From Text to Card: Transforming Design Implications with Generative AI

DONGHOON SHIN, University of Washington, USA

LUCY LU WANG, University of Washington, USA

GARY HSIEH, University of Washington, USA

Communicating design implications is common within the HCI community when publishing academic papers, yet these papers are rarely read and used by designers. One solution is to use design cards as a form of translational resource that communicates valuable insights from papers in a more digestible and accessible format to assist in design processes. However, creating design cards can be time-consuming, and authors may lack the resources/know-how to produce cards. Through an iterative design process, we built a system that helps create design cards from academic papers using an LLM and text-to-image model. Our evaluation with designers ($N = 21$) and authors of selected papers ($N = 12$) revealed that designers perceived the design implications from our design cards as more inspiring and generative, compared to reading original paper texts, and the authors viewed our system as an effective way of communicating their design implications. We also propose future enhancements for AI-generated design cards.

CCS Concepts: • **Human-centered computing** → **Systems and tools for interaction design**; **HCI design and evaluation methods**.

Additional Key Words and Phrases: design card, translational science, generative AI

1 INTRODUCTION

One of the key goals for research in an applied field such as human-computer interaction (HCI) is to generate insights that can directly impact practice—in our case, design practice. Many authors attempt to achieve this through the *design implication* section in an academic paper. However, despite best efforts in communicating these insights, studies have found that design implications, as presented in academic papers, are rarely consumed and used by design practitioners [21]. Designers describe difficulties reading the often jargon-filled papers and not finding academic papers useful [21].

This has led to several calls to develop better translational resources to bridge the gap between research and practice [21, 23]. Instead of relying on academic papers to communicate valuable research findings to design practitioners, what may be needed are other types of resources to support this translation. One approach is to distill the insights into *design cards*. Design cards are tools that facilitate the design process, and can contain several types of design knowledge—from human insights to material & domain knowledge [34]. There is a long tradition of creating design cards to communicate research insights [4, 19].

However, while design cards may be more accessible and provide more prescriptive insights to practitioners, one challenge with relying on using design cards to communicate academic findings is how these cards should be generated. It is unrealistic to expect researchers to do it, as HCI researchers describe the lack of time and competing responsibilities that prevent them from doing more outreach work [72]. They also lack the proper incentive to do this type of translational work [23]. Further, even if researchers are motivated, they may have limited understanding of what designers are seeking in design implications, and lack the skills needed to generate these cards so that they are inspiring [62]. A potential solution may be to leverage crowdwork—ideally crowds of designers—to perform this translational work; this approach has been explored for the creation of videos for research talks [70]. Nonetheless, changes in incentives are required for long-term sustainability. Furthermore, reliance on volunteer contributors can lead to spotty coverage and a biased focus on certain topics over others [26].

Title — Reducing visual complexity for improved search efficiency.

Description — Designers should strive to reduce the visual complexity of their websites to a minimum, while taking into account user preferences, in order to improve search efficiency and information recall for the users.

Image

Paper summary

About this paper

The authors conducted an online study to investigate the impact of visual complexity on search efficiency and information recall.

The results showed that the search efficiency of those who preferred simple websites was more negatively affected by highly complex websites than those who preferred high visual complexity.

Evidence

Which part of the paper did the design guideline come from?

"Our analysis revealed that website complexity has a significant negative effect on search efficiency, confirming our first hypothesis (see Table 2). A linear mixed-model ANOVA confirmed the significant effect of complexity on search response time ($F_{(2,8.99)}$ = 4.44, $p < .05$). Participants were fastest on simple websites (M = 4.2s, SD = 4.7s) and slowest on highly complex websites (M = 7.3s, SD = 7.0s), regardless of their preferred complexity level (...)" (Section 4)

Citation

Baughan, A., August, T., Yamashita, N., & Reinecke, K. (2020). Keep it Simple: How Visual Complexity and Preferences Impact Search Efficiency on Websites. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.
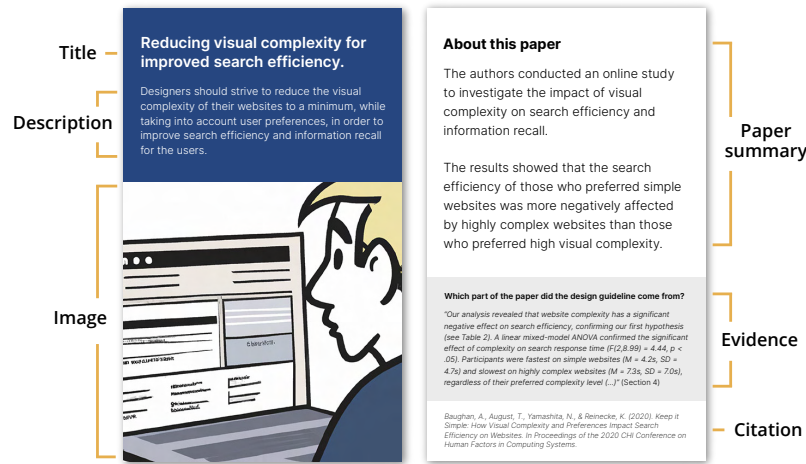
Fig. 1. A design card built on Baughan *et al.* (2020) [8] using our system

In this paper, we propose the use of generative AI models to support this important translational work. Generative AI models, such as large language models (LLMs) and text-to-image models, have recently shown remarkable capabilities in producing high-fidelity texts and images, thus being adapted for various creative tasks [20, 40, 66]. Motivated by this line of research, we believe generative AI models could provide a scalable and efficient way of translating academic findings into a prescriptive format such as design cards, without requiring significant time or effort from researchers and designers.

To test our hypothesis, we develop and evaluate an end-to-end system to automatically generate design cards from HCI publications. We first ran a preliminary interview study with design practitioners ($N = 12$) to inform the design of design cards built on academic papers. We presented participants with mock-ups of design cards that we created from three HCI papers, each with several design elements and alternate formats. Based on our findings, we iterated on our design and built a system that automatically generates a design card from the design implication text extracted from a paper using LLMs (*i.e.*, GPT-3 [13]) and text-to-image models (*i.e.*, DALL-E 2 [54]).

To test our system and explore future enhancements of AI-generated design cards, we evaluated a set of generated cards through a survey study with designers ($N = 21$) and the authors of HCI papers ($N = 12$). We found that designers perceived content from AI-generated design cards as more inspiring and generative compared to the text of design implications sections, and did not consider the generated content to be different in validity, generalizability, or originality. The paper authors further reported that the contents in a design card were overall clear and accurate, and highlighted the potential of cards in strengthening the delivery of their contributions in HCI. In addition to this positive feedback, conversations with designers and authors also surfaced future directions for enhancing AI-generated design cards.

Our work offers several important contributions:

- Surfacing design elements of what design practitioners want from design cards as a form of translational resource
- Implementing a system that helps streamline the generation of design cards from HCI papers using generative AI models

- Survey results with designers and HCI paper authors demonstrating the potential added inspirability and generativity of design cards generated by LLMs, as well as the potential of our system as a delivery tool for authors to present their contributions
- Insights for the future enhancement of AI-generated design cards

2 RELATED WORK

Generative AI models are powerful tools that have been incorporated into many creative and practical tasks. Large language models (LLMs) such as GPT are trained on large amounts of unlabeled text data and human feedback, and are capable of generating highly consistent and contextually relevant text across a variety of domains and tasks, such as robot journalism [53], customer service [67], and healthcare [65]. Moreover, the generative capabilities of LLMs have been featured as a potential way to inspire and support creative, in addition to practical, tasks. For example, prior studies have explored the use of LLMs in supporting creative writing [15, 20], idea generation [28], and generating song lyrics [71].

Text-to-image models are another type of generative AI model that has been widely used for a range of applications [31]. Using deep learning techniques to generate or modify images, text-to-image models allow users to create various illustrations, from realistic depictions of landscapes to abstract art. A well-known example of a text-to-image model is DALL-E 2 [54], a transformer-based diffusion model that produces images based on descriptive textual prompts. Other latent diffusion models such as Stable Diffusion [60] and Midjourney [36] have also been shown to effectively generate images based on textual prompts. These models have been applied to various pragmatic and creative application areas, such as fine-art painting [55, 56], fashion design [3], and webtoon sketching [39].

