The Infinite Index: Information Retrieval on Generative Text-To-Image Models

Niklas Deckers
Leipzig University and ScaDS.AI

Maik Fröbe
Friedrich-Schiller-Universität Jena

Johannes Kiesel
Bauhaus-Universität Weimar

Gianluca Pandolfo
Bauhaus-Universität Weimar

Christopher Schröder
Leipzig University

Benno Stein
Bauhaus-Universität Weimar

Martin Potthast
Leipzig University and ScaDS.AI

ABSTRACT

Conditional generative models such as DALL-E and Stable Diffusion generate images based on a user-defined text, the prompt. Finding and refining prompts that produce a desired image has become the art of prompt engineering. Generative models do not provide a built-in retrieval model for a user's information need expressed through prompts. In light of an extensive literature review, we reframe prompt engineering for generative models as interactive text-based retrieval on a novel kind of "infinite index". We apply these insights for the first time in a case study on image generation for game design with an expert. Finally, we envision how active learning may help to guide the retrieval of generated images.

CCS CONCEPTS

• **Information systems → Search engine indexing**; **Users and interactive retrieval**; **Image search**; **Novelty in information retrieval**; **Search engine architectures and scalability**; **Users and interactive retrieval**; **Image search**.

KEYWORDS

case study, evaluation, generative models, image retrieval

ACM Reference Format:
Niklas Deckers, Maik Fröbe, Johannes Kiesel, Gianluca Pandolfo, Christopher Schröder, Benno Stein, and Martin Potthast. 2023. The Infinite Index: Information Retrieval on Generative Text-To-Image Models. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '23), March 19–23, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3576840.3578327

1 INTRODUCTION

Conditional generative models allow the generation of a desired output based on a user-specified condition. For generative text-to-image models such as DALL-E [68] or Stable Diffusion [72], this means that the model generates images conditional on a text description known as a prompt. For a user, the prompt is the primary means of controlling the generated image. If an ad hoc prompt does not produce a satisfactory result, the user usually interacts with the model by adjusting the prompt until they get one, or they give up after a few tries. Since such systematic refinement of prompts is often necessary to achieve a satisfactory result, writing prompts has evolved into the art of *prompt engineering* [53, 63, 71], for which users exchange best practices in new communities. But even using examples from others, it's often not obvious how to change a prompt to steer image generation in a particular direction.

As a new perspective on the use of conditional generative models in general, we interpret them as a search engine index. Under this interpretation, the prompt is a request that represents a user's need for information. Prompt engineering can then be considered a form of interactive text-based retrieval, in which a user interacts with the model by modifying their prompt as if to refine their query to find a result that meets their needs. This raises a number of new challenges: When using a generative model, the initiative currently lies solely with the user, without support from the model as a "retrieval system". There is no intermediary retrieval model to help users produce satisfactory images fast(er), if not ad hoc. The manual refinement of prompts is not supported by system-side log analysis and query expansion. There is no operationalization of the concept of image relevance, which is needed for ranking images, and thus essential when many images are generated.

A striking difference from traditional retrieval is that when generative models are used as an index, new results are generated rather than existing ones retrieved.[1] A non-empty result is returned for every conceivable query prompt. This includes query prompts for which a traditional retrieval system would return no results. Also, the number of different results that can be generated per query prompt is not conceptually limited, but only by the available computational capacity for model inference. Thus, a generative model is effectively an "infinite index".

Our contribution is to explore this perspective on generative models as indexes in four ways, focusing on text-to-image generation: (1) Section 2 presents a literature survey on image generation,

[1] Generative model occasionally reproduce parts of their training data [46, 84].

CHIIR '23, March 19–23, 2023, Austin, TX, USA

This work is licensed under a Creative Commons Attribution International 4.0 License.
ACM ISBN 979-8-4007-0035-4/23/03.
https://doi.org/10.1145/3576840.3578327

text-based image retrieval, retrieval for creative tasks, and interactive retrieval. (2) Section 3 conceptualizes generative text-to-image models as an index integrated into a retrieval system: from the user perspective, the query language and interaction methods are presented, and from the system perspective, retrieval technologies capable of supporting retrieval are examined. Requirements for the evaluation of retrieval systems based on generative models are also presented. (3) Based on these findings, Section 4 presents a case study of image generation. For creative tasks in game design, we observe an expert and highlight several issues related to currently available technology. (4) Finally, based on the insights gained, Section 5 discusses an active learning approach to interactive retrieval to guide image generation using generative models.

2 BACKGROUND AND SURVEY

We review the relevant literature to place retrieval on generative models in the context of established concepts.

2.1 Image Generation

In image synthesis, Brock et al. [10] and Goodfellow et al. [28] have achieved promising results with generative adversarial networks (GANs) that allow images to be generated from the distribution of given training images. Autoregressive transformer models as per Razavi et al. [69] and Ramesh et al. [68] have proven to be effective for high-resolution image synthesis. Dhariwal and Nichol [21] has recently shown that diffusion models [83] are capable of outperforming traditional models such as GANs in image synthesis. In addition, Rombach et al. [72] have shown how to condition the generated images on text. This forms the basis for text-to-image models, which are often trained on datasets of text–image pairs [81].

Table 1 provides an overview of relevant text-to-image models, starting with diffusion models such as DALL·E by Ramesh et al. [68] and Imagen by Saharia et al. [74]. Most models are only accessible via a web interface. Their code and model weights are not publicly available. Stable Diffusion by Rombach et al. [72] achieved great impact not only because of its impressive results, but also because the model itself was made publicly available. As a result, it was rapidly adapted and now serves as the basis for numerous new applications. More recent approaches pursue other research goals: eDiff-I by Balaji et al. [4] introduces an ensemble of export-denoising networks that allow different behavior at different noise levels. This increases the number of parameters, but also improves the results. Muse by Chang et al. [13] uses a discrete token space instead of a pixel space to increase efficiency.

2.2 Image Retrieval

While text-to-image models are relatively new, image retrieval has a long history of research. Two cases are distinguished in the literature: In content-based image retrieval, the user enters an image as a query, while in text-based image retrieval, the user makes a textual query. Content-based image retrieval systems aim to bridge the gap between the semantic meaning of images and their quantified visual features through sophisticated image representations [50]. Once a collection of images is represented and indexed, the representation of the query image is used for similarity-based search and ranking. Text-based image retrieval has often focused on retrieval based on image metadata and tags in the past, which is why it is sometimes referred to as annotation-based, concept-based, or keyword-based image retrieval. Some approaches also generate textual representations for unannotated images, e.g., using optical character recognition [90], clustering images with and without annotations [52], or using image captioning methods [33].

Some studies have examined users' search interactions with a text-based image retrieval system. Choi [16] analyzed the search logs of 29 students and found that participants changed their textual queries more frequently to refine their results. Hollink et al. [31] studied the image search behavior of news professionals and showed that they often modified their queries by following semantic relationships of query terms, e.g., searching first for images about a person and then for images about their spouse.

Cho et al. [15] took a closer look at why people search for images. In their study of 69 papers, they identified seven information need categories (1) entertainment, (2) illustrations (explanation or clarification of details, e.g., creating presentation slides or preparing study material), (3) images for aesthetic appreciation (e.g., for desktop backgrounds), (4) knowledge construction (four sub-categories: information processing, information dissemination, learning, and ideation), (5) eye-catchers (e.g., to grab audiences' attention), (6) inspiring images, and (7) images for social interactions (e.g., images to trigger emotions). They also found seven categories of problems that could affect a user's ability to find the images they were looking for: (a) semantic issues, i.e., related to employed terminology, (b) content-based issues, i.e., related to describing content of images, (c) technical limitations of retrieval systems, (d) lacking aboutness or relevance of retrieved images, (e) lacking inclusivity with regard to cultural or linguistic aspects of the user, (f) lacking skills in handling search technology, and (g) cognitive overload. As we discuss in Section 3, most of these requirements and issues are also relevant to retrieval from text-to-image models.

2.3 User Feedback for Image Generation

Based on GANs, Ukkonen et al. [89] have proposed and implemented systems for relevance feedback and Liu et al. [55] for exploratory search. This was to overcome the lack of prompts in GANs to condition image generation, leaving users with little control over the generated images. Similar techniques to incorporate relevance feedback could be considered for text-to-image models.

2.4 Retrieval for Creative Tasks

Text-to-image models are particularly suited to artistic and creative applications, raising the question of whether there are parallels between such applications and the literature on creative task search. Interestingly, text-to-image models have quickly led to the formation of communities dedicated not only to the use of these tools, but also to prompt engineering and the sharing of successful image generation techniques.[2] This development is consistent with the formation of creative communities by artists in other art genres [29]. On the other hand, such strong community building is somewhat surprising, since artisans generally rely less on human sources [47].

Several studies have already specifically analyzed user behavior and goals in creative tasks. Chavula et al. [14] investigated the information behavior of 15 graduate students in creative web search

[2]The Midjourney Discord server has more than 8 million members (as of January 2023).

Table 1: Overview of the most relevant text-to-image models (web link; * replicated; † includes the text encoder).

