Artificial Intelligence in Design: User Experience optimisation using a Large Language Model

University of Turku
Department of Computing
Master of Science Thesis
Computer Science
May 2025
Rasmus Vainio

Supervisors:
Jaakko Järvi

UNIVERSITY OF TURKU
Department of Computing

RASMUS VAINIO: Artificial Intelligence in Design: User Experience optimisation using a Large Language Model

Master of Science Thesis, 67 p.
Computer Science
May 2025

Artificial intelligence (AI) is already used within numerous applications for creative work. Subsequently, interest in utilising AI in supporting user experience (UX) workflows has increased. Research into how this can be achieved has focused on bespoke AI tools and on AI models trained for a specific task. Research has not focused on evaluating the viability of off-the-shelf large language models (LLM) for such tasks. UX serves a fundamental role in ensuring that a product is efficient and easy to use. Therefore, the efficacy of more accessible AI tools for UX work should be investigated more. To this objective, two website designs for an e-commerce store were created. One made entirely by a human, and another one that is an improved version of it using an LLM. These two designs were then tested against each other by having test subjects complete tasks on them. The results show significant improvement in user error rate and perceived aesthetic quality in the LLM-improved version.

Keywords: User experience, UX, UX design, artificial intelligence, AI, generative AI, large language model, LLM

TURUN YLIOPISTO
Tietotekniikan laitos

Rasmus Vainio: Artificial Intelligence in Design: User Experience optimisation using a Large Language Model

Master of Science Thesis, 67 s.
Tietojenkäsittelytiede
Toukokuu 2025

Tekoälyä käytetään jo lukuisissa luovan työn sovelluksissa. Tästä johtuen on myös syntynyt kiinnostusta tekoälyn soveltamiseen käyttäjäkokemuksen työnkulkuihin. Tutkimus tämän saavuttamiseksi on ollut rajoittunut räätälöityihin tekoälytyökaluihin, sekä tiettyihin tehtäviin koulutettuihin tekoälymalleihin. Tutkimus ei ole keskittynyt sen arvioimiseen, miten etukäteen koulutetut suuret kielimallit soveltuvat tällaisiin tehtäviin. Käyttäjäkokemuksella on keskeinen rooli tuotteen tehokkuuden ja helppokäyttöisyyden takaamisessa. Tämän vuoksi on oleellista tutkia enemmän tekoälytyökalujen, joihin useammilla on pääsy, tehokkuutta käyttäjäkokemuksen kehittämisen tehtävissä. Tämän vuoksi kehitettiin kaksi käyttöliittymää verkkokaupalle. Yksi käyttöliittymä oli kokonaan ihmisen suunnittelema, ja toinen käyttöliittymä oli ensimmäisestä käyttöliittymästä suurta kielimallia käyttäen paranneltu versio. Näitä kahta käyttöliittymää verrattiin pyytämällä testihenkilöitä suorittamaan niitä käyttäen tehtäviä. Näiden testien tulokset osoittavat, että suurella kielimallilla paranneltu käyttöliittymä oli vähemmän altis käyttäjävirheille, sekä esteettisesti miellyttävämpi.

Asiasanat: Käyttäjäkokemus, käyttäjäkokemussuunnittelu, tekoäly, generatiivinen tekoäly, suuri kielimalli

Contents

List of acronyms

AI Artificial Intelligence

CNN Convolutional Neural Network

GAN Generative Adversarial Network

HCI Human-Computer Interaction

LLM Large Language Model

ML Machine Learning

NLP Natural Language Processing

UI User Interface

UX User Experience

1 Introduction

Artificial intelligence is rapidly evolving with the introduction of generative AI [1], [2]. Generative AI is a type of AI that is capable of generating multimodal output, such as text, images, and audio. Large language models are one example of generative AI. The recent increase in applications of generative AI has been fueled in large part by the adoption of foundation models [1]. These are AI models, that are trained on large and diverse datasets, and they can be specialised and applied across many different use cases.

We are experiencing an era where AI is applied to many new areas. Indeed, AI has potential for aiding various tasks, such as those within UX design [2], [3]. The advancement of AI has initiated discussions regarding how it could be harnessed to enable designers to design better products, faster, and with lower costs [4]. Design tools that integrate AI into them have displayed capability for creative endeavours, such as ideation and sketching [5]. It's becoming increasingly common to augment the design process using such tools. However, the incorporation of AI into the tools themselves is still rudimentary [6].

It seems inevitable that AI's impact on design tools continues to increase, though attitudes towards AI are still cautious. Designers have had generally positive reactions to including AI tools into parts of their workflows, with some caveats [1], [6]. AI's limitations in understanding user-centered design and potential for skill degradation, especially for less experienced designers, are concerns [1]. Designers are also apprehensive about sharing certain types of data with AI tools, which could affect these tools' performance [6]. Designers would wish for the development of AI tools to continue in a collaborative and intuitive direction, where such tools would take care of repetitive tasks and take on an assistive role [1].

What may hinder a large-scale adoption of AI in design is that designers often have trouble working with AI effectively, as a typical UX designer's education does not give enough technical knowledge regarding machine learning (ML) [3], [7]. As a result, UX designers may not have a solid understanding of what ML models can learn and how they function. Prototyping designs using AI may also be a challenge for designers, in particular applying AI for the rapid iteration that UX design often employs [3].

The goal of this thesis is to evaluate the viability of off-the-shelf LLMs for optimising UIs. The central question is whether AI models can improve designs, possibly sketches or otherwise unpolished, that UX designers have created. We hypothesize that large AI models do have these kinds of capabilites, and set out to test this hypothesis. To this end, we designed and developed an e-commerce website. The design was given to an LLM that was instructed to improve it. The improved design was then compared against the original design. Statistically significant improvements were found in users' error rate and perceived aesthetic quality.

2 Research questions and methodology

Many experiments and inquiries have been made into the potential of AI technologies within different design domains [2], [8]–[11]. However, we are not aware of research into the viability of an off-the-shelf LLM in UI design tasks. The AI tools that research has shown as having potential for improving UX design processes are ones that researchers have created themselves, added additional features to, or otherwise modified to fit the task at hand. As such, they are tools that the average designer would not have access to. An off-the-shelf LLM is likely the only kind of AI that most designers and developers would have ready access to, which merits research into its capabilities. This serves as the main motivation for this thesis.

To evaluate the performance of off-the-shelf LLMs in the task of UI optimisation, the following research questions (RQ) were devised:

- RQ1: Is a user interface that an LLM has been asked to improve more efficient to use?
- RQ2: Is a user interface that an LLM has been asked to improve less prone to user errors?
- RQ3: Is a user interface that an LLM has been asked to improve more aesthetically pleasing?

In all these research questions, the comparison is to a user interface that has not been subjected to any modifications by an LLM.

To explore and answer these research questions we performed an experiment and conducted a survey using a questionnaire. For the experiment a simple website, effectively a high-fidelity prototype, for an e-commerce clothing shop was developed. Enough functionality was implemented for this mockup website so that it was possible to navigate around it and interact with simple elements. The HTML and CSS code for the website was then given as input to an LLM (GPT-4-0125-preview) with the instruction of improving certain aspects of the website. The HTML and CSS outputted by the LLM were then used to create another version of the same website. These two versions of the website were compared by having people complete tasks on both versions. The results of this experiment were used to answer research questions 1 and 2.

3 Current status of Artificial Intelligence

Artificial intelligence has already had a profound impact on many fields and industries and it has the potential to affect numerous others [1], [12]–[16], such as medicine, finance, education, and various creative professions. This list is far from exhaustive but it serves to highlight how far-reaching AI technologies are. The adoption is happening fast, but it is not entirely surprising: already back in 2011 *McKinsey Global Institute* made the prediction that the next wave of innovation would be driven by machine learning technologies [12]. By now they're present, for example in many modern facets of internet usage, like web searching and content filtering on social media sites [16]. AI has become a part of both consumer and enterprise applications [17]. Businesses have largely recognised that AI is not simply an advantage, but a necessity for competing in the modern digital era [18].

Not all of AI's impacts have been positive, naturally. A notable recent example is the Hollywood screenwriters' strike [2], where among the points of contention was the regulation of AI within creative projects. The strike stemmed from a fear of job losses due to the potential of workforce being replaced in part by AI. Concern over potential harms that AI may have on various industries are also echoed by researchers [1], [4]. Among these harms are copyright infringement, plagiarism, job losses, increased stress, creativity exhaustion, and possibly decreased quality of end products due to abundant automation within the design process.

Pertinent to this thesis, artificial intelligence has already seen some adoption within the realm of user interface (UI) design [13], [14]. Researchers believe AI has potential for improving user experiences [3], and that even the implementation of *weak AI*, one that handles simple tasks that humans traditionally would, is beneficial for designers [4]. As a result of AI acquiring a more prevalent role within design, new concepts for describing its function within design have emerged. Examples include *artificial design intelligence*, which refers to AI being trained to gain knowledge of design so that it can generate designs and predict trends, and *computational creativity*, which refers to AI that performs actions thought of as creative if performed by humans (e.g. art, writing, music) [4]. Designers are also aware of the advantages that AI can bring and believe that the intermingling of UI design and AI will only increase in the future [1], [7]. They recognise AI's strength as a tool to assist in some aspects of their work, and are somewhat optimistic about AI's future potential, though the general attitude towards adopting it in the workflow remains cautionary [1].

Despite AI having found its way into UI design, most designers report having difficulties effectively making use of it [3], [15]. Research groups from Google and Microsoft agree, and consider there to be a lack of guidance for designers on how to design applications with AI [17]. To remedy this dearth, both Google and Microsoft have published guidelines for designing with AI [19], [20]. Many designers have only a surface-level understanding of various AI technologies and as such they are not aware of the capabilities and limitations of these technologies. Designers freely admit to knowing very little about machine learning, but don't consider it an obstacle for collaborating with data scientists or for designing UIs with ML components. They treat AI as one of the tools they use and do not necessarily need to understand the intricacies of. Yang et al. [3] likened the designers' relation to ML as if it were a black box to them. They, in contrast to the designers' own feelings, find that some technical literacy is required to work with AI efficiently. Stige et al. [4] predict that such a lack of understanding may lead to designers not comprehending the effects that AI can have on the design process and its resulting solution. They don't consider this lack of understanding a failing on the designers' part, but rather a consequence of AI in parts being a black box technology that gives no insight as to how it produces its outputs.

Song et al. [14] describe AI's role within UI design as being that of automation: performing rudimentary tasks and lowering error-proneness while letting the human designers focus on innovation and creativity. Stige et al. [4] mostly agree with this assessment and predict that AI will be used to automate the repetitive tasks of design, for example within low-fidelity prototyping, where the prototypes themselves can be used as input for an AI to generate a mostly functional version of the prototype. They also believe that collaboration between UX designers and AI is the likely scenario in how AI will be utilised in the future. Zhang et al. [13] also see the current status of AI as a synergistic collaborator rather than a tool that is used every now and then and then put away until it's needed again.

User experience designers interviewed by Li et al. [1] mostly agree with these assessments, but disagree in that they believe the usefulness of AI in UX design manifests only if the designers themselves already possess sufficient skill in their craft. Another survey of UX professionals was conducted by Knearem et al. [6], consisting of employees of a "large multinational technology company" (the researchers themselves are all employees of Google), with the goal of gaining an understanding of their general views toward AI in design tools. From the respondents nearly all (97%) reported having familiarity with AI concepts and 73% felt that AI has a positive contribution to society. However they were apprehensive about the prospect of having AI assist them in design work. In another group of UX professionals, 73%

responded positively to having an AI help them in finding usability problems [21]. While they agree on AI having strong potential for automating repetitive and formulaic tasks they believe there are some tasks that are not possible to do without a human, such as work that requires collaboration between humans, designing user-centered functionalities, and work requiring user inputs.

Stige et al. [4] believe similarly and emphasise the importance of identifying what are the best possible solutions for a given problem, instead of attempting to solve them all through using AI. Particularly with creative design work the focus should be on identifying what are the tasks that designers find rudimentary and repetitive and if AI could be used to automate them. They further note that not every problem can be automated and not every problem even should be automated, since some tasks are such that humans may be more capable in them. Such tasks are for example ones involving extrapolation; humans excel at making predictions using human attributes like creativity, intuition, and domain knowledge [1]. AI on the other hand may then have a "natural" aptitude towards tasks that are interpolative in nature; AI models are trained on a finite amount of data that they then use to produce their output, and as such AI is more suited for tasks involving known patterns of data.

It was also important to several designers that verification by humans be applied to every stage of the design process, so that if and when AI is used it's done in a controlled manner [1]. Some interviewees mentioned creativity and originality as being very human qualities that would be difficult, though not impossible, to be replaced by AI. The interviewees also uniformly believe humans to be unique in exhibiting empathy towards users and deriving joy from designing, which they find to be abilities outside the realm of AI competencies.

Abbas et al. [7], however, note that among designers artificial intelligence has been underutilised and has not yet seen full integration into the design process.

They find that this is largely due to the various tools in development still being in prototyping stages and due to the common machine learning problem of not having enough training data for producing a sufficiently competent model. In opposition to these findings, many more studies showcase various ways in which AI has already been included in the design process. Zhang et al. [13] describe *FashionQ*, which is a model that analyses pieces of fashion imagery to be used during the ideation phase, and *Rico*, which can support UI layout design by analysing design and interaction data from popular Android applications. This discrepancy between these studies' findings hints at there being differences in how succesfully various AI technologies have been adopted across the industry. This thought is further reinforced by the findings of Yang et al. [15], who found there to be a significant difference in how successfully designers work with machine learning based on their organisation's structure. Large, "AI-focused" organisations could provide their designers with the services of data scientists who did demonstrations with AI and who the designers could consult and collaborate with. In comparison, designers in smaller organisations with less access to such services fared noticeably worse when it came to working with AI — they considered fewer alternatives to their designs and AI innovation was limited to the larger AI-focused organisations. Further evidence is that when businesses consider integrating AI into their workflows, ease of use, access, and transparency are all highly desirable features in a prospective AI, since all these result in interaction with the AI being more fluid [18]. Alternatively, considering the speed at which artifical intelligence is advancing it is not improbable for the findings of Abbas et al. [7] to have been true back in 2022 when their paper was published, and to have since become outdated through industry practitioners becoming more familiar with using AI in their work and learning to integrate it within their workflows more effectively.

The integration of artificial intelligence within businesses is still a new enough phenomenon that there is often uncertainty in how exactly it could benefit them [18]. Subsequently, AI that is used to augment decision-making faces resistance from stakeholders regardless of whether its output decisions match those of the humans using it. If aligned with the humans the AI may be seen as ineffectual since its decision was one that was already thought of by humans and is as such unneeded. When aligned against humans it is especially scrutinised, due to going against the already established beliefs of people who already have a wealth of experience and knowledge in their respective fields. It is therefore imperative to be able to fluidly query the AI, through it being sufficiently transparent and explainable, on how it arrived at a particular conclusion. Focusing more on these aspects is also beneficial for designers, for it makes designers less likely to implement questionable solutions suggested by AI; designers can review what sort of logic these solutions are based on [4]. Today the lack of AI's transparency is clearly an issue among designers and an obstacle when it comes to working with AI technologies. The process of querying an AI should be akin in easiness to being able to ask a human consultant the same questions one would from the AI.

The explainability of an AI system has already had its usefulness proven in healthcare and recommender systems [18]. It stands to reason that a similar effect could be seen in general when experts are working with AI in their workflows, such as within UI design; when the AI recommends a change that an expert disagrees with, it's likely to face greater resistance than if a novice were to be given the same recommendation by the AI. Stige et al. [4] make note of this possible regression in the role of designers: without a deeper understanding of design principles they may concede to the AI in conflicts rather than trust their own judgment.

Outside of user interface design and decision-making, AI has been applied in e-commerce to analyse user behaviour and various other data of the user in order to recommend products to the user [14], [16]. For instance Amazon is a well-known e-commerce platform that uses such a system [22]. These types of applications of AI are commonly known as *recommendation systems* or *recommender systems*, and are a common and important application of AI within businesses [18]. They are commonly seen within online streaming platforms, such as Netflix, YouTube, and TikTok where they recommend videos and TV series. Social media is also no exception: Facebook serves its users curated content through recommender systems [14], [22], and X (formerly Twitter) utilises machine learning within its recommender system that provides the 'For You' -timeline, which serves the user posts recommended by the system in addition to content the user has themselves followed [23].