Generative AI models have been used to make scientific literature more accessible and efficient to consume. As generative AI models are capable of producing contextually relevant information (*e.g.*, summaries, visual representations), researchers have highlighted their use in transforming or augmenting the reading of scholarly articles to allow a wider audience to gain knowledge from research. For instance, previous studies have demonstrated the use of generative AI models for facilitating paper consumption, on tasks such as paper summarization [14, 27, 43, 48], generation of research highlights [57], and automatic generation of posters or slides [30, 64, 75]. Not limited to these approaches of supporting general paper-reading practices, generative AI models have also been shown to be effective in augmenting domain-specific papers. For example, August *et al.* proposed a tool called Paper Plain [6], which leverages an LLM to lower the barrier to consuming medical knowledge by augmenting medical papers with features such as section summaries and passage-specific question-answering.

Motivated by this research, we highlight in this study the potential of generative AI models in communicating design implications presented in HCI papers. Sas *et al.* [62] suggested several evaluation criteria for assessing the quality of design implications: *theoretical/empirical validity* (*i.e.*, the level of being academically grounded), *generalizability* (*i.e.*, applicability beyond the fieldwork), *originality* (*i.e.*, ensuring that the design implication is original), *generativity* (*i.e.*, ability to open up new design spaces), *inspirability* (*i.e.*, ability to stimulate designers to explore further), and *actionability* (*i.e.*, ability to be acted upon); that is, communicating design implications requires the conveyance of both creativity and academic value. As previous literature has shown the feasibility of generative AI models in generating both contextually relevant and creative outputs, we believe that these characteristics will help augment the utility (*i.e.*, generativity, inspirability, actionability) of the academic paper without losing its scholarly robustness (*i.e.*, validity, generalizability, originality) [52] when generating design cards, thus assisting designers to more easily consume design insights from academic papers.
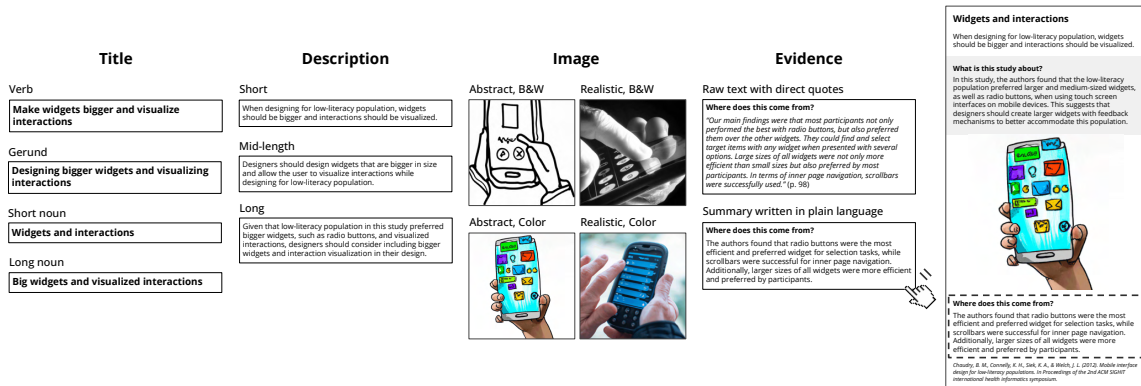
Fig. 2. An illustration of the Miro board participants used during the preliminary study. Each participant was asked to choose one among several options for each component, and drag & drop it to a card template to make a design card.

3 PRELIMINARY STUDY

To understand what designers want to see from design cards for gaining design insights from academic papers and guiding design, we first conducted an interview study with 12 designers.

Although there are no fixed rules for designing design cards, several common components have been used in previous design cards [1], presumably due to the limited size of design cards, their physical format, and the emphasis on the visibility of each component. Thus, we began by designing an initial set of design cards that contain several frequently occurring components to guide participants in offering feedback.

3.1 Initial Design of Design Cards

To choose papers for our initial design cards, we searched in the ACM Digital Library using keywords related to design implications (*e.g.*, design implication, design guideline, future design). From the results, we filtered for full-paper publications from ACM-sponsored conferences/journals that include design implication sections. Among these, we chose three papers with non-overlapping topics that offer implications across a range of design practices (*i.e.*, automotive UX design [63], online security & safety for children [50], and mobile UI design [18]).

First, we performed a cursory exploration of the capabilities of generative models (*i.e.*, summarization, image generation, question answering, instruction-based text completion, in-context learning) as well as rule-based techniques (*e.g.*, deriving a bibliographic reference using a paper's DOI) to determine if each component of the design card can be generated using these capabilities. We then prepared a set of initial design cards consisting of potentially feasible components that are commonly found in existing design cards [16, 22, 25, 35, 46, 76]. Our initial design cards included a title, description, image, source text of the design implication, paper summary (concise overview of the whole paper), and citation. At this stage, each component was manually generated and prepared by the researchers except for images, which were created using a text-to-image model (DALL-E 2). For some components, we offered multiple options to let participants explore diverging styles (see examples in Figure 2). Descriptions for each component are as follows:

Title. The title of the design card is meant to provide an easy-to-understand, high-level description of the design implication. To understand the designer's preference around the title's actionability, length, and grammatical format, we set up four types of titles for each paper: verb-format, gerund-format, long-noun, and short-noun. Verb- and

gerund-format were designed to describe an action of a design guideline (*e.g.*, `Make widgets bigger and visualize interactions` / `Designing bigger widgets and visualizing interactions`), while the goal of long- and short-noun was mainly to convey the design target and context in a keyword (*e.g.*, `Widgets and interactions` / `Big widgets and visualized interactions`).

Description. The description of the design card aims to offer detailed guidance on following the design implication. To probe the details that designers want to see in the description of a design implication, we decided to let designers explore the formats of this component by preparing three options for the description: `short` (approx. 10-15 words), `mid-length` (approx. 20 words), and `long` (approx. 30 words) descriptions, each of which varies in the level of detail.

Image. Image is a visual illustration that is included to inspire and to help designers gain a sense of the context of a design implication. Prior literature suggested that the use of color in images has been shown to capture the attention and interest of viewers, yet it can also be distracting and hinder comprehension of information [42]. Similarly, although photorealistic images offer a higher level of believability, abstracting an image is frequently employed to effectively communicate image cues and enhance the viewer's perception of the image's subject. [32]. To test these two dimensions in design cards, we prepared images by varying two stylistic dimensions: *abstractness* and *color*.

Specifically, we decided to use a text-to-image model instead of retrieving from existing image repositories, because academic papers often introduce new concepts and use cases for which there is no suitable existing illustration. We set up combinations of two specifications for each dimension, resulting in four styles: `abstract colored`, `abstract black & white (B&W)`, `realistic colored`, and `realistic B&W`. Using several prompts involving the paper title and specified style, we generated images for each option in our study using DALL-E 2 [54].

Paper summary. As each design implication is rooted in the research context, we believe it is important to present the study context in the design card. Thus, we summarized the abstract of the source paper in two sentences and included it in our design. Unlike other components, we did not present multiple options for the paper summary.

Evidence of the design implication. The purpose of evidence is to let readers review the foundation that underpins the design guideline. From a part of the paper that contains the result of the study, we identified a paragraph providing evidence that supports the design guideline. Particularly, we wanted to determine whether designers would feel the design card to be more grounded when given the original evidence text from the paper versus a summary of the paragraph written in plain language. Thus, we prepared two options that can potentially be used as evidence: `raw text of the evidence with direct quotes` and a `summary of the evidence written in plain language`.

Citation. In cases where designers want to more closely examine the paper, we included a reference to the paper at the bottom of the design card. This could be designed as a clickable link to the paper if the card is presented in digital format. Similar to the paper summary, we also provided only one version of the citation in APA format.

3.2 Recruitment & Participants

To recruit designers, we posted our study recruitment in two design-focused online communities at our university. We recruited 12 participants: 7 of the participants had more than 3 years of prior design work experience, 3 participants had 1–3 years of experience, and 2 participants had no prior work experience. Of these participants, 7 were working at a company as a designer, and 5 were pursuing a design degree at the time of the study. The average age of participants was 25.3 ($SD = 3.4$); 9 self-identified as female, 2 non-binary, and 1 male.

3.3 Study Procedure & Analysis

We prepared a Miro board (a collaborative diagramming tool) for each paper option, which contains (i) the options for each component and (ii) a card template where participants can drag & drop options to build a design card.

Studies were conducted remotely on Zoom. We first consented participants to the study and then briefly explained the goal of the study and the design cards to each participant. Using the Zoom chat functionality, we asked participants to choose one among the three paper titles used in this study that is most relevant to their design interests. We sent participants the paragraph of the paper that contains the design implication and asked them to read it.