Text-to-image model		Training data		Open Source			Reference		
Name	Parameters	Size	Source	Code	Data	Model	Publication	Link	Month / Year
DALL·E	12 B	n/a	Custom web crawl	*	–	*	Ramesh et al. [68]		01 / 2021
DALL·E 2	3.5 B	n/a	Custom web crawl, licensed sources	*	–	*	Ramesh et al. [67]		04 / 2022
Imagen	4.6 B	860 M	400 M [81] from Common Crawl	*	–	*	Saharia et al. [74]		05 / 2022
Midjourney	n/a	n/a	n/a	–	–	–	Salkowitz [77]		07 / 2022
Stable Diffusion	0.9 B	400 M	Common Crawl; cf. Schuhmann et al. [81]				Rombach et al. [72]		08 / 2022
eDiff-I	9.1 B†	n/a	n/a	–	–	–	Balaji et al. [4]		11 / 2022
Muse	3 B	460 M	n/a; cf. Saharia et al. [74]	–	–	–	Chang et al. [13]		01 / 2023

tasks using questionnaires and the think-aloud method. They identified four creative thinking processes that participants switched back and forth between: planning creative search tasks (i.e., deciding on a vague idea), searching for new ideas, synthesizing search results, and organizing ideas. Palani et al. [64] use log analyses and self-reports in a study of 34 design students. They observed three main goals of the students: To get an overview of the information space, to discover design patterns and criteria, and to get inspired and develop ideas. In the study, special attention was paid to the fact that participants initially had difficulty finding appropriate terms to describe their information needs, but then arrived at appropriate terms by quickly querying and reformulating queries. They also note that participants typically go through a divergent exploration phase before a convergent synthesis phase. Based on a previous online survey and study [103, 104], Li et al. [51] examine the information behavior in a diary study of 11 university students on self-selected creative tasks. They use Sawyer's eight-step creativity framework [78] and focus specifically on the use of information resources (search, images, Q&A, social sites, videos). They grouped them into five categories: Searching for specific information, supporting creative processes, learning definitional domains, learning procedural knowledge, and managing (organizing) found information. Especially with images, they distinguish specific uses (e.g., as on Pinterest, Instagram, Tumblr, Flickr, and image search): Support ideation and other creative processes, see finished examples, find out what one likes or dislikes, and manage and overview found information. They found that image search engines were primarily used to search for a wide range of images, while image sites like Pinterest and Instagram were often used to search for high-quality images by specific artists or professionals. In summary, we identify three common topics when searching for creative tasks: Searching to learn, to get inspired, and to get an overview. We also observe these behaviors in our case study (Section 4).

2.5 Interactive Retrieval

Interactive retrieval explores users' information behavior during and beyond search, as well as the development of new interaction methods to assist them [73]. In relation to our work, we review relevant research on query understanding based on query logs as a source of user interaction data.

Query Log Analysis. Joachims and Radlinski [39] introduced query log analysis for web search, which has since become a valuable tool, e.g., for improving retrieval effectiveness and studying user behavior [11, 35–37]. Broder [11], for example, established a taxonomy for web search queries showing that web search queries are divided into informational, navigational, and transactional queries, which is still the case today [1]. A further categorization derives from Jansen et al.'s [35] work on query reformulation: queries are either generalizations (subset of words), specializations (superset of words), synonyms, or other topics. Today, query logs are used for creating large training datasets for retrieval models based on transformers [60, 70] and remain an important asset.

Query Reformulation. Query reformulation approaches aim to improve the effectiveness of retrieval by replacing the original query with substituted or extended reformulations [20]. Here, the reformulation of a query can be either precision-oriented (when a term is replaced by a more specific one) or recall-oriented (when the query is expanded). Jansen et al. [37] shows that searchers do not start with perfect queries but reformulate them instead: more than 50% of searchers reformulate at least one query during a search. Approaches to automatic query expansion, such as RM3 [34], can use (pseudo) relevance feedback to add new (weighted) terms to the original query, thus solving the vocabulary mismatch problem that occurs in text retrieval. However, it is not yet clear which reformulations are helpful in which situations when working creatively with generative text-to-image models (i.e., precision-oriented or recall-oriented reformulations).

Query Suggestion. Search engines assist their users and offer a list of suggested queries for an input query [7], which is called query auto-completion [12] if the query is incomplete. Query suggestions are important; according to Feuer et al. [22], 30% of queries in a commercial query log are suggested to users beforehand. Likewise, Cucerzan and Brill [19] notes that spelling corrections are required for 10-15% of queries with spelling errors. In addition, query suggestions often aim to assist users by displaying related terms [32], where Jansen et al.'s [36] analysis shows that suggested related terms are also heavily used. However, it is important not to overwhelm users and rather show fewer alternatives for suggestions than many [96]. Overall, users value the interaction methods used in "traditional" search engines, and we believe that offering similar ones for retrieval interfaces built on generative text-to-image models will provide benefits to users with creative tasks.

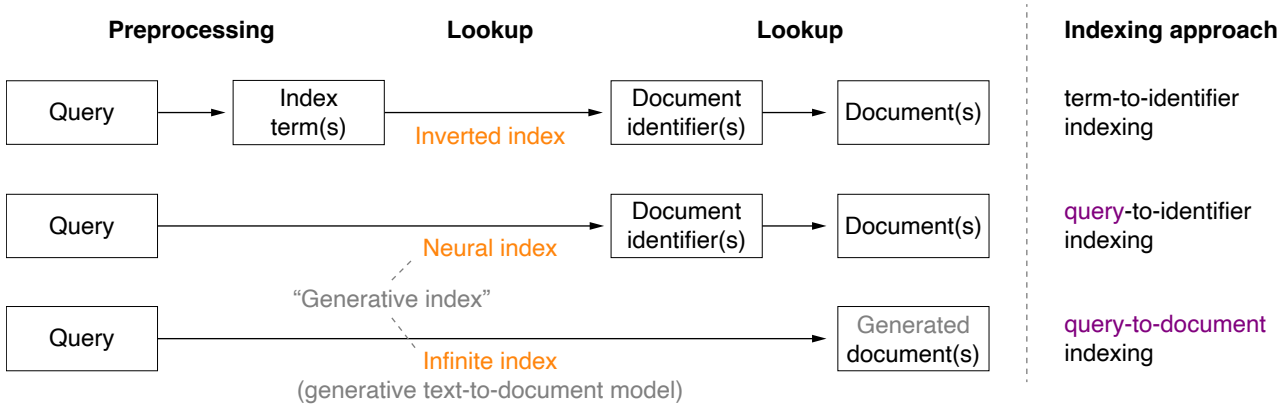

Figure 1: Overview of indexing approaches in information retrieval. The top row shows the classic term-to-identifier indexing approach, the middle rows the recent query-to-identifier indexing approach, and the bottom row the new query-to-document indexing approach introduced in this paper.

3 TEXT-TO-IMAGE GENERATION AS SEARCH

Considering a text-to-image model as a virtually infinite index, a prompt as a query, and prompt engineering as a form of user-driven query refinement yields a rudimentary retrieval system (Section 3.1). In the following, the interaction methods (Section 3.2) that are (potentially) available to users and the retrieval technologies (Section 3.3) that are (potentially) applicable to such a retrieval system are examined in detail. Subsequently, requirements for the evaluation of such a system are formulated (Section 3.4).

3.1 Classification of the "Infinite Index" in IR

Figure 1 shows how we place the concept of an infinite index in the context of known information retrieval concepts. The basic and most widely used concept of an (inverted) index was defined by Anderson [2] as "a systematic guide designed to indicate topics or features of documents or parts of documents." The topics or features of documents are represented by (index) terms. In modern information retrieval, these index terms correspond to the vocabulary of an indexed document collection. Anderson [2] further explains that "[t]he function of an index is to provide users with an effective and systematic means for locating documentary units (complete documents or parts of documents) that are relevant to information needs or requests." Specifically, the documents that can be looked up in an index are stored elsewhere, with an index lookup providing the necessary information that identifies the storage location of the matching documents within the filing system.

This concept of indexing, invented long before the days of computers, is still used today, in the form of data structures that fulfill the definition and function of an index in the above sense. Most importantly, the inverted index data structure implements a mapping of index terms to so-called postlists, where each postlist is a list of "postings" containing, among other things, a document identifier for locating the document within a file system or document store. Recently, index data structures have been revisited in the context of research on neural information retrieval [58, 87]: The neural index (the authors call it "transformer-based generative indexing") [6, 86, 95] has been proposed as a new type of index that mimics the function of a classical index by mapping queries directly to document identifiers. This mapping is trained based on a given document collection. Using an approach to predict queries that users might make to retrieve a given document, such as Doc2Query [61], it is straightforward to generate training examples consisting of a triple of query, document, and the document's identifier, or even just tuples of identifiers and synthetic queries [105]. The goal of the model is to predict the identifiers of the relevant documents given a query.