The adoption of AI within consumer electronics is aided by companies like NVIDIA and Intel, who design AI chips that enable less powerful devices like smartphones and cameras to utilise AI [16]. Within video games AI has been used to dynamically alter aspects of games such as the difficulty and story to better accommodate different types of players [14].

Considering all the advantages that AI can bring it's no surprise that researchers are almost in unison in believing that AI will bring great change to how users interact with online platforms in the future and to how designers build them [1], [4], [7], [14], [15], and even enable entirely new modes of interaction [3]. Generative AI specifically has already been shown to produce outputs in text, code, and forms of media that rival in quality that of skilled professionals [4]. Whether or not a similar shift in user-platform interaction has already happened or not is not as agreed upon. Some researchers have a moderate view of the subject, placing emphasis on understanding the effects that increased usage of AI will have, both positive and negative [17]. According to the moderate view, AI will not completely substitute the design process of UX [4]. Instead, the belief is that AI will result in the design process being more approachable. Researchers have also called for generative AI development to be handled responsibly: with the aim of human well-being and with appropriate oversight to ensure that creative professionals are safeguarded [1]. Another group of researchers calls for further inquiries into the real-world effects of AI and regulations to ensure creators are treated fairly [2]. Technology giants appear to agree with the merits of responsible use of AI. As early as 2015 companies like Google, Microsoft, and IBM were conducting research into deep learning, and have since published guidelines for how AI design should be conducted [17].

3.1 Overview of current language models

Language models are a specific type of machine learning model, and machine learning is itself a subset of artificial intelligence. The defining feature of language models is that they can be communicated with using natural language (i.e., normal human language), and that they can output natural language.

Large language models are a specific type of language model that are trained on large amounts of data and typically require tremendous amounts of computational power for their training [24]. The data that these models are trained on are massive datasets consisting of natural language texts. Because of this online forums, news websites' comment sections and articles, and other discussion-focused websites are of special interest. For example, sites like *Reddit* and *StackExchange*, neatly categorise their discussion topics. This is advantageous for the fine-tuning of models, so that they can be endowed with domain knowledge.

"Garbage in, garbage out" is a common saying regarding AI, and it refers to the quality of the data that the model is trained on affecting the quality of the model's output. Hoffman et al. [24] find that the quality and size of datasets are a limiting factor of scaling language models further. Attempts have been made to use LLMs to generate training data to overcome this lack of data of a sufficient quality, due to LLMs' success in other areas [25]. This does, of course, have the danger of propagating any biases and errors present in the model used to generate training data onto the models trained using said data.

LLMs are the premier AI technology for natural language processing (NLP), and they have performed exceptionally well in numerous NLP tasks [25]. LLMs also serve as good starting points from which to start fine-tuning specialised models. This is of interest due to the substantial computation and energy costs of training large language models from scratch [24].

Multimodal AI models are ones that are capable of taking as input or outputting multiple different modalities (e.g. text, video, speech, images). The training data required for training such models is naturally more varied than for a model of a single modality, and the amount of data can be expected to grow with the number of modalities. The specific datasets that commercial AI models are trained on are typically not public information; they tend to be trade secrets. The type of data used in training, however, can be ascertained based on the modalities of the model; text models require text as training data, image models require images as training data. Multimodal models aren't necessarily required to be trained from scratch, and can instead be fine-tuned from models trained for only a single modality. Lu et al. [26] find that models trained for natural language tasks can be generalised for other modalities with fine-tuning. Their fine-tuned models achieved similar performance to models trained specifically for the modalities being tested on.

While combining any two modalities is possible, text is of particular interest since a model that understands text is one that can communicate with humans using natural language and vice versa. For this reason multimodal models often combine text with another modality. Examples include text combined with images for a text-to-image model where the user describes to the model the image they want it to generate, or sound-to-text for a speech recognition model that takes sound as input and outputs the captioning for it.

The GPT series of large language models by *OpenAI* are likely the most well-known language models currently. Initially these models only covered text, but have later also been trained for other modalities. For instance, the GPT-4o model can process audio and video in addition to text. OpenAI's chatbot, *ChatGPT*, uses LLMs from this series.

The data used to train OpenAI's models has varied with each individual model. OpenAI published their first GPT model, GPT-1, in a paper where they also detail the kind of data that they use for the training of the model [27]. They use the *BooksCorpus* dataset, which is a collection of unpublished books, from many different genres. They point out that the dataset containing long, uninterrupted segments of text is crucial for enabling the model to handle context better. OpenAI has not made the training data of the later models public. GPT-3 is the last model for which the training data is officially made public by OpenAI. For training GPT-3 they use a mixture of different datasets [28]: *Common Crawl* makes up 60% of the training mix, and *WebText2*, which is a dataset of web-crawled data, 22% of the training mix. The rest of the training data is composed of book datasets and English Wikipedia articles.

At the time of writing this thesis, the latest model published by OpenAI is GPT-4.5 [29]. They claim the model feels more natural, has greater ability to follow user intent, and has fewer hallucinations. However, the model's performance in various benchmarks is mixed — it performs better than their other models in multilingual and multimodal benchmarks, but worse in science and math benchmarks [29].

Google has their own series of LLMs called *Gemini*, which are also used on Google's chatbot of the same name. The training data used for the training of this series' models is not public, but the models are multimodal, so the training data would be accordingly varied as well.

However, Google has also published another series of LLMs called *Gemma*, which are lightweight versions of Gemini that are free and open-source. Unlike the Gemini series of models, these lightweight models are not multimodal and are not specifically trained to be multilingual [30]. *Gemma 2 27B* is a model of this series and was trained on a primarily English mixture of web documents, code, and science articles. Versions with fewer parameters (2 billion and 9 billion) than the original 27 billion were also trained. All three models were tested against other models in a blind side-by-side evaluation, where humans pick the better answer after asking a question that two different models answer. The Gemma 2 models outperformed all other open models in the same parameter range [30].

Llama is the name of a series of LLMs by *Meta AI*. The initial model was published in a paper in February of 2023, that also detailed the training data used [31]. Similarly to OpenAI's GPT-3, the majority of the training data (67%) consists of Common Crawl, though in this case specifically English Common Crawl. In addition, for 15% of the training data they use a pre-processed Common Crawl dataset called *C4*. The rest of the training data was composed of public GitHub projects, multilingual Wikipedia pages, book corpora, cleaned LaTeX files from arXiv, and Stack Exchange dumps [31]. Subsequent Llama models have not had the specifics of their training data made public.

3.2 Ethics of Artificial Intelligence

As mentioned previously, ethical concerns of AI are very much a point of discussion among AI researchers and industry practicioners. Considering the tremendous potential that artificial intelligence has, it's unsurprising that this is so. Many ethical concerns to consider are concrete issues that might affect peoples' daily lives. Broad topics include unemployment, amplifying data biases, equal access to AI, privacy, transparency, and copyright infringement [1], [6], [32], [33]. A particular example is concern among UX design professionals that the value of human UX design work might drop through certain parts of the design process being replaced by AI [1].

This could potentially then lead to skill degradation, unemployment, and limiting the designers' creativity. AI is a difficult thing to compete with as a creative professional, since expecting a human to keep pace with AI in producing outputs is unrealistic. Designers also feel that biased data and a lack of equal access to AI could further inequality due to questionable outputs and through limiting who can and who can not take advantage of AI [1]. Some designers have expressed concerns over data biases carrying over to their designs, and over potential privacy issues stemming from using AI outputs in design work [6]. Another concrete concern is AI not serving its user's best interests. Consider, for instance, an assistive system for the elderly or other vulnerable groups that makes product purchasing recommendations. It would be quite simple for the system's owner to give it a bias for recommending the products of a specific manufacturer more often, instead of always recommending the products that would be most useful to the user [17].

An example outside the domain of design makes it clear that AI's ethical concerns are real and present. In 2023 a company called *Beamery* launched the world's first "human resources generative AI" called *TalentGPT* [22]. It's an AI system that can be used for talent searching, recruitment automation, and automatic candidate screening. The claimed benefits are time and labour cost reductions for the screening process along with more informed decision-making through better analytics. TalentGPT is given a list of requirements for a potential new hire that it then tries to find matching candidates for [22]. Considering the lack of transparency in how AI makes decisions, lack of legislation in their application to processes where fairness is supposed to be ensured, and undeniable existence of biases in the data that is used to train AI models and in the final models themselves, this has the potential to be quite problematic.

Aside from the ethical dilemmas stemming from the application of AI there is no small amount of discourse regarding the training of machine learning models. A noteworthy example of a legal dispute relating to this is *Getty Images* suing *Stability AI*, where Getty Images accused Stability AI of training its models using more than 12 million Getty photos without compensation [34]. Another example is a class action lawsuit on behalf of a group of artists against Stability AI, *Midjourney*, and *DeviantArt*. This lawsuit, too, is about photos and other images being used for the training of AI models without permission [35]. AI has faced pushback from creative communities such as that of *ArtStation*, where artists uploaded images protesting AI-generated art appearing on the website [36]. ArtStation promptly removed such protest posts, citing them breaking the website's Terms of Service. Many creators argue that it is unethical to train AI models using original works; regardless of whether permission is given, the AI model will still be a direct competitor to the industry that its training data is sourced from [2].

Indeed, most organisations behind prominent AI models are reluctant to publish specifics of the data used in the training of their models, which is unsurprising considering the extreme values of these acquisitions. An example of this is the online forum website Reddit's deal with Google in early 2024, where Reddit gave Google's AI models access to its vast amounts of natural language data in exchange for $60 million USD per year [37]. A notable exception to this trend is *Adobe Firefly*, which is Adobe's own collection of generative AI models that have been trained with Adobe's own licensed content and public domain content with expired copyright [38]. This is in stark contrast to the likes of OpenAI that have been repeatedly targeted by lawsuits on the basis of using data for training their models without permission. Among these lawsuits against OpenAI are by *The New York Times* for allegedly using millions of articles for model training [39], and by *The Authors Guild* and a group of authors for using their written works to train models [40]. The latter lawsuit was later consolidated with two similar lawsuits by an author and a pair of journalists. OpenAI has also used Common Crawl in the training of their models [28], which is an open repository of crawled web data containing copyrighted data. Zhong et al. [33] find that current techniques for protecting the intellectual property rights of AI models work well, but that techniques for protecting the rights of the training data sets are not as effective.

The copyright protection of the AI models themselves is the other side of the same coin in this issue — generative AI, such as Generative Adversarial Networks (GAN), can be used to generate output, whether it be images, text or sound, with no surefire way of knowing what model was used to generate it and nothing necessitating that the owners of the model be credited. This problem is apparent with so called 'black box APIs', where a user only provides the input and is then given the output, generated by an undetermined AI model(s), by the service. Techniques to counteract the illicit use of models have been proposed: embedding watermarks within the model that are then included in the images generated by the model, and attributing generated images to specific models through either spectral or contrastive analysis [33].

Ownership of the content that is generated by AI models is also not at all clear-cut [1], [14]. The previously mentioned interviewed UX designers are split on how attribution should be handled; some feel that ownership should always remain with a human, since humans are the ones leading the process through which the content is created, and the AI is in a supportive role. Others consider attributing the parts of the work where AI was used to the AI as appropriate. Some considered the alternative of labeling the entire content as AI-contributed and mentioning the AI tools that were used, akin to how researchers mention their sources. In any case the designers considered it important that the original creators of artworks and styles be credited when used in UX work, regardless of how much they've been modified from their original forms (i.e., used as training data for an AI model) [1].

The consumer-facing facet of AI interaction is not to be disregarded either —

it brings with it significant deviations from traditional digital product interaction. For instance, a user's interaction with a digital system can be partially controlled by the AI depending on how autonomous the AI is. An AI controlling the interaction between the user and a digital system could be observed in text editing as an AI system predicting text that the user is writing and automatically predicts and fills in words or autocorrects words that it interprets as being misspelled [17]. How autonomous the AI is would then affect how its information is relayed to the user: does it automatically make changes to the text or does it simply notify the user and instead give suggestions.

Due to a combination of a lack of understanding, pop culture representation, and media sensationalism, the average person may have a somewhat fantastical idea of what AI is. Therefore the attitude towards systems where AI has any significant amount of autonomy will likely be met with apprehension. For conversational assistant AIs a proactive system is more helpful for older adults than a passive, reactive AI, and may be more effective in professional UX evaluation use as well [21]. Kliman-Silver et al. [17] highlight explainability and trust as integral parts of "AI-driven experiences" that serve to alleviate the user's uncertainty towards AI. Explainability is fundamental for achieving trust, since if the user is able to understand how the system makes a judgment it's easier for the user to vet the AI's output and to build trust towards it. Knearem et al. [6] also recognise that trust is required for an end user to accept assistance from an AI system. They too emphasise explainability as a crucial aspect of trust, along with the accuracy of the system. Systems that are used as tools to help creative pursuits may have more leeway when it comes to accuracy, since even if from a handful of the system's outputs, such as website layouts, only one or two are useful, the system may still be a useful tool. In contrast, a system that has a critical role and from which every output is significant must have high accuracy in its outputs to achieve trust.

3.3 Future role of Artificial Intelligence

AI has already seen a good deal of research in the domain of UX and UI. Chiou et al. foresee similar research being done in many different areas of design and architecture, and that AI will have its value proven through AI-assisted designs being used in "rigorous industrial processes" [2]. Businesses will likely require some overhauling internally, resulting from the increased usage of AI in its various processes [4]. Of particular interest are those relating to design processes, where AI has already seen adoption. The exact nature of these required changes is still nebulous, but researchers have already found that changes are necessary in order to effectively include both humans and AI within the business [4].

UX designers would like for AI in the future to be capable of assisting them in the design process [1]. Specifically mentioned tasks where assistance was seen as valuable were, for example, graphics design, prompt writing, and handoff automation to developers. Such wishes weren't limited only to creative tasks, but also extended to knowledge summarisation, data processing, and to offer guidance with regard to specialist knowledge. The ultimate goal for human-AI collaboration in UX design is to enable designers to work on things they truly want to, letting AI handle the undesired parts of the design work. Attaining this, both productivity and job satisfaction would be improved [1]. Clearly then among creative professionals the increased usage of AI isn't necessarily seen only as inevitable, but as desirable as well.

Some designers see AI adoption leading to a three-way collaboration in which designers, users, and AI work together, with AI improving based on human feedback and humans learning from the AI's insights. Stige et al. [4] also concur with this, and believe that in the future, collaboration between UX designers and AI will be the optimal way to include AI in the design process. This partnership will require designers to gain a more comprehensive understanding of AI technologies, to identify the places in the design process where applying AI would be beneficial. AI may not fit neatly into existing design processes and some researchers have suggested that AI may require its own, AI-specific, design processes [3].

Li et al. [22] believe that through AI continuing to improve so too will UI design increasingly focus more on personalisation and intelligence with AI becoming a more central component of it. The value of UI design won't diminish, but instead it will continue to be crucial in delivering UIs for users, whose expectations can be expected to increase. Most researchers agree with this, and don't believe that AI will bring with it a complete shift to how UX design works or that it'll supplant UX designers [1]. Instead, it's believed that AI will be assigned to roles that can be described as "creativity support tools" for the designers. UX designers themselves generally aren't worried about being replaced either, as they consider themselves to be "ambassadors" for human needs, where they understand what users need within the product, and are the final decision-makers within the design process. This sentiment isn't shared by artists, who have concerns that with AI systems improving in quality they themselves could be relegated to being essentially editors whose job is to act as quality control for the AI's outputs [1].

Since AI is by its nature probabilistic, so too will UX design change to be probabilistic in products where AI is a part of the final product [17]. The designers are then forced to look at design from the perspective of what are the likely outputs of the AI and how the user might react to them. This also includes preparing for the AI outputting anything unwanted by mistake. This will be apparent during design when common techniques like "user stories" and "scenarios" need to be modified to accommodate this newfound uncertainty. Designers will also need to be cognisant of the average user's unfamiliarity with AI and design accordingly, since the goal is for a product to be easily usable [17].