Following, we sent participants a web link that directed them to the Miro board of the paper they chose. Once they were on the Miro board, for each design card component, participants were asked to drag & drop one among several options they prefer the most to the card template, and explain their reason for choosing it (Figure 2). We then asked them to evaluate the design card based on metrics for evaluating design implications [62] (*i.e.*, empirical/theoretical validity, generalizability, originality, generativity, inspirability, and actionability), and discuss any suggestions for enhancing the design cards based on each metric. Subsequently, we showed participants the design cards for the other two papers with the same configurations they chose, and asked them to evaluate the other cards as well, to help enrich our evaluation.

Each interview lasted approximately 45 minutes. Once complete, each participant was compensated 20 USD for their participation. All the procedures of our study were reviewed and approved by the IRB of our university's human subjects division.

To analyze the data, we first transcribed the voice recordings. Then, we analyzed the results by using thematic analysis [11], where the researchers first identified and categorized the significant themes from the responses. Then, we discussed the initial themes and refined them until we reached a consensus on the final themes.

3.4 Results

We first outline the themes related to the overall design of design cards. Following that, we report the themes categorized by each design card component.

3.4.1 Overall structure and design of design cards. Overall, participants were satisfied with the types and structure of information in our design card: *"I like the structure (of design cards). I like how title, description, and all the components make sense with each other."* (P7) At the same time, we identified potential improvements around the visual representation of the design card:

Distributing textual information to avoid information overload. First, displaying every component on a single side of the design card made it difficult for participants to read. As our initial design card template puts all content on a single page, this was reported to affect the visibility of each component, ultimately making the readers overwhelmed and hesitant to dig deeper: *"(On a single page) there are a lot of texts, and that's something that I am overwhelmed with before reading (each component)."* (P3)

To resolve this, we explored the potential of using the back side of design cards to distribute the textual information. Participants mentioned that it would make the cards less overwhelming to put important components on the front and the rest on the back side: *"Let's say that we have another side of this card, and we can include some stuffs in another side of the design card."* (P8)

Importance of visual aesthetics for augmenting inspirability. In addition, participants described the importance of visual aesthetics (*e.g.*, vivid background color) for inspiring designers. They reported that the visually pleasing design

of cards is a major factor that induces them to take a closer look at contents and think through them: *"I want the cards to be more visually attractive and it will make me feel creative and inspired (...) I once saw very pretty design cards and they made me want to use them."* (P2)

3.4.2 Title.

Title as a means of guiding action without being directive. As the title is the first component that designers read while glancing at the design cards, participants spoke to the importance of the title in delivering actionable insights in a concise way. In other words, instead of simply offering keywords (nouns) related to the design context or subject, they preferred to read which action they should take from the title. For these reasons, 10 out of 12 participants preferred either gerund or verb format, as these grammatical constructs are perceived as actions: *"Seems like they (verb and gerund) are telling me what I'm supposed to do so they are clearer."* (P6)

At the same time, participants mentioned that titles starting with a verb are too directive, potentially constraining the scope and generalizability of the use of design cards in their future designs. Instead, gerund format was mentioned as a way of delivering guidance to act, without forcing the readers to follow the action: *"Verb seems like it's directing me to must follow a certain thing like as expecting (...) so I think gerund format would make more sense."* (P5)

3.4.3 Description.
Designers mainly viewed the role of description as a 'tool for understanding the title more in a detailed way.' Specifically, we elicited from designers several necessary elements of a description that they wished to see:

Desire to know about the target users. Participants reported that the target user of the design implication (*e.g.*, low-literacy population, drivers, children) should be clearly noted in the description, unless it is general users. This is reported to help them identify the scope and limitation of the design implication, and determine if the design implication is applicable to their current designing target: *"The target user should be included because I would want to know who I'm doing it for."* (P10)

Identifying the context (i.e., design subject) in which the design implication can be applied. Similarly, design subject (*i.e.*, specific object or system that the design implication can be applied to) was reported as an essential component that should be clearly indicated in the design implication description. It could be a physical object (*e.g.*, widget, ride), or an intangible system (*e.g.*, monitoring systems): *"(I prefer mid-length because) it has a context of which system designers are designing."* (P7)

Justification for following the design implication. Lastly, they would like to see 'why' they should follow a certain design implication. Even though citations make the design implication believable, having a reasoning of why the design should be a certain way (*e.g.*, *Designers should AAA,* **in order to BBB**) helps them perceive the implication as valid and gain more confidence in following the recommendation: *"(in a short description) There is no justification for why you should follow it (...) I won't believe in following the guideline without justification."* (P2)

3.4.4 Image.

Representing a potential usage context. As an image is the only visual illustration among the components, participants mentioned the importance of images in imagining the overall context of the design implication. Thus, they wanted to see how the potential users might be using the design where the design implication can be applied: *"If someone's not*
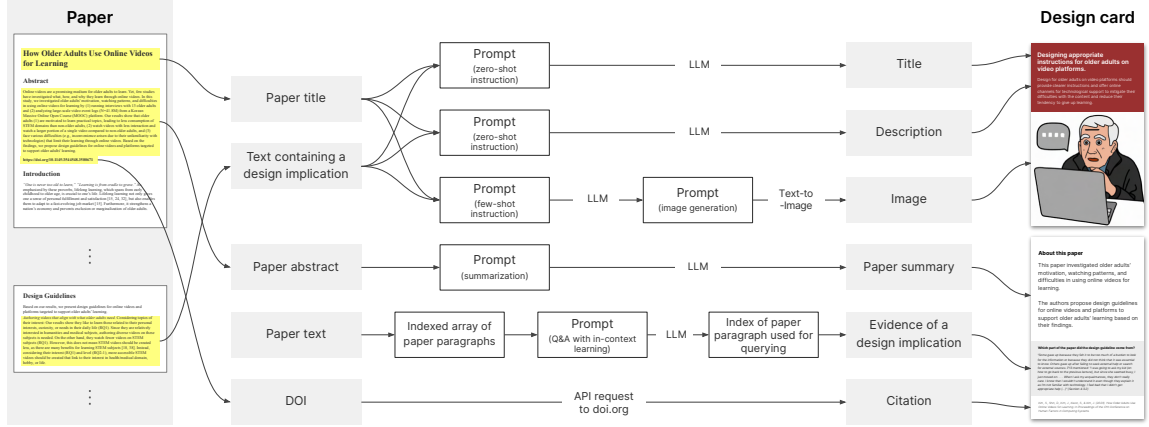
7

Fig. 3. Overview of the pipeline for generating design cards using generative AI models. In this case, we use GPT-3 as the LLM and DALLE-2 as the text-to-image model.

immediately reading all of the text, just looking at the image should give them a context or like they're able to imagine a picture themselves in this scenario where the design implication is applied." (P10)

Color to inspire creativity, but without distorted representation. Except for one participant, everyone responded that they preferred colored illustrations for images, by speaking out to the importance of using color in an image to stimulate inspirability: *"It (colored image) makes it visually appealing (...) the most amount of my attention is very clear when it's in color in some way."* (P6) However, participants also expressed concern about the potential for distorted elements in realistic images after noticing distorted representations of human faces, which is a known issue in AI-generated images [10, 33]. Therefore, they preferred colored images for the inspirational value, but more abstract as to not compromise image authenticity: *"I noticed those faces (distorted).. maybe an abstract one is better if you can't do realistic."* (P3)

3.4.5 Paper summary, evidence of the design implication, and citation. Overall, participants found the paper summary and citation presented in the design card to be necessary. First, a summary of the paper was reported to help them better understand the overall context of the paper that the design implication is built on: *"I really like the paper summary because I can actually see what the study is about."* (P4) Also, the citation helped them perceive that the design implication was from a believable source: *"(By seeing the citation) I feel like it's grounded and build trust on the contents."* (P6)

For the evidence of design implication, a majority of the participants (8 out of 12) preferred the raw text with direct quotes instead of a summary in plain language, due to the credibility of the format: *"It (direct quotes for the evidence) makes the design implication more formal and credible."* (P1)

4 DESIGN ITERATION & IMPLEMENTATION

From the preliminary study, we uncovered insights on how to better structure and design the cards to support designers' understanding and potential use of the design implications. Based on these results, we iterated on the design card layout and content. We also implemented a pipeline to automatically generate design cards from academic papers, which we discuss below. An example design card generated by our system is illustrated in Figure 1.

