards Howard users. It combines several useful tools for statistical research in a single app. The users can look up statistical terms and find detailed explanations for procedures they don't understand. They can make their search entries in their natural language. The pocket toolkit will understand their intents nonetheless and show results matching the level of their respective inputs.





[FIG19] - [FIG26] Screenshots from the lo-fi prototypes

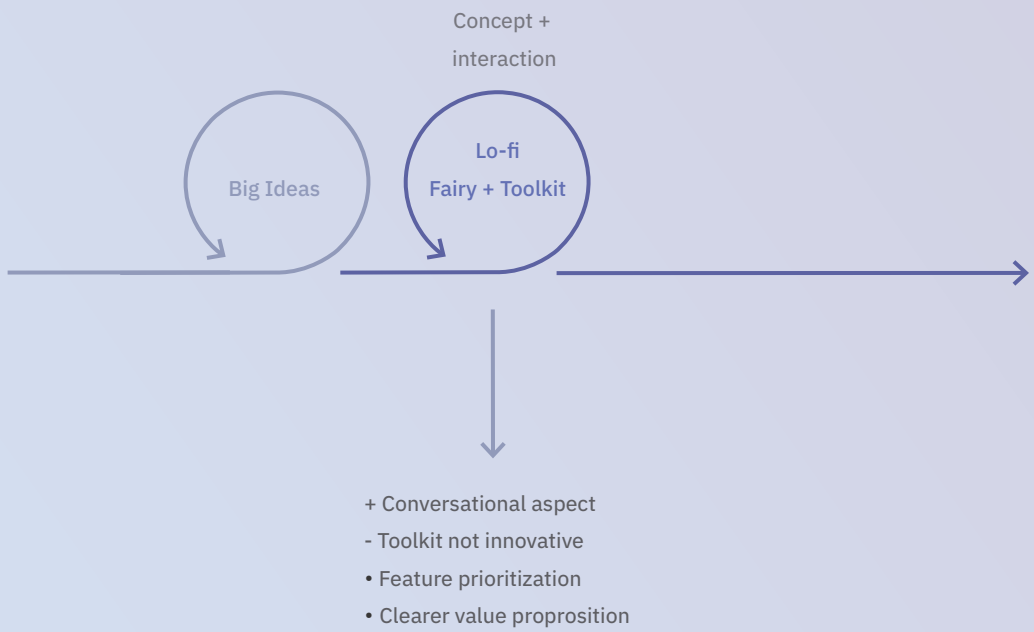
Feedback

User tests

The two low-fidelity prototypes were presented to both Emma and Howard users to get their feedback on the general concept and find out where to improve.
The Emma test users especially enjoyed the conversational style of the *magical fairy* concept. According to one, it made statistics as a subject more interesting. The *pocket toolkit* was less attractive to them, because „it just felt like a statistics book as an app“. It might be more useful and effective in terms of education compared to the magical fairy but if they wanted that they could just open a book.
The Howard test users generally liked both approaches. One asked if the two concepts could just be combined instead. Having the useful features of the pocket toolkit and the fun characteristics of the magical fairy would be a possible compromise.

Design critique

A designer from the IBM Studios Böblingen tried the prototypes and shared his view as an expert. A merge of the two concepts was suggested but the features should be laid out against each other and prioritized. A conversational and personal companion proposes a higher value compared to a mobile statistics toolkit. It would be more beneficial to the user to have one aspect expanded more instead of having a lot of features included.



[FIG27] Feedback for the initial concepts

3.3

Iteration 1

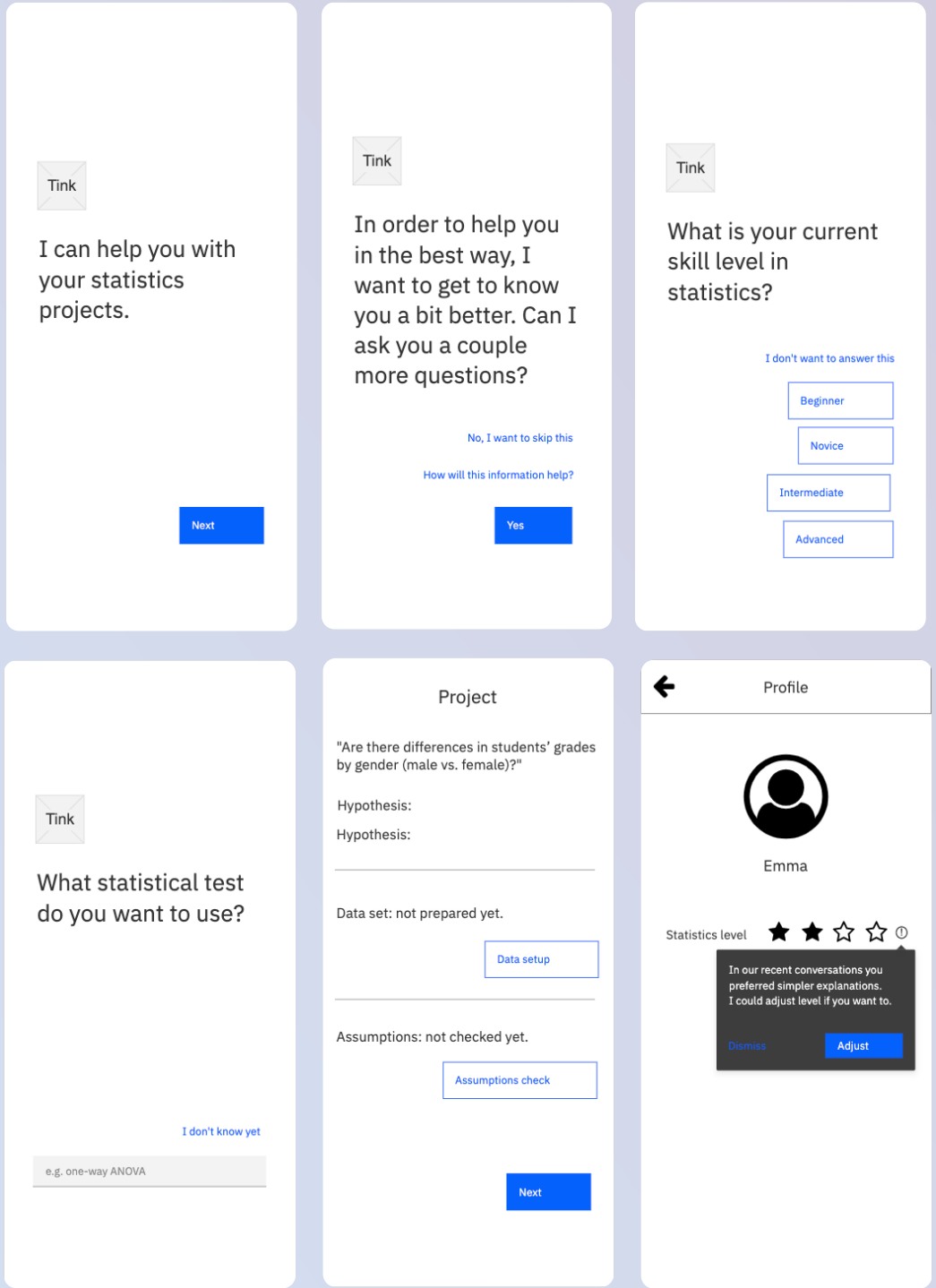
Feedback and critique are processed into an iterated version of the previous low-fidelity prototype.

Low-Fidelity Prototype

Low-fidelity (lo-fi) prototypes focus on quick testing of the concepts and less on the visuals. (BHOWMICK, 2018, [W60])

The low-fidelity prototype was created in Axure RP, a wireframing and prototyping tool. It does not have detailed functionality. The process of setting up a research with the appropriate statistical test is explored. Features like the profile are portrayed superficially.

Tink, the statistics companion is an expanded version of the magical fairy. It incorporates the fun aspects from the earlier concept and has a more concrete structure in its conversations. Tink has a personality that the users should be able to relate to and feel comfortable with. They are addressed by a self-defined name.
The first use begins with optional questions about the user's circumstances, such as their statistical knowledge level, their reason for using Tink and educational background. Through these, Tink will try to tailor the experience towards the user.
Explanations differ from user to user depending on what statistical knowledge level was chosen in the beginning. Beginners will receive more simple and short texts that are easier to understand. Over time, Tink can analyze the user's behavior and speed of understanding. The content will be adjusted accordingly. The users can view and edit their current statistics level in the profile section. The user's educational background can also help Tink customize the usage.
Tink guides the user through the first steps of a statistical research study such as preparing the data set and choosing the appropriate statistical test for their research goal.