AI will be a central component of UX design and the designers themselves are clearly aware of it. In interviews with UX experts 35 of 49 respondents believed that that interaction between UX and machine learning will grow, and that they will "converge to some extent" [7]. Therefore, several designers hope for increased understanding of AI among designers, whether junior or seasoned, since understanding the technology will help designers leverage it more effectively and to not be caught off guard when changes stemming from its adoption happen [1]. They also hope for AI to become more transparent and explainable in its working and output, so that it may be adjusted in an informed manner by humans when necessary. This desire for transparency isn't limited to just people working with AI, but is shared by users as well, who are keen to know why and how the AI systems around them make the decisions that they do [18].

Researchers have also recognised that in order for AI to become more widely adopted in businesses, transparency is critical. In larger enterprises an AI engineer may be necessitated to act as a middleman for all interactions between the AI system and other key personnel who might wish to interact with it, due to said lack of transparency. Implementation of AI in a business then requires for these key personnel to be empowered in how they can interact with an AI system. For this purpose Chander et al. [18] put forward their "four pillars of transparent AI": accessibility, explainability, interactivity, and tunability. They are as follows:

1. AI should be accessible enough that key personnel can ask questions of it without having an AI engineer do it for them.

2. The answers that the AI provides should be understandable by people other than the AI engineer as well.

3. When getting an answer non-AI engineers should have the option to engage in a dialogue with the AI where the end result is either the person's beliefs changing or the AI's dataset changing or both.

4. Building off the previous point, AI should identify and readily offer end-users "tuners" that they can use to enact said changes.

Chiou et al. [2] call for investigation into AI to understand the impacts that it has on social dynamics through things such as: exacerbating class differences, social justice, and who has the capability to use AI.

Generative AI is now somewhat of a tool of democratisation: it's enabled people who may be lacking the skillset or specialisation required to work in a domain the traditional way to still dabble in it. For instance, people without programming expertise can generate code with a reasonable chance that it will work, or they can transcribe a picture from their thoughts to text and have it be generated [4]. UX design is no exception and some parts of the process that would have been completely handled by expert UX designers could potentially be partially delegated to anybody.

A problem that some UX designers foresee is defining metrics for gauging the quality of an AI's output [1]. They already face issues with choosing between AI outputs, so being able to reference standardised industry-wide metrics to make these decisions would make their job easier. Regarding this, technology companies have put forward suggestions for principles and guidelines for human-AI interaction [3]. Researchers agree with the importance of building a framework of understanding and rules for how AI operates [1], [4]. This would help with making AI more predictable, which would make it more useful in projects with multiple people working with the same AI. A framework of understanding would also enable designers to more critically judge the outputs that AI provides.

The reliability of outputs is a voiced concern among UX professionals. In a survey of 38 such professionals [6] almost half (49%) said that they do not trust the quality of an AI's output, and 70% responded in an either neutral or negative way to the statement "AI is a reliable partner in the work that I do". There was also concern over the types of data that would have to be provided to an AI system in order to increase its reliability. Only 24% of respondents were willing to share personal data, anonymised or not, work-related data ranged from 54% to 66% depending on the specific data, but only 9% refused to share any data.

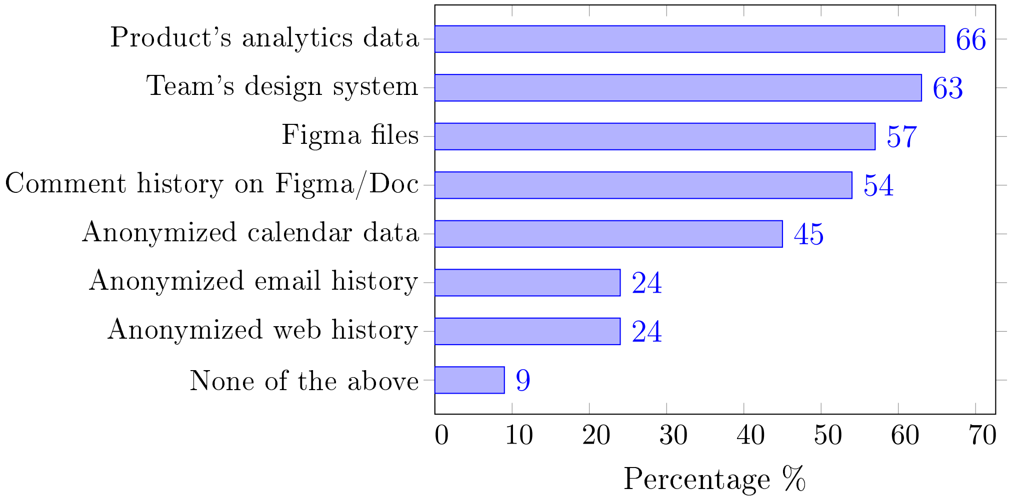


Figure 3.1: Proportion of participants that were willing to provide different design-related personal data [6].

Design professionals disagree on if and how AI should be attributed in content utilising it [1], and consistency in this regard is crucial to ensure fairness. Researchers recognise the need for a framework for attributing authorship of different parts of the output.

Some researchers have identified the current paradigm of interaction with AI systems as being a problem [22]. Chat-based interaction carries the burden of requiring users to be able to express themselves clearly enough using text that the AI can produce outputs of a good enough quality. The researchers believe that many users are not articulate enough for this, and therefore another mode of interaction is needed [22]. Current generative AI has already shifted to such a mode, and uses an intent-based paradigm in which the user only describes the output to the system, in contrast to chatbots where the user instructs them in what to do and then receives an output based on that.

On page 9 we covered the findings of Yang et al. [15], in which they stated that there were significant differences in how succesfully designers work with AI systems across the industry. The key difference was that the better-performing designers worked in enterprises with greater focus on AI and access to data scientists. With that, Yang et al. [15] recommend developing design tools and methodologies to help bridge the gap between designers with access to these resources and those without. In addition they call for a shift in the curriculum of UX design toward including data science; specifically understanding and manipulating quantitative data, since that's what telemetry data consists of, and qualitative user study methods. Lomas et al. [41] believe a lack of knowledge regarding data science may also prove to be a hindrance when designers are required to interpret the meaning behind quantitative data. They also recognise this need for designers to update their understandings in the face of increasingly data-driven design tasks. Designers specifically need to understand optimisation metrics, so that they may understand why a design that is optimised for a metric may still not be an ideal design.

Within the field of human-computer interaction (HCI) researchers have generally agreed that the technical literacy of UX designers needs to be improved in order for them to productively work with AI [3]. The specifics of what type of knowledge of AI would enable this development are still unclear. Providing UX designers with this knowledge serves to make them be better equipped to work in their industry that is becoming increasingly reliant on such technical knowledge.

The need for better learning resources is also noted; the UX designers interviewed by Yang et al. [15] had very individualistic interpretations of various machine learning concepts, and they differred from textbook definitions. UX designers should also have easier access to tinkering with AI so that they can build a general familiarity with such systems and for them to gain a better understanding of what AI can do [3]. The researchers highlighted the designers' tendency to employ exemplars in an effort to understand machine learning concepts and qualities [15]. As such, they suggest that building a repository of these exemplars of different applications of machine learning could be an effective way of communicating limitations and capabilities of ML to UX practicioners.

To summarise the views of UX designers on AI, they hope for AI tools to be developed into more intuitive, visually-oriented, and collaborative intelligent systems [1]. They look forward to the continued improvement of AI while simultaneously expecting that humans have a place in the design process, especially by having control of the creative process and being the final arbiter in decision-making.

4 User Interface development optimisation

Designing and developing good user interfaces is undoubtedly a laborious process and is a cornerstone of a successful software system. This is particularly true for websites, which have become ubiquitous ways of providing information or attracting customers for practically all organisations [9]. The specfic goal of a website may vary across organisations, but for many organisations, their websites serve as the digital facade of the organisation. As such, a wealth of tools and techniques have been created for the purpose of easing the creation of good user interfaces [10]. AI is a new avenue for developing such aids: sometimes by revamping already existing tools by incorporating AI within them, and sometimes by enabling entirely new ways of conducting UI design [7].

User experience design seeks to ensure that a product is efficient and pleasing to use, and is therefore a critical part of the work that makes a UI good. Due to this, UX design is an area of particular interest for optimisation using AI, whether it be within user data analysis [21], layout optimisation [42], or other parts of the process. Machine learning has been used for such purposes, and is becoming increasingly common [7]. Many of the AI tools developed for UX development are focused on the evaluation phase of design [43]. Fewer tools are focused on assisting designers during ideation or the implementation of design.

A weakness of non-AI optimisation, specifically model-based optimisation, is that it completely relies on the predictive model being comprehensive and valid [44]. Then again, the output of optimisers using AI models is likewise reliant on their respective models having similar qualities, though perhaps to a lesser degree. Non-AI optimisers can mitigate this necessity by putting a human in the loop, allowing them to guide the optimiser [44]; but there's no apparent reason optimisers utilising AI couldn't implement such measures as well. Several contemporary AI tools are already synergistic with humans, relying on humans to provide guidance in order to direct the AI's output.

4.1 Non-AI methods of development optimisation

Various methods for optimising the design of user interfaces have been developed and used before the harnessing of AI for such purposes, and their potential for this purpose has been recognised by researchers [42]. Non-interactive optimisers are ones in which the designer has little or no control over the process once it has started. Many such optimisers have been used for optimising specific aspects of UI, such as menus, accessibility, widget and dialogue layouts [44]. Predictive models are an efficient way of building non-AI optimisation tools: because such models are a way of encapsulating scientific knowledge from across the literature, they are a cheap way of incorporating expert knowledge into a designer's workflow [45].

Algorithmic methods are often the way of approaching the problem of user interface optimisation, when AI isn't the chosen tool. Algorithmic approaches serve well as complementary tools for designers, enabling them to explore design spaces more thoroughly and with less biases [42]. This is a difficult task, however, as UI optimisation is a multi-objective task where the goal is to achieve quality from both a usability and an aesthetic standpoint [42]. In addition, due to the wealth of permutations that any given UI design can have, UI optimisation is a combinatorially complex process resulting in algorithmic approaches being especially time-consuming.

For instance, interactive genetic algorithms are a technique employed by Quiroz et al. [46]. Interactive genetic algorithms differ from standard genetic algorithms in that they use user evaluation in place of a fitness function, which in this case enables designers to impart, to the algorithm itself, their own domain knowledge and other skills necessary for designing a good UI. They achieve this by encoding ready UIs as individuals for the algorithm, and then iterating over generations to get different UI designs from the algorithm. At certain points the designers are shown nine of the best designs within population, from which they choose the best and the worst design. A weakness of this type of genetic algorithm is that due to the subjectivity of user evaluation the human element may cause noise by having the evaluator change their goals during a run of the algorithm [46].

One of the earlier examples is *SUPPLE* by Gajos and Weld [47], which is a system for interface generation. Their system approaches the problem of UI generation from an optimisation standpoint: they search for the best rendition of a UI by seeking to minimise a cost function that estimates the required user effort to manipulate a user interface, while simultaneously satisfying constraints given by the user. The process of generating a UI using the system, from start to finish, took around 10 seconds after the algorithm had been optimised. Though the system was capable of generating UIs that were deemed functional, the researchers recognised that their system had only been tested on two relatively simple applications, and that more complex applications would need the system to be more robust. They also predicted that UI designers would likely be skeptical of such a system, citing discomfort with functional specification and a preference for fine control of the UI. In retrospect, this prediction seems quite correct considering the current prevalent attitudes toward AI among designers.

DesignScape is another graphic design system created by O'Donovan et al. [48].

It provides its user layout suggestions for a UI that the user creates using a UI editor included in the system. DesignScape generates designs that encode various design principles in order to improve on the UI that was given to it as input. The system can generate both UIs that are slight improvements to the original (suggestive mode), and UIs that significantly differ from the original (adaptive mode) in an attempt to brainstorm new layout ideas. Both types of UIs generated by the system were received well by users; when prompted to choose between a UI generated by the system and a baseline UI, users voted for adaptive UIs 63.4% of the time and for suggestive UIs 57.7% over baseline UIs.

MenuOptimizer by Bailly et al. [44] is a similar toolset, designed to interactively assist in UI menu design and is implemented as a set of interactions for the already existing design tool *QtDesigner*. MenuOptimizer, like DesignScape, includes two modes of operation: one that updates an existing menu where it infers user behaviour from logs or interviews, and one that suggests completely new layouts based on the required commands and predicted user behaviour inputted by the designer, from which the system makes assumptions.

For both of MenuOptimizer's modes it's possible for the designer to modify the commands within menus that are then taken into account for the next iteration of the generated layout. Like SUPPLE [47], MenuOptimizer also takes an optimisation angle to this problem, seeking to find designs that minimise or maximise an objective function. Quality ratings between MenuOptimizer and QtDesigner menus were not statistically significant, but the users made around 37.5% fewer edits while using MenuOptimizer than when using QtDesigner [44].

These types of tools are ones that could be improved using AI quite naturally, since the fundamental idea of them is to apply underlying UI design principles to a UI given to the system. These principles deal with element positioning, size and alignment, and as such they could likely be learned by an AI, since they're going to appear in a lot of "good" UIs.

Similar tools exist for UI animations as well. *P2A* is a tool developed by Natarajan and Csallner [11] for the purpose of developing animated mobile applications. P2A accomplishes this by reverse engineering a set of images given to it as input and constructing view hierarchies from them, from which the user then selects pairs that represent the start and end states of an animation, and finally the system generates appropriate source files that build a UI and accompanying animation(s).

Development toolkits and libraries are typical tools for speeding up development of UIs. They provide basic elements of UIs as ready-made components, such as buttons and menus [46], so that designers aren't forced to create them themselves from scratch for each project. They are low-level tools and don't impose any rules or restrictions on the designers as to how to conduct UI design. Because of this, using them alone may enable designers to create UIs quickly, but the UI may still be unwieldy or otherwise poorly made [46].

For this purpose designers use style guidelines and design principles, which are higher-level tools meant to deter designers from creating UIs that would be considered "bad" [46]. The guidelines define how UIs are to be structured with regard to the placement and properties of components, and the use of visual elements, such as colours, fonts, and spacing. Style guidelines are ubiquitous. Many organisations have their own, which helps to ensure the uniformity of an organisation's UI designs. Since these tools are purposefully abstract enough that they could be applied to most UIs, accurately applying them is up to the interpretation of the designer [46]. Therefore, as with interactive genetic algorithms, the human element is a source of some unpredictability for the effectiveness of such tools.

Researchers working on MenuOptimizer [44] garnered insights from its development that can be generalised to other interactive optimisation systems: information regarding the design should stay simple enough so as to not confuse the user; when generating novel designs interactive optimisation systems should prioritise producing a minimally viable first design from which to begin iterating, since users are impatient; and finally, designers may favour the quality of UI element item groupings over that of the UI's overall performance in the early stages of design.

4.2 Utilising Artificial Intelligence in development optimisation

For the purpose of optimising user interface development AI has some advantages in comparison to non-AI methods. In the early stage of UI development designers are coming up with designs and refining them, and these tasks can be aided using tools. Non-AI methods for generating designs and layouts have been created, such as MenuOptimizer [44] and DesignScape [48]. These methods are somewhat rigid, usually being limited to moving components around or adding predefined elements. AI, instead, could be utilised in a multitude of ways as creativity support for designers. In such roles AI could assist designers by brainstorming and providing them with completely new (though not necessarily novel) designs that serve as a starting point from which to iterate, or by generating variations of existing designs [6]. AI could also be used to generate templates for common screens, such as login and account screens, and for automatically adapting designs to different form factors.

AI's role as creativity support isn't limited to the initial drafting of a design, it could encompass further tasks as well. Collecting and organising information, providing suggestions, and idea refinement are some other ways than generating designs that AI could optimise development [6]. As discussed in section 3.3, designers tend to prefer that AI takes such assistive roles and leaves the creative design work to humans.

Outside of creativity support AI could be utilised during the implementation of a design. Much of the work in design is predictable and repetitive, which are the kinds of tasks AI excels at [6]. Routine tasks, such as consistency checking (colour, spacing, etc.) and grammar checking, could have most of their workload handled by AI. Wireframe generation, layout variation generation, and design prototypes are also tasks for which AI is well-suited for [1]. Especially prototyping is a task where repetition is common, as designers often make low fidelity prototypes that they then may have to create high fidelity prototypes of, if the proposed solution is good enough. AI has been proposed as a way of efficiently moving between different levels of fidelity, which could hasten the process of prototyping designs considerably [4]. UX evaluation is a crucial task for ensuring smooth user experience, and has seen some integration of AI into its processes as well [21]. Many of the aforementioned tasks are also time-consuming, making such tasks of particular interest for automation [6], [21].