In this paper, we propose a different way of indexing by using generative text-to-document models as indexes. Although we focus on images as documents, this type of indexing is in principle applicable to all types of documents. In this scenario, the "index" is trained using documents and texts describing the document as training examples. Unlike the indexing approaches mentioned above, the resulting model does not necessarily retrieve the documents that were part of the document collection used to train the generative model, but rather generates new documents. Thus, this indexing approach is different from the other two, while it can be considered as a kind of independent neural indexing approach.

Altogether, we classify the three indexing approaches as follows:

- Term-to-identifier indexing: building a lookup table that maps index terms to document identifiers.
- Query-to-identifier indexing: training a model to predict identifiers of relevant documents for a query.
- Query-to-document indexing: training a model to generate relevant documents for a query.

To find a technical name for these indexes, the following alternatives are suitable: "generative index", or "neural index", or, "query-to-identifier index" vs. "query-to-document index", respectively.

3.2 Interaction Methods

Although the characteristic way of interacting with generative text-to-image models is the text prompt, other features have been rapidly added to the interfaces to support the process of image

generation. To illustrate the possibilities, we give here a brief snapshot of interaction methods based on the most common models (as of October 2022). Related generative text-to-video or text-to-3D models are not considered [30, 65].

Prompting. For generative text-to-image models, prompting the model is the primary interaction method. This interaction method serves as the initial point of contact with the model during image generation, much like a query in a standard web search. The interaction method is identical for both: the user sends a short text and receives images in response. Some interfaces of generative models allow images to be included in the prompt to steer the generated images in a particular direction, much in the same way that content-based image retrieval is used to find similar images. Unlike content-based image retrieval, model interfaces typically require that the prompt also contains text. Another aspect of prompting in some interfaces is the specification of model parameters along with the prompt, e.g., the size of the image to be generated or whether to generate tiled images, which is similar to filters (e.g., by size) in regular web search. Moreover, the negation operator allows to exclude certain terms from the generated image. The widely used Stable Diffusion model provides only a command line interface, but the community has implemented several graphical interfaces for it, for example one maintained by AUTOMATIC1111[3] (cf. Figure 2a).

In addition, several services have emerged in the larger text-to-image model generation ecosystem to assist users with prompt engineering. Specialized search engines allow users to search for images created with generative text-to-image models. The search engines then reveal the prompts used to generate the images they find, allowing prompts to be reused. Images are indexed either by their prompt or by the image content (e.g., with CLIP [66]). Examples of such search engines include the "community feed" of the Midjourney web app or the independent search engine Lexica, which indexes images from the Stable Diffusion Discord server (cf. Figure 2b). According to the developer, 1.4 million queries were made in a week, the index contained 12 million images in September 2022, and 5 million USD was earned, which clearly indicates the need for such systems. Other services enable (social) prompt engineering in a click interface[4] or even to buy prompts that supposedly provide consistent results.[5] Other projects carefully analyze how the prompt affects the result and create extensive lists of examples.[6] Although these services have a similar goal as query suggestions in web search, namely to help with prompt engineering, their interaction pattern is different. We discuss the implications in Section 5.

Variations. When generating an image, the variations interaction method allows to change parts of the image composition. This is useful when a generated image is broadly satisfactory but needs improvement in certain aspects. We distinguish three ways of generating variations: (1) the user does not change the prompt, which causes the composition to change only slightly and randomly (cf. Figure 2d); (2) the user changes the prompt and gives the model a new target as it continues from a generation checkpoint of the original image; (3) the user specifies semantic processing of the image, changing elements of the original image while preserving its original characteristics [40]. This interaction method, especially in the case of (1), is similar to the "show similar results" button in regular image retrieval. However, (2) and especially (3) allow a clearer specification of the need.

In- and Out-Painting. When generating an image, in- and out-painting allows to limit the generation of variations to user-defined areas of the image. This is useful when the user wants to change a certain area of the generated image (in-painting; cf. Figure 2c) or expand an image (out-painting), where the model tries to fill the region to match both the prompt and the parts of the original image at the edge of the region. This interaction method goes beyond the capabilities of regular search interfaces, and in most cases one would expect finite indexes to contain no matching results. For an infinite index, this interaction method can be extremely useful to finding images that satisfy multiple requirements.

Quality Enhancements. If the user is satisfied with the composition of an image, quality enhancement allows improving the image quality in one or more ways without changing the composition. The most common way to improve quality is to upscale the image to a higher resolution. There are often various upscaling methods that create new versions from a source image that look sharp or soft, realistic or artistic, without losing the original composition. Choosing a specific upscaling algorithm is useful to generate different images that should look similar in terms of their composition. Another type of enhancement is the use of image-to-image models trained specifically for correcting faces [94]. We anticipate that other image-to-image models specializing in specific operations will be integrated in the future. As with the variations tool, the closest counterpart to this method in regular search is the "show similar results" function, which can be quite effective for finding higher resolution images. However, quality enhancements allow a much clearer specification of what is needed by comparison.

Image-to-Text. If the user wants to rephrase the prompt but also use parts of the generated images, image-to-text models can be used to obtain a textual description of the image that reads like a prompt. We are not yet aware of any regular image search engine that integrates image-to-image models on the user page, although we believe that major image search engines such as Google Images will use them to index images.

3.3 Relevance in Text-To-Image Generation

As the above overview of interaction methods shows, the text-to-image generation community develops support for a variety of common search problems, but so far used information retrieval concepts only as search facets supported by external tools. This section reviews relevance as a core information retrieval concept that needs to be operationalized to steer the generation.

As with regular image retrieval, also generative models the concept of result relevance depends on the information needs of users, of which seven different categories have been identified in the literature (cf. Section 2). Generating images rather than finding them can, at least in theory, satisfy most of these needs, and is particularly useful for the needs of entertainment, illustration, aesthetic appreciation, engaging others, inspiration, and social interaction.

[3] https://github.com/AUTOMATIC1111/stable-diffusion-webui
[4] E.g., https://phraser.tech
[5] https://promptbase.com
[6] E.g., https://github.com/willwulfken/MidJourney-Styles-and-Keywords-Reference

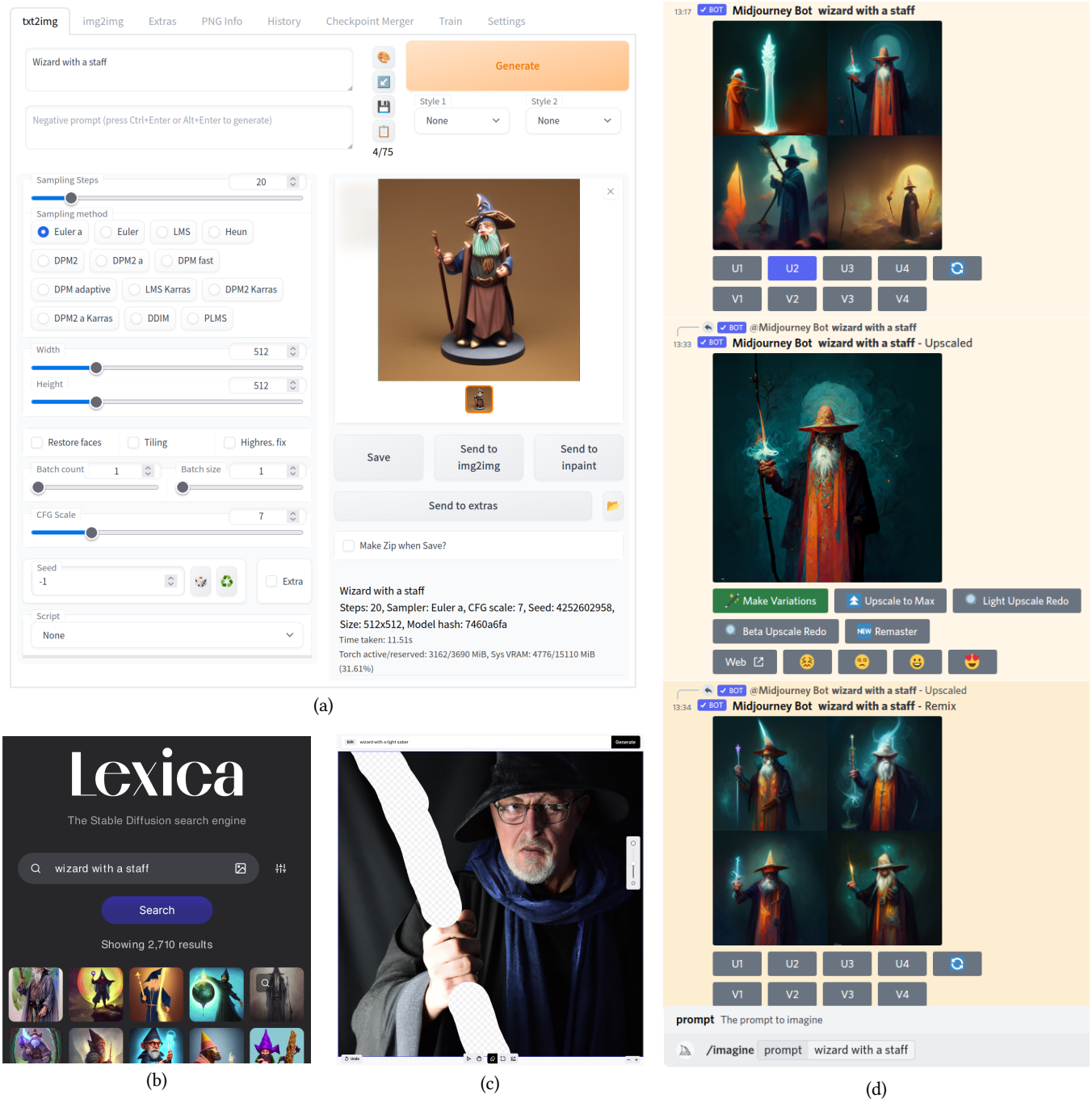

Figure 2: Screenshots illustrating the interfaces and interaction methods discussed in Section 3.2: (a) prompting in a community-maintained stable diffusion web interface; (b) Lexica search engine for generated images along their prompts; (c) in-painting in DALL·E 2 on an image originally created for the "Wizard with staff" prompt: the staff was manually masked (shown in white) to produce a modified prompt; (d) upscaling and variation generation in Midjourney.