4.1 Overall Structure and Design

First, to alleviate information overload stemming from displaying all information on a single side of a card, we extended our design to use both sides of a card. Specifically, we put core design card components (*i.e.*, title, description, image) on the front side, and all the other components (*i.e.*, paper summary, evidence, citation) on the backside.

In addition, to improve upon aesthetics and achieve consistent styling, we programmed our system to match the background color of the front side with the most dominant color of the generated image. Specifically, our system obtains a dominant color from the generated image using the modified mean cut quantizer technique [9], adjusts the color to a high level of saturation and low level of value in the HSV color dimension, and sets this as the background color. This results in a vivid and visually engaging front side with a clearly readable card title and description.

4.2 Generating Card Components

4.2.1 Title. Given participants' input, we focused on making the title actionable. As such, we decided to generate the title using an LLM with an instructional prompt specifying that the output include an action verb in the form of a gerund phrase.

While engineering the prompt, we noticed that some paper authors included the target context of their design (*e.g.*, user, design subject) in the paper title while omitting them in their design implication text. Simply relying on the design guideline text may omit salient keywords from the paper, so we also included the paper title in the LLM prompt to encourage centering of the output around the keywords in the title. We use the following prompt:

```
The following text snippet is part of an academic paper (title: {PAPER TITLE}), which contains
a guideline for a certain design practice. From this text, extract a concise title that delivers
what action designers should take in gerund-format (i.e., verb+ing): {DESIGN IMPLICATION TEXT}
```

To verify that the output is formatted as intended, we programmatically detect that the starting word of the output is in gerund format.

4.2.2 Description. For the description, we focused on surfacing the target user, design subject, and justification for following the design implication. Thus, similar to the title, we set up a prompt instruction for an LLM to generate a one-line description for designers such that the output includes these elements:

```
The following text snippet is part of an academic paper (title: {PAPER TITLE}), which contains
a guideline for a certain design practice. From this text, extract a one-line argument for
designers that contains (i) who the design is targeting, (ii) what the design subject is, and
(iii) the justification for the design: {DESIGN IMPLICATION TEXT}
```

To check if the output satisfies the requirements we included in the prompt instruction, we used the LLM to evaluate its own output (inspired by work such as [7, 45, 73]). Specifically, we programmed our system to verify the results as another classification task using an LLM. For example, to ensure that the justification for design is included in the output description, the system asks: `Does the following statement contain the justification for designers to follow a certain design direction? Answer in YES or NO.` If one of our three requirements are not met, we prompt the system to generate three additional outputs and select from among them the best output based on the same criteria.

4.2.3 Image. To generate an image, we used both an LLM and a text-to-image model. Specifically, (i) we first prompt the LLM to generate an image prompt that illustrates the context of the design implication; (ii) we then use the text-to-image model with the prompt output from (i) to generate an image.

For implementation, we engineered a few-shot prompt that contains three training pairs of source design implication text and the corresponding text-to-image prompt describing its usage context [37, 41, 59]. We noticed that including jargon (*e.g.*, MOOC; massive-open online course) was more likely to cause inaccurate representations in the generated images, so we added instructions to the prompt to avoid jargon in the output of (i). An example prompt is:

```
Illustrate a one-line usage context for the following design guideline. The output should not
contain any jargon:

Design guideline: Since older adults are unfamiliar with video medium, (...)
Usage context: an elderly person looking at a laptop with a video playing on it
..
{FEW-SHOT EXAMPLES}
..
Design guideline: {DESIGN IMPLICATION TEXT}
```

We augmented the output of (i) with a styling prompt instructing the text-to-image model to produce an abstract colored image. Specifically, we decided to have our system generate an image without including complex, realistic details by presenting an image in a *cartoon* style. As the model generates a colored image by default, we designated the styling of a colored image by prepending `"A cartoon of"` to the output prompt from (i). Finally, we generated an image by feeding the complete prompt as input into the text-to-image model.

4.2.4 Paper summary. To generate the summary of the paper, we instructed an LLM to summarize the abstract of the paper in two sentences using the following prompt. Here, as the papers are often written in the first person, we engineered the prompt to summarize in the third person:

```
Summarize the following text in two sentences, describing the authors in the third person:
{ABSTRACT}
```

4.2.5 Evidence. The system converts the XML structure into an array that contains every paragraph of the paper, matched to its section title. Then, we prompted an LLM to run a Q&A task by asking for evidence for the design implication text (*i.e.*, `What can be the evidence for this design implication: {DESIGN IMPLICATION TEXT}`), and find out which source paragraph was used to generate the output. We used the LlamaIndex [44] library to overcome the prompt-length limitation of an LLM and enable in-context learning. Lastly, we programmed the system to use the selected paragraph as evidence for the design implication, appending the section number for reference.

4.2.6 Citation. We used the `doi.org` API to construct a citation for each paper. We designated APA as the citation format in the API request header. For some papers that do not have a DOI, our system includes the paper title as well as the author list as a citation.

4.3 Implementation

Our system is implemented as a web interface, where the user can upload a paper PDF and choose the design implication text from which to generate a design card. The interface is built on a Javascript framework (SvelteKit), which is

connected to a Python backend server deployed on AWS EC2 that generates and returns the components of the design card.

More specifically, the server takes as input (i) a PDF of the paper and (ii) a selected design implication text, and outputs a design card in HTML format. First, the system uses the Grobid [47] library to parse the paper PDF file into the structured XML format, and the system uses the BeautifulSoup 4 library [58] to process the XML structure and extract components (*i.e.*, title, abstract, DOI, paragraphs) used to generate the design cards. The overall pipeline for generating design cards is illustrated in Figure 3.

For the system implementation tested in our evaluation, we used GPT-3 as our LLM for generating all textual components, and DALL-E 2 for images. These generative AI models are connected to the backend server through an API. For all GPT-3 API requests, we used the *text-davinci-003* model with the following parameters: `temperature: 0.5`, `top p: 1`, `frequency penalty: 0`, `presence penalty: 0`, `best of: 1`. We selected these models as they are strong state-of-the-art models for text and text-to-image generation, though we make no claims that they are the optimal models for these tasks. Other models could be easily adapted for use in our modular pipeline.

5 EVALUATION

To evaluate the feasibility of communicating design implications using design cards generated by our system, we conducted a survey study with designers ($N = 21$) and authors of HCI papers ($N = 12$).

5.1 Study Setup & Procedure

5.1.1 Survey study with designers. In order to understand the perception of designers toward the design cards generated by our system, we set up a within-subjects survey study with designers as follows:

First, we used the same approach as we did in the preliminary study to identify HCI papers in the ACM DL containing design implications. We chose 4 HCI papers with varying themes and research methods for our survey study with designers (Table 1; paper 1 – 4). For each paper, we selected a segment of paper text that contains a design implication; if there are multiple implications, we randomly chose one among them. Then, we fed the segment into our system and generated a design card for each paper.

Each participant in our study conducted a survey in our online survey interface. Here, each participant was asked to evaluate two papers that were randomly chosen, and for each paper, was presented two formats for communicating the design implication: (i) raw text format (a part of the paper that contains a design implication) and (ii) design card format. Each evaluation contains six questions for evaluating the perceived qualities of the design implication [62] (*i.e.*, empirical/theoretical validity, generalizability, originality, generativity, inspirability, and actionability) on a 7-point Likert scale. To ensure participants' consistent understanding of the evaluation metrics, our questionnaire incorporated the definitions outlined in the original paper [62] (*e.g.*, generativity – to create and open up new design spaces; p. 1978). In addition, we randomized the order of both papers and formats for every participant to avoid the ordering effect.

Once participants finished evaluating formats for the two papers, they were directed to a screen where they were asked their preferred format for consuming design implications from academic papers. They were also asked to provide the reasons for their preference, as well as future enhancements for the design cards.