As for what kind of usage AI sees within UX design, Stige et al. [4] divide the process of user-centered design into five steps. Of interest from these steps are the third and fifth steps: producing the design solution and developing the solution, respectively. Their analysis of the literature showed that 35% of all AI usage within UX design has focused on the third step and 31% on the fifth step. The remaining portion of AI usage has focused on evaluating design solutions, and on understanding the proposed design solution's context of use. None of the AI usage within their analysis was focused on requirement specification for the design.

The task of acting as a subject expert could also be handled by an AI [6]. It could provide information and guidance for any subject it's trained for (e.g. accessibility or internationalisation). Considering the uncertain nature of AI output there are clear limits to this (cybersecurity experts likely shouldn't be outsourced entirely to AI). The non-AI method of accomplishing such a task would be to find an actual expert in the subject, or for the designer to learn it themselves. Both of these options likely eclipse in effort that of asking an AI for guidance.

Outside of bespoke tools created by researchers, AI has been adopted into numerous commercial creative work tools, such as *Photoshop*, *Canva*, and *Figma*. Although AI's implementation into said tools may be in part rudimentary there is interest in utilising AI further within UX workflows, as its potential has been demonstrated in the literature [1], [6].

During a workshop themed around design ideation with AI, designers identified several practical ways of using AI within design: generating alternative designs, brainstorming initial design ideas, helping to identify design problems, among others [5]. Each identified task was time-consuming, but relatively simple to perform, which are the kinds of tasks discussed in section 3.3 that UX designers wish to automate with AI as well. In another survey of UX professionals, a similar desire for having AI handle such tasks was expressed [6]. When handling such tasks, AI-aided design tools may lower the barrier for testing out new ideas and designs by speeding up the process and lowering costs for doing so. A shorter design cycle then allows designers to explore a greater number of designs, ultimately resulting in a better user experience [1].

During the workshop, AI was identified as particularly useful for designers working alone, since AI could act as a stand-in for a human coworker [5]. Regardless of the quality of the AI's responses, the responses would still prompt the designer to engage in reflection and reasoning, akin to humans working together and suggesting ideas to each other. One designer felt that AI could also be used for getting feedback on ideas, and on developing them further.

A lack of context and memory for the AI was a voiced concern for several of the designers; the AI treated every problem as if it existed in a void [5]. The designers felt that this led to the AI being unsuitable for creating any more complex designs, or for further developing designs it had already created. Again similar to the feelings expressed by UX designers in section 3, designers in the workshop group felt that AI couldn't be trusted to work on its own to create high-quality designs. Instead, they felt that AI was more suited for providing them with different designs to work with — exploring the design space widely, rather than deeply.

The non-AI tool P2A accomplishes generating animations and the accompanying UI using reverse engineering [11]. On the AI side of the task of UI optimisation *pix2code* fulfils a similar goal, using neural networks to accomplish the reverse engineering [8]. The model outputs code based on a given GUI screenshot. This given code will then produce the said GUI. It's not designed for any one specific target platform, but is instead purposefully made to be platform agnostic and such that the model can be trained on different datasets to support different platforms. Tests on the system have given succesful results on Android, IOS, and other mobile platforms.

The researchers who created pix2code acknowledge that the model sometimes has issues replicating the exact styles used or long lists of components, but overall performs satisfactorily. They additionally consider it feasible to improve the model's performance substantially by increasing the parameter count and the amount of training data [8]. With the ongoing AI spring the amount of data that has been harnessed for training datasets has increased. It is not clear, however, that this is the case for data that consists of GUI screenshots and associated source code, which is the kind of data that was used to train pix2code's model.

Aşıroğlu et al. [9] created a tool for the purpose of automatically generating HTML code from mock-up images of webpages. The process of website design is iterative in nature and highly repetitive due to designs being changed based on received feedback. This then requires developers to rewrite the code for each subsequent iteration of the website's design, so that it can in turn be tested and be provided feedback for. This tool could then serve the purpose of expediting the process of website design by reducing the amount of rewriting that developers need to do. This is achieved by allowing them to generate a ready webpage from a draft of the design and merely fix inconsistencies and mistakes that the tool may make.

The tool is similar to pix2code in its purpose, but differs considerably in how prominent AI is in the process, which is divided into three parts. The initial step is to detect objects within the draft of the design that is given to the tool as an input image. This is done through blurring the image, performing contour detection, and morphological transformations on the image. The detected objects are cropped and passed to a convolutional neural network (CNN) that was trained to perform object recognition, categorising the detected objects into different types of webpage components. Finally, the tool runs a HTML builder algorithm using coordinates from the object detection phase and component types from the object recognition phase. With these, the algorithm outputs ready HTML code for a webpage with the correct types of components in their appropriate places. No estimate for how succesful the tool is was given, but their CNN model achieved a validation accuracy of 73%.

Lomas et al. [41] suggest that multi-armed bandits offer insight for effective ways of using AI to aid in interface design. The *Multi-Armed Bandit problem* is the theoretical problem of maximising one's winnings from a selection of slot machines where payoffs are different and unknown. In the case of UX design each variation of a given design would then represent the "arm" of a single slot machine with an unknown payoff. How one solves this problem is a selection strategy: seeking to maximise winnings by figuring out which slot machines give the best returns based on previous pulls and their payoff. Multi-armed bandits have already seen application for optimisation purposes within interface design and other areas: within research, news article recommendations, and advertising [41].

The researchers' first experiment tested if bandit algorithms will automatically search for the best designs in a given design space, and if bandit algorithms reduce the cost of experimentation (i.e., are less costly for user engagement). Two bandit algorithms were compared against each other and against fully random selection. This was done by having users play a game with six permutations. The algorithms decided which permutation was given to the user after each session, after which the user decided if they wanted to keep playing. Both of the bandit algorithms outperformed fully random selection in regard to user engagement — they gave users fewer low-engagement permutations and as a result users played for longer. Additionally the experiment confirmed that bandit algorithms do explore the design space for the best design; they preferentially assigned different permutations of the game to users, rather than equally as the fully random selection did.

Lomas et al. identify a caveat when using bandits for UI design in that a human is required to be part of the process [41]. This is due to automated systems potentially optimising for the wrong metric, as they theorise may have happened to them. This is especially true for AI where optimising for arbitrary goals is a possibility [49]. The need for a human in the process is in line with opinions of UX specialists and researchers discussed in section 3, where human oversight into AI systems is seen as a necessity. Lomas et al. recognise that neither humans nor AI are infallible, and that both have their own strengths. Therefore, systems should be designed so that they can facilitate the effective cooperation of humans and AI.

Aside from producing complete UIs or otherwise aiding in design, AI has also been utilised in UX analysis [21]. Analysing UX is a time-consuming process for which resources are often strained due to its breadth. Supplementing usability testing with AI also sees promise, as AI can be used to make predictions on how users may react to a proposed design [4]. Because of usability testing's irreplacability for the purpose of identifying hindrances and barriers, fully replacing user testing with AI has been argued against [4]. Likewise here as well, the proposed solution is one that builds off the strengths of both designers and AI — in preference to a complete substitution by AI. Fully automated UX analysis by AI has seen some success: two-thirds of usability problems that were detected during manual testing could also be detected using automated methods [21]. The undetected problems were such that they had no overt signs in the analysed user data and were very obfuscated for an AI. Contrastingly, problems undetected by UX evaluators can be banal usability problems that an AI has no difficulty detecting, or problems outside of the scope of preplanned tests.

With this in mind, Kuang [21] suggests that an interactive assistant may be a good way of integrating AI into UX evaluation. The UX evaluator would be able to focus their efforts on discovering the more elusive usability problems while letting the AI discover the rudimentary problems. This is similar to how many UX designers wish to collaborate with AI: letting it handle the parts of design they themselves don't wish to work on. The previously discussed issue of distrust from UX specialists toward fully automatic applications of AI is present here as well, and an assistant type AI would also help to alleviate it and let UX evaluator build trust towards the system. Further supporting Kuang's advocacy for an AI assistant is that UX evaluators would work more reliably and comprehensively as teams rather than individually, but the cost in time, resources, and effort to actually do so is often prohibitive.

4.2.1 Challenges in working with Artificial Intelligence

While the prospect of utilising AI within UI development workflows is promising, there are notable roadblocks to furthering this goal. Especially within UX design, designers have issues working with AI effectively [7]. The lack of technical knowledge regarding AI among designers is severe, which then results in designers being unable to conceptualise how to make use of AI effectively within the design workflow [15].

Attempts to remedy, or to mitigate the effect of this lack of knowledge have been made, such as creating design heuristics to aid UX designers to utilise AI during the conceptual phase of design [43].

Adapting technical information into a more easily digestable format for designers has also been suggested. In the literature designers have been found to communicate design ideas and machine learning topics among each other, using exemplars and abstractions [7]. Due to the designers' familiarity with these modes of communication, researchers have suggested using them to communicate the capabilities and limitations of AI to designers [15]. This would be easier for designers to comprehend than technical literature, and would still enable designers to communicate with data scientists on AI topics.

When working with AI some designers may face challenges that are fundamental in nature, because of the designers' particular design philosophy. "Fail early, fail often" is a common adage that some designers have adopted as a kind of strategy. This motto is, however, contradictory to the goal of utilising AI within UX design — projects using AI are longer time-wise to both plan and to execute, and prototyping them is harder [3], [7], [15]. It is the probabilistic nature of AI that makes systems incorporating it so uniquely challenging to prototype for [3]. Predicting all the ways an AI system might adapt to each individual user or every kind of error it might make are both impossible tasks. This then necessitates a shift from deterministic product design to probabilistic, where a system's outputs have a certain likelihood of being correct, and user experience techniques like "user stories" and "scenarios" additionally need to include information on the probabilities of the related experience [17].

Collaboration between UX designers and data scientists has been observed to help the designers work more effectively with machine learning [7]. Designers did, however, have trouble working with data scientists proactively [3], [7]. This is likely partially due to the designers' lack of knowledge regarding ML and the resulting inability to communicate about their desired topic. A lack of a shared workflow or boundary objects, that are information used for cross-disciplinary communication, are also contributing factors to this difficulty [3]. Clearly then some minimal amount of technical knowledge of AI for designers is required, so that data scientists can mitigate the issue of the designers themselves lacking deeper understanding. This inability to communicate with experts is inconvenient, as there doesn't currently appear to be clear alternatives for working with AI effectively, aside from the designers themselves becoming AI experts. Whether this collaboration is a necessity after the institutional knowledge among UX designers matures more regarding AI is unclear.

The issues of being unable to make effective use of AI, and being unable to communicate AI topics effectively with experts, have noticeable overlap — both stem from a lack of understanding of many attributes of AI: its capabilities, limitations, and how it functions, among others. The main reason for this is that UX design education does not prepare designers for working with AI nor does it inspire them to learn more about it [7]. Both of the aforementioned issues are such that they will likely diminish over time. That is if design education shifts towards data science: teaching designers basic data literacy, analytics, and interpretation, along with some basic knowledge of AI technologies [15]. With such changes the future designers entering the workforce would be equipped with the knowledge for effectively making use of AI within their workflows, for continuing to learn about AI as it develops, and for keeping up with design that is growing to be more data-driven [3].

In the future, AI will likely see further integration into design processes, once designers become more comfortable and proficient in working with it. This is due to the valuable contributions that AI can provide, that are recognised by researchers [14], [43] and design professionals [1], [21] alike.

5 Comparing User Interfaces created by humans and by language models

An experiment and a survey were conducted in order to test how successfully an off-the-shelf LLM can improve a user interface designed by a human. For the experiment, we designed and implemented a website for an e-commerce clothes store. The website has limited functionality, but enough so that it is possible to complete a set of tasks using it. The HTML and CSS for the website were given to an LLM (GPT-4-0125-preview), and it was instructed to improve the website using different prompts. The HTML and CSS output by the LLM were then used to create a variation of the website.

Prior to building the e-commerce website and attempting to improve it, a suitable LLM needed to be chosen. This was done by having numerous LLMs generate a small webpage using only instructions on what the webpage should include, as opposed to having it improve an existing design. Several different large language models were tested this way to identify the most promising ones. The best performing ones were Google's *Bard* (which has since been rebranded as *Gemini*), Microsoft's *Bing Chat* (which has since been rebranded as *Microsoft Copilot*), and OpenAI's *ChatGPT*.

Of these, ChatGPT was ultimately chosen, as the university already had access to more robust models from OpenAI than just the ones offered for free users. OpenAI's well-documented API also affected this choice: it enabled developing a simple workflow for improving the original website design using an LLM. The specific model that was used was *GPT-4-0125-preview*, which was at the time the latest one that OpenAI had released.

While the HTML and CSS that the LLM output was often of an otherwise acceptable quality, in many cases it wasn't immediately functional when added to the website's code base. At times this was due to LLM hallucinations, which often manifested as the LLM referring to user interface elements that did not exist. With a programming language with static typing, this kind of code would result in syntax errors. With HTML and CSS, which are dynamically typed, these hallucinations did not completely break the design, but of course the intended improvements were ineffective due to not functioning. At other times the problem was the LLM inventing new functionality, but leaving it completely abstract and instructing the user to implement it instead. The LLM, for instance, recognised that it's working on an e-commerce website and added front-end components that it then instructed for the user to connect to a non-existent backend shopping cart software. These issues were such that they could in many cases be fixed with little effort. Since the purpose for this thesis was to evaluate how successfully an LLM can optimise a UI, as opposed to whether an LLM can independently create a UI from start to finish, these shortcoming weren't considered fatal. Attempts at using AI to optimise UI layouts have similarly seen researchers been forced to tweak the output of an AI to get satisfactory results [10].

5.1 Experiment and survey test setting

Both the experiment and the survey were conducted using Hallway Testing around the University of Turku campus. The process of improving the website's design using an LLM was done using a bespoke setup, in which *OpenAI*'s API was used to facilitate communication with the LLM. The experiment was completed by 42 people, and the survey had 23 respondents.

For the experiment, test subjects were randomly assigned to either the human-made website or the LLM-improved website, where they were then asked to complete two tasks. The tasks were designed to be typical actions that a user would likely wish to perform on an e-commerce website:

1. Change your account's data sharing settings.

2. Within the store's catalogue, filter for shoes of size L that are made of faux leather.

The instructions for both tasks were minimal, so that they forced the test subject to use their own knowledge of how such websites typically work in order to complete the task. Once a test session was started, the website started recording information about the performed clicks: When the click happened, its coordinates, and the webpage it was performed on were recorded. At the end of the test these clicks were then saved as the data for a single test session. This test was completed in total 42 times. In these tests, 19 subjects were randomly assigned to the human-made website, and 23 to the LLM-improved website. Sessions lasting over two minutes were dropped from analysis, as they were the result of either errors or subjects choosing to cease the session. There were five such sessions, leaving in total 37 sessions for analysis. The five dropped sessions included two sessions with the human-made website and three with the LLM-improved website.