In social interactions, for example, it is very useful for generative models to take into account the general moods mentioned in the prompt, providing a clear path to generating images that evoke specific emotions. The one information need category for which image generation is unsuitable is the need for knowledge construction, since generated images are not tied to real world knowledge.

When generating images, a distinction must be made between two different intentions. First, the user may already have a clear idea of the target image, for example, in an illustration. A user with this intention iteratively refines their prompt until the system generates an image that approximates their ideas, which we call a *descriptive approach*. Second, they may not have a clear vision or goal, just a set of constraints. With this intent, the user iteratively refines their prompt in a feedback-loop with random elements introduced by the system, loosely steering the system toward an image that they like and that meets the constraints, which we call the *creative approach*. Although they are very different from the user's point of view, both approaches are more or less indistinguishable for the system in terms of query log analysis: a general prompt is extended with details to become more specific.

With respect to text-to-image model-based retrieval, the research in interactive information retrieval is highly related (cf. Section 2). Query log analysis will be important to identify keywords in prompts that generally produce satisfactory results, to model user intent at a finer level, and to identify search queries and early abandonments that may indicate problems in the model. We assume that query suggestion methods will be very helpful, especially to assist inexperienced users. However, automatic query reformulation for prompts is more challenging because such changes have a generally more unpredictable impact on the generated images. In our case study (Section 4), the creative professional therefore refrained from optimizing the prompt and instead tried completely new ones. We see here a clear lack of user support in terms of retrieval in the current interfaces. External tools such as prompt search engines attempt to compensate for this shortcoming, but cannot match the effectiveness of integrated solutions that are widely used in search engines today (see Section 5 for a discussion of possible remedies).

With these considerations in mind, the notion of relevance and thus retrieval methods such as query suggestions can be transferred from information retrieval to text-to-image model generation, and thus retrieval evaluation measures can be adopted.

3.4 Evaluating Retrieval on Text-To-Image Models

Framing text-to-image generation as a retrieval problem implies measuring the effectiveness of generated rankings of generated images according to standard experimentation practices in information retrieval. However, we show that the infinite index in the form of a text-to-image model has far-reaching consequences for the design and evaluation of experiments, since the set of relevant documents is not closed and can thus not form the basis to calculate recall. We also discuss the challenges this poses for creating reusable benchmark collections and speculate on approaches to overcome these challenges. We focus on measuring ranking effectiveness because other aspects, such as user interface design and layout, are not considered in Cranfield-style evaluations.

Impact of the Infinite Index on IR Evaluation Measures. Effectiveness measures can be divided into utility-oriented (based on a ranking only) and recall-oriented (normalized by a "best possible" ranking) evaluation measures [56] so that an appropriate measure can be used depending on the nature of the information need. However, the virtually unending stream of alternative images that can be generated leads to problems with recall-oriented evaluation measures. An infinite number of images that can be generated allows for the subset of highly relevant images is to also be infinite. For recall-oriented measures like nDCG [38], this means that their normalization term can default to a ranking that is completely filled with highly relevant images. In practice, a human will still only search a query up to a certain rank k, so an nDCG@k can still be computed in this way, since a specific retrieval model requesting a text-to-image model may still deviate more or less from actually providing only highly relevant images. Utility-based measures (such as Precision@k, MRR, RBP [59], etc.) are not affected by this problem because they measure the effectiveness of a ranking based only on the images available in the ranking.

Another problem is that an infinite number of near-duplicate images of high relevance can be generated. Retrieval models could therefore rank many/exclusively (near-)duplicate images highly. If evaluated in isolation, each one would be considered highly relevant. Evaluation measures that operate on rankings with (near-)duplicates overestimate their effectiveness [5, 24], and learning-to-rank approaches learn suboptimal ranking models as well when trained on redundant data [23]. Therefore, it is important to deduplicate the rankings before evaluation. For the development of retrieval models, this means that ensuring diversity of images in the top ranks can be instrumental for users.

Overall, utility-based measures (such as RBP) on deduplicated rankings with judgments for the top-k images allow theoretically grounded evaluations when using text-to-image models as index.

Evaluations with Active Judgment Rounds. Experimental evaluation of retrieval systems usually follows the Cranfield paradigm [17, 18], which assumes that all documents are judged for all information needs. The original Cranfield experiments [17, 18] were conducted on a collection of 1,400 documents and complete relevance judgments for 225 topics. However, complete judgments became impracticable almost immediately thereafter as the size of collections increased significantly. The current best practice for shared tasks in IR is to create pools of the top-ranked documents from the submitted systems for each topic and then score each topic's pool [92], assuming that unjudged documents are not relevant. However, the assumption that judgment pools are "essentially complete" is likely incorrect when text-to-image models are used as index, especially if query expansion approaches are involved. As a result, rigorous evaluations must include manual rounds of judgments of unjudged images to reestablish "completeness" (e.g., for the top-k results), at least for utility-oriented measures, which hinders fully automated evaluations.

Evaluations without Active Judgment Rounds. IR research has benefited largely from the availability of robust and reusable test collections created during shared tasks [91]. However, these collections are robust only if most of the unjudged documents are irrelevant, which is not the case for text-to-image models. Consequently,

creating robust and reusable test collections is a major challenge that requires experience from several different shared tasks and subsequent post hoc experimentation (e.g., some robustness checks for traditional test collections are not performed until years after their creation [93]). Therefore, any post-hoc experiments based on an infinite index would need to include appropriate handling of unjudged images. Traditionally, unjudged documents are either simply removed (where a system's result lists are condensed to the included judged documents in their relative order) [75], classified as not relevant (default setting) or highly relevant (lower/upper bound) [56], or their relevance labeling can be predicted [3]. While these approaches are well studied for conventional retrieval experiments (e.g., condensed lists often overestimate the effectiveness values [76] and the gap between lower and upper bounds can be very large [56]), it is not yet clear whether they are suitable for an infinite index. As a result, it is not yet clear how to construct robust and reusable test collections, but we speculate that techniques from machine translation (e.g., measuring the similarity of an unjudged document via phash [100] to judged reference images) or relevance prediction may be appropriate.

3.5 The First Index for The Library of Babel?

At the beginning of the 20th century, Kurd Laßwitz, a German writer, scientist, and philosopher who became the first German science fiction author, introduced "The Universal Library" [45] as part of a series of short stories published in a newspaper around that time. The Universal Library contains every conceivable book with a length of 1 million characters. Assuming an alphabet of 100 Latin letters, numerals, and punctuation marks, each combination of these characters in a book of 1 million characters yields $10^{2,000,000}$ books, virtually everything that can been written in every language (assuming an appropriate transliteration). The only problem with such a library is that it is extremely unlikely to find a book by chance that contains a plausible sentence. This idea was taken up by Jorge Luis Borges, a well-known Argentine author, and made widely known under the name "The Library of Babel" [9]. He imagines this library as a universe of its own and invents stories about various tribes of humanity that might develop in such a place, always looking for scraps of knowledge among the many books of incomprehensible gibberish. In an earlier work called "The Total Library" [8], Borges traces the history of this concept back to Laßwitz and even to Aristotle and Cicero, who formulated what is now known as the "Infinite Monkey Theorem" [98], which states that a monkey hitting a typewriter at random will eventually type every text, including the complete works of William Shakespeare.

Given this fictional concept, generative text-to-document models can be understood as an index and a search engine for the library of Babel: By entering a short phrase as a query, the model is prompted to search the library for a document that matches the query. This completely circumvents the problem outlined by Laßwitz and Borges, since a document returned by a generative text-to-document model is very likely to be related to the query, and as long as the query itself is not gibberish, the retrieved documents will not be gibberish either.