To recruit designers for our survey study, we posted our recruitment in two design-focused online communities at our university. In this posting, we provided an explanation and goals of our study, as well as a link that directs to our online survey. As such, 21 designers completed the survey—13 were working as designers, 7 were pursuing a degree in

Table 1. Papers used for generating design cards and running the evaluation

	Theme	Venue	Research method	
1	Accessibility, game	CHI '22	Qualitative	[49]
2	Healthcare, voice agent	TiiS, '22	Mixed-methods	[12]
3	Web search UI design	CHI '20	Quantitative	[8]
4	Chatbot	DIS '21	Quantitative	[69]
5	Healthcare	CHI '20	Mixed-methods	[38]
6	Online survey	CHI '19	Quantitative	[5]
7	Accessibility, interaction technique	MobileHCI '21	Quantitative	[2]
8	AI fairness toolkit	FAccT '22	Qualitative	[24]
9	Online community	CSCW, '20	Quantitative	[74]
10	Authoring tool	UIST '22	Mixed-methods	[68]
11	Social VR, online interaction	CSCW, '21	Qualitative	[29]
12	Accessibility, voice-based CA	CHI '21	Qualitative	[17]

design, and 1 was unemployed. The average age of the participants was 26.5 ($SD = 5.7$); 14 of them were female, 6 male, and 1 preferred not to say. Of all participants, 10 responded that they had heard of the concept of a design card before.

5.1.2 Survey study with the authors of HCI papers. To understand if our system can be an effective tool for the paper authors to communicate design implications, we contacted the authors of HCI papers asking them several open-ended questions via email. In addition to the four papers that we used for the survey study with designers, we decided to contact the authors of the additional set of papers to gain richer insights into their perception of the design cards. In the email, we included a design card that we generated from their paper, as well as the part of the paper that we used to generate the design card. Based on these contents, we asked each author to (i) assess the accuracy and clarity of the contents of our design card and (ii) provide any feedback for future improvement to make AI-generated design cards better support communication of design implications. We received responses from the 12 papers in total, including from the authors of all 4 papers that were used for the survey study with designers along with 8 other respondents (Table 1; paper 1 – 12).

5.2 Analyses

5.2.1 Quantitative analysis. In our survey study, each participant evaluated the design implications and their corresponding design cards for two (randomly assigned) papers. Thus, to analyze our quantitative results while controlling for the random effect and the paper ID, we used a linear mixed-effects model to test for a significant difference across
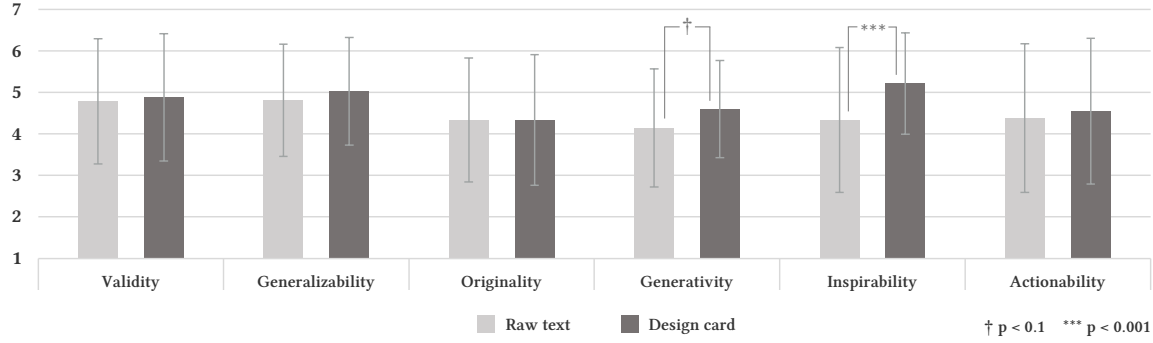
Fig. 4. Perceived qualities of design implication for each format using metrics from Sas *et al.* (2014) [62]. Significance level is based on the generalized linear mixed-effects modeling.

the perceived qualities of design implications from reading the two formats. Specifically, we modeled by including the paper ID as a control variable and the participant ID as a random variable.

5.2.2   Qualitative analysis. Once the studies were complete, we then followed the same approach as we did with the preliminary study data to develop themes. In Subsection 5.4, we refer to each participant as D1 – D21 (designers) and A1 – A12 (authors), with participant order randomized.

5.3   Quantitative Results

5.3.1   Preference on the format. Our survey revealed that a majority of designers preferred to consume a design implication using design cards. In survey results, 15 designers responded that they preferred to consume design implications using a design card format, while 3 preferred the raw text format; the other 3 designers had no preference.

5.3.2   Perceived qualities of design implication. The linear mixed-effects modeling of our survey results revealed that designers perceived the content from the design card format as more inspiring ($M = 5.21$, $SD = 1.22$) and generative ($M = 4.60$, $SD = 1.17$), compared to the raw text format ($M = 4.33$, $SD = 1.75$; $M = 4.14$, $SD = 1.42$) (Figure 4). The analysis revealed that there was a significant difference in the perception of inspirability ($t = 3.53$, $p < 0.001$) and a marginally significant difference in the perception of generativity ($t = 1.74$, $p < 0.1$).

At the same time, the results showed that design cards were able to deliver a design implication to designers without sacrificing its validity ($M = 4.88$, $SD = 1.53$ vs. $M = 4.79$, $SD = 1.51$; $t = 0.35$, $p = 0.73$), generalizability ($M = 5.02$, $SD = 1.30$ vs. $M = 4.81$, $SD = 1.35$; $t = 0.81$, $p = 0.42$), or originality ($M = 4.33$, $SD = 1.57$ vs. $M = 4.31$, $SD = 1.47$; $t = 0.09$, $p = 0.93$). There was also no significant difference in actionability between design cards and raw texts ($M = 4.55$, $SD = 1.76$ vs. $M = 4.38$, $SD = 1.79$; $t = 0.48$, $p = 0.63$).

5.4   Qualitative Results

5.4.1   Visual hierarchy and visual elements made design cards appealing and easy to consume. Participants reported the card format to be well-structured, with a clear visual hierarchy that facilitated information processing. They perceived the information to be well-organized, improving the readability of the contents: *"I liked the cards mainly due to their visual hierarchy as I could know what is important and what is not."* (D2) Specifically, having the key information (*i.e.,*

title) upfront and background contexts on the opposite side, design cards allowed participants to quickly gain a sense of a design implication. At the same time, they can optionally consume other information if required: *"I could easily digest by getting an at-a-glance summary of the insight with the title and then clear divisions of a longer explanation of the insight with factual references below."* (D18)

In addition, designers from our study evaluated the design of design cards (*e.g.*, color, image styling) as visually clear and appealing, which was cited as the reason for making them want to read the card: *"I loved the colors (of design cards)"* (D16); *"I enjoyed seeing a visual graphic, which led me to prefer design cards over plain texts."* (D11) Also, the images of design cards were reported to successfully play a role in assisting the designers to easily grasp the whole context of a design implication. Designers mentioned that the context depicted in an image was catchy, allowing them to easily consume the design implication: *"The art on the cards successfully served its purpose of illustrating what information was being shared (in the paper)."* (D14) Authors also highlighted the image's potential use of assisting readers in comprehending the connection between the target user and the design context of the paper: *"The picture describes the relationship and interaction between [target user of the paper] and [design artifact] well. The most eye-catching part."* (A9)

5.4.2 Our design cards afforded digestibility and ease of understanding through concise language. The raw text format was perceived as more difficult to read. With a long and dense text block, raw text format was reported to make it harder for participants to follow the contents, making them easily lose interest and get bored: *"I found it harder to absorb a large bit of text from the raw text."* (D14); *"I felt the raw text format is too boring to read."* (D7) In contrast, participants found the card format easy to understand and digest compared to the raw text format. They appreciated the use of concise language enabled by automated summarization, which helped them easily identify key points and main takeaways: *"Design insights were given to me without me having to read too much or decipher from the source text."* (D21)

Aligning with the perception of designers, authors of the paper who responded to our questions similarly reported the clarity and effectiveness of design cards to be satisfactory, as well as the idea of automatically generating design cards. They mentioned that the design card reflects their design implication well: *"(Design card) correctly reflects what we stated in [section] and the wording of the [description] is relatively similar to the text in the paper - it's good."* (A8); *"The idea is very cool and I think your system does a good job in summarizing the design implication from [section] in a crisp and concise way."* (A6) With such well-aligning contents presented in a handy format, authors viewed design cards as an effective way of communicating their design implications: *"The generated design card communicates the design implication listed in [section] of our paper effectively."* (A2)

Based on the clarity and effectiveness of the format, the authors highlighted the potential of design cards for augmenting their contributions around the design implications. By collaborating with our design card system, authors believed that our system could be a great authoring tool for augmenting the delivery of their paper contributions: *"If researchers were to go through one or two revisions based on this approach, it could indeed greatly enhance the delivery of their contribution much strongly and in a more meaningful way."* (A3)