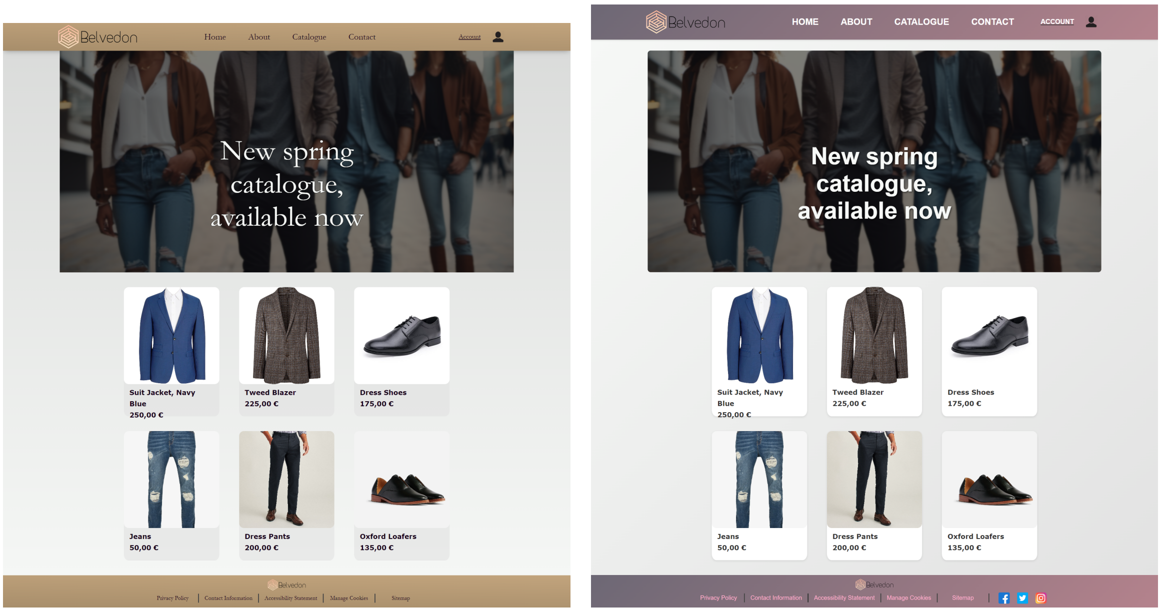
The questionnaire for the survey contained five questions that all pertained to the websites that were developed for the experiment. Before seeing the questions the respondent was first shown a comparison of the two websites by comparing each webpage of both websites against each other. The questions were related to the aesthetic appeal of the websites' designs. Each of the questions asked the respondent to choose which of the two websites' designs performed better in regard to the aesthetic quality presented within the question. All the questions were presented as five-point Likert items.

5.2 Changes made to the LLM-improved UI

Since the aim of the experiment was to compare a human-made UI and an LLM-improved version of the same UI, rather than a completely new design by an LLM, the two UIs are very similar. As the improved version is based on the original UI, they are structurally almost identical. Their main differences have to do instead with the layouts of elements and their aesthetics. Both UIs consist of four pages: the index page (Figure 5.1), the catalogue page (Figure 5.2), the account management page (Figure 5.3), and the data sharing page (Figure 5.4). The LLM often added comments into the code that it output where it described what changes it had made and why. The only instances of the LLM attempting to add new functionality to the UI were in the index page and the catalogue page.

Within the index page (Figure 5.1), the LLM attempted to include social media links within the footer of the webpage. As the LLM wasn't explicitly instructed to add any specific elements, this then indicates that the LLM would consider the presence of such links an aspect of good design in the context of website design. This is likely due to social media being an important part of marketing for most public-facing organisations, and as a result such links are present within many websites in a similar location.

The other attempts at adding functionality were within the catalogue page (Figure 5.2). The first item that the LLM attempted to add was a product search bar. As the website in question is a clothes store, this is a component with obvious use. This is, again, a component present in most clothes stores, and so it's unsurprising that the LLM made an attempt to include such an item to this UI as well. This is likely another characteristic typical of UIs that the LLM considers examples of good



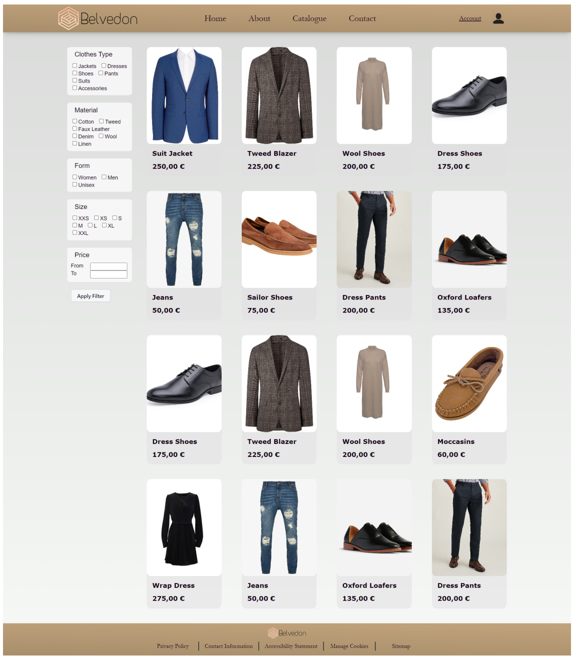
(a) Index page, human-made.

(b) Index page, LLM-improved.

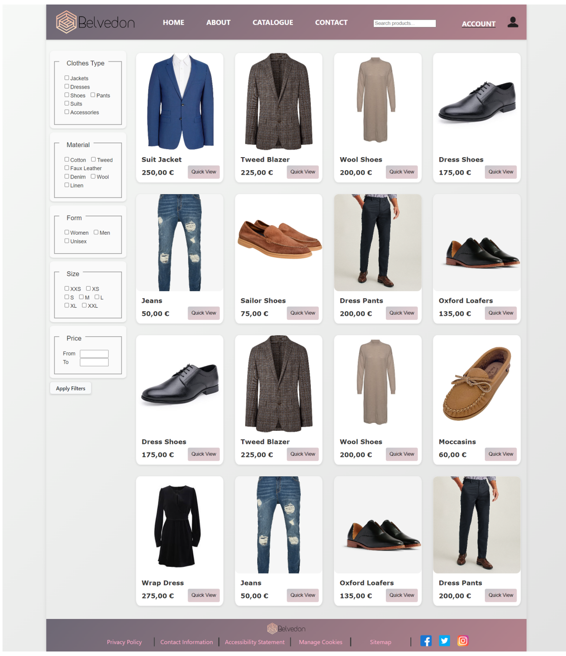
Figure 5.1: Index pages of both UIs.

design, and therefore due to being instructed to improve the design it added such a component here as well. The final attempt at including functionality was a "Quick View" button on each individual product card within the catalogue. Typically clicking on a product card opens the product's own webpage, in which the user can see all the available information regarding the product. A Quick View button opens a pop-up window within the webpage that allows the user to quickly preview the product in greater detail, without exiting the catalogue page. These pop-up windows usually contain images and information regarding the product beyond what is shown on the product card and are often used in e-commerce websites.

In addition to the attempts at adding functionality, the LLM made changes to the UI's aesthetics and the layout of elements. The LLM's changes to the design included a lot of additional whitespace. This is apparent in every page of the website; within the index page the banner image had padding added above it, so that it's not directly attached to the page header. In the catalogue page, whitespace was added around the entire design, while in the original design the content of the webpage



(a) Catalogue page, human-made.

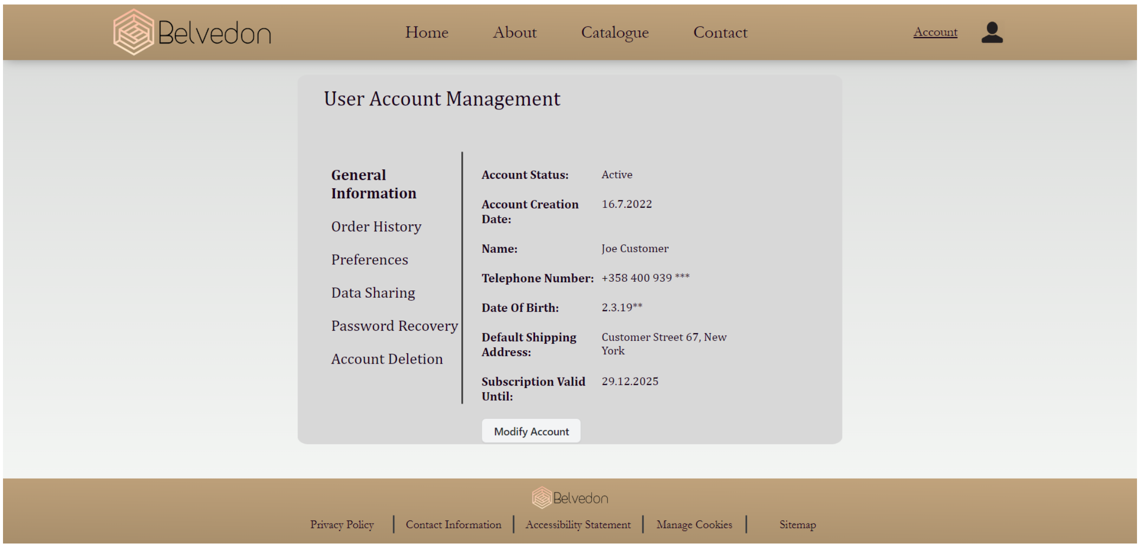


(b) Catalogue page, LLM-improved.

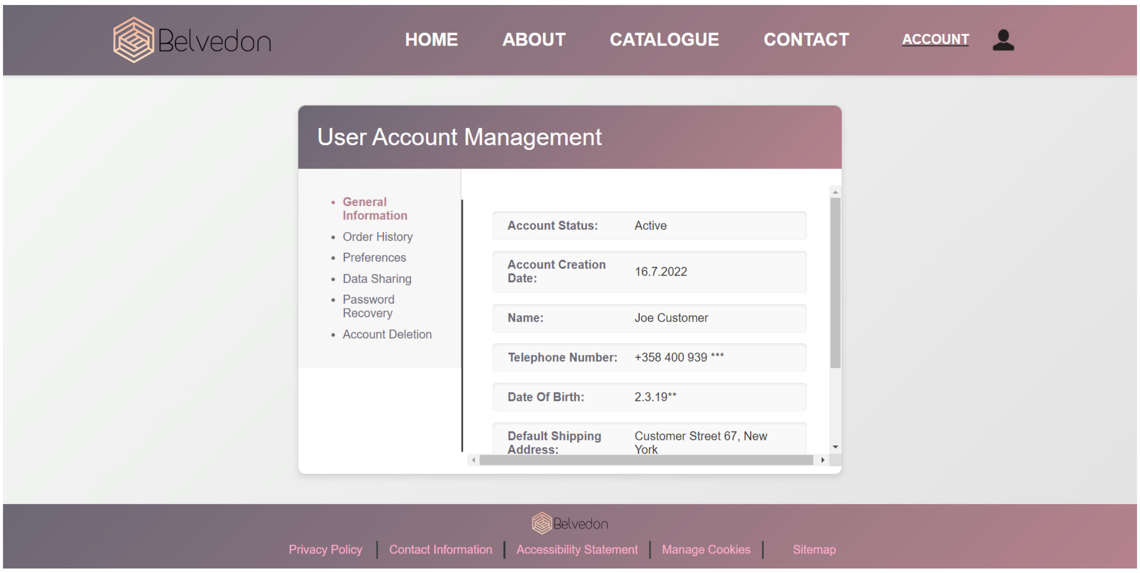
Figure 5.2: Catalogue pages of both UIs.

extended to the left and right sides of the page. The user account (Figure 5.3) and data sharing pages (Figure 5.4) both had whitespace added around the middle element of the page, which contains the actual user account settings, so that they're further from the top of the screen.

The importance of including an appropriate amount of whitespace in a design is well-understood by UX designers; having too little of it can make a design feel claustrophobic [50]. Coursaris and Kripintris [51] tested the effect of the amount of whitespace in a design on its usability. They found that the perceived effectiveness of a design increased as the proportion of whitespace to the rest of the content approaches 50%. They additionally found that an overabundance of whitespace (occupying more than 50% of a design) was potentially more harmful towards the perceived usability of websites than having too little whitespace. It is thus unsurprising that the LLM increased the amount of whitespace in the design, as it is likely another thing that the LLM associated with good design.



(a) Account management page, human-made.



(b) Account management page, LLM-improved.

Figure 5.3: Account management pages of both UIs.

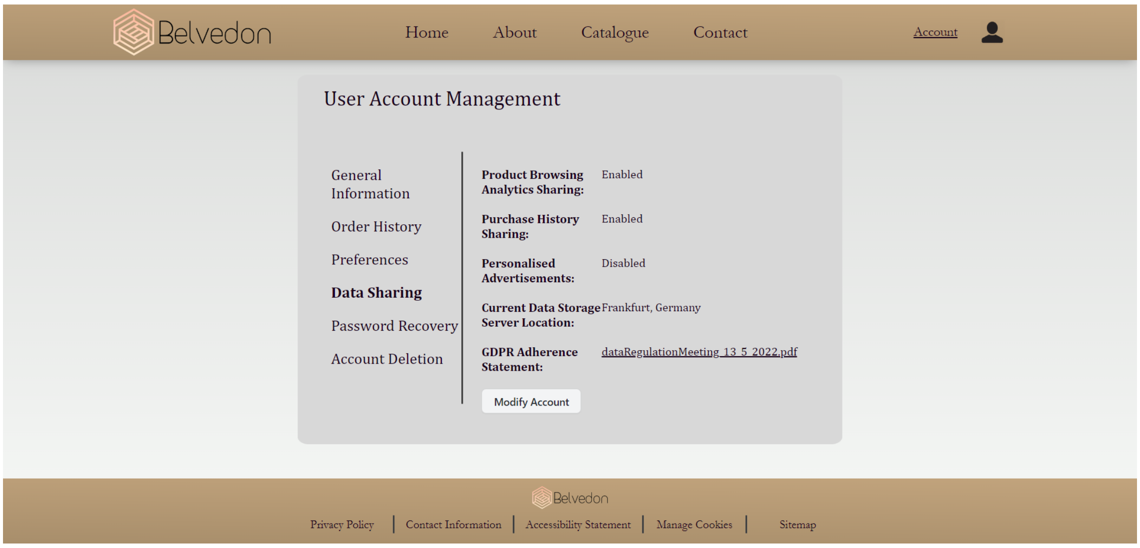
For aesthetic changes, the LLM additionally changed the font in a few places: the header and the footer, the index page banner, and the user account and the data sharing page. The original font used in these places was Garamond, which was chosen as it's one of the more elegant looking fonts that has serifs, and the intent was for the design to be for a clothes store selling higher-end items. The font that the LLM replaced this with was Roboto, which it described as a "more modern and legible" font. It also made the text in the header white for "better contrast", along with capitalising it and adding boldness to it. The text in the footer was instead changed to be the colour that the header's navigation text is on hover, which is a light pink; the footer links' colour when hovering them was changed to be the colour of the header's navigation links when **not** hovering them. The reasoning for these changes was cited as "consistency with the header hover state".

In addition to font colours, the LLM also changed the colour gradient present in the header and the footer. The new banner colour was considered to be more vibrant by the LLM, which might be another attempt at increasing the modernity of the design, as more muted colours are less favoured in modern web design. This colour scheme also carried over the to the user account and data sharing pages, where the same colour gradient was added to the title bar of both pages. In addition, in these pages the LLM attempted to increase clarity by adding an inset shadow to each individual setting within the page. This resulted in each setting being clearly separated from the others. This also made the design bulkier — causing that section of the account management panel to become scrollable due to all the content not fitting on it at once. Also the different categories on the left side of the account management panel had also their clarity improved; they were stylised as a bullet point list, which allowed them to be grouped more tightly without it being unclear where one category ended and another began. A highlight colour was also added to the currently selected category. This is an additional level of feedback for the user in addition to the bolding that was already present in the original design.

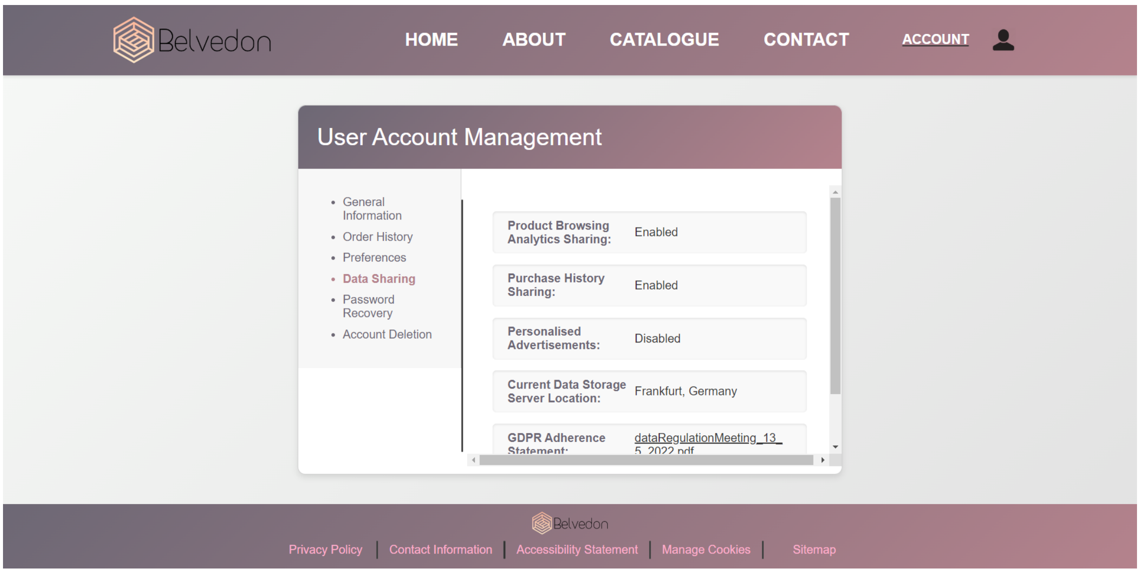
The LLM added shadows to many of the elements in the design. The settings panel in user account and data sharing webpages had a shadow added around its border, the product cards in the catalogue and index webpages had shadows added around their borders, and the filter categories in the product page had shadows added around their borders as well. These all serve to make the design feel more three-dimensional, and they help the user to distinguish different elements in the webpage. The colours of the user account settings panel and products cards were also made noticeably lighter. In the original design the colours for these elements were of a darker shade than their backgrounds, but now they are clearly lighter than their backgrounds. With the contrast of the lighter colours and the dark background, the elements are more easily distinguished.