4 CASE STUDY: GAME ARTWORK SEARCH

To illustrate retrieval using a query-to-document index for images, we report on an observational case study in which a text-to-image model is used for a creative task. First, we describe the study setup and the exemplary creative task, generating graphics for an online card game (Section 4.1). Subsequently, the main observations of the study are summarized (Section 4.2). A full report on the study is available as supplementary material.[7]

4.1 Setup of the Case Study

For the case study, we recruited a creative professional through personal contacts who allowed us to observe him as he explored the use of generative text-to-image models in his creative process. The professional described himself as a game designer and developer with the experience of five major game releases and as a lecturer in game development at a university. Prior to the case study, he described himself as very intrigued by generative text-to-image models he had come across in his Twitter feed, and had also seen some online videos on this technology ("2 minute papers"). Moreover, he had already generated about 50 images in DALL·E 2, about 20 in Midjourney, and less than 10 with Stable Diffusion on his own hardware, but none of them as part of a project. He anticipated, however, that generative text-to-image models will become very useful for the video game industry.[8]

Based on his experience, the professional decided to investigate the use of generative text-to-image models in the creation of graphics for an online card game for the study. Specifically, he was interested in developing a "deck-building online card game like Magic the Gathering set in a fantasy universe." In this game, each playing card has its own artwork that visually links it to the fantasy universe. Moreover, the cards belong to different "factions" that must be visually distinguishable. The professional opted for a "concept art-like style" from the outset. In the five hours we provided for the study, the professional expected to first create a "mood board" of images to capture the artistic style of the desired artwork [48], and then create the artwork itself for some cards. Based on his own testing, he decided to use Midjourney for this task. This choice reflects Midjourney's concept, which emphasizes "painterly aesthetics" and aims to help creatives "converge on the idea they want much more quickly" [77], especially at the beginning of a project.

The case study was conducted using the think-aloud method, asking additional questions while the professional waited for the images to be generated. Since the study did not focus on search interface design, one of the authors used Midjourney extensively to prepare for the study and provide technical support to the professional. To record observations, we took extensive notes as well as video and audio recordings and used the logging capabilities of the Midjourney web app. Following Li et al. [51], we used forms to structure our notes for various events, in our case for queries, problems, and shifts in design goals. A report on the study with all generated images is available as supplementary material.

[7] Case study report: https://doi.org/10.5281/zenodo.7221434

[8] Video games account for about 57% of digital media market revenue in 2022, or US$197 billion [85]. Meanwhile, other game developers have also published reports on their experiments with text-to-image models, e.g., https://www.traffickinggame.com/ai-assisted-graphics/

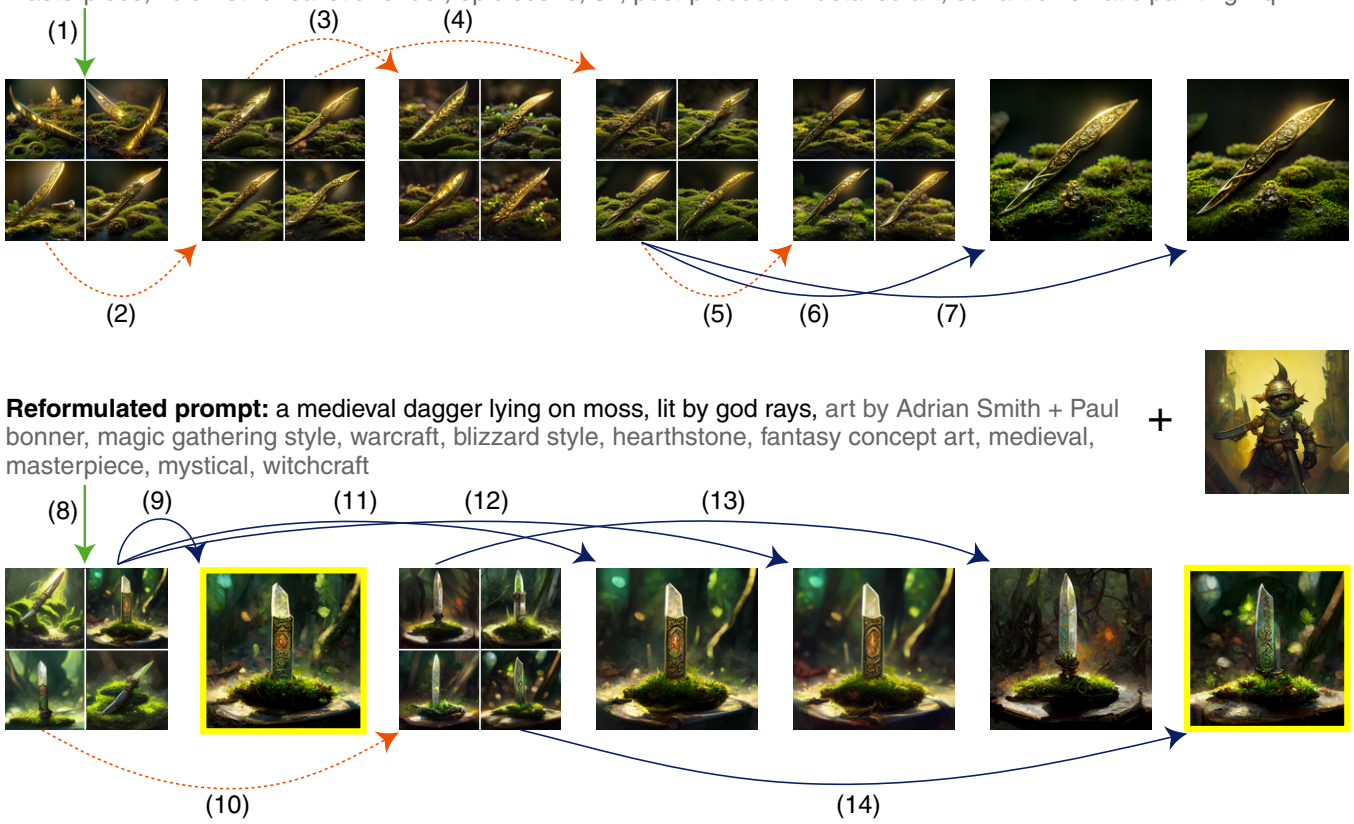

Figure 3: Exemplified search for a generated image from the case study, consisting of 14 steps in 22 minutes. Gray prompt text is copied from the prompt of another image in the mood board. For the reformulated prompt—after the first series of images has been abandoned as "leading nowhere"—, it is copied from the image that is part of the prompt. Interactions are text-to-image generation of four images (⟶), generating four variations of one image (same prompt, ⇢), and upscaling one image (⟶). The "beta" upscaling method is used in Step 12, the "light" method in Step 13, and the default method ("detailed") in all other cases. The professional kept the two images with the yellow border. Although the image generated in Step 9 did not show a dagger as intended, he found it intriguing and said that it evoked a story, especially in combination with the kept image.

4.2 Main Observations from the Case Study

This section summarizes the insights from the case study into three main observations. We found that the mood board is a key tool for professionals and analyze its use based on the five reasons for using information resources [51]. To analyze the mental state of the professional, we use Kuhlthau's [44] model of the information search process. And based on the professional's comments during the study, we identified the lack of control he mentioned as the main problem that needs to be addressed by future tools.

The mood board as prompt library. Lemarchand [48] defines a mood board as "a single page or screen of pictures arranged around a certain idea or theme" that serve two main purposes: first, to inspire new ideas by juxtaposing images (supporting creative processes), and second, to communicate a concept quickly and effectively (managing found information). After creating the mood board from images in Midjourney's community feed, however, the professional immediately began using the mood board as a source for his prompts as well. When creating a new image, he selected from the mood board the image that came closest to his ideas in terms of artistic style, and then copied the "style part" of that image's prompt for his own creation (cf. the gray text in Figure 3). Thus, he additionally used the mood board to learn domain knowledge (style names, rendering engines, etc.) and procedural knowledge (parameters such as "`--q 2`" to increase image quality). Only once did the professional search for the artists of the "Magic the Gathering" cards using an external search engine and was pleased to find that they were already included in the prompts he copied. Learning happened only on a superficial level, copying entire style sections of a prompt and using it like an atomic unit. This behavior is so widespread in the text-to-image generation community at the moment that commercial services have emerged for them.[9]

[9] E.g., https://promptbase.com

Uncertainty never fully ceases. In Kuhlthau's [44] model of the information-seeking process, the seeker moves from uncertainty to understanding as the search progresses. During the case study, we were able to identify clear parallels to this model and its phases, particularly the selection, exploration, formulation, and collection phases. In the selection phase, the professional uses the mood board as inspiration to choose content and style for a new image. In the exploration phase, he created and modified the prompt: he mentioned that he was very unsure about the results he would get and how he could modify the prompt to achieve what he envisioned. Once he found something he thought was promising, he moved into the formulation phase, focusing on generating variations over and over again and figuring out certain aspects that the final image should have. With a clear sense of direction, he would then upscale matching images in the acquisition phase and test the various upscaling algorithms as necessary. As accounted for in the model, the professional also regressed to earlier stages, especially when he saw an impasse (cf. Figure 3). Kuhlthau, however, mentions two "types of uncertainty," and although uncertainty about the concept (what he is looking for) decreases as described above, uncertainty about the technical process (how to get there) remains high, with the AI remaining largely unpredictable to him.