5.4.3 Our system supported communication of design implications without loss of validity and originality. From our quantitative analysis, we identified that design cards delivered a design implication without sacrificing any validity and originality. Aligning with these results, design card and raw text formats were both reported to be valid and believable, even if the design card was programmatically generated from the original paper text. In particular, we identified that the inclusion of extracted evidence and the paper citation contributed to making the design card a reliable source for design insights. Such cues implying that the results were verified led participants to perceive the design card as more legitimate: *"Seeing the parts on the bottom (evidence, citation) made it look professional."* (D19)

At the same time, the survey with the authors revealed the need to refine the pipeline for extracting evidence for a design implication. To identify evidence from the paper, we prompted an LLM to run a Q&A task and programmed our system to use the paragraph of a paper that was used to derive the answer, which was reported to be an effective way to deliver the author's reasoning of the design guideline: *"I like (evidence)... It helps effectively communicate the rationale for the design implications."* (A11) Yet, four authors mentioned that the evidence included in the design card of their paper could be potentially enhanced, by pointing out that it was partially correct, not the only evidence present in the paper, or misleading: *"The part about where the guideline came from seems off."* (A5) As a potential mitigation strategy, one author offered that having a rule-based benchmarking tool to verify how much the evidence overlaps with the design implication might help: *"Some checks based on word overlap between the generated guideline and the text in the referenced paper paragraph could be helpful to get a proxy for the accuracy."* (A2)

5.4.4 Opportunities for additional images. At the same time, although the image of our design card system depicting the general context assisted designers in grasping a context easily, they also wished to have an additional image in a design card for gaining 'more practical' information. For instance, designers mentioned that adding additional infographics that directly show the detailed results of the paper would further help them to understand the study details: *"Styling of an image is straightforward but I also want to see a pure infographic for the results."* (D13) Similarly, one author suggested using a figure directly driven from the original paper to clearly illustrate the design guideline in a way the author envisioned: *"I think showing a representative figure of the paper can be helpful, especially if that figure/artifact exemplifies the design mentioned in the design implications or can further clarify."* (A11)

5.4.5 Potential way of engaging authors in the pipeline. Our analysis revealed the opportunity of having paper authors engage in the process of generating design cards. Authors perceived the use of auto-generated images to be a handy way of authoring and communicating their design implications: *"The approach of using an AI-generated image for authoring the design implication is interesting and will be super helpful (when communicating design implications)."* (A3), but at the same time two authors described misalignments in the expectation that could potentially be mitigated by involving authors in the design card generation process.

For example, one author mentioned that there are multiple key messages they were intending to communicate in their design implications section, but that the context of the image reflects only one of them: *"I feel that the image only covers a subset of the design implication presented in the text and may not necessarily highlight all the key messages."* (A4) Another author pointed out the misalignment between their expected emotion toward potential users and that of the user illustrated in the design card image: *"The valence is a little negative to me (as the user looks mad)."* (A1) These examples indicate the potential for engaging authors in verifying the alignment of output cards with their original intention, to more accurately represent the intent of the authors.

6 DISCUSSION

In this study, we examined the feasibility of AI-generated design cards as a tool for communicating design insights between paper authors and designers. Our system automatically produces design cards based on a design implication from a paper, and demonstrated its potential to assist authors in communicating their design implications. We expect that a more effective process for presenting design implications could help alleviate the burden on paper authors to communicate their findings to design practitioners. Bridging this translational gap can further enhance the impact of valuable design insights from academic papers.

Results from our evaluation suggest that designers perceived our AI-generated design cards as more inspiring and generative compared to traditional paper formats. We also found that these gains were had without sacrificing the validity, originality, and generalizability of the design finding. A key reason for this that we identified in our qualitative analysis is that our system lets designers view the original text via evidence quotes and citations, which ultimately assured them that design card contents were legitimate. However, our survey with paper authors revealed that extracted evidence for a design implication can sometimes be inaccurate or not perfectly aligned with the part of the paper authors used to derive the design implication. As such a misalignment might affect the overall credibility of generated design cards, future systems should focus on validating the accuracy of represented information. Involving authors in the process of generating design cards could help guarantee the validity and appropriateness of outputs.

Despite the effectiveness of our system, one remaining question is how to make the content of design cards generated by our system more actionable. Although we designed our title to involve action, our survey with designers revealed no significant difference in actionability between the design card format and raw text format. One possible reason is that the papers' design implication sections (our study comparisons) were already well-written and were calling for designers to take a particular design action. Thus, designers did not perceive a difference in actionability. However, our study does suggest some changes that we can explore to make the design card even more actionable, such as having an additional component that gives more detailed guidance (*e.g.*, step-by-step guidance for designing) or examples of connecting the design implication to real-world artifacts (*e.g.*, how a well-known design was enhanced by applying this implication).

Our study results also suggest that there are opportunities to involve the paper authors in the card generation process. For example, while the AI-generated images were found to be effective in providing an overall context relevant to the design implication, a few authors reported its misalignment with their own intention (*e.g.*, emotion of the user depicted in the image, selection of one among multiple messages to deliver). To address this, we believe that including authors in the loop may help; for instance, the system may offer multiple image prompt candidates, and let authors choose the one that best aligns with their intended message. The system can also allow authors to specify keywords, and prompt an LLM to incorporate these keywords into the generated image prompts. Similarly, by having paper authors verify the alignment of the AI-generated/detected content (*e.g.*, image, title, description, evidence) with their original intentions, we believe that the system could produce more accurate and effective design cards.

While our system has the potential to improve designers' use of the abundance of rich insights in academic papers, one challenge that will need to be studied is how to make our card-based translational resource available to designers. Recent research suggests that designers often lack awareness of the wealth of design cards that exist and their potential benefits, despite a set of existing design cards that can be readily used [34, 61]. This awareness problem could worsen if our system helps generate a large influx of cards. One possible solution is to integrate our system into digital libraries (*e.g.*, ACM Digital Library) and present the generated design cards alongside each paper. Other more designer-friendly repositories should also be explored to help the dissemination of this type of resource.

7 LIMITATION & FUTURE WORK

There were several assumptions that we made in building our system and running evaluation studies. First, although we evaluated the qualities of design cards generated by our system, we did not test how generated design cards would be used in the real-world designing contexts of designers. In future work, we should further explore this question by investigating how AI-generated design cards affect the design process and outputs of designers.

Our system also assumes a context where the authors of papers manually localize paper segments containing design implications. While design implication sections are frequently included in HCI publications, future work could investigate ways to automatically identify text corresponding to design implications, as well as segment these sections into multiple distinct implications. These remaining steps are needed to further streamline the generation of design cards from HCI papers.

Additionally, although our focus was to demonstrate the potential of AI-generated cards by choosing reasonable methods, there are many alternate models, techniques, and prompts that could be used to build our system. Since our system is modular, each component can easily accommodate alternate techniques or models to test performance, and we believe this can be expanded on in future work.

Lastly, future work should strive to mitigate AI biases. As utilizing AI-generated images may retain biases from the datasets they are trained on, it is an open concern [13, 51] that the outputs manifest gender and racial stereotypes (*e.g.*, higher likelihood of featuring white men rather than representative users; negative associations for certain racial/ethnic groups). To overcome this, it is necessary to develop methods to identify and mitigate such biases in future systems.

8 CONCLUSION

In this paper, we propose an approach of using generative AI models to help create design cards that distill academic findings into a more accessible and prescriptive format. Based on the preliminary interview sessions with design practitioners, we developed a system that automatically generates a design card using generative AI models. Results from our evaluation with designers and HCI paper authors demonstrated the potential of AI-generated design cards in communicating design implications, as well as their future enhancements.