Finally, the LLM-improved design attempted to add reactivity to the webpages in many places. In the LLM-improved design the navigation bar had the text of the links change when hovering with the cursor. In the original design such reactivity was not present. This is an example of the LLM attempting to adhere to conventional design standards. Reactivity when hovering interactable elements is commonplace in design; many users may not notice the change of the cursor icon from a pointer into a hand when hovering such elements. In addition, the LLM added a transformation to the products cards present on the catalogue page — when they're being hovered they move slightly upwards and have a slight shadow added around their borders. This makes it easier for a user to track their cursor's movement within the page. This too is reactivity that was not present in the original design. There's a slight inconsistency in this LLM-improved design, as within the index page only the upwards movement of the card is present, but not the added shadow.

When making changes to the design the LLM frequently injected comments into the HTML and CSS files, which it used for giving reasons for the changes it made.



(a) Data sharing settings page, human-made.



(b) Data sharing settings page, LLM-improved.

Figure 5.4: Data sharing settings pages of both UIs.

The LLM justified most of the changes it made on the grounds of modernity and usability. For instance, using the Roboto font, making the backgrounds lighter, and making elements more rounded were all made with the aim of creating a more modern design. Usability was the most common justification; most of the reactivity added, colour changes, and font size changes were rationalised as improving the usability of the design. For many changes the LLM did not give an explanation as to why it made a change, only commenting on what the intended effect of the change is.

5.3 Differences in performance

We conducted an experiment to determine whether a UI improved by an LLM is more performant than a UI made entirely by a human. The differences in performance between UIs that are improved using LLM and those created entirely by humans are the target of the research questions 1 and 2 presented in Chapter 2. The metric for measuring the efficiency of the UI (RQ1) was chosen to be the time taken to complete both tasks of the test. The metric for user error proneness (RQ2) was chosen to be the number of clicks. Both of these metrics are often used in the context of evaluating the usability of a UI [52], and as such they're well suited for the metrics of our research questions too.

The test setting for the experiment was such that the test subject was given a laptop on which there was a short preamble explaining why this experiment was being conducted, and instructions for the test. After the subject had confirmed that they've read and understood the instructions they could begin the test session. Once the subject had completed the previously described tasks, the test would automatically end. At any point during the test, it was possible for the subject to review what their current task is. It was also possible for the subject to stop the test at any point.

The arithmetic mean of the time taken to complete the test for subjects on the human-made UI was 1 min and 6 s. For subjects on the LLM-improved UI the arithmetic mean to complete the test was 59 s. This discrepancy did not reach statistical significance, the p-value for these means was 0.3264. The LLM did therefore not succeed in improving the original design enough so that it would have performed better time-wise. This is not entirely unsurprising, as the original design wasn't particularly complex. Because of this, the amount of improvement that the LLM could realistically achieve through modifying the HTML and CSS code of the website is quite low.

It could be possible that more significant improvements were able to be achieved if the LLM was capable of more comprehensively altering the design — changing the entire website, rather than individual webpages. Being able to alter the structure of the website, or more sophisticatedly move elements around would also likely see improvement in this front. Duan et al. [10] demonstrate that it is possible to optimise UI layouts using deep learning. They created a technique for automatically optimising layouts of mobile UIs, but it required for them to train their own AI model that predicts task performance, and to create an optimisation algorithm that makes use of the model's predictions. In their study they used a task performance prediction model and an algorithm that iteratively makes changes to the UI. Their method resulted in task performance improvements that were statistically significant. This suggests that while current off-the-shelf LLMs alone might not be capable of optimising designs for the metric of task completion time it is possible to do so with a more purpose-built setup.

The average number of clicks also varied between designs. The arithmetic mean of the number of clicks during a single test session for subjects on the human-made design was 15.4 clicks. For subjects on the LLM-improved design the arithmetic mean was 12.5 clicks. This discrepancy did reach statistical significance with a p-value of 0.0491. Therefore, the LLM improved the original design in such a way that it was less prone to user errors. As previously mentioned, the LLM justified many of the changes it made to improve the usability of the design. Notably, the changes made it easier to distinguish between different elements of the webpages, which could have resulted in fewer missed clicks by the test subjects. These aspects of the design were likely easier to change, given that the LLM could only modify the HTML and CSS code of the webpages.

To pinpoint any particular areas in which a change in clicking behaviour could possibly be seen, heatmaps were constructed by plotting all the recorded clicks onto their respective webpages. For constructing the heatmaps each webpage was split into a two-dimensional array of cells for grouping the clicks into discrete areas. The colour of each cell was determined by the number of clicks recorded within it.

The clearest difference between the two designs in regard to where the clicks took place was within the catalogue page (Figure 5.6), which is also the page that received the most clicks during any given test session. The areas of interest on both designs are clearly the different filters on the designs' left sides. With the original design, the distribution of the clicks is fairly uniform on the filters, but with the LLM-improved design only the "form" category remains as intense. This suggests that test subjects made fewer errors when selecting the filters required for the task, which were "shoes, of size L, that are made of faux leather". The "form" category remaining a hotspot also suggests that test subjects did often, in both designs, mistakenly select one of the options in that category even though it wasn't instructed, which they then unselected, requiring an additional click.

Aside from the catalogue page the differences in hotspots for clicks were not as noticeable. In the index page (Figure 5.5) test subjects often tried clicking the navigation links within the page's header. Other than the "catalogue" and "account" links these were not implemented and didn't work. Those were also the only links

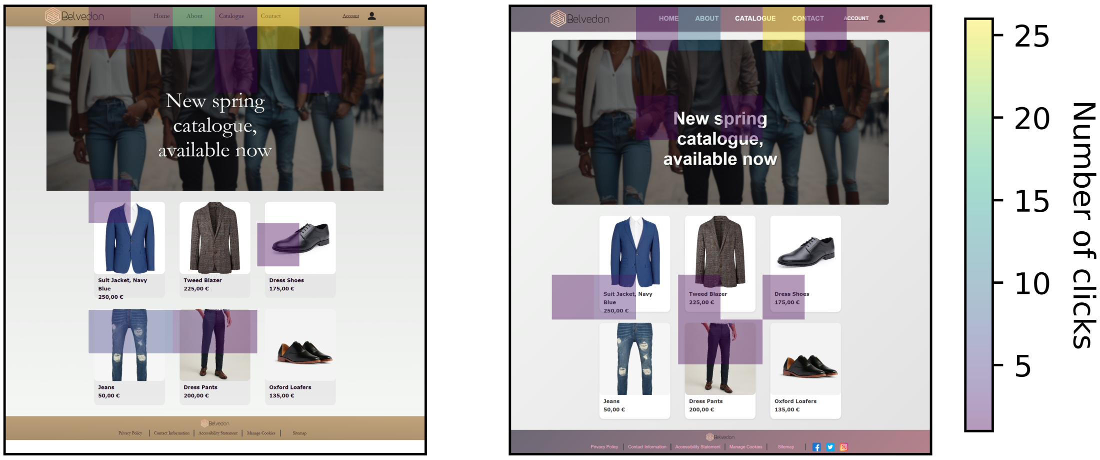


Figure 5.5: Index page heatmaps of both UIs.

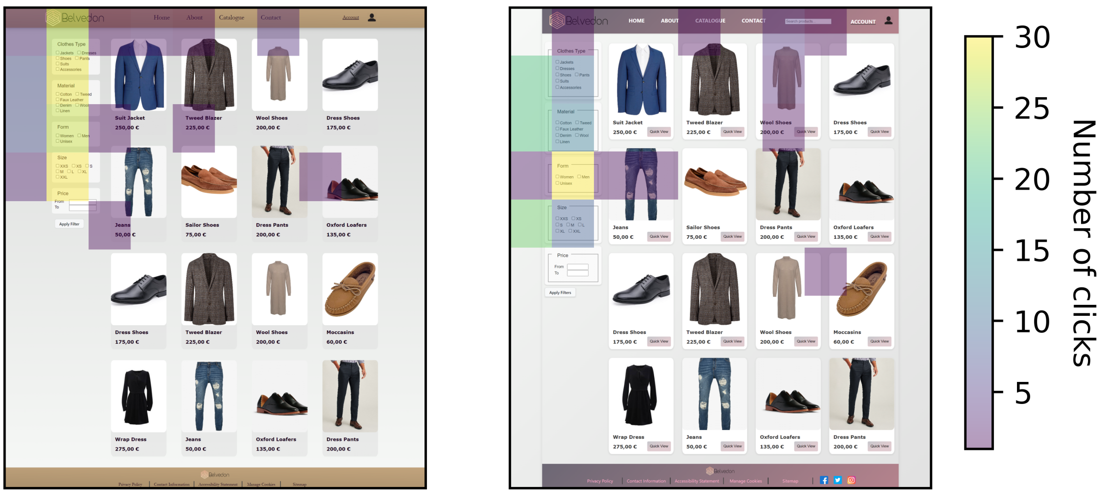


Figure 5.6: Catalogue page heatmaps of both UIs.

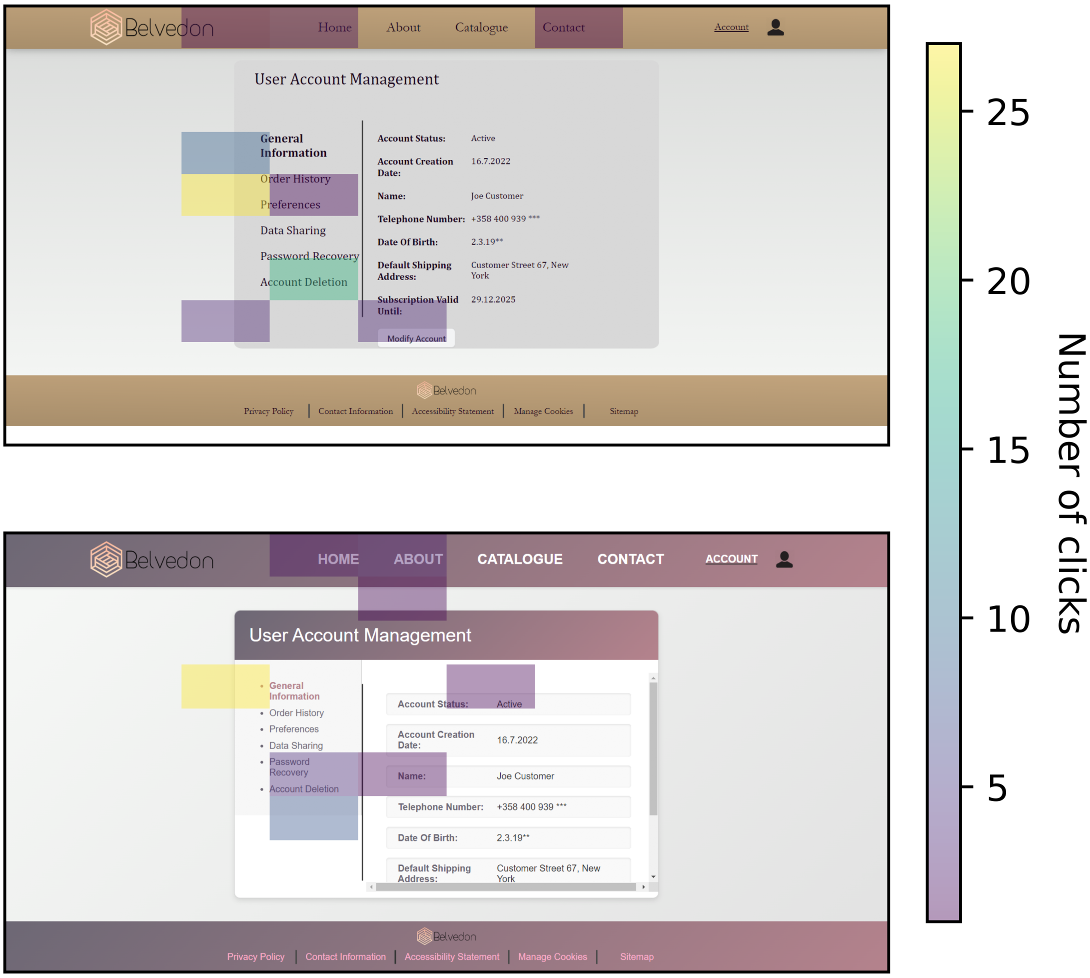


Figure 5.7: Account management page heatmaps of both UIs.

that could lead the test subjects closer to the completion of either task. The "about" and "contact" links were often clicked, which was unexpected as the first task during the test is "change your account's data sharing settings", and neither of these links would lead to a settings page in a typical website. It could be then that this clicking of links in the navigation bar is just natural curiosity, as it wasn't repeated in the other webpages and only happened at the beginning of the test session.

The differences in the concentration of clicks between both designs in the account management page (Figure 5.7) and the data sharing page (Figure 5.8) were quite subdued. The account management page is the first page that opens when the test subject clicks on the "account" link in the navigation bar. After that, the test subject needs to select the "data sharing" category from the left side of the panel in order to change those settings, which is the goal of the first task. There was some difference in the concentration of clicks on the account management page, specifically in the settings categories on the left side of the panel. In the original design there were more clicks around this area, which indicates that the test subjects may have had some difficulties selecting the correct category. This seems plausible, as the fact that the items were categories wasn't made all that clear in the original design, and they may seem like a random assortment of individual settings or just text instead. In the LLM-improved version they were stylised as a bullet point list, which may better communicate that they are interactable elements. That, along with the addition of a highlight colour indicating which category is currently selected, possibly made the webpage more understandable for the test subject, which resulted in fewer unnecessary clicks.

The data sharing page also had little difference between the two designs in their concentrations of clicks. In both designs the test subjects attempted to click the different settings categories in the left side of the panel, even though the data sharing page was already opened and contained the button that they could use to complete the task. It may be that it wasn't clear enough after they had clicked the "data sharing" category in the account management page that the transition to the correct webpage had already happened. Both the account management page and the data sharing page are very similar, so perhaps the bolding on the current category and the highlight colour on the LLM-improved design weren't enough to distinguish these two pages from each other. As a result, this may have then resulted in the test subjects attempting to open other categories to see if they would work instead.