Sense of direction, but lack of control. Although in some situations the professional noted that the unpredictability inherent in the process was appealing ("I also wanted to be surprised"), he also mentioned that the process was very exhausting, which we related to the fact that he often went back in the history of his generated images to keep checking which interactions yielded good results and which image he should continue with. An interface that supports the user in organizing generated images therefore seems necessary. The professional noted that he was developing a sense of the direction the image variations would take, but also felt he had no control. He decided whether to continue down one path or try another, but did not feel he could change direction. After the case study, Midjourney introduced the ability to modify a prompt when generating variations, but the professional says this does not solve the problem of choosing the right words. Uncertainty about how to change the prompt to achieve the desired results therefore has a major negative impact on the user's sense of control.

Indeed, the case study showed clear parallels between text-to-image generation and image search. In particular, we found that existing theoretical models of the (creative) search process are broadly applicable. The main difference lies in the never-ending uncertainty about how to get to a particular result—although the user must assume, because of the index being virtually infinite, that there is a path that leads to the goal. Based on our observations, we believe that tools that provide the user with more intuitive ways to control the generation process are needed to bridge this gap.

5 DISCUSSION

Based on our conceptualization of text-to-image models as search and the case study, we next explore the limitations of text-to-image generation (Section 5.1). Then we discuss how active learning might help (Section 5.2), and address ethical concerns (Section 5.3).

5.1 Limitations of Text-To-Image Generation

While the functionality of text-to-image models is already of sufficient quality to be used in real-world applications [62, 77], we identified the following two main limitations related to the workflow or capabilities of the current methods—the same workflow that was used in the case study.

Prompt Engineering. Although prompt engineering has been successfully applied to other generative tasks such as co-writing screenplays and theater scripts with a large language model [57], the need to engineer the prompt compromises the intuitiveness of the prompt interface. Users quickly realized that iteratively adding modifiers to the prompt (as in Section 4), causing the model to apply the desired result styles to the generated image, is the most effective way to control the image generation process [53, 63]. This has given rise to a whole new subfield of text-to-image prompt engineering [53], where prompts are increasingly becoming long strings of keywords instead of text descriptions. These manipulated prompts resemble highly optimized search engine queries where users select and fill in keywords—so users have learned to adapt to the algorithm rather than the other way around.

Influence of the Training Data. A fundamental limitation of current models is that both text encoders and diffusion models generate new data by merging concepts learned from large datasets and are thus limited to those concepts. Writing a prompt that contains a concept that does not appear in either the text corpora or the image datasets is likely to result in sub-par generation of images. One possible remedy is that unknown terms can be described as paraphrases. If the training data does not contain images of a centaur, a prompt such as "`a mythical creature with the body of a horse and the torso of a human`" might still produce the desired result.

5.2 Active Learning for Text-To-Image Generation

From the case study, it appears that targeted text-to-image generation is already surprisingly effective. As described in Section 3.2, the current way of working amounts to iterative prompt engineering, which in turn is a fundamental limitation, as stated in Section 5.1. We propose active learning as a solution to this problem and outline how it can be integrated as a feedback mechanism in an image generation workflow that uses text-to-image models.

Active learning [49, 102] is an iterative approach to classification that involves a feedback loop involving a user and a (semi-)supervised machine learning model. It is intended for scenarios where training data is not available to minimize the effort required to obtain a suitable labeled training dataset while maximizing model quality. According to Schohn and Cohn [79], an active learning setting consists of (1) a model that is trained for a specific task, (2) a query strategy that selects data from an existing resource or generates new data to be labeled, and (3) a stopping criterion that indicates at what point continuing the process is unlikely to sufficiently improve the result any further. At each iteration, the query strategy selects the examples it deems most informative for the model, for example, based on the prediction uncertainty of the model [80]. These examples are then annotated by the user according to the task at hand. A new model is then trained on all previously marked data, and the loop is repeated until an objective stopping criterion is met or the user stops.

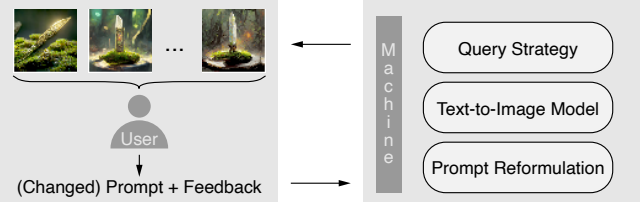

Figure 4: A conceptual overview of the active learning loop for the guided text-to-image generation use case.

For text-to-image generation, the whole structure of active learning is shown in Figure 4. The process begins with the user and an initial prompt. The active learning model learns to reformulate prompts, which in turn are passed to the text-to-image model. The model is trained with user feedback as target values, so that the resulting images should become increasingly appealing to the user. Subsequently, the query strategy decides which images are displayed to the user. It strikes a balance between exploration and exploitation, a well-known trade-off in information retrieval: exploration selects images that are different from the current best candidates, and exploitation selects images that are close to the current best solutions. Finally, the stopping criterion is the user who stops the process as soon as his information need is satisfied. In this setup, active learning uses relevance feedback [88, 99, 102].

Information retrieval systems can let users explicitly specify relevant documents (explicit relevance feedback) or learn from passive observations (implicit relevance feedback) [97], though this discussion focuses on explicit feedback to guide active learning for image retrieval. There are different types of explicit relevance feedback for the user: (1) binary relevance feedback [27], where the user rates each image as "unappealing" or "appealing" with respect to the target concept; (2) graded relevance feedback [27], in which the user rates each image from "unappealing" to "appealing" on a multilevel scale (e.g., from 0 to 5); (3) ranking, where the user rates each image (possibly including images from previous iterations) from unappealing to appealing. Users can provide feedback on the entire image or on individual parts (e.g., the background) or aspects (e.g., the color scheme). Similar to query customization during a regular search, the user can change the prompt in each iteration.

The main challenge for this feedback mechanism is to convert the images into a textual representation that preserves the specifics of each image, which can then be used to learn how to reformulate the prompt. For example, a prompt like "`wizard with staff`" could generate images with different poses and backgrounds. To learn reformulations from relevance feedback, it is necessary to obtain a textual representation that includes these differences. One could, of course, try to learn to reformulate based only on latent image representation and relevance feedback, but this would solve the problem exclusively in the image space and largely ignore the text embedding space. This could also be a useful approach, but is outside the realm of natural language processing and information retrieval. Although the reverse step of image-to-image generation required for this has recently attracted increasing attention [25, 26], it remains a challenge, and moreover, multiple images are required to generate one text [25]. Once this reverse direction is improved, the full spectrum of natural language processing and information retrieval can be applied to effectively process user feedback to improve prompts during the reformulation step.

When text-to-image generation is viewed as a retrieval problem (as in Section 3), the process of trying different prompts until a satisfactory image is generated is similar to traditional image retrieval, and thus the inclusion of active learning as a relevance feedback mechanism is an obvious choice of a well-established method. We anticipate that active prompt generation will be a strong interface competitor for generative text-to-image models once image-to-text models are sufficiently mature (apart from editing options such as in-painting or out-painting, which are orthogonal to this approach).

5.3 Ethical Concerns

A computational approach powerful enough to generate documents such as images, text, and other media types at a quality difficult to distinguish at times from human-made illustrations naturally raises ethical concerns. We discuss the most important ones below.

Will algorithms replace artists? We begin with the obvious question: will generative text-to-image models threaten artists' jobs? First, based on our experience in the case study, it is currently difficult to get text-to-image models to generate a desired result. The decision whether the generated images represent the desired scene with sufficient quality still has to be made by the user. Therefore, we believe that these new models will be a powerful tool, but will not replace the human illustrator in the foreseeable future—even if the image quality should eventually reach human levels. This is corroborated by others such as Liu et al. [54], who developed and evaluated a system that assists users in generating images for news articles, noting that artistic knowledge is still beneficial to the generated result, explicitly saying "generative AI deployment should [...] augment rather than [...] replace human creative expertise". We support this view: instead of an autonomous AI that acts on its own, we want to emphasize the benefits of a "supportive AI" that inquires about and incorporates the decisions of its users.

Who is the author of a generated image? And who owns the rights? This is currently an unresolved situation that leads to uncertainties regarding the use of AI-generated images. For this reason, major platforms such as the well-known image provider Getty Images have recently banned all AI-generated content.[10] Stakeholders may include the user, the creators, and the artists who created the images used for model training. Ultimately, this decision must be made by policy makers and by the courts, where many legal precedents have been set in the past through copyright litigation.

Text-to-image models for generating misinformation? Generated misinformation is already a pervasive problem and is widely discussed in the context of so-called "deep fakes" and AI-generated text [43, 82, 101]. To mitigate this problem in text-to-image models such as Stable Diffusion, an image is watermarked to identify it as artificially generated.[11] Although watermarks are not easy to remove, this may not be enough if they are not checked on virtually all devices. However, this requires that policymakers legally oblige device manufacturers to detect fakes and warn users. In addition,

[10] https://voicebot.ai/2022/09/23/getty-images-removes-and-bans-ai-generated-art/
[11] https://github.com/CompVis/stable-diffusion

watermarking images itself raises privacy concerns. As for text-to-text models, fully generated documents can be useful, provided they are not used to generate factual knowledge, which is currently woefully inadequate. Therefore, the use of such models as an infinite index must at least be subjected to post-processing in the form of fact checking or the like. This is exactly what is happening at present, after OpenAI recently introduced ChatGPT[12] with lots of publicity: The search engines You[13] and Neeva[14] have already integrated facsimiles of ChatGPT into their search interfaces and check the generated documents against traditional search results. Whether this proves to be a good idea remains to be seen.