REFERENCES

[1] Tessa Aarts, Linas K Gabrielaitis, Lianne C De Jong, Renee Noortman, Emma M Van Zoelen, Sophia Kotea, Silvia Cazacu, Lesley L Lock, and Panos Markopoulos. 2020. Design Card Sets: Systematic Literature Survey and Card Sorting Study. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 419–428. https://doi.org/10.1145/3357236.3395516

[2] Dragan Ahmetovic, Daniele Riboli, Cristian Bernareggi, and Sergio Mascetti. 2021. RePlay: Touchscreen Interaction Substitution Method for Accessible Gaming. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*. 1–12. https://doi.org/10.1145/3447526.3472044

[3] Kenan Emir Ak, Joo Hwee Lim, Jo Yew Tham, and Ashraf Kassim. 2019. Semantically Consistent Hierarchical Text to Fashion Image Synthesis with an Enhanced-Attentional Generative Adversarial Network. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 3121–3124. https://doi.org/10.1109/ICCVW.2019.00379

[4] Khadeeja Alkhuzai and Alena Denisova. 2021. Evaluating the Use of Persuasive Design Cards for Novice Designers. *Journal of usability studies* 16, 2 (2021), 91–120. https://uxpajournal.org/persuasive-design-cards-novice-designers

[5] Tal August and Katharina Reinecke. 2019. Pay Attention, Please: Formal Language Improves Attention in Volunteer and Paid Online Experiments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11. https://doi.org/10.1145/3290605.3300478

[6] Tal August, Lucy Lu Wang, Jonathan Bragg, Marti A Hearst, Andrew Head, and Kyle Lo. 2022. Paper Plain: Making Medical Research Papers Approachable to Healthcare Consumers with Natural Language Processing. *arXiv preprint arXiv:2203.00130* (2022). https://doi.org/10.48550/arXiv.2203.00130

[7] Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, et al. 2022. Constitutional AI: Harmlessness from AI Feedback. *arXiv preprint arXiv:2212.08073* (2022). https://doi.org/10.48550/arXiv.2212.08073

[8] Amanda Baughan, Tal August, Naomi Yamashita, and Katharina Reinecke. 2020. Keep it Simple: How Visual Complexity and Preferences Impact Search Efficiency on Websites. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376849

[9] Dan S Bloomberg. 2008. Color quantization using modified median cut. *Leptonica* (2008).

[10] Ali Borji. 2022. Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. *arXiv preprint arXiv:2210.00586* (2022). https://doi.org/10.48550/arXiv.2210.00586

[11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706QP063OA

[12] Robin Brewer, Casey Pierce, Pooja Upadhyay, and Leeseul Park. 2022. An Empirical Study of Older Adult's Voice Assistant Use for Health Information Seeking. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 12, 2 (2022), 1–32. https://doi.org/10.1145/3484507

[13] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems* 33 (2020), 1877–1901. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[14] Isabel Cachola, Kyle Lo, Arman Cohan, and Daniel Weld. 2020. TLDR: Extreme Summarization of Scientific Documents. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, 4766–4777. https://doi.org/10.18653/v1/2020.findings-emnlp.428

[15] Alex Calderwood, Vivian Qiu, Katy Ilonka Gero, and Lydia B Chilton. 2020. How Novelists Use Generative Language Models: An Exploratory User Study. In *HAI-GEN+ user2agent@ IUI*. https://ceur-ws.org/Vol-2848/HAI-GEN-Paper-3.pdf

[16] Mafalda Casais, Ruth Mugge, and Pieter MA Desmet. 2016. Using symbolic meaning as a means to design for happiness: The development of a card set for designers. (2016). https://doi.org/10.21606/drs.2016.424

[17] Inha Cha, Sung-In Kim, Hwajung Hong, Heejeong Yoo, and Youn-kyung Lim. 2021. Exploring the Use of a Voice-based Conversational Agent to Empower Adolescents with Autism Spectrum Disorder. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15. https://doi.org/10.1145/3411764.3445116

[18] Beenish M Chaudry, Kay H Connelly, Katie A Siek, and Janet L Welch. 2012. Mobile Interface Design for Low-Literacy Populations. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. 91–100. https://doi.org/10.1145/2110363.2110377

[19] David Chung and Rung-Huei Liang. 2015. Understanding the Usefulness of Ideation Tools with the Grounding Lenses. In *Proceedings of the Third International Symposium of Chinese CHI*. 13–22. https://doi.org/10.1145/2739999.2740002

[20] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19. https://doi.org/10.1145/3491102.3501819

[21] Lucas Colusso, Cynthia L Bennett, Gary Hsieh, and Sean A Munson. 2017. Translational Resources: Reducing the Gap Between Academic Research and HCI Practice. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. 957–968. https://doi.org/10.1145/3064663.3064667

[22] Lucas Colusso, Tien Do, and Gary Hsieh. 2018. Behavior Change Design Sprints. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 791–803. https://doi.org/10.1145/3196709.3196739

[23] Lucas Colusso, Ridley Jones, Sean A Munson, and Gary Hsieh. 2019. A Translational Science Model for HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3290605.3300231

[24] Wesley Hanwen Deng, Manish Nagireddy, Michelle Seng Ah Lee, Jatinder Singh, Zhiwei Steven Wu, Kenneth Holstein, and Haiyi Zhu. 2022. Exploring How Machine Learning Practitioners (Try To) Use Fairness Toolkits. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 473–484. https://doi.org/10.1145/3531146.3533113

[25] Ying Deng, Alissa N Antle, and Carman Neustaedter. 2014. Tango Cards: A Card-Based Design Tool for Informing the Design of Tangible Learning Games. In *Proceedings of the 2014 Conference on Designing Interactive Systems*. 695–704. https://doi.org/10.1145/2598510.2598601

[26] Peter Denning, Jim Horning, David Parnas, and Lauren Weinstein. 2005. Wikipedia Risks. *Commun. ACM* 48, 12 (2005), 152–152. https://doi.org/10.1145/1101779.1101804

[27] Jay DeYoung, Iz Beltagy, Madeleine van Zuylen, Bailey Kuehl, and Lucy Lu Wang. 2021. MSˆ2: Multi-Document Summarization of Medical Studies. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 7494–7513. https://doi.org/10.18653/v1/2021.emnlp-main.594

[28] Giulia Di Fede, Davide Rocchesso, Steven P Dow, and Salvatore Andolina. 2022. The Idea Machine: LLM-based Expansion, Rewriting, Combination, and Suggestion of Ideas. In *Creativity and Cognition*. 623–627. https://doi.org/10.1145/3527927.3535197

[29] Guo Freeman and Divine Maloney. 2021. Body, Avatar, and Me: The Presentation and Perception of Self in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–27. https://doi.org/10.1145/3432938

[30] Tsu-Jui Fu, William Yang Wang, Daniel J. McDuff, and Yale Song. 2021. DOC2PPT: Automatic Presentation Slides Generation from Scientific Documents. In *AAAI Conference on Artificial Intelligence*. https://doi.org/10.1609/aaai.v36i1.19943

[31] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. 2022. An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. *arXiv preprint arXiv:2208.01618* (2022). https://doi.org/10.48550/arXiv.2208.01618

[32] Paul Haeberli. 1990. Paint by Numbers: Abstract Image Representations. In *Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques*. 207–214. https://doi.org/10.1145/97879.97902

[33] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *NIPS*. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf

[34] Gary Hsieh, Brett A Halperin, Evan Schmitz, Yen Nee Chew, and Yuan-Chi Tseng. 2023. What is in the Cards: Exploring Uses, Patterns, and Trends in Design Cards. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3544548.3580712

[35] IDEO. 2023. *Method Cards*. https://www.ideo.com/post/method-cards

[36] Midjourney Inc. 2022. *Midjourney*. https://www.midjourney.com

[37] Seoyoung Kim, Donghoon Shin, Jeongyeon Kim, and Soonwoo Kwon. 2023. How Older Adults Use Online Videos for Learning. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3544548.3580671

[38] Taewan Kim, Mintra Ruensuk, and Hwajung Hong. 2020. In Helping a Vulnerable Bot, You Help Yourself: Designing a Social Bot as a Care-Receiver to Promote Mental Health and Reduce Stigma. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3313831.3376743

[39] Hyung-Kwon Ko, Subin An, Gwanmo Park, Seung Kwon Kim, Daesik Kim, Bohyoung Kim, Jaemin Jo, and Jinwook Seo. 2022. We-toon: A Communication Support System between Writers and Artists in Collaborative Webtoon Sketch Revision. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–14. https://doi.org/10.1145/3526113.3545612

[40] Hyung-Kwon Ko, Gwanmo Park, Hyeon Jeon, Jaemin Jo, Juho Kim, and Jinwook Seo. 2023. Large-Scale Text-to-Image Generation Models for Visual Artists' Creative Works. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 919–933. https://doi.org/10.1145/3581641.3584078

[41] Leah Kulp, Aleksandra Sarcevic, Megan Cheng, Yinan Zheng, and Randall S Burd. 2019. Comparing the Effects of Paper and Digital Checklists on Team Performance in Time-Critical Work. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3290605.3300777

[42] Richard J Lamberski and Francis M Dwyer. 1983. The Instructional Effect of Coding [Color and Black and White] on Information Acquisition and Retrieval. *ECTJ* 31, 1 (1983), 9–21. https://doi.org/10.1007/BF02765207