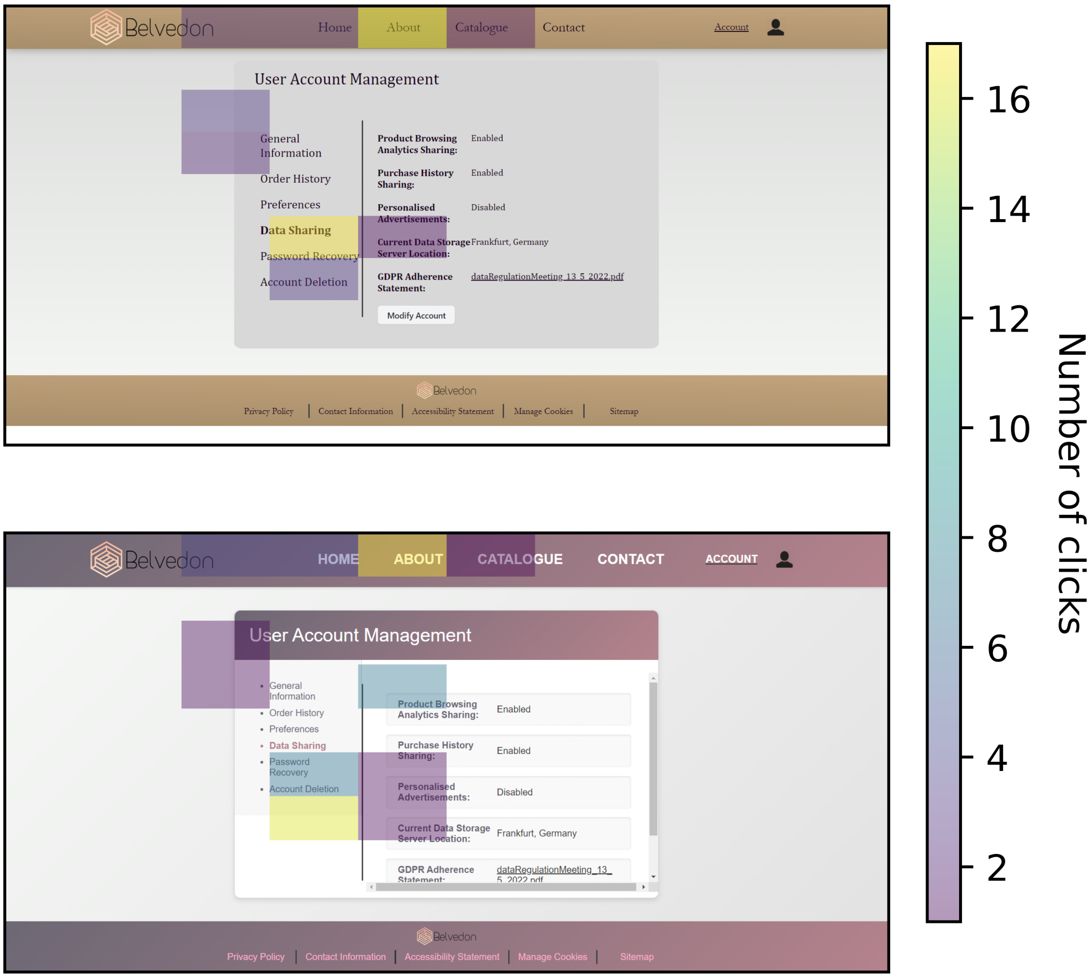


Figure 5.8: Data sharing settings page heatmaps of both UIs.

5.4 Differences in aesthetics

For the purpose of answering the research question 3 both designs needed to be evaluated with regard to their aesthetic appeal. To achieve this a questionnaire consisting of five questions was created. These questions weren't asked on a per webpage basis, but rather for the entire design. The original, human-made design was called "Design A", and the LLM-improved design was called "Design B". At the beginning of the questionnaire the respondent was shown both designs. After this the respondent was given two sets of questions. The first set of questions was themed around the aesthetic appeal of the designs:

1. Which website design uses colours in a more pleasing way?
2. Which website design has a more pleasing layout of elements?
3. Which website design looks more pleasing generally?

The second set of questions was themed around the visual clarity of the designs:

1. In which website design is it easier to distinguish between different parts of the page?
2. In which website design is it easier to distinguish the interactable elements of the page?

The results of the first set of questions (Figure 5.9) show that the respondents thought that the LLM-improved design used colours more pleasingly, had a more pleasing layout of elements, and looked more pleasing generally as well. The difference in how well the designs used colours wasn't as noticeable as in the latter two questions. This may be because the only big difference in the colour scheme of the two designs is in the header and the footer. Here the LLM changed the very subtle brown to light brown gradient to a more divergent light purple to light pink-red.

Other colours were changed as well, but they underwent changes in tone, rather than to entirely different colours. This lack of large change in the colours may be why respondents didn't feel as strongly about the differences between the designs.

The designs' layout of their elements received a stronger difference in opinion. The LLM-improved design was clearly preferred regarding this attribute. The catalogue page provides the largest difference in respect to the layout of elements. Here, the LLM-improved design has noticeably less "wasted" space, and it groups elements together more tightly. The product cards for instance have less empty space between them in the LLM-improved design, due to them being larger than in the original design. This increase in size isn't purely about aesthetics, as it enables the cards to show more of the product's preview image. The filter categories in the left side of the page are also larger than in the original design. While each individual filter category is less compact in the LLM-improved design, their overall grouping appears tighter. This is, again, due to the LLM reducing the empty space between each element just as it did with the product cards. Here the size increase is only vertical, which doesn't result in them taking space from other elements of the page, as the space beneath them was unused.

The account management and the data sharing pages also had some changes in their layouts. The settings categories on the left side of the user account management panel are clearly more tightly grouped. In the original design doing so would have been impossible without making it difficult to distinguish between each individual cateogory, but by accompanying the categories with bullet points it was possible to move them closer to each other without this happening. The information on the right side of the panel was also made clearer by adding an indent around each individual piece of information.

The final question for the first set wasn't focused on any specific attribute of the designs, but rather on the general aesthetic appeal. With this question, the preference among respondents was clearly for the LLM-improved design. No specific change by the LLM is likely the sole cause for this, but instead the combined effect of individual changes. The colour scheme of the website is more modern in the LLM-improved design, which might have been more appealing to the respondents: the design separates its elements more clearly, which results in a cleaner look, and the design wastes less space in its layout. All of these changes, and likely other less noticeable ones, affect the design's general aesthetic appeal. As stated earlier, the LLM often justified the changes that it made to the original design on the basis of modernising the design. With that and the clear preference for the LLM-improved design by the respondents, it seems that a more modern design is more aesthetically pleasing, and that an LLM can enact changes onto an existing design that makes it more modern.

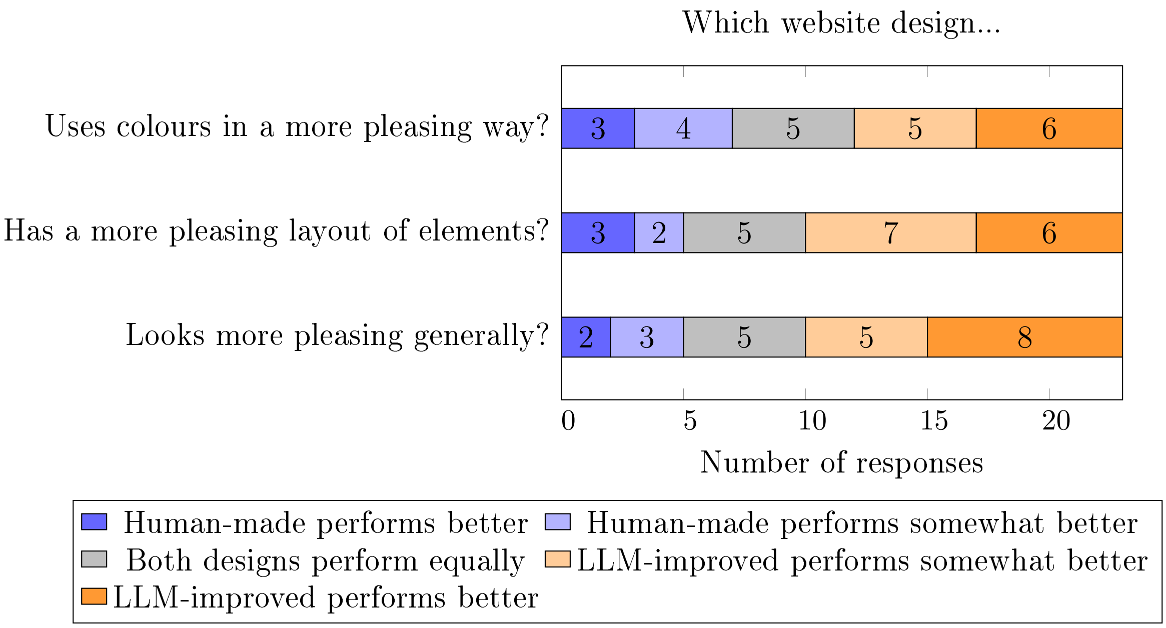


Figure 5.9: Responses for the first set of questions.

The second set of questions (Figure 5.10) focused on the visual clarity of the designs. The first question asked the respondents which design makes it easier to distinguish between different parts of their webpages. The preference was clearly for the LLM-improved design. This is highlighted well in the account management and data sharing pages. The LLM improved visual clarity by separating the settings categories on the left side of the panel more clearly by introducing a bullet point list with each category serving as a single item in the list. In addition, the information on the right side of the panel is more clearly separated as well by the addition of an indent for each individual piece of information. The title of the panel is clearly separated in the LLM-improved design by giving it a background colour gradient of the same colour as in the header and the footer.

In the catalogue page the LLM-improved design improved visual clarity of the product cards by adding a border and a shadow around them, adding an element of three-dimensionality that causes them to stand out from the background. In the original design such borders didn't exist, and the product cards blended into the background if the colour of the image within the card was too similar to the background's. A similar addition was made to the contents of the webpage; in the LLM-improved design the catalogue page was limited to take only as much space as it needed, which meant empty space on the left and right side of the webpage. To separate the actual content from the empty space a border and a shadow were added here as well. This empty space is likely not purely for aesthetics either — such spaces are often used to add a background image to the webpage in many e-commerce websites.

The second question was in which design it is easier to distinguish the interactable elements of the designs from the static elements. Compared to the previous question, here more respondents felt that the designs were equal, though the preference was still in favour of the LLM-improved design. The interactable elements were quite few in each webpage, with the majority being within the catalogue page. The addition of the search bar and "quick view" buttons may have also led the respondents to consider the LLM-improved design better in this regard, even though simply having more interactable elements wouldn't necessarily make them easier to distinguish.

The product cards themselves are interactable elements and are also clearly easier to distinguish in the LLM-improved design as previously discussed.

The navigation bar is present in every webpage and contains interactable elements that may be clearer in the LLM-improved design. As the font colour was changed from black to white, it may be easier to distinguish them as clickable links, since they contrast against the rest of the black text within the designs, which is primarily used for non-interactable elements.

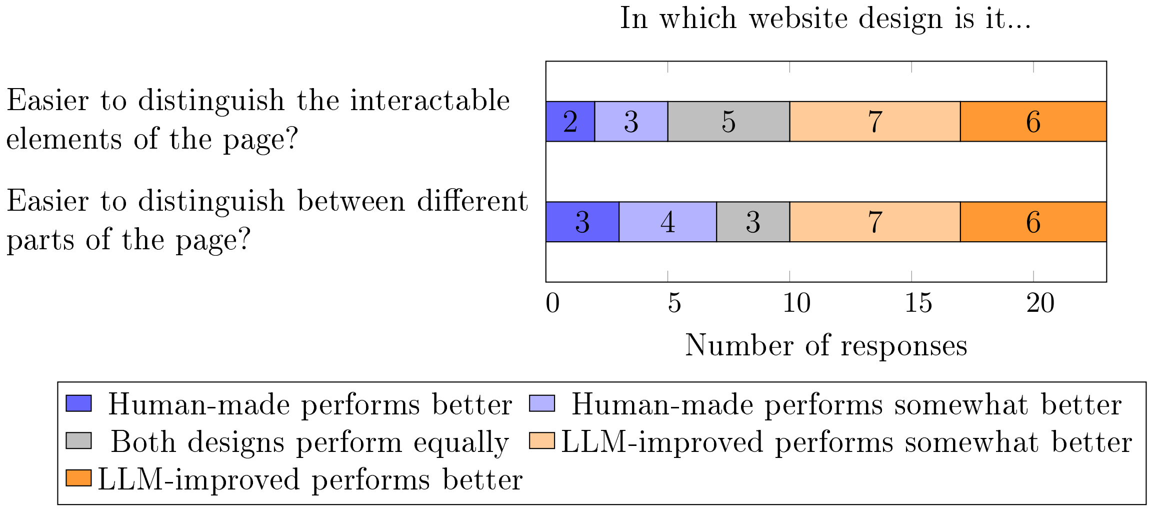


Figure 5.10: Responses for the second set of questions.

The results of both sets of questions suggest that a user interface improved by an LLM is indeed more aesthetically pleasing. Whether the improvements to aesthetics would have been as succesful if the LLM had been instructed to change the design according to a certain style, rather than just to enact general improvements, is unclear. It could be that the LLM has an affinity for modern designs, as its training data likely had a proportionally large amount of such designs. If instructing it to make improvements while still preserving an elegant style, in accordance with what the human-made design originally was, the results could differ.

6 Conclusions

The role of user experience in improving the usability and interactivity of a product is critical [7]. User expectations for the quality of UX have also increased over time, and designing UX that meets these increasing demands is a time-consuming process [4]. Due to the labour cost of this process the desire for making it more efficient is high [9]. To this end, artificial intelligence has great potential, and machine learning technologies in particular are becoming increasingly popular for improving the quality of UX [4], [7].

In this thesis a literature review of current practices regarding the usage of AI for optimising UX design was performed, and an experiment was conducted to measure AI's potential benefits in these kinds of uses. In the experiment, a mockup e-commerce clothes store was designed and implemented. Based on this design a separate LLM-optimised design was created using an off-the-shelf LLM. Differences between these two designs in task completion time and user error rate were measured, and aesthetic quality was assessed. The experiment demonstrated how an off-the-shelf LLM can be utilised for the purpose of optimising an existing UI with significant improvements to user error rate and aesthetic quality. The LLM's output did sometimes need minor refinement to get it working as intended. This is in line with many researchers' predictions that collaboration between human and AI will be the optimal way of working with AI [1], [4], [10]

The amount of improvement that one could see for a given UI likely depends heavily on the level of quality of the original UI. Here the original design wasn't particularly advanced and had some easy mistakes to fix, which the LLM mostly performed. It's probable that given a more expertly crafted UI the improvements that an LLM could achieve would be more limited. The designers creating a more advanced UI would also likely rather use AI for a more assistive role [1]. Therefore, it may be that experts wouldn't even wish to use AI to optimise their designs, and would rather use it as an ideation tool and implement the suggested optimisations themselves.

Aside from having fewer easy fixes for an LLM to implement, a more advanced UI could also pose challenges for an LLM because of its complexity. The LLM used here was already prone to sometimes making errors that resulted in its output being unusable or not working as intended without fixing it first. This would likely become more common if the HTML and CSS used were complex. For the purposes of AI optimisation, HTML and CSS have the advantage of being quite robust; if there is a line with syntax errors the browser will just ignore said line and continue onto the next line. JavaScript doesn't have this luxury as it's a scripting language and will instead just crash if it encounters such an error. JavaScript is also utilised to some extent by most modern websites, and is often dependent on HTML elements within the website. As LLMs have a tendency to hallucinate it's possible that AI optimisation would break these dependencies between the JavaScript code and the HTML elements of the website.

LLMs by themselves are not sophisticated enough tools that one can reliably use to generate a functional UI from scratch. However, they can be a good tool for people with no access to purpose-built AI tools for their specific needs. LLMs can be useful for those seeking to improve an already existing UI, and who have a sufficient level of technical knowledge so that they can tweak the LLM's output to fix their common mistakes.

6.1 Limitations and further research

A limitation of this thesis is that only a single UI was created, improved, and tested against. Creating additional UIs that are similarly improved by an LLM and compared against the original design would have improved the reliability of the results and enabled making more informed conclusions. In addition, the testing methods used for testing the UIs and assessing their aesthetic quality imposed limitations on the allowed complexity and length of the tests. Test subjects that have in advance agreed to participate in the study would allow for more robust test sessions with more complicated UIs.

Further research into how well off-the-shelf LLMs perform when attempting to emulate specific design styles, or improving specific aspects of the UI could prove insightful, rather than general optimisations for the entire design. Another interesting research avenue would be to test how succesfully an LLM performs, when it has context understanding of the entire website (HTML, CSS and JavaScript), rather than just the part of the website that it's currently improving. Finally, testing UIs of different qualities would give insight as to how well an LLM performs when the initial quality of the UI being optimised is better.