[FIG28] - [FIG33] Screenshots from lo-fi Tink

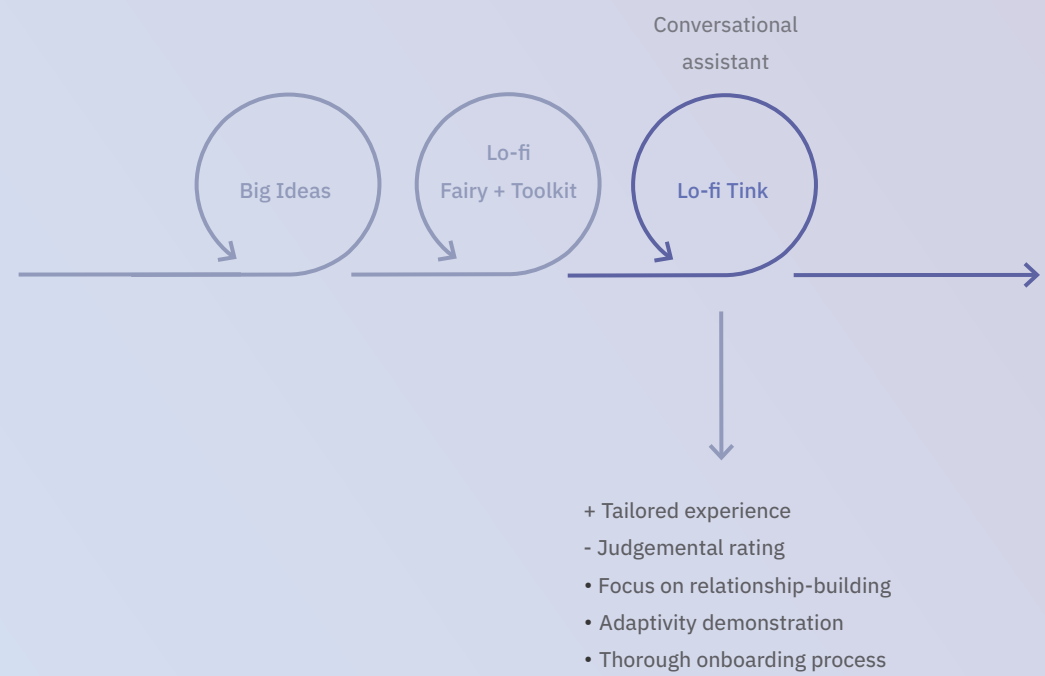
Feedback

User tests

The revised low-fidelity prototype was presented to test users. The text adjustment feature received mixed responses. It was generally a positive thing to have content tailored to one's preferences and needs but some test users express that it felt like being assessed. The experience was similar to teachers ranking their students according to their grades.

Design critique

Advice on the relationship-building was given. One designer emphasized the importance of the first phase of Knapp's relationship model, the initiating part. This is where the users decide whether they want to proceed with the application or not. Clarifying the application's value should take part in this phase. So a fun onboarding experience would be worth investing more time in. One designer recommended looking into how the language learning platform Duolingo onboards their users. They go right into a case of translating a sentence to demonstrate their value. The users experience a success moment very early. This motivates them to continue.
Another aspect which could be improved is the statistics level assessment. While it is good to be transparent about the personalized adjustments, it might be better to take an approach that is less judgemental.



[FIG34] Feedback from the first iteration

3.4

Iteration 2

Another iteration is done based on the received feedback. Certain aspects are expanded in more detail to leverage the capabilities of AI.

Mid-Fidelity Prototype

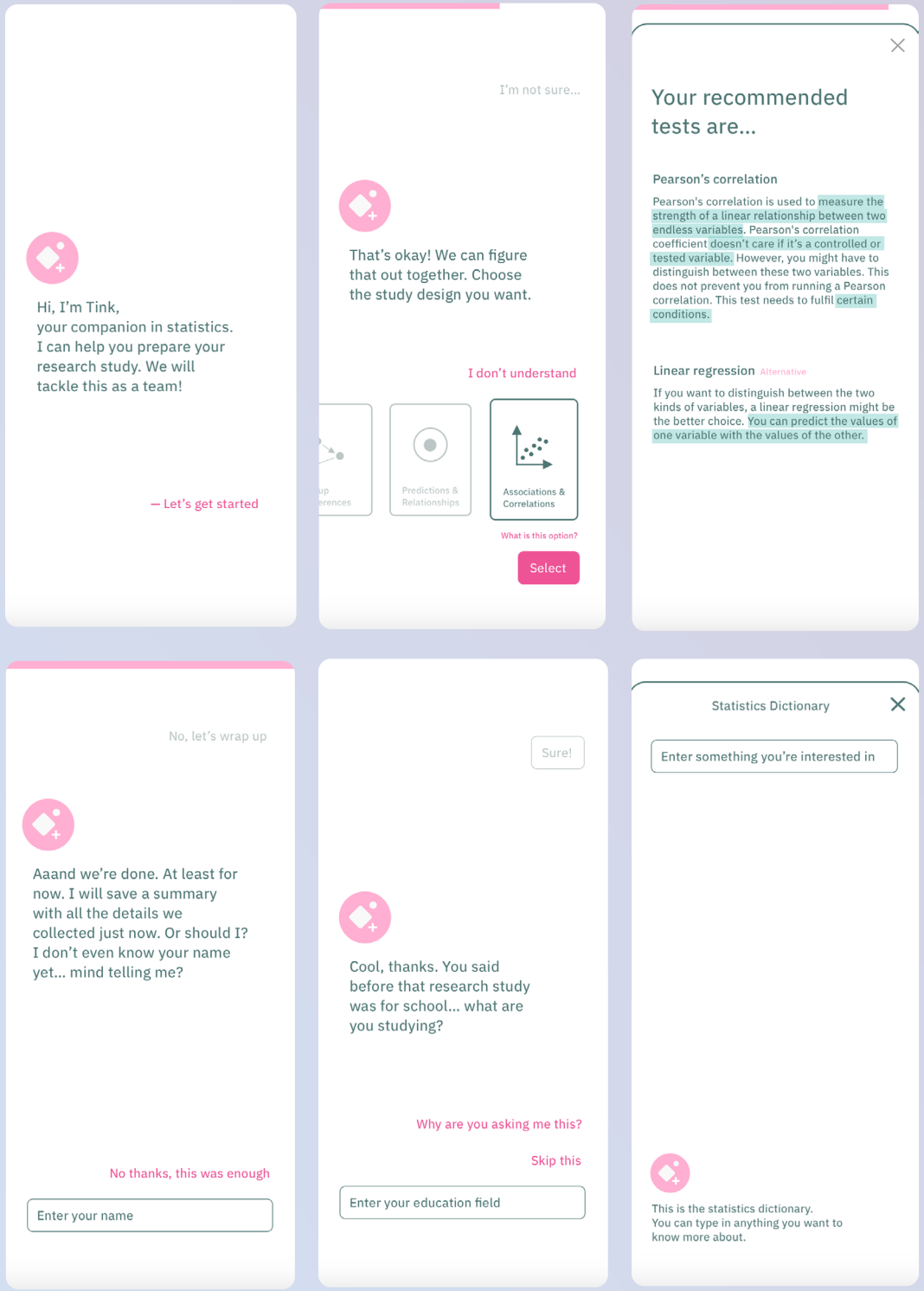
To make it more tangible, the concept went into a mid-fidelity state in this iteration.

Mid-fidelity prototypes have more depth than low-fidelity prototypes but are not as detailed as high-fidelity prototypes. (BHOWMICK, 2018, [W60])

The mid-fidelity prototype here was created with the design toolkit Sketch. It further expands the onboarding process.

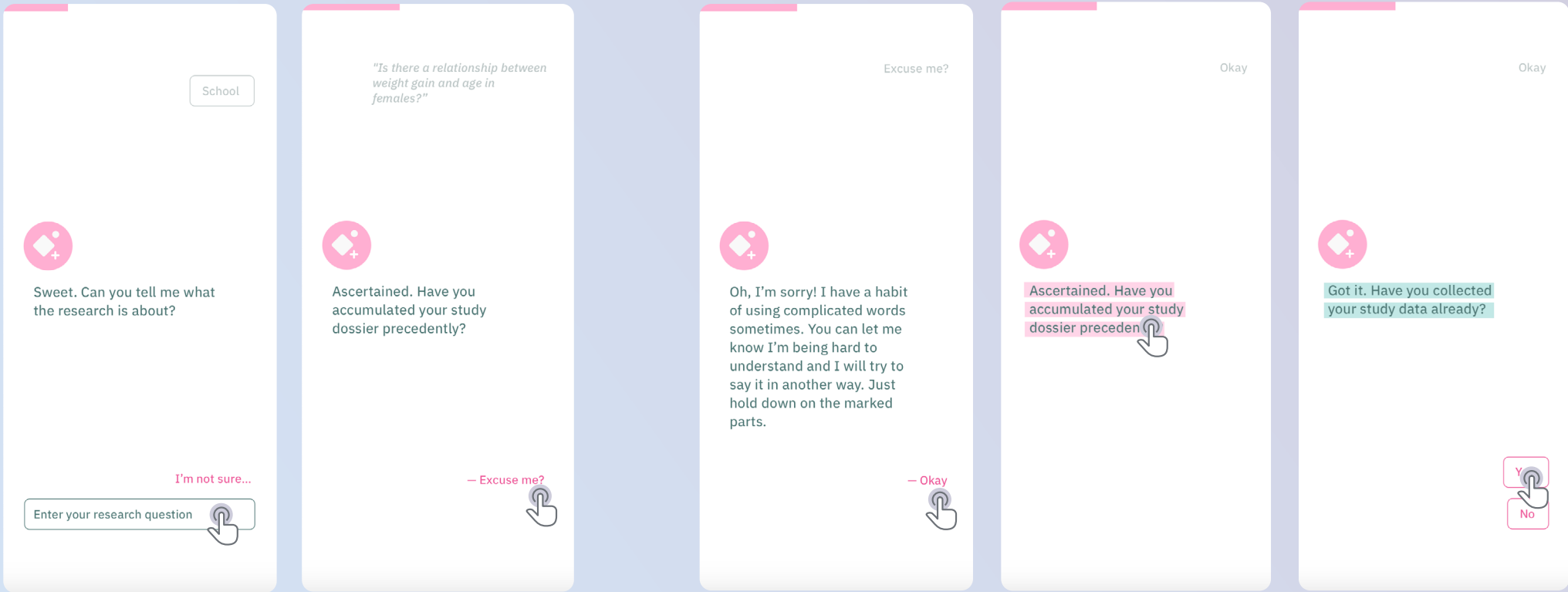
Tink, the statistics companion now has a more distinct personality. Noticeable is the fun and friendly tone of voice, a trait that was praised and very popular in the past iterations.
Every person learns and understands at their own pace. Having alternative versions of a text could help the users be more confident. The statistical level rating was omitted. In this iteration, there are two levels of complexity. The pink highlighted text indicates a more advanced phrasing and vocabulary. The green highlighted parts display a simplified version of the pink text.
After the introduction, Tink transitions directly into a research study case. The users start setting up a new project for the research study and Tink helps them in the process. In order to familiarize the user with the text adapting feature, Tink demonstrates it very early in the onboarding.
There is also a dictionary where the user can look up everything related to statistics. The content can also be adjusted here.

The application's visuals are going in a minimalistic direction, having only a few accent colors.



[FIG35] - [FIG40] Screenshots from mid-fi Tink

To teach the user about the text adjustment feature, Tink has a trap set at the beginning of the onboarding.



[FIG41] -[FIG45] Screenshots of the mid-fidelity prototype

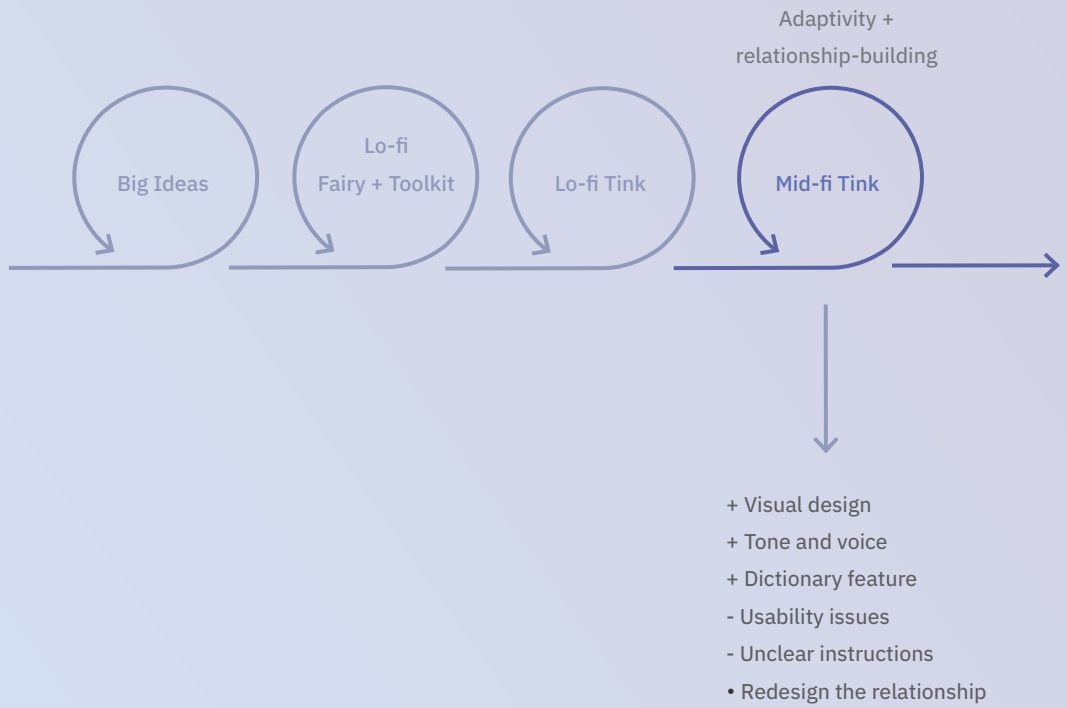
Feedback

User tests

The test users liked this iteration‘s visuals. They amplify the application‘s overall image which one Emma test user described as whimsical. The playful voice and tone became better compared to the last iteration. A Howard user voiced their delight about having the dictionary component included. While the person does need support in the research study setup, looking up certain terms would be very helpful.

Design critique

The medium-fidelity prototype was demonstrated to several designers in IBM Studios Böblingen. They went through the complete flow of the mid-fidelity prototype.
The tone and voice were well received in this iteration as well. It made the experience fun and engaging. This contributed to building a relationship successfully but Tink should not move too quickly. Rushing the initiating phase or overstepping the users‘ personal boundaries can make them uncomfortable.
In addition, the text adapting and dictionary were good approaches to help the users be confident in their statistical studies. However, the instructions are insufficient to explain how the feature works and when it is available. They need to be more clear.
There were also a couple of usability problems. The progress bar indicated at the top did not have enough affordance and did not inform the users about where they are headed. Button designs were not very consistent and the text was sometimes not readable.
The onboarding and statistical research setup process are not visually distinguished from the „idle“ state. One designer said a „hub“ was necessary, a space the user returns to after completing a task.
It was suggested that Tink could take up different states depending on the situation. The users need to understand how the application differentiates itself from conventional tools.



[FIG46] Feedback from the second iteration

4

Prototype

The outcome of this chapter is a high-fidelity prototype that will be tested and evaluated.

4.1

Structure

Before the final prototype is introduced, the scope and structure are explained and the tools used are listed.

Functionality scope

Focus

The feedback received from the preceding conception phase steered the prototypes into a certain direction. The feature which was focused on the most in order to help the test users set up their statistical research study was Tink's adaptive functionality. In the presented concept, Tink can assess the users' statistics skills and knowledge level through interacting with them. Tink offers alternative versions of the text shown to accomodate to the users' preferences and help them understand the content better. A tap on a button switches to an alternative text and it's up to the user to decide whether it is more understandable than the other version. Tink can analyze this interaction and predict for the next instance which level of content might be best suited for the user. This is the aspect that should make the users realize that this application is intelligent.
Adaptive experiences allow the user to work at their own pace and level. Not googling everytime they don't understand something but rather getting more understandable content automatically enables them to feel like they are in control of the situation. In case of uncertainty they can find explanations in the dictionary section of the application.
To construct a relationship between the user and Tink, the user needs to be convinced of this capability in the first few moments of usage (the initiating phase). Hence, the prototype will focus on the introduction and onboarding part of the app. Whether these aspects contribute to an increased confidence in the user, will be evaluated later.

Dimensions

In the preceding prototypes, the fidelity levels increased from iteration to iteration. As stated before, the focus manifested itself in the onboarding phase to demonstrate the adaptive aspect. While the last iterations validated the general concept, interaction, and features on a broader level, the final prototype dives deeper for the onboarding in terms of detail.

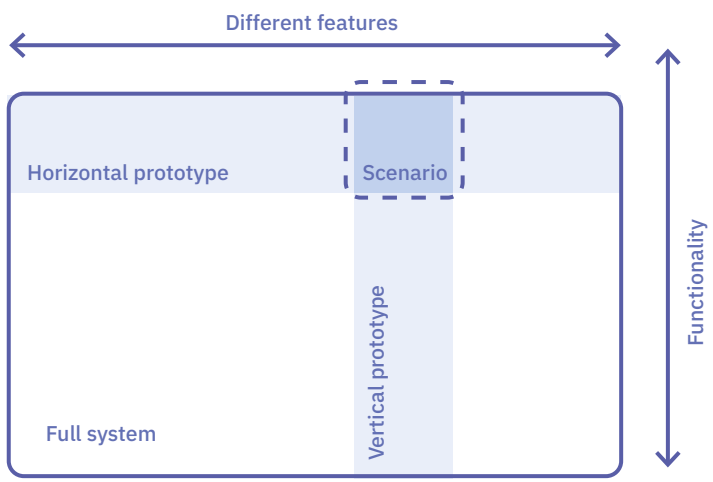
There are two ways one can implement a prototype.

Horizontal prototypes demonstrate a wide range of features that are not fully implemented. They show the entire surface level of a system without any underlying functionality. This type of prototype is useful to explore a system‘s infrastructure. (NIELSEN, 1993, [L7], P. 95)

Vertical prototypes cut down on the number of features. They demonstrate in-depth functionality for a few selected features. Because of their limited scope they can only test parts of a full system but tests realistic circumstances are possible. (NIELSEN, 1993, [L7], P. 95)

Scenarios limits the complexity of a prototype even further. They reduce both the number of features and the level of functionality by directing the users to follow a specific path (NIELSEN, 1993, [L7], P. 95)

The final prototype will be a vertical prototype with a predefined scenario, exploring the adaptive functionality of Tink in the onboarding and its effect on the user‘s confidence.



[FIG47] Two Dimensions of Prototyping according to Jacob Nielsen

Tools

Sketch

The design toolkit Sketch was used to modify icons and create the Tink logo. The previous iteration was implemented in Sketch so a lot of the screens could be reused for improvement. Most of the visual components were imported into Proto.io.

Proto.io

Proto.io is used as the main prototyping tool for the final iteration. Designers can create fully-interactive high-fidelity prototypes with this online tool. The process is quick in intuitive where code is not necessary. Interactions can be implemented via drag and drop and can be chained on top of each other so several actions can be done with one trigger.
Proto.io has a built-in component and icon library which can be utilized for quick prototyping. The prototype can be exported and shared through a link that can be opened in a browser or the dedicated Proto.io mobile app. The latter allows the prototype to be tested directly on its target device. (PROTO.IO, N.D., [W61])

4.2

Final Iteration

The prototype is moving into its final phase. Feedback from the earlier iterations is processed and incorporated as well as possible.

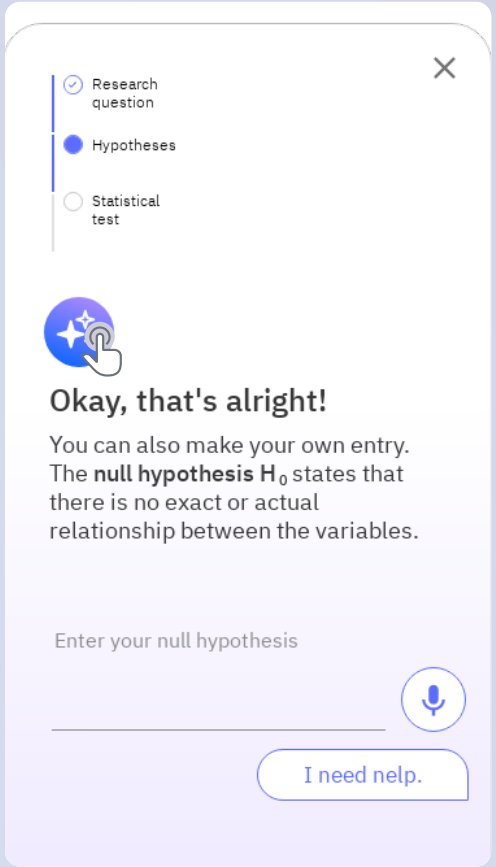
High-Fidelity Prototype

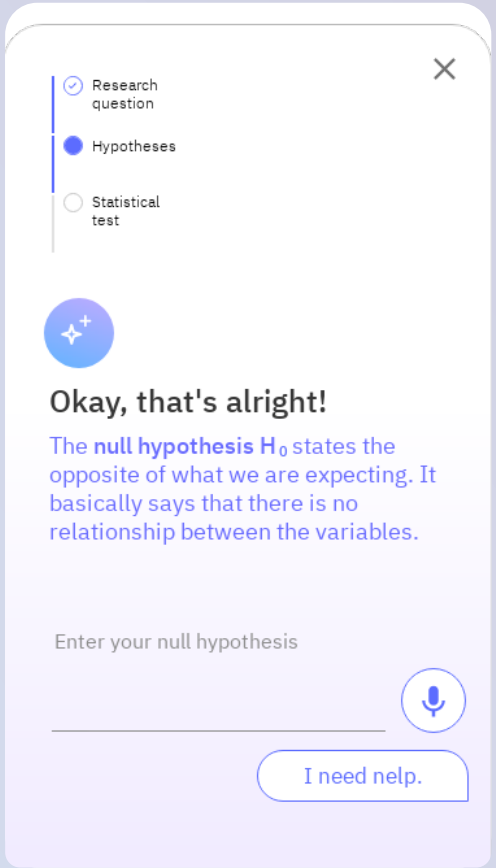
High-fidelity (hi-fi) prototypes try to be as detailed as possible to simulate the experience of a real product. (BHOWMICK, 2018, [W60])

The high-fidelity prototype here was created on the prototyping platform Proto.io. It focuses on the onboarding process.

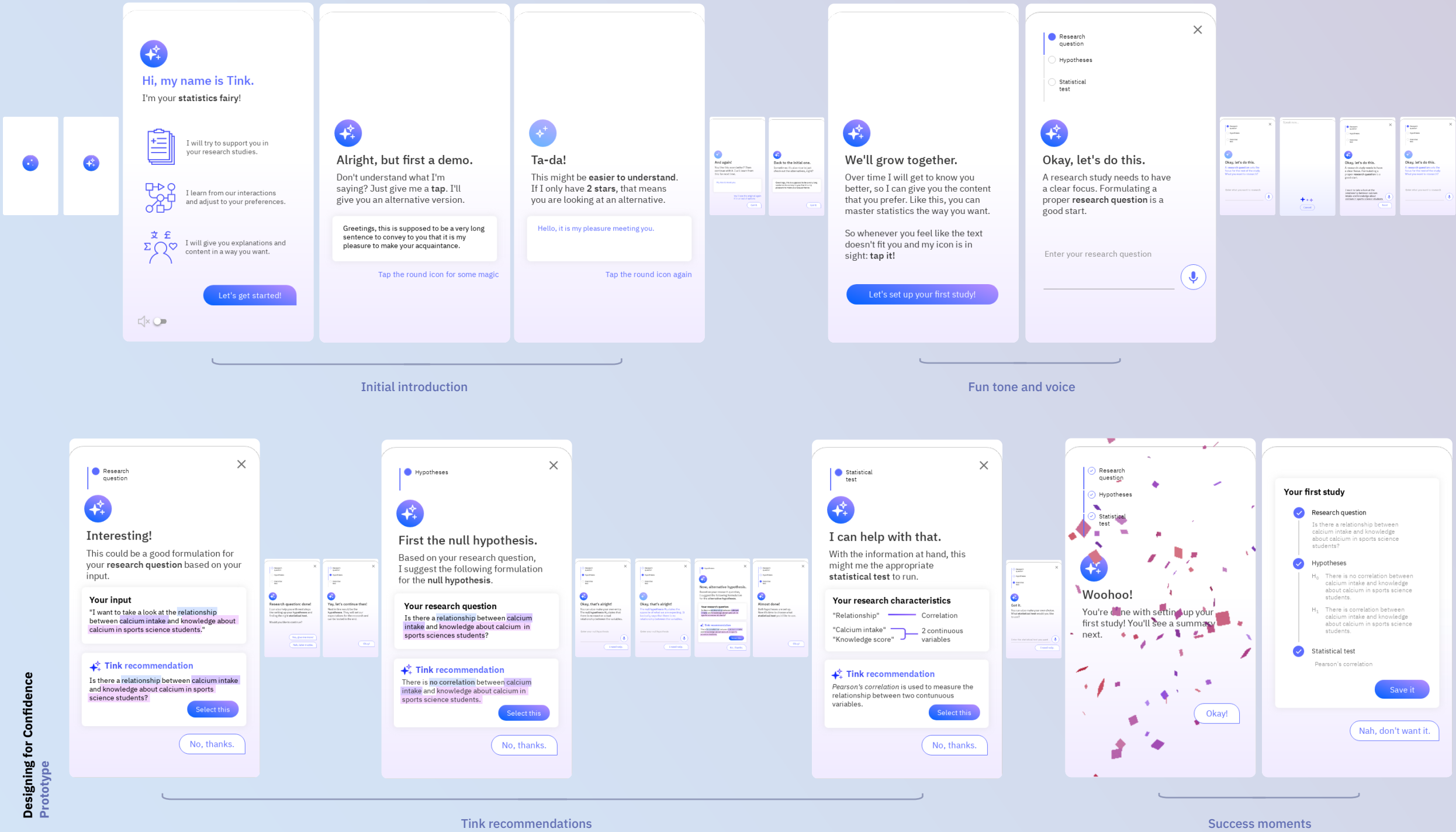
Tink, the statistics fairy is a concept for a statistics app that supports the user in the beginning phase of their statistical research study. The user can get help with tasks like formulating a proper research question, setting up useful hypotheses and selecting an appropriate statistical test. Tink understands the user's input and can give recommendations on how to proceed. One example is detecting the characteristics of the research data and suggesting a fitting statistical test. Beyond the basic research preparation, Tink could also help with data setup and assumption checks. The assistance occurs in the form of a conversation, utilizing natural language processing and voice recognition.
The main feature of Tink is adaptivity. Tink offers several formulations for a text. The users can switch to alternative versions by tapping on Tink's icon in case the content shown is not comprehensible to them. Tink analyses parameters of the preferred text form such as length, tone and voice and complexity, to try to calculate the optimal kind of content. This way the user should be able to grasp matters faster, making them feel more competent in their tasks. With regular use Tink can try to elevate the content level, taking the user's progress into consideration.

The prototype implemented the onboarding process in detail, introducing the adapting and recommendation features thoroughly. Besides those, other features like a statistics dictionary and study overview are also included on a horizontal scale.





[FIG48] - [FIG49] Tink's text adapting feature



[FIG50]-[FIG74] Screenshots from the Tink prototype

Design Critique

The high-fidelity prototype was presented to a few design experts from IBM Studios Böblingen before the evaluation. The aspects that were pointed out contributed to the preparation of the evaluation.

This design critique phase mostly detected usability problems.
The adaptive feature of the prototype still needed more clarity in terms of instructions and functionality. There were not enough visual cues to guide the user towards the feature. In general, the pattern for changing the text could be redesigned.
In addition to that, it was recommended to revise the wording in the conversations. Important words should be mentioned first and need more attention directed towards them to establish a semantic hierarchy. They should be more distinct throughout the application. For example, one designer pointed out that it was not immediately clear to them what Tink meant by „name“ when it was asked for. The person was confused about whether Tink wanted to name the research study file or the user.
The recommendation features were a good approach but needed to be structured better in terms of the visuals. The recommendation should the visual main component. In the prototype, the screen was cluttered with other content.

Not all the feedback gathered in the sessions could be incorporated into the final prototype. However, it was helpful to get another round of feedback to reflect on the iteration and future possibilities.

[FIG75] Outcome of the final iteration

5

Evaluation

The final prototype is tested on its user experience and fulfilment of chosen design principles and user needs. The question whether an AI-powered solution can elevate the users' confidence in the beginning phase of a statistical research can potentially be answered.

5.1

Strategy

This chapter describes what the objective of the evaluation is and in what manner the evaluation will be executed.

Outline

The final prototype is evaluated with real test users who match the personas defined earlier. To have a structured plan on how to proceed and what exactly to aim for, one needs to set certain aspects beforehand. In the following, the key data around the evaluation procedure are defined.

Why is there an evaluation?

The goal is to increase a beginner's confidence when entering a statistical research on his own. The design process created an AI-powered solution. A proof of concept will be executed through an evaluation.

Who will test the system?

The concept is targeted towards beginners in statistics, especially in academics. Students from different educational backgrounds with minimal knowlegde in statistics will test the prototype.

Which UX principle is explored?

In order to analyze a potential increasement of the users' confidence, the psychological need „competence" is focused on. (→ USER EXPERIENCE DESIGN) The fulfillment of that need will be aimed for. „Stimulation" and „security" will also be explored.

What is evaluated?

The outcome of the final iteration, a high fidelity prototype of the „Tink" concept, will be evaluated. The focus of this prototype is the onboarding phase so that will be the main target.

What type of data will be collected?

For the evaluation mostly qualitative data will be collected through interviews, usability tests and questionnaires.

What constraints are there?

The final prototype is a vertical one, so certain features like the statistics dictionary can only be tested on a superficial level. In addition to that, simulating an intelligent application is difficult. The prototype might not be conceived as smart in the tests.

Test Users

According to Jacob Nielsen, it is sufficient to test a prototype with five test users. He recommends conducting several small tests with this number of participants instead of investing resources and time into an elaborate test with more users. With an increasing number of test users, the number of new insights gained from the tests will start to stagnate. In general, testing with five users will uncover 85% of the usability problems. Another test round will mostly like pinpoint the remaining 15%. (NIELSEN, 2000, [W62])

For this evaluation session, eight people are invited to take part. Around half of them were already involved in the inital interviews and user tests from earlier iterations. Five of them match the Emma persona, the other three fall into Howard's spectrum. All of them are beginners in statistics majoring in different studies like economics or tourism. They were asked beforehand to rate their statistical research skills and knowledge on a scale from 1 to 5, with 1 being not very skilled and 5 very skiled. None of them chose anything higher than 3. Half ot them settled with 2. With the relatively low rating results, it is indicated that there is room for improvement in terms of their confidence.

Testing Methods

Remote testing

The majority of the test participants were not able to test the prototype in person, so a lot of the tests will be done remotely. The communication tool Skype and its screensharing feature are used for almost all the tests, except for one instance where it is not available for the test participant. The video conference platform Zoom will be used in this case which also offers screensharing. Both tools enable screen recording so the sessions can be reviewed afterwards.
The users either downloaded the Proto.io app beforehand or tested the prototype in a browser if it was not possible. Through the screensharing function, one could follow what the test participant was doing.

Thinking Aloud

The „Thinking Aloud“ method enables the test moderator to look into the test participant‘s soul. While testing the prototype, the test participants are asked to articulate their thoughts and emotions. The moderator should try to say as little as possible and let the users do as they want. Through this method, one can discover what the users really think about the design. Misconceptions and unfulfilled expectations on the user‘s side lead to actionable redesign suggestions. (NIELSEN, 2012, [W63])

Interviews

The test participants were interviewed before and after the prototype testing. Before the testing, the general topic of the thesis was introduced and the participants were briefed on the structure of the test. Preceding questions were about the person‘s opinion on statistical research studies and their expectations on the prototype. Following the test, positive and negative aspects were uncovered in another set of questions. The participants could make suggestions in the end.

Questionnaire

Following the test sessions, an online questionaire was sent to the participants where general demographic information, their educational background and statistical knowlegde level were inquired. The participants were asked to rate their confidence in statistical research before and during the test. After the general and open questions, the questions from the User Need Questionnaire (UNeeQ) are included.

The UNeeQ from the Fraunhofer Institute is used to measure the fulfillment of human needs in a product or prototype. (KOMPETENZZENTRUM USABILITY, N.D., [W64]). The questionnaire evaluates the five psychological needs defined by Hassenzahl and Diefenbach that were described in the beginning of this thesis: competence, relatedness, popularity, stimulation and security.
(→ USER EXPERIENCE DESIGN) The Fraunhofer Institute added three more to this list which are tested in the UNeeQ. Individual fulfillment, influence, and competition. (HOFFMANN, 2016, [L9], P. 19)

5.2

Results

The concrete results from the evaluation are described and visualized. Through the interview and questionnaire, feedback was collected from the test participants.

Observations

During the prototype test, the participants were asked to verbalize their thought processes. In addition to this, they shared their screen so the participants‘ actions could be followed in real time. Certain aspects could be observed through these methods.

Adaptive feature

- 5 of the test participants recognized that the text changed
- 4 of the test participants liked the text adapting feature
- 4 of the test participants took longer to figure out where to tap in order to see the alternative text
- 4 of the test participants tapped Tink‘s icon several times and toggled between the different versions of the text
- 3 of the test participants did not understand the text adapting feature
- 2 of the test participants temporarily forgot about the text adapting feature

Recommendations

- 6 of the test participants were pleasantly surprised by Tink‘s recommendations
- 2 of the test participants expected Tink to have multiple options for a recommendations

Comprehension

- 5 of the test participants did not understand individual terms during the statistical study setup
- 3 of the test participants preferred a German version
- 3 of the test participants liked the dictionary option
- 2 of the test participants skimmed over the text quickly

Other

- 5 of the test participants liked the visual design
- 5 of the test participants were confused about the e-mail question
- 4 of the test participants liked the tone and voice
- 3 of the test participants liked the confetti effect upon success

User Feedback

What did you like about Tink?
it‘s cute, the idea is nice
I like it it‘s modern and easy to understand
very useful tool when fully developed
Aesthetics, mostly easy to understand, that it helped me with my task, didn‘t have to think about my task at all, that it gave me the confidence to fulfill my task properly
I liked the design (simple and clear). In addition, there was not too much text or icons. Even if I forgot that Tink can simplify the texts, I think it‘s a good function. I was amazed how quickly and easily Tink came up with a test proposal.
I liked that Tink had different possible explanations you could click through, since I often have the problem of not understanding an explanation and then needing to search for other explanations. Also, I found Tink really smart because she ‚understood‘ which of the words were the variables and how many of them were there. Tink might be really helpful for students who have troubles seeing the important factors of a bigger problem
Intuitive, simple, beautiful, intelligent
Easy usage, convenience, really helpfu

[FIG76] Positive feedback

Where could Tink improve?
easier descriptions
language varieties (e.g. in german too ;))
more options in the offer
Possibility to pause mid-way without losing progress (and if that is possible then a notification that that is possible), possibility to return to the home menu, in the end when the whole menu is shown maybe a short optional introduction what there is, what you can do, where I can find the summary of my test
I should be reminded of the function of text simplification more often. I would also like a German version.
-
can‘t think of anything
privacy

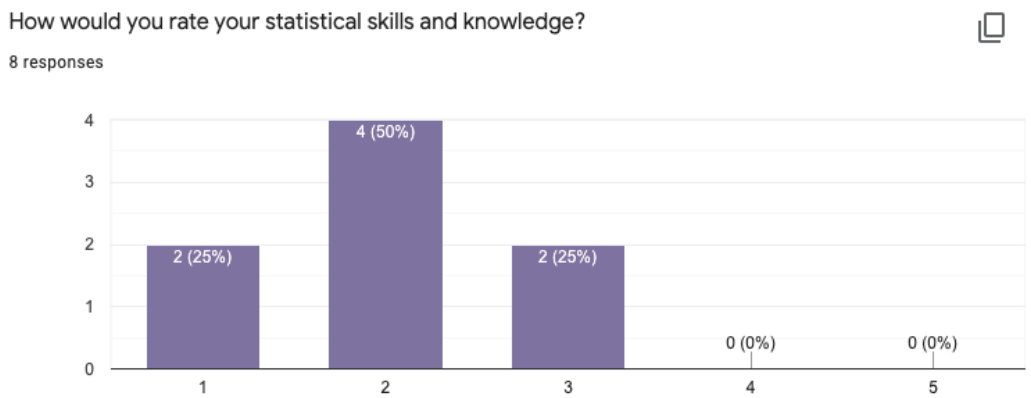
[FIG77] Negative feedback

Are you more confident in doing statistical research with Tink compared to using other sources of help?
seems easier than doing research on multiple sources
yes I think so.
no
Yes probably. It‘s easy and I don‘t have to actively search for tips on how to do this.
I would have to search for the right solution in books or scripts for a long time. Tink provides a solution very quickly - that gives me safety.
Yes, googling yourself through the process takes really long and is not always helpful
yes definitely. Very intelligent and smart assistant.
yes

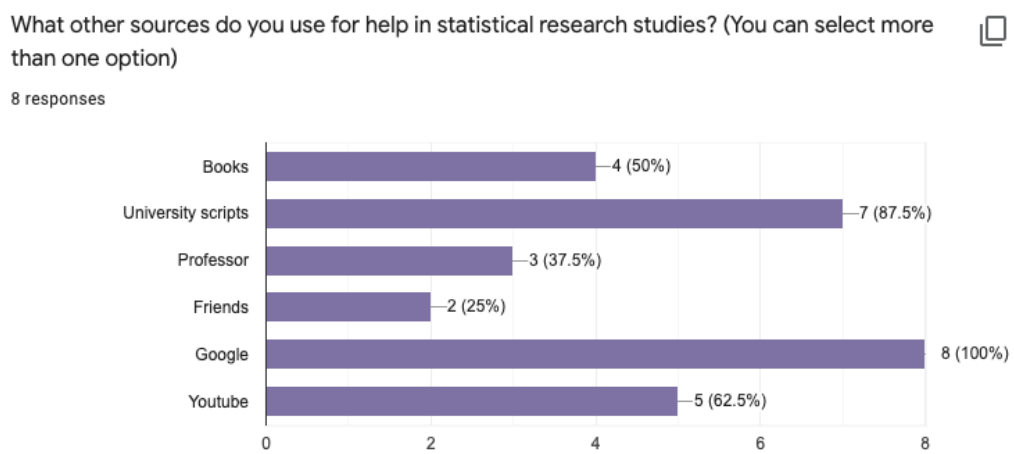
[FIG78] User feedback

Could you solve the tasks given to you? Please give explanations.
yes, the app was easy to handle
yess. there was a help assistant and easy using buttons
yes
yes, but didn‘t see that the app already suggested „statistical hypothesis“ probably because of the color grey which indicates to me that something is not that important
Setting up my research study and looking up for statistical hypothesis was easy! No in-depth knowledge of statistics was required.
Yes, with the help of tink I could start on evaluating the actual statistical numbers.
yes, very self explanatory and simple
I could because I am currently working on quantitative studies. However, a month ago I probably would only be able to set up the research question.

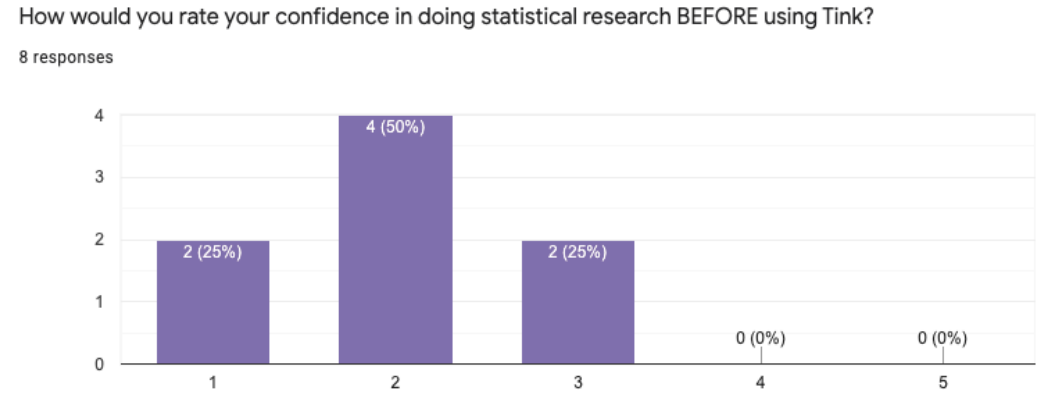
[FIG79] User feedback



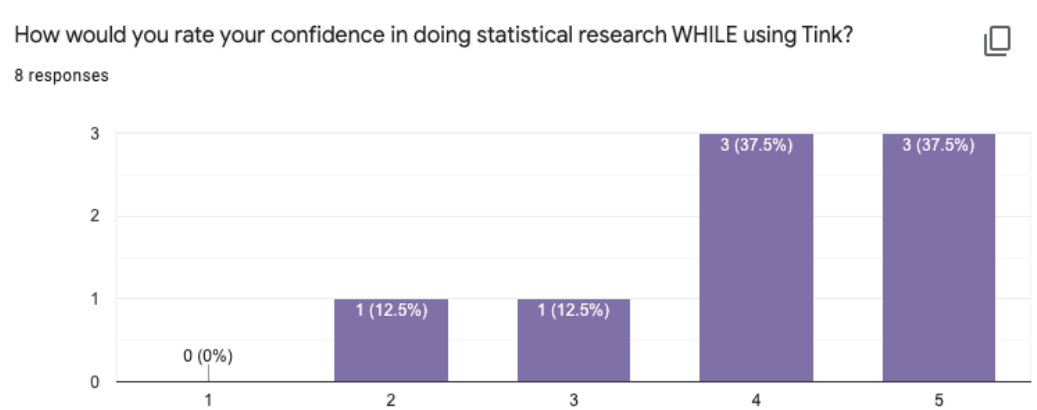
[FIG80] User feedback



[FIG81] User feedback



[FIG82] User feedback

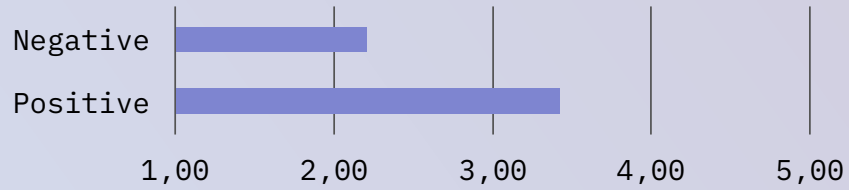


[FIG83] User feedback

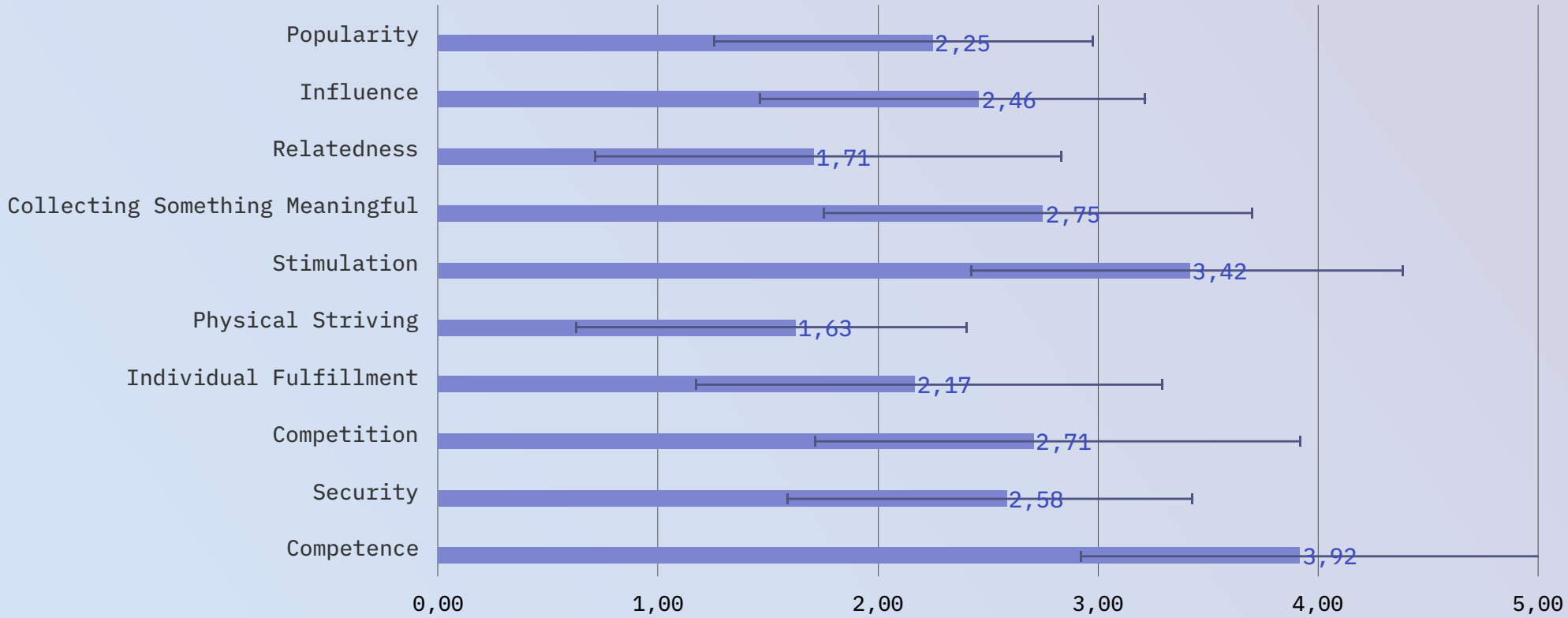
UNeeQ Results

The UNeeQ Questionnaire results were inserted in an Excel spreadsheet provided by another designer from IBM Studios Böblingen. The spreadsheet summarized and visualized the results.

The results of the UNeeQ questionnaire could be visualized in bar charts. They make it easier to gain insights on the spot before taking a closer look at the numbers. The overall user experience was received positively. (→ [FIG84]) On first sight, one can recognize certain tendencies in the fulfillment of psychological needs. The needs „competence“ and „stimulation“ reached the highest scores, with the former accounting to 3,92 out of 5. (→ [FIG85])



[FIG84] OVERALL USER EXPERIENCE



[FIG85] USER NEEDS FULFILLMENT

5.3

Discussion

The results that have been laid out in the previous chapter will be discussed and analyzed.

Design Problems

Usability

After taking a look at the evaluation results, several usability problems could be detected. These insights were mostly collected through the thinking aloud activity and through observing the participants‘ actions. Only a couple of participants elaborated on usability issues in the questionnaire.

Low learnability

Some of the participants struggled to understand how the text adapting feature works. They had to search the screen for a while until they realized they had to tap the Tink icon. This happened with 3 test users although there was a hint text on the screen, explaining the procedure to them. The icon and text did not have enough affordance. To make the feature more noticeable, stronger visual cues are needed. The hint text could be placed closer to the action item or have arrows pointing to the icon. One could even implement a pulsating or jumping animation of the icon to catch the user‘s attention. In this case, a hint text might not even be needed.

Lacking memorability

4 of the test users tapped the icon several times to compare the alternative versions of the text. They said they wanted to check what exactly changed after the tapping as they could not remember what the original version looked like. A way to improve this would be changing the overall mechanic. Instead of replacing the entire text with every tap, only certain parts could be highlighted and switched out upon the user‘s selection. One could also consider expanding the text parts instead of replacing them. The mid-fidelity prototype explored this before and it could be the better approach after all.
Towards the end of the flow, after the participants finished their task, they could explore the rest of the Tink app and try out the dictionary feature. Two of the test users forgot that the text adapting feature was available, despite having the Tink icon on the screen. Similarly to the case described before, the Tink icon should pull more attention to itself to invite the user to tap it. Clearer and more frequent explanations could also be useful for internalizing the feature.

Like the UNeeQ questionnaire outcome depicts, the user experience was perceived mostly in a positive way. However, a major part of the usability issues concerned the adaptive feature of Tink. As the main feature of the prototype, it should perform better. Several additional iterations are likely needed to bring forward a more usable and user-friendly state.

Interaction

Throughout the test, minor interaction problems could be observed. They did not really obstruct the test participants from continuing the flow but they expressed their wishes for improvement.

Difficult wording

More than half of the participants wanted to look up individual statistical terms during the process of setting up a study. This is possible in the dictionary but not in a setup process. Tink only offers the alternative texts in that case. A possible solution to this could be the approach described before that lets the user selectively switch out or expand parts of a text.
Even though the tone and voice were well-received, general improvements in the wording could help elevate the overall user experience.
In addition to this, 3 of the participants requested a German version of the prototype as they struggled to process Tink's instructions and comments.

Lack of visuals

Another aspect that is also related to the wording issue is the lack of visualizations. The app's visual design is very simple and minimalistic but more visuals could majorly improve the user's rate of understanding something. 2 of the participants skimmed over the texts very quickly and often did not properly read them. One of them explained, that images would help them get into the content easier and quicker.

Ethics

The concept of having artificial intelligence supporting oneself in a statistical research study was generally well-received. The capabilities introduced like adaptive content and smart recommendations have the potential to become inherent parts of a student's statistical work routine. The user tests revealed weaknesses in the concept which have to be improved in order to make that possible.

Insufficient explainability

Tink tries to give recommendations based on the user's input. It shows how a recommendation came to be, by displaying the factors used to come up with it. The current prototype includes this in an attempt to be explainable, however, the several users requested a more detailed explanation. Properly understanding where the recommendation came from would help internalize the process, contributing to a learning effect. One user suggested including a small „Learn why" button in the recommendations, making it an optional action.

User Data Rights

An important component of an AI-powered application is the transparent handling of user data. Tink asked for the user's e-mail address at some point to register the user and save the work done so far. Having an account will allow Tink to track the user's progress and preferences and potentially use the data to construct a database for collaborative filtering. Comparing data from many different users will enable more precise predictions and recommendations for the individual user. Tink attempted to simplify this part and stated the „best experience possible" as the reason for collecting the e-mail address. This piece of information was not sufficient, as the majority of the test participants were confused at this step. For next iterations, it will be inevitable to redesign and expand this part of the concept extensively considering the user's data.

UNeeQ Analysis

The UNeeQ questionnaire and its accompanying spreadsheet delivered measurements of the user's needs fulfillment. (→ [FIG]) As the concept tried to increase the user's confidence during the setup of statistical research, it was decided to especially focus on the psychological need for competence. The more competent users feel in their tasks, the higher their confidence level will be. In order to reach its goal, Tink tried to stimulate the competence need.

Competence

According to the UNeeQ spreadsheet's outcome, competence turned out to be the need with the highest fulfillment rate. As a reminder, the competence need is defined through the following aspects: managing to do something on their own, facing challenges and mastering them and lastly, feeling competent. One can deduce from the result that users did feel competent with the help of Tink. Even if Tink is theoretically an external source of help, using an app could give the users a stronger sense of independence than asking a professor for example.

Stimulation

The confidence could also be increased through the fulfillment of the stimulation need, defined by curiosity, joy and the senses. It was the second-highest need on the fulfillment scale. In order to make the process of a statistical research study more fun and welcoming, Tink aimed to be a friendly and relatable companion app. The light tone and voice were supposed to make it less intimidating for statistics beginners to dive into their research. This way, they could embrace the work more confidently compared to starting off with books and Google.

Medium fulfilled needs

The majority of the remaining needs were fulfilled to a satisfying extent. „Security" and „competition" could have been relevant for achieving Tink's goal. Not having to fear obstacles and thus feeling secure and protected also contributes to a user's confidence. The competition need received a lower score than expected. Solving a task successfully and competently could feed into a person's sense of competition. Users could perceive themselves as better compared to others. However, the results state otherwise.

Least fulfilled needs

„Relatedness" and „physical striving" were the least fulfilled needs. The test participants did not sense a connection to other people around them while using Tink. This need could be stimulated in future iterations through more community-based approaches. Tink's recommendations could stem from other users' interactions and choices. The physical striving need was not a target for the fulfillment, so it was expected to receive low scores.

While some needs were not as fulfilled as others, the most relevant ones got a satisfying result. If competence is considered a major source of confidence, then it can be established that Tink did indeed elevate the user's confidence in a statistical research study.

Goal Fulfillment

The goal was to find out whether a solution with AI could increase the user's confidence in setting up statistical research studies. A prototype aiming to have a pleasant UX was created to achieve that goal.

It was determined that the competence need had to be stimulated in order to have an effect on the confidence level, so that need was focused on. As the UNeeQ results indicated, that aspect was fulfilled in a satisfactory manner.
Tink's evaluation general results turned out to be positive. In the open questions that were posed in the questionnaire, the majority of the test participants stated that they were indeed more confident with Tink supporting them than without. (→ [FIG83]) It was convenient for them to have a smart assistant that fits in their pocket as an app and has the answers they need on demand. In comparison to their traditional way of finding answers which were books and Google, they preferred this by far. The intelligent recommendations removed their doubts about decisions and the fact that it could be adapting to their needs was a pleasantly received idea.
These outcomes answer the question whether an AI solution could elevate a person's confidence in statistical research. Tink, a solution of such nature, did manage to increase the test participants' confidence during usage.

Thus, the goal set in the beginning and defined further throughout this thesis was fulfilled.

Can I use this for my Master's thesis, please?

Test participant during the evaluation

6

Conclusion

The design process and outcome will be summarized and a future outlook is given.

Closing statement

This thesis explored a concept to improve the confidence of statistics beginners. As the SPSS design research provided and user research conducted in the course of the thesis have uncovered, students with little to no statistics experience feel overwhelmed during the preparation and execution of a statistical study. Insecurities prevail throughout the whole process and several sources of help are needed to reach their goal.

In order to explore solutions for the problem area in question, it was necessary to have an understanding of all the domains involved. The theory section laid out a foundation in statistics, AI, and design on which the final concept "Tink" could be built upon.

As the final outcome is designed for the end-user, test users who matched the target group were involved throughout a big portion of the process. Talking to and listening to the students who have struggled with statistical research studies before established empathy for the target group. The user needs were researched and defined. The psychological need "competence" was focused on to induce a feeling of confidence. Ideas and concepts could be generated based on the theoretical foundations and research findings.

Several ideas were gathered, refined into tangible concepts and turned into prototypes that could be tested. The initial concepts went through several iterations until they reached their final form for this thesis. AI approaches were explored and validated on their applicability. With AI's capabilities came the complexity to prototype them. It was difficult to simulate intelligent behavior. That task was only achieved to a certain extent.

The evaluation showed that the test participants' confidence was increased with the help of the presented prototype. The competence need was fulfilled the most and the prototype received an overall positive rating in terms of UX. However, the test sessions uncovered several usability and design issues, which would have to be improved in future iterations.

Nevertheless, the prototype fulfilled its purpose which was proving or disproving the effectivity of an AI solution for raising the user's confidence.

Future Outlook

The prototype proved to be effective for the goal that was set for this thesis. One could research in future iterations whether it was a combination of the intelligent features or individual ones that raised confidence in the students. As the user feedback shows, the recommendations were very well received although there were still a couple of usability issues. Tink's adaptive feature was the main one, however it did not receive significantly more attention or praise than the recommendation feature. This could pose the foundation for further research that investigates which aspect of the intelligent statistics assistant was the decisive one in elevating the confidence. Whether it was the adaptive feature or the recommendations, or a combination of both, is still unclear. Factors like the fun tone and voice and the conversational approach could also have played a major role.

However, that will extend the scope of this thesis. The work done was sufficient to show that confidence in statistical research could be increased with the proper use of design principles and AI.

List of References

[W1] **Elements of AI (n.d.)** *How should we define AI?* https://course.elementsofai.com/1/1 (Accessed 12 December 2019)

[W2] **Stanford University (2019)** *The AI Index 2019 Annual Report* https://hai.stanford.edu/sites/g/files/sbiybj10986/f/ai_index_2019_report.pdf. (Accessed 21 January 2020)

[W3] **Google Design (2019)** *AI and Design: Putting People First* https://design.google/library/ai-design-roundtable-discussion/ (Accessed 30 January 2020)

[W4] **Stauber, D. (2017)** *Announcing the Future of SPSS Statistics* https://developer.ibm.com/predictiveanalytics/2017/10/23/newuipreviewspss/ (Accessed 30 December 2019)

[W5] **CrashCourse (2018)** *What Is Statistics: Crash Course Statistics #1* https://www.youtube.com/watch?v=sxQaBpKfDRk (Accessed 29 December 2019)

[W6] **Taylor, C. (2018)** *Understanding Statistics* https://www.thoughtco.com/what-is-statistics-3126367 (Accessed 30 December 2019)

[W7] **Farnsworth, B. (2019)** *The Top 7 Statistical Tools You Need to Make Your Data Shine* https://imotions.com/blog/statistical-tools/ (Accessed 30 December 2019)

[W8] **Stauber, D. (2018)** *SPSS: 50 Years of Innovation* https://developer.ibm.com/predictiveanalytics/2018/04/05/spss-50-years-innovation/ (Accessed 30 December 2019)

[W9] **Van den Berg, R. B. (n.d.)** *SPSS – What Is It?* https://www.spss-tutorials.com/spss-what-is-it/ (Accessed 2 January 2020)

[W10] **Pezzullo, J. (n.d.)** *Levels of Measurement for Biostatistics Data* https://www.dummies.com/education/science/biology/levels-of-measurement-for-biostatistics-data/ (Accessed 2 January 2020)

[W11] **StatisticsSolutions (n.d.)** *Statistical Data Analysis* https://www.statisticssolutions.com/statistical-data-analysis/ (Accessed 29 January 2020)

[W12] **IBM (n.d.)** *Artificial intelligence* https://www.ibm.com/design/ai/basics/ai (Accessed 2 January 2019)

[W13] **Ghahramani, Z. (2019)** *Uber AI in 2019: Advancing Mobility with Artificial Intelligence* https://eng.uber.com/uber-ai-2019/ (Accessed 28 December 2019)

[W14] **Chandrashekar, A., Amat, F., Basilico, J., Jebara, T. (2017)** *Artwork Personalization at Netflix* https://netflixtechblog.com/artwork-personalization-c589f074ad76 (Accessed 18 December 2019)

[W15] **Elements of AI (n.d.)** *Related fields* https://course.elementsofai.com/1/2 (Accessed 12 December 2019)

[W16] **IBM (n.d.)** *Machine learning* https://www.ibm.com/design/ai/basics/ml (Accessed 2 January 2019)

[W17] **Elements of AI (n.d.)** *The types of machine learning* https://course.elementsofai.com/4/1 (Accessed 2 January 2020)

[W18] **Elements of AI (n.d.)** *The nearest neighbor classifier* https://course.elementsofai.com/4/2 (Accessed 4 January 2020)

[W19] **Elements of AI (n.d.)** *Regression* https://course.elementsofai.com/4/3 (Accessed 4 January 2020)

[W20] **IBM (n.d.)** *Data* https://www.ibm.com/design/ai/basics/data (Accessed 2 January 2020)

[W21] **Hayes, A. (2019)** *Simple Random Sample* https://www.investopedia.com/terms/s/simple-random-sample.asp (Accessed 30 December 2019)

[W22] **Glen, S. (2019)** *Difference Between Stratified Sampling, Cluster Sampling, and Quota Sampling* https://www.datasciencecentral.com/profiles/blogs/difference-between-stratified-sampling-cluster-sampling-and-quota (Accessed 30 December 2019)

[W23] **Aalto University (2019)** *Learning with Artificial Intelligence* https://www.youtube.com/watch?v=Dr499PgPB7A (Accessed 10 January 2020)

[W24] **Norman, D., & Nielsen, J. (n.d.)** *The Definition of User Experience (UX)* https://www.nngroup.com/articles/definition-user-experience/ (Accessed 23 January 2020)

[W25] **Interaction Design Foundation (n.d.)** *User Experience (UX) Design* https://www.interaction-design.org/literature/topics/ux-design (Accessed 23 January 2020)

[W26] **Lyonnais, S. (2017)** *Where Did the Term "User Experience" Come From?* https://theblog.adobe.com/where-did-the-term-user-experience-come-from/ (Accessed 23 January 2020)

[W27] **Komninos, A. (2020)** *An Introduction to Usability* https://www.interaction-design.org/literature/article/an-introduction-to-usability (Accessed 25 January 2020)

[W28] **Nielsen, J., (2012)**, *Usability 101: Introduction to Usability* https://www.nngroup.com/articles/usability-101-introduction-to-usability/ (Accessed 26 January 2020)

[W29] **Interaction Design Foundation (n.d.)** *User Interface (UI) Design* https://www.interaction-design.org/literature/topics/ui-design (Accessed 27 January 2020)

[W30] **Teo, Y. S. (2020)** *What is Interaction Design?* https://www.interaction-design.org/literature/article/what-is-interaction-design (Accessed 2 February 2020)

[W31] **Silver, K. (2007)** *What Puts the Design in Interaction Design*
https://www.uxmatters.com/mt/archives/2007/07/what-puts-the-design-in-interaction-design.php
(Accessed 3 February 2020)

[W32] **IBM (n.d.)** *Fundamentals*
https://www.ibm.com/design/ai/fundamentals (Accessed 28 January 2020)

[W33] **Lovejoy, J. (2018)** *The UX of AI*
https://design.google/library/ux-ai
(Accessed 30 January 2020)

[W34] **Communication Theory (n.d.)** *Knapp's Relationship Model*
https://www.communicationtheory.org/knapps-relationship-model/
(Accessed 11 February 2020)

[W35] **IBM (n.d.)** *Everyday Ethics for AI*
https://www.ibm.com/design/ai/ethics/everyday-ethics
(Accessed 29 January 2020)

[W36] **IBM (n.d.)** *Accountability*
https://www.ibm.com/design/ai/ethics/accountability
(Accessed 29 January 2020)

[W37] **IBM (n.d.)** *Value alignment*
https://www.ibm.com/design/ai/ethics/value-alignment
(Accessed 29 January 2020)

[W38] **IBM (n.d.)** *Explainability*
https://www.ibm.com/design/ai/ethics/explainability
(Accessed 29 January 2020)

[W39] **IBM (n.d.)** *Fairness*
https://www.ibm.com/design/ai/ethics/fairness
(Accessed 29 January 2020)

[W40] **IBM (n.d.)** *User data rights*
https://www.ibm.com/design/ai/ethics/user-data-rights
(Accessed 29 January 2020)

[W41] **Burmester, M. (2016)** *Design Thinking – die neue alte Kreativität*
https://www.uid.com/de/aktuelles/hcd-design-thinking
(Accessed 15 February 2020)

[W42] **IBM (n.d.)** *The Framework*
https://www.ibm.com/design/thinking/page/framework
(Accessed 16 February 2020)

[W43] **IBM (n.d.)** *Statistics Coach*
https://www.ibm.com/support/knowledgecenter/SSLVMB_subs/statistics_project_statcoach_ddita/spss/statcoach/statcoach_main.html
(Accessed 7 February 2020)

[W44] **de Vries, R. M. (n.d.)** *About statkat*
https://statkat.com/about.php
(Accessed 7 February 2020)

[W45] **de Vries, R. M. (2018)** *Statkat – find the right statistical technique for your data and research question*
https://blog.jamovi.org/2018/06/25/statkat.html
(Accessed 7 February 2020)

[W46] **Tableau (n.d.)** *What is Tableau*
https://www.tableau.com/products/what-is-tableau
(Accessed 8 February 2020)

[W47] **Markas, R. (2018)** *Ask Data: Simplifying analytics with natural language*
https://www.tableau.com/about/blog/2018/11/ask-data-simplifying-analytics-natural-language-98655#fDB1A22XW3kIrIhA.99
(Accessed 8 February 2020)

[W48] **JASP (n.d.)** *Our Goals*
https://jasp-stats.org/about/
(Accessed 8 February 2020)

[W49] **Goss-Sampson, M. A. (2019)** *Statistical Analysis in JASP: A Guide for Students.*
2nd Edition, JASP v0.10.2 J
https://static.jasp-stats.org/Statistical%20Analysis%20in%20JASP%20-%20A%20Students%20Guide%20v0.10.2.pdf
(Accessed 8 February 2020)

[W50] **Intellectus Statistics (n.d.)** *Statistics Platform for Non-Statisticians*
https://www.intellectusstatistics.com/
(Accessed 8 February 2020)

[W51] **Intellectus Statistics (2018)** *Intellectus Statistics Demo*
https://www.youtube.com/watch?v=8bMfRW_IrrM&feature=youtu.be&t=119
(Accessed 8 February 2020)

[W52] **Intellectus Statistics (2019)** *Decision Tree*
https://www.youtube.com/watch?v=QAJUAaQv0cc
(Accessed 8 February 2020)

[W53] **Statistics Solutions (n.d.)** *How decision trees can help you select the appropriate statistical analysis*
https://www.statisticssolutions.com/how-decision-trees-can-help-you-select-the-appropriate-statistical-analysis/ (Accessed 8 February 2020)

[W54] **Nikolova, S. (n.d.)** *Is AI already good enough to transform healthcare? – Interview with Daniel Nathrath, Founder and CEO at Ada Health*
https://research2guidance.com/is-ai-already-good-enough-to-transform-healthcare-interview-with-daniel-nathrath-founder-and-ceo-at-ada-health/ (Accessed 16 February 2020)

[W55] **Ada (n.d.)** *Manage your health with Ada*
https://ada.com/app/
(Accessed 16 February 2020)

[W56] **Wikipedia (2020)** *Google Flights*
https://en.wikipedia.org/wiki/Google_Flights
(Accessed 2 February 2020)

[W57] **Google Design (2019)** *Designing Human-Centered AI Products (Google I/O'19)*
https://www.youtube.com/watch?v=rf83vRxLWFQ
(Accessed 4 February 2020)

[W58] **Holden, R. (2019)** *Want the best prices for your trip? Google can help.*
https://blog.google/products/flights-hotels/best-prices-for-your-trips/
(Accessed 4 February 2020)

[W59] **IBM (n.d.)** *Big Idea Vignettes*
https://www.ibm.com/design/thinking/page/toolkit/activity/big-idea-vignettes/ (Accessed 16 February 2020)

[W60] **Bhowmick, Arin (2018)** *The design fidelity conundrum*
https://medium.com/design-ibm/the-design-fidelity-conundrum-63769bbc3e40
(Accessed 19 February 2020)

[W61] **Proto.io (n.d.)** *Prototypes that feel real.*
https://proto.io/
(Accessed 22 February 2020)

[W62] **Nielsen, J. (2000)** *Why You Only Need to Test with 5 Users*
https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ (Accessed 22 February 2020)

[W63] **Nielsen, J. (2012)** *Thinking Aloud: The #1 Usability Tool*
https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/
(Accessed 22 February 2020)

[W64] **Kompetenzzentrum Usability (n.d.)** *Werkzeugkasten*
https://www.kompetenzzentrum-usability.digital/angebote/materialien/werkzeugkasten?pagenr=42#compilationObject_1876
(Accessed 22 February 2020)

[W65] **Faulkner, R. T. (2018)** *Personas: They actually do work*
https://medium.com/design-ibm/personas-they-actually-do-work-6e86a58542b
(Accessed 24 February 2020)

[L1] **Minsky, M. (1988)** *The Society of Mind*
Pages Bent ed., New York: Simon & Schuster.

[L2] **Norman, D. (2013)** *The Design of Everyday Things*
Revised & expanded ed., New York: Basic Books

[L3] **Hassenzahl, M., & Diefenbach, S. (2012)** *Well-being, need fulfillment, and Experience Design*
https://www.academia.edu/2880303/Well-being_need_fulfillment_and_Experience_Design
(Accessed 23 January 2020)

[L4] **Anderson, S. P., (2011)** *Seductive Interaction Design: Creating Playful, Fun, and Effective User Experiences*
Berkeley: New Riders

[L5] **Shneiderman, B., & Plaisant, C. (2005)** *Designing the User Interface*
Fourth ed., Addison Wesley

[L6] **Moggridge, B. (2007)** *Designing Interactions*
First ed., [s.l.]: The MIT Press

[L7] **Nielsen, J. (1993)** *Usability Engineering*
Academic Press.

[L8] **Hoffmann, D. (2016)** *Using Interactive Data Visualization on Tablets*

List of Figures

Thank you...

Dimitri Hoffmann, for being my fairy godmother in times of self-doubt and loss of focus.

Tuyen Cao, for being my partner in crime and achieving this with me together.

Büsra Akay, for proofreading and always cheering me up.

Oliver Nguyen, for proofreading and giving me advice when I needed it most.

Lina Urbat, Minh Nghi Bui Le, Chloe Lee, Katharina Schippert, Jan Hassel, Robin Auer and Tim Reiser for your time and valuable insights.

... to all the people who have helped and supported me throughout this thesis.