[43] Lei Li, Yang Xie, Wei Liu, Yinan Liu, Yafei Jiang, Siya Qi, and Xingyuan Li. 2020. CIST@CL-SciSumm 2020, LongSumm 2020: Automatic Scientific Document Summarization. In *Proceedings of the First Workshop on Scholarly Document Processing*. Association for Computational Linguistics, 225–234. https://doi.org/10.18653/v1/2020.sdp-1.25

[44] Jerry Liu. 2022. *LlamaIndex*. https://github.com/jerryjliu/gpt_index

[45] Yang Liu, Dan Iter, Yichong Xu, Shuohang Wang, Ruochen Xu, and Chenguang Zhu. 2023. GPTEval: NLG Evaluation using GPT-4 with Better Human Alignment. *arXiv preprint arXiv:2303.16634* (2023). https://doi.org/10.48550/arXiv.2303.16634

[46] Dan Lockton, David Harrison, and Neville A Stanton. 2010. The Design with Intent Method: A design tool for influencing user behaviour. *Applied Ergonomics* 41, 3 (2010), 382–392. https://doi.org/10.1016/j.apergo.2009.09.001

[47] Patrice Lopez. 2009. GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications. In *Research and Advanced Technology for Digital Libraries: 13th European Conference, ECDL 2009, Corfu, Greece, September 27-October 2, 2009. Proceedings 13*. Springer, 473–474. https://doi.org/10.1007/978-3-642-04346-8_62

[48] Yao Lu, Yue Dong, and Laurent Charlin. 2020. Multi-XScience: A Large-scale Dataset for Extreme Multi-document Summarization of Scientific Articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 8068–8074. https://doi.org/10.18653/v1/2020.emnlp-main.648

[49] Liam Mason, Kathrin Gerling, Patrick Dickinson, Jussi Holopainen, Lisa Jacobs, and Kieran Hicks. 2022. Including the Experiences of Physically Disabled Players in Mainstream Guidelines for Movement-Based Games. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3501867

[50] Brenna McNally, Priya Kumar, Chelsea Hordatt, Matthew Louis Mauriello, Shalmali Naik, Leyla Norooz, Alazandra Shorter, Evan Golub, and Allison Druin. 2018. Co-designing Mobile Online Safety Applications with Children. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–9. https://doi.org/10.1145/3173574.3174097

[51] Pamela Mishkin, Lama Ahmad, Miles Brundage, Gretchen Krueger, and Girish Sastry. 2022. DALL·E 2 Preview - Risks and Limitations. (2022).

[52] ACM CHI Conference on Human Factors in Computing Systems. 2023. *Guide to a Successful Submission*. https://chi2023.acm.org/for-authors/presenting/papers/guide-to-a-successful-submission

[53] John V Pavlik. 2023. Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *Journalism & Mass Communication Educator* (2023). https://doi.org/10.1177/10776958221149577

[54] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv preprint arXiv:2204.06125* (2022). https://doi.org/10.48550/arXiv.2204.06125

[55] Reddit. 2023. *r/aiArt*. https://www.reddit.com/r/aiArt/

[56] Reddit. 2023. *r/ImagenAI*. https://www.reddit.com/r/ImagenAI

[57] Tohida Rehman, Debarshi Kumar Sanyal, Prasenjit Majumder, and Samiran Chattopadhyay. 2022. Named Entity Recognition Based Automatic Generation of Research Highlights. In *Proceedings of the Third Workshop on Scholarly Document Processing*. Association for Computational Linguistics, 163–169. https://aclanthology.org/2022.sdp-1.18

[58] Leonard Richardson. 2007. Beautiful Soup Documentation. https://buildmedia.readthedocs.org/media/pdf/beautiful-soup-4/latest/beautiful-soup-4.pdf

[59] Yann Riche, Jonathan Dodge, and Ronald A Metoyer. 2010. Studying Always-On Electricity Feedback in the Home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1995–1998. https://doi.org/10.1145/1753326.1753628

[60] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10684–10695. https://doi.org/10.1109/CVPR52688.2022.01042

[61] Robin Roy and James P Warren. 2019. Card-based design tools: A review and analysis of 155 card decks for designers and designing. *Design Studies* 63 (2019), 125–154. https://doi.org/10.1016/j.destud.2019.04.002

[62] Corina Sas, Steve Whittaker, Steven Dow, Jodi Forlizzi, and John Zimmerman. 2014. Generating Implications for Design through Design Research. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1971–1980. https://doi.org/10.1145/2556288.2557357

[63] Tobias Schneider, Joana Hois, Alischa Rosenstein, Sabiha Ghellal, Dimitra Theofanou-Fülbier, and Ansgar RS Gerlicher. 2021. ExplAIn Yourself! Transparency for Positive UX in Autonomous Driving. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3411764.3446647

[64] Athar Sefid, Jian Wu, Prasenjit Mitra, and C. Lee Giles. 2019. Automatic Slide Generation for Scientific Papers. In *SciKnow@K-CAP*. http://ceur-ws.org/Vol-2526/paper2.pdf

[65] Emre Sezgin, Joseph Sirrianni, and Simon L Linwood. 2022. Operationalizing and Implementing Pretrained, Large Artificial Intelligence Linguistic Models in the US Health Care System: Outlook of Generative Pretrained Transformer 3 (GPT-3) as a Service Model. *JMIR Medical Informatics* 10, 2 (2022), e32875. https://doi.org/10.2196/32875

[66] Hanieh Shakeri, Carman Neustaedter, and Steve DiPaola. 2021. SAGA: Collaborative Storytelling with GPT-3. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. 163–166. https://doi.org/10.1145/3462204.3481771

[67] Rashi Shrivastava. 2023. *ChatGPT is coming to a customer service chatbot near you*. https://www.forbes.com/sites/rashishrivastava/2023/01/09/chatgpt-is-coming-to-a-customer-service-chatbot-near-you/

[68] Sangho Suh, Jian Zhao, and Edith Law. 2022. CodeToon: Story Ideation, Auto Comic Generation, and Structure Mapping for Code-Driven Storytelling. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. https://doi.org/10.1145/3526113.3545617

[69] Ekaterina Svikhnushina, Alexandru Placinta, and Pearl Pu. 2021. User Expectations of Conversational Chatbots Based on Online Reviews. In *Designing Interactive Systems Conference 2021*. https://doi.org/10.1145/3461778.3462125

[70] Rajan Vaish, Shirish Goyal, Amin Saberi, and Sharad Goel. 2018. Creating Crowdsourced Research Talks at Scale. In *Proceedings of the 2018 World Wide Web Conference*. 1–11. https://doi.org/10.1145/3178876.3186031

[71] Kento Watanabe, Yuichiroh Matsubayashi, Satoru Fukayama, Masataka Goto, Kentaro Inui, and Tomoyasu Nakano. 2018. A melody-conditioned lyrics language model. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. 163–172. https://doi.org/10.18653/v1/N18-1015

[72] Spencer Williams, Ridley Jones, Katharina Reinecke, and Gary Hsieh. 2022. An HCI Research Agenda for Online Science Communication. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–22. https://doi.org/10.1145/3555591

[73] Ning Wu, Ming Gong, Linjun Shou, Shining Liang, and Daxin Jiang. 2023. Large Language Models are Diverse Role-Players for Summarization Evaluation. *arXiv preprint arXiv:2303.15078* (2023). https://doi.org/10.48550/arXiv.2303.15078

[74] Yan Xia, Haiyi Zhu, Tun Lu, Peng Zhang, and Ning Gu. 2020. Exploring Antecedents and Consequences of Toxicity in Online Discussions: A Case Study on Reddit. *Proceedings of the ACM on Human-computer Interaction* 4, CSCW2 (2020), 1–23. https://doi.org/10.1145/3415179

[75] Sheng Xu and Xiaojun Wan. 2022. PosterBot: A System for Generating Posters of Scientific Papers with Neural Models. In *AAAI Conference on Artificial Intelligence*. https://doi.org/10.1609/aaai.v36i11.21738

[76] JungKyoon Yoon, Pieter MA Desmet, and Anna E Pohlmeyer. 2016. Developing Usage Guidelines for a Card-Based Design Tool: A Case of the Positive Emotional Granularity Cards. *Archives of Design Research* 29, 4 (2016), 5–18. https://doi.org/10.15187/adr.2016.11.29.4.5