Do these models express or even amplify bias? Bias in training data is a known problem for both image data [41] and language models [42]. Therefore, text-to-image models must also be systematically screened for social and other types of bias. In information retrieval, for example, fair ranking is now a widely studied problem. A retrieval process built on generative models could be designed to mitigate their inherent biases. Image search engines based on generative models must post-process and re-rank their results to compensate for bias, just like their traditional counterparts. However, the technologies developed for traditional search engines can also be applied to search engines based on generative models.[15]

6 CONCLUSION

Supporting systems and services are needed for the use of generative text-to-image models. Their integration into existing systems is already in full swing, as has been seen for years in generative models for writing assistance and translation systems, but now also in more creative areas. However, integration with end-user software to create slide presentations or artwork will not meet all the needs of those looking for inspirational images. Given the recent moves by You and Neeva, specialized search engines based on generative text-to-image models as indexes, with user interface for formulating information needs and customized retrieval models, are probably already being developed. However, the development of a search engine is not trivial, and the information retrieval community faces the renewed challenge of developing an understanding and technological foundation for such search engines. This includes the development of new retrieval models and relevance scores as well as the adaptation of evaluation methods for benchmarking search engines based on generative models. Moreover, because results can vary widely from one day to the next (cf. Figure 5), users cannot rely on things like remembering specific queries to search for known items. Therefore, to effectively use generative image models as a search index, it may be necessary to maintain a history of search results with appropriate model parameters. Finally, what is true for generative text-to-image models is likely to be true other kinds of text-to-document models, opening up a whole new world of exciting new research directions and promising high impact.

[12]https://openai.com/blog/chatgpt/
[13]https://blog.you.com/a9e05080c8ea
[14]https://neeva.com/blog/introducing-neevaai
[15]For example, compare the result from lexica.art (https://lexica.art/?q=nurse) with that from Google Images (https://www.google.com/search?tbm=isch&q=nurse).

Figure 5: Results for the prompt "wizard with a staff" in Midjourney: (left) version 3, default at the time of our case study; (right) version 4, the default three months later.

REFERENCES

[1] Daria Alexander, Wojciech Kusa, and Arjen P. de Vries. 2022. ORCAS-I: Queries Annotated with Intent using Weak Supervision. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*, Enrique Amigó, Pablo Castells, Julio Gonzalo, Ben Carterette, J. Shane Culpepper, and Gabriella Kazai (Eds.). ACM, 3057–3066. https://doi.org/10.1145/3477495.3531737

[2] James D. Anderson. 1997. Guidelines for Indexesand RelatedInformationRetrieval Devices.

[3] Javed A. Aslam and Emine Yilmaz. 2007. Inferring document relevance from incomplete information. In *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, CIKM 2007, Lisbon, Portugal, November 6-10, 2007*, Mário J. Silva, Alberto H. F. Laender, Ricardo A. Baeza-Yates, Deborah L. McGuinness, Bjørn Olstad, Øystein Haug Olsen, and André O. Falcão (Eds.). ACM, 633–642.

[4] Yogesh Balaji, Seungjun Nah, Xun Huang, Arash Vahdat, Jiaming Song, Karsten Kreis, Miika Aittala, Timo Aila, Samuli Laine, Bryan Catanzaro, Tero Karras, and Ming-Yu Liu. 2022. eDiff-I: Text-to-Image Diffusion Models with an Ensemble of Expert Denoisers. *CoRR* abs/2211.01324 (2022). https://doi.org/10.48550/arXiv.2211.01324 arXiv:2211.01324

[5] Yaniv Bernstein and Justin Zobel. 2005. Redundant documents and search effectiveness. In *Proceedings of the 2005 ACM CIKM International Conference on Information and Knowledge Management, Bremen, Germany, October 31 - November 5, 2005*, Otthein Herzog, Hans-Jörg Schek, Norbert Fuhr, Abdur Chowdhury, and Wilfried Teiken (Eds.). ACM, 736–743. https://doi.org/10.1145/1099554.1099733

[6] Michele Bevilacqua, Giuseppe Ottaviano, Patrick Lewis, Wen-tau Yih, Sebastian Riedel, and Fabio Petroni. 2022. Autoregressive Search Engines: Generating Substrings as Document Identifiers. *CoRR* abs/2204.10628 (2022). https://doi.org/10.48550/arXiv.2204.10628 arXiv:2204.10628

[7] Sumit Bhatia, Debapriyo Majumdar, and Prasenjit Mitra. 2011. Query suggestions in the absence of query logs. In *Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2011, Beijing, China, July 25-29, 2011*, Wei-Ying Ma, Jian-Yun Nie, Ricardo Baeza-Yates, Tat-Seng Chua, and W. Bruce Croft (Eds.). ACM, 795–804. https://doi.org/10.1145/2009916.2010023

[8] Jorge Luis Borges. 1939. *La Bibliotheca Total (The Total Library).* Buenos Aires. https://www.gwern.net/docs/borges/1939-borges-thetotallibrary.pdf

[9] Jorge Luis Borges. 1941. *La Bibliotheca de Babel (The Library of Babel).* https://maskofreason.files.wordpress.com/2011/02/the-library-of-babel-by-jorge-luis-borges.pdf

[10] Andrew Brock, Jeff Donahue, and Karen Simonyan. 2019. Large Scale GAN Training for High Fidelity Natural Image Synthesis. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net. https://openreview.net/forum?id=B1xsqj09Fm

[11] Andrei Z. Broder. 2002. A taxonomy of web search. *SIGIR Forum* 36, 2 (2002), 3–10. https://doi.org/10.1145/792550.792552

[12] Fei Cai and Maarten de Rijke. 2016. A Survey of Query Auto Completion in Information Retrieval. *Found. Trends Inf. Retr.* 10, 4 (2016), 273–363. https://doi.org/10.1561/1500000055

[13] Huiwen Chang, Han Zhang, Jarred Barber, AJ Maschinot, Jose Lezama, Lu Jiang, Ming-Hsuan Yang, Kevin Murphy, William T Freeman, Michael Rubinstein, et al. 2023. Muse: Text-To-Image Generation via Masked Generative Transformers. *arXiv preprint arXiv:2301.00704* (2023).

[14] Catherine Chavula, Yujin Choi, and Soo Young Rieh. 2022. Understanding Creative Thinking Processes in Searching for New Ideas. In *ACM SIGIR Conference on Human Information Interaction and Retrieval* (Regensburg, Germany). ACM, New York, NY, USA, 321–326. https://doi.org/10.1145/3498366.3505783

[15] Hyerim Cho, Minh TN Pham, Katherine N. Leonard, and Alex C. Urban. 2021. A systematic literature review on image information needs and behaviors. *Journal of Documentation* 78, 2 (2021), 207–227.

[16] Youngok Choi. 2013. Analysis of image search queries on the web: Query modification patterns and semantic attributes. *Journal of the American Society for Information Science and Technology* 64, 7 (2013), 1423–1441.

[17] Cyril W. Cleverdon. 1967. The Cranfield tests on index language devices. In *Aslib proceedings*. MCB UP Ltd. (Reprinted in Readings in Information Retrieval, Karen Sparck-Jones and Peter Willett, editors, Morgan Kaufmann, 1997), 173–192.

[18] Cyril W. Cleverdon. 1991. The Significance of the Cranfield Tests on Index Languages. In *Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Chicago, Illinois, USA, October 13-16, 1991 (Special Issue of the SIGIR Forum)*, Abraham Bookstein, Yves Chiaramella, Gerard Salton, and Vijay V. Raghavan (Eds.). ACM, 3–12.

[19] Silviu Cucerzan and Eric Brill. 2004. Spelling Correction as an Iterative Process that Exploits the CollectiveKnowledge of Web Users. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing , EMNLP 2004, A meeting of SIGDAT, a Special Interest Group of the ACL, held in conjunction with ACL 2004, 25-26 July 2004, Barcelona, Spain*. ACL, 293–300. https://aclanthology.org/W04-3238/

[20] Van Dang and W. Bruce Croft. 2010. Query reformulation using anchor text. In *Proceedings of the Third International Conference on Web Search and Web Data Mining, WSDM 2010, New York, NY, USA, February 4-6, 2010*, Brian D. Davison, Torsten Suel, Nick Craswell, and Bing Liu (Eds.). ACM, 41–50. https://doi.org/10.1145/1718487.1718493

[21] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems* 34 (2021), 8780–8794.

[22] Alan Feuer, Stefan Savev, and Javed A. Aslam. 2007. Evaluation of phrasal query suggestions. In *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, CIKM 2007, Lisbon, Portugal, November 6-10, 2007*, Mário J. Silva, Alberto H. F. Laender, Ricardo A. Baeza-Yates, Deborah L. McGuinness, Bjørn Olstad, Øystein Haug Olsen, and André O. Falcão (Eds.). ACM, 841–848. https://doi.org/10.1145/1321440.1321556

[23] Maik Fröbe, Janek Bevendorff, Jan Heinrich Reimer, Martin Potthast, and Matthias Hagen. 2020. Sampling Bias Due to Near-Duplicates in Learning to Rank. In *43rd International ACM Conference on Research and Development in Information Retrieval (SIGIR 2020)*. ACM, 1997–2000. https://doi.org/10.1145/3397271.3401212

[24] Maik Fröbe, Jan Philipp Bittner, Martin Potthast, and Matthias Hagen. 2020. The Effect of Content-Equivalent Near-Duplicates on the Evaluation of Search Engines. In *Advances in Information Retrieval. 42nd European Conference on IR Research (ECIR 2020) (Lecture Notes in Computer Science, Vol. 12036)*, Joemon M. Jose, Emine Yilmaz, João Magalhães, Pablo Castells, Nicola Ferro, Mário J. Silva, and Flávio Martins (Eds.). Springer, Berlin Heidelberg New York, 12–19. https://doi.org/10.1007/978-3-030-45442-5_2

[25] 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. *CoRR* abs/2208.01618 (2022). https://doi.org/10.48550/arXiv.2208.01618 arXiv:2208.01618

[26] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H. Bermano, Gal Chechik, and Daniel Cohen-Or. 2022. Imagen Video: High Definition Video Generation with Diffusion Models. *CoRR* abs/2208.01618 (2022). https://doi.org/10.48550/arXiv.2208.01618 arXiv:2208.01618

[27] Gregory Gay, Sonia Haiduc, Andrian Marcus, and Tim Menzies. 2009. On the use of relevance feedback in IR-based concept location. In *25th IEEE International Conference on Software Maintenance (ICSM'09)*. IEEE Computer Society, 351–360. https://doi.org/10.1109/ICSM.2009.5306315

[28] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.

[29] William S. Hemmig. 2008. The information-seeking behavior of visual artists: a literature review. *Journal of Documentation* 64, 3 (2008), 343–362. https://doi.org/10.1108/00220410810867579

[30] Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey A. Gritsenko, Diederik P. Kingma, Ben Poole, Mohammad Norouzi, David J. Fleet, and Tim Salimans. 2022. Imagen Video: High Definition Video Generation with Diffusion Models. *CoRR* abs/2210.02303 (2022). https://doi.org/10.48550/arXiv.2210.02303 arXiv:2210.02303

[31] Vera Hollink, Theodora Tsikrika, and Arjen P. de Vries. 2011. Semantic search log analysis: A method and a study on professional image search. *J. Assoc. Inf. Sci. Technol.* 62 (2011), 691–713.

[32] Chien-Kang Huang, Lee-Feng Chien, and Yen-Jen Oyang. 2003. Relevant term suggestion in interactive web search based on contextual information in query session logs. *J. Assoc. Inf. Sci. Technol.* 54, 7 (2003), 638–649. https://doi.org/10.1002/asi.10256

[33] Sethurathienam Iyer, Shubham Chaturvedi, and Tirtharaj Dash. 2017. Image Captioning-Based Image Search Engine: An Alternative to Retrieval by Metadata. In *Soft Computing for Problem Solving (SocProS'17) (Advances in Intelligent Systems and Computing, Vol. 817)*, Jagdish Chand Bansal, Kedar Nath Das, Atulya Nagar, Kusum Deep, and Akshay Kumar Ojha (Eds.). Springer, 181–191. https://doi.org/10.1007/978-981-13-1595-4_14

[34] Nasreen Abdul Jaleel, James Allan, W. Bruce Croft, Fernando Diaz, Leah S. Larkey, Xiaoyan Li, Mark D. Smucker, and Courtney Wade. 2004. UMass at TREC 2004: Novelty and HARD. In *Proceedings of the Thirteenth Text REtrieval Conference, TREC 2004, Gaithersburg, Maryland, USA, November 16-19, 2004 (NIST Special Publication, Vol. 500-261)*, Ellen M. Voorhees and Lori P. Buckland (Eds.). National Institute of Standards and Technology (NIST). http://trec.nist.gov/pubs/trec13/papers/umass.novelty.hard.pdf

[35] Bernard Jansen, D. Booth, and A. Spink. 2009. Patterns of Query Reformulation During Web Searching. *J. Assoc. Inf. Sci. Technol.* 60, 7 (2009), 1358–1371. https://doi.org/10.1002/asi.21071

[36] Bernard Jansen, Amanda Spink, and Sherry Koshman. 2007. Web searcher interaction with the Dogpile.com metasearch engine. *J. Assoc. Inf. Sci. Technol.* 58, 5 (2007), 744–755. https://doi.org/10.1002/asi.20555

[37] Bernard Jansen, Amanda Spink, and Jan Pedersen. 2005. A temporal comparison of AltaVista Web searching. *J. Assoc. Inf. Sci. Technol.* 56, 6 (2005), 559–570. https://doi.org/10.1002/asi.20145

[38] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.* 20, 4 (2002), 422–446.

[39] Thorsten Joachims and Filip Radlinski. 2007. Search Engines that Learn from Implicit Feedback. *Computer* 40, 8 (2007), 34–40. https://doi.org/10.1109/MC.2007.289

[40] Bahjat Kawar, Shiran Zada, Oran Lang, Omer Tov, Huiwen Chang, Tali Dekel, Inbar Mosseri, and Michal Irani. 2012. Imagic: Text-Based Real Image Editing with Diffusion Models. *CoRR* abs/2210.09276 (2012). arXiv:2210.09276 http://arxiv.org/abs/2210.09276

[41] Aditya Khosla, Tinghui Zhou, Tomasz Malisiewicz, Alexei A. Efros, and Antonio Torralba. 2012. Undoing the Damage of Dataset Bias. In *Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 7572)*, Andrew W. Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid (Eds.). Springer, 158–171. https://doi.org/10.1007/978-3-642-33718-5_12

[42] Hannah Rose Kirk, Yennie Jun, Filippo Volpin, Haider Iqbal, Elias Benussi, Frederic Dreyer, Aleksandar Shtedritski, and Yuki Asano. 2021. Bias Out-of-the-Box: An Empirical Analysis of Intersectional Occupational Biases in Popular Generative Language Models. In *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (Eds.), Vol. 34. Curran Associates, Inc., 2611–2624. https://proceedings.neurips.cc/paper/2021/file/1531beb762df4029513ebf9295e0d34f-Paper.pdf

[43] Sarah Kreps, R. Miles McCain, and Miles Brundage. 2022. All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation. *Journal of Experimental Political Science* 9, 1 (2022), 104–117. https://doi.org/10.1017/XPS.2020.37

[44] Carol Collier Kuhlthau. 1993. A Principle of Uncertainty for Information seeking. *J. Documentation* 49, 4 (1993), 339–355. https://doi.org/10.1108/eb026918

[45] Kurd Laßwitz. 1897. *Bis zum Nullpunkt des Seins und andere Science-Fiction-Erzählungen (Kapitel 10: Die Universalbibliothek.* Schlesische Zeitung; Neuauflage auf Projekt Gutenberg 2017. https://www.projekt-gutenberg.org/lasswitz/nullpunk/titlepage.html Erschienen zwischen 1871 und 1908.

[46] Jooyoung Lee, Thai Le, Jinghui Chen, and Dongwon Lee. 2022. Do Language Models Plagiarize? *CoRR* abs/2203.07618 (2022). https://doi.org/10.48550/arXiv.2203.07618 arXiv:2203.07618

[47] Lo Lee, Melissa G. Ocepek, Stephann Makri, George Buchanan, and Dana McKay. 2019. Getting creative in everyday life: Investigating arts and crafts hobbyists' information behavior. *Proceedings of the Association for Information Science and Technology* 56, 1 (2019), 703–705. https://doi.org/10.1002/pra2.141 arXiv:https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/pra2.141

[48] Richard Lemarchand. 2021. *A Playful Production Process: For Game Designers (and Everyone).* MIT Press, Cambridge, MA.

[49] David D. Lewis and William A. Gale. 1994. A Sequential Algorithm for Training Text Classifiers. In *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*,

W. Bruce Croft and C. J. van Rijsbergen (Eds.). Springer, ACM/Springer, 3–12. https://doi.org/10.1007/978-1-4471-2099-5_1

[50] Xiaoqing Li, Jiansheng Yang, and Jinwen Ma. 2021. Recent developments of content-based image retrieval (CBIR). *Neurocomputing* 452 (2021), 675–689.

[51] Yuan Li, Yinglong Zhang, and Robert Capra. 2022. Analyzing Information Resources That Support the Creative Process. In *Proceedings of the 2022 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR'22)* (Regensburg, Germany). ACM, 180–190. https://doi.org/10.1145/3498366.3505817

[52] Wen-Cheng Lin, Yih-Chen Chang, and Hsin-Hsi Chen. 2004. From Text to Image: Generating Visual Query for Image Retrieval. In *Multilingual Information Access for Text, Speech and Images, 