The challenge of effectively integrating AI into the design process is not trivial. A notable difficulty is enabling designers to make effective use of AI tools [1], [53]. Due to a lack of technical knowledge many UX practicioners face challenges in understanding the capabilities and limitations of machine learning [7], [15]. Issues with ethicality regarding the use of AI technologies may also be an obstacle, as they are often perceived as stealing intellectual property [53]. Finally, most designers see AI systems as being unreliable, and many don't trust the quality of these systems' outputs [6].

Users' demands for the quality of digital products will continue to rise [22], which will in turn require improvements in the tools used to create said products. This will drive further development in utilising AI in the design process, due to AI's potential that is already recognised by researchers and practicioners alike [7], [22]. With every advancement to AI its profound impact on many fields will continue to increase, and UX design is not an exception.

References

[1] J. Li, H. Cao, L. Lin, Y. Hou, R. Zhu, and A. El Ali, "User Experience Design Professionals' Perceptions of Generative Artificial Intelligence", en, in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, May 2024, pp. 1–18, ISBN: 9798400703300. DOI: 10.1145/3613904.3642114. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3613904.3642114.

[2] L.-Y. Chiou, P.-K. Hung, R.-H. Liang, and C.-T. Wang, "Designing with AI: An Exploration of Co-Ideation with Image Generators", en, in *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, Pittsburgh PA USA: ACM, Jul. 2023, pp. 1941–1954, ISBN: 978-1-4503-9893-0. DOI: 10.1145/3563657.3596001. Accessed: Nov. 11, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3563657.3596001.

[3] Q. Yang, A. Steinfeld, C. Rosé, and J. Zimmerman, "Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design", en, in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–13, ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376301. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3313831.3376301.

[4] Å. Stige, E. D. Zamani, P. Mikalef, and Y. Zhu, "Artificial intelligence (AI) for user experience (UX) design: A systematic literature review and future re-

search agenda", en, *Information Technology & People*, vol. 37, no. 6, pp. 2324–2352, Sep. 2024, ISSN: 0959-3845. DOI: 10.1108/ITP-07-2022-0519. Accessed: Nov. 6, 2024. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/ITP-07-2022-0519/full/html.

[5] J. Tholander and M. Jonsson, "Design Ideation with AI - Sketching, Thinking and Talking with Generative Machine Learning Models", en, in *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, Pittsburgh PA USA: ACM, Jul. 2023, pp. 1930–1940, ISBN: 978-1-4503-9893-0. DOI: 10.1145/3563657.3596014. Accessed: Nov. 11, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3563657.3596014.

[6] T. Knearem, M. Khwaja, Y. Gao, F. Bentley, and C. E. Kliman-Silver, "Exploring the future of design tooling: The role of artificial intelligence in tools for user experience professionals", en, in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, Hamburg Germany: ACM, Apr. 2023, pp. 1–6, ISBN: 978-1-4503-9422-2. DOI: 10.1145/3544549.3573874. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3544549.3573874.

[7] A. M. H. Abbas, K. I. Ghauth, and C.-Y. Ting, "User Experience Design Using Machine Learning: A Systematic Review", en, *IEEE Access*, vol. 10, pp. 51 501–51 514, 2022, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2022.3173289. Accessed: Nov. 11, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/9770797/.

[8] T. Beltramelli, "Pix2code: Generating Code from a Graphical User Interface Screenshot", en, in *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, Paris France: ACM, Jun. 2018, pp. 1–6, ISBN: 978-1-4503-5897-2. DOI: 10.1145/3220134.3220135. Accessed: Nov. 6,

2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3220134.3220135.

[9] B. Asiroglu *et al.*, "Automatic HTML Code Generation from Mock-Up Images Using Machine Learning Techniques", en, in *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, Istanbul, Turkey: IEEE, Apr. 2019, pp. 1–4, ISBN: 978-1-72811-013-4. DOI: 10.1109/EBBT.2019.8741736. Accessed: Nov. 6, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/8741736/.

[10] P. Duan, C. Wierzynski, and L. Nachman, "Optimizing User Interface Layouts via Gradient Descent", en, in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–12, ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376589. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3313831.3376589.

[11] S. Natarajan and C. Csallner, "P2A: A tool for converting pixels to animated mobile application user interfaces", en, in *Proceedings of the 5th International Conference on Mobile Software Engineering and Systems*, Gothenburg Sweden: ACM, May 2018, pp. 224–235, ISBN: 978-1-4503-5712-8. DOI: 10.1145/3197231.3197249. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3197231.3197249.

[12] P. Domingos, "A few useful things to know about machine learning", en, *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, Oct. 2012, ISSN: 0001-0782, 1557-7317. DOI: 10.1145/2347736.2347755. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/2347736.2347755.

[13] H. Zhang, P. Chen, X. Xie, Z. Jiang, Z. Zhou, and L. Sun, "A Hybrid Prototype Method Combining Physical Models and Generative Artificial Intelli-

gence to Support Creativity in Conceptual Design", en, *ACM Transactions on Computer-Human Interaction*, p. 3 689 433, Sep. 2024, ISSN: 1073-0516, 1557-7325. DOI: 10.1145/3689433. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3689433.

[14] T. Song, X. Li, B. Wang, and L. Han, "Research on Intelligent Application Design Based on Artificial Intelligence and Adaptive Interface", en, *Journal of Social Science and Humanities*, Apr. 2024. DOI: 10.53469/wjimt.2024.07(02).01. Accessed: Nov. 6, 2024. [Online]. Available: https://adwenpub.com/index.php/wjimt/article/view/234.

[15] Q. Yang, A. Scuito, J. Zimmerman, J. Forlizzi, and A. Steinfeld, "Investigating How Experienced UX Designers Effectively Work with Machine Learning", en, in *Proceedings of the 2018 Designing Interactive Systems Conference*, Hong Kong China: ACM, Jun. 2018, pp. 585–596, ISBN: 978-1-4503-5198-0. DOI: 10.1145/3196709.3196730. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3196709.3196730.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", en, *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature14539. Accessed: Nov. 6, 2024. [Online]. Available: https://www.nature.com/articles/nature14539.

[17] C. Kliman-Silver, O. Siy, K. Awadalla, A. Lentz, G. Convertino, and E. Churchill, "Adapting User Experience Research Methods for AI-Driven Experiences", en, in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–8, ISBN: 978-1-4503-6819-3. DOI: 10.1145/3334480.3375231. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3334480.3375231.

[18] A. Chander, J. Wang, R. Srinivasan, K. Uchino, and S. Chelian, "Working with Beliefs: AI Transparency in the Enterprise", en, Mar. 2019.

[19] S. Amershi *et al.*, "Guidelines for human-ai interaction", in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19, Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–13, ISBN: 9781450359702. DOI: 10.1145/3290605.3300233. [Online]. Available: https://doi.org/10.1145/3290605.3300233.

[20] [Accessed 13-2-2025], 2019. [Online]. Available: https://pair.withgoogle.com/.

[21] E. Kuang, "Crafting Human-AI Collaborative Analysis for User Experience Evaluation", en, in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, Hamburg Germany: ACM, Apr. 2023, pp. 1–6, ISBN: 978-1-4503-9422-2. DOI: 10.1145/3544549.3577042. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3544549.3577042.

[22] X. Li, H. Zheng, J. Chen, Y. Zong, and L. Yu, "User Interaction Interface Design and Innovation Based on Artificial Intelligence Technology", en, *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 03, pp. 1–8, Mar. 2024, ISSN: 2790-1505. DOI: 10.53469/jtpes.2024.04(03).01. Accessed: Nov. 6, 2024. [Online]. Available: https://centuryscipub.com/index.php/jtpes/article/view/507.

[23] *Twitter's Recommendation Algorithm*, https://blog.x.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm, [Accessed 13-11-2024].

[24] J. Hoffmann *et al.*, "Training compute-optimal large language models", in *Proceedings of the 36th International Conference on Neural Information Process-*

ing Systems, ser. NIPS '22, New Orleans, LA, USA: Curran Associates Inc., 2022, ISBN: 9781713871088.

[25] Y. Yu *et al.*, "Large language model as attributed training data generator: A tale of diversity and bias", vol. 36, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., pp. 55 734–55 784, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/ae9500c4f5607caf2eff033c67daa9d7-Paper-Datasets_and_Benchmarks.pdf.

[26] K. Lu, A. Grover, P. Abbeel, and I. Mordatch, "Frozen pretrained transformers as universal computation engines", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 7, pp. 7628–7636, Jun. 2022. DOI: 10.1609/aaai.v36i7.20729. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/20729.

[27] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving Language Understanding by Generative Pre-Training", en,

[28] T. B. Brown *et al.*, *Language Models are Few-Shot Learners*, en, arXiv:2005.14165 [cs], Jul. 2020. DOI: 10.48550/arXiv.2005.14165. Accessed: Nov. 26, 2024. [Online]. Available: http://arxiv.org/abs/2005.14165.

[29] [Accessed 3-3-2025], 2025. [Online]. Available: https://openai.com/index/introducing-gpt-4-5/.

[30] G. Team *et al.*, *Gemma 2: Improving open language models at a practical size*, 2024. arXiv: 2408.00118 [cs.CL]. [Online]. Available: https://arxiv.org/abs/2408.00118.

[31] H. Touvron *et al.*, *Llama: Open and efficient foundation language models*, 2023. arXiv: 2302.13971 [cs.CL]. [Online]. Available: https://arxiv.org/abs/2302.13971.

[32] N. Bostrom and E. Yudkowsky, "The ethics of artificial intelligence", en, in *The Cambridge Handbook of Artificial Intelligence*, K. Frankish and W. M. Ramsey, Eds., 1st ed., Cambridge University Press, Jun. 2014, pp. 316–334, ISBN: 978-0-521-87142-6 978-0-521-69191-8 978-1-139-04685-5. DOI: 10.1017/CBO9781139046855.020. Accessed: Nov. 6, 2024. [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9781139046855A027/type/book_part.

[33] H. Zhong *et al.*, "Copyright Protection and Accountability of Generative AI: Attack, Watermarking and Attribution", en, in *Companion Proceedings of the ACM Web Conference 2023*, Austin TX USA: ACM, Apr. 2023, pp. 94–98, ISBN: 978-1-4503-9419-2. DOI: 10.1145/3543873.3587321. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3543873.3587321.

[34] B. Brittain, *Getty Images lawsuit says Stability AI misused photos to train AI*, https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/, [Accessed 25-11-2024].

[35] B. Brittain, *Lawsuits accuse AI content creators of misusing copyrighted work*, https://www.reuters.com/legal/transactional/lawsuits-accuse-ai-content-creators-misusing-copyrighted-work-2023-01-17/, [Accessed 25-11-2024].

[36] J. Weatherbed, *Artstation is hiding images protesting ai art on the platform*, https://www.theverge.com/2022/12/23/23523864/artstation-removing-anti-ai-protest-artwork-censorship, [Accessed 20-01-2025].

[37] A. Tong and E. Wang, *Exclusive: Reddit in ai content licensing deal with google*, https://www.reuters.com/technology/reddit-ai-content-licensing-deal-with-google-sources-say-2024-02-22/, [Accessed 23-01-2025], 2024.

[38] *Adobe Firefly - Free Generative AI for creatives — adobe.com*, `https://www.adobe.com/products/firefly.html`, [Accessed 25-11-2024].

[39] https://www.nytimes.com/by/ryan-mac, *The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work — nytimes.com*, `https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html`, [Accessed 26-11-2024].

[40] *The Authors Guild, John Grisham, Jodi Picoult, David Baldacci, George R.R. Martin, and 13 Other Authors File Class-Action Suit Against OpenAI - The Authors Guild — authorsguild.org*, `https://authorsguild.org/news/ag-and-authors-file-class-action-suit-against-openai/`, [Accessed 26-11-2024].

[41] J. D. Lomas *et al.*, "Interface Design Optimization as a Multi-Armed Bandit Problem", en, in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose California USA: ACM, May 2016, pp. 4142–4153, ISBN: 978-1-4503-3362-7. DOI: `10.1145/2858036.2858425`. Accessed: Nov. 6, 2024. [Online]. Available: `https://dl.acm.org/doi/10.1145/2858036.2858425`.

[42] K. Todi, D. Weir, and A. Oulasvirta, "Sketchplore: Sketch and Explore with a Layout Optimiser", en, in *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, Brisbane QLD Australia: ACM, Jun. 2016, pp. 543–555, ISBN: 978-1-4503-4031-1. DOI: `10.1145/2901790.2901817`. Accessed: Nov. 6, 2024. [Online]. Available: `https://dl.acm.org/doi/10.1145/2901790.2901817`.

[43] X. Jin, M. Evans, H. Dong, and A. Yao, "Design Heuristics for Artificial Intelligence: Inspirational Design Stimuli for Supporting UX Designers in Generating AI-Powered Ideas", en, in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, May 2021, pp. 1–8, ISBN: 978-1-4503-8095-9. DOI: 10.1145/3411763.3451727. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3411763.3451727.

[44] G. Bailly, A. Oulasvirta, T. Kötzing, and S. Hoppe, "MenuOptimizer: Interactive optimization of menu systems", en, in *Proceedings of the 26th annual ACM symposium on User interface software and technology*, St. Andrews Scotland, United Kingdom: ACM, Oct. 2013, pp. 331–342, ISBN: 978-1-4503-2268-3. DOI: 10.1145/2501988.2502024. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/2501988.2502024.

[45] G. Bailly, A. Oulasvirta, D. P. Brumby, and A. Howes, "Model of visual search and selection time in linear menus", en, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto Ontario Canada: ACM, Apr. 2014, pp. 3865–3874, ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557093. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/2556288.2557093.

[46] J. C. Quiroz, S. J. Louis, A. Shankar, and S. M. Dascalu, "Interactive Genetic Algorithms for User Interface Design", en, in *2007 IEEE Congress on Evolutionary Computation*, Singapore: IEEE, Sep. 2007, pp. 1366–1373, ISBN: 978-1-4244-1339-3 978-1-4244-1340-9. DOI: 10.1109/CEC.2007.4424630. Accessed: Nov. 6, 2024. [Online]. Available: http://ieeexplore.ieee.org/document/4424630/.

[47] K. Gajos and D. S. Weld, "SUPPLE: Automatically Generating User Interfaces", en, in *Proceedings of the 9th International Conference on Intelligent*

User Interfaces, ser. IUI '04, Funchal, Madeira, Portugal: ACM, 2004, pp. 93–100. DOI: 10.1145/964442.964461. [Online]. Available: https://doi.org/10.1145/964442.964461.

[48] P. O'Donovan, A. Agarwala, and A. Hertzmann, "DesignScape: Design with Interactive Layout Suggestions", en, in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul Republic of Korea: ACM, Apr. 2015, pp. 1221–1224, ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702149. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/2702123.2702149.

[49] N. Bostrom, "Ethical Issues in Advanced Artificial Intelligence", en, in *Machine Ethics and Robot Ethics*, W. Wallach and P. Asaro, Eds., 1st ed., Routledge, Sep. 2020, pp. 69–75, ISBN: 978-1-00-307499-1. DOI: 10.4324/9781003074991-7. Accessed: Nov. 6, 2024. [Online]. Available: https://www.taylorfrancis.com/books/9781000108934/chapters/10.4324/9781003074991-7.

[50] M. Soegaard, *The power of white space in design*, [Accessed 19-3-2025], 2021. [Online]. Available: https://www.interaction-design.org/literature/article/the-power-of-white-space/.

[51] C. K. Coursaris and K. Kripintris, "Web Aesthetics and Usability: An Empirical Study of the Effects of White Space", en, *International Journal of E-Business Research*, vol. 8, no. 1, pp. 35–53, Jan. 2012, ISSN: 1548-1131, 1548-114X. DOI: 10.4018/jebr.2012010103. Accessed: Mar. 19, 2025. [Online]. Available: https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/jebr.2012010103.

[52] J. Nielsen, *Usability metrics*, [Accessed 18-3-2025], 2001. [Online]. Available: https://www.nngroup.com/articles/usability-metrics/.

[53] L. Hegemann, Y. Jiang, J. G. Shin, Y.-C. Liao, M. Laine, and A. Oulasvirta, "Computational Assistance for User Interface Design: Smarter Generation and Evaluation of Design Ideas", en, in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, Hamburg Germany: ACM, Apr. 2023, pp. 1–5, ISBN: 978-1-4503-9422-2. DOI: 10.1145/3544549.3583960. Accessed: Nov. 6, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3544549.3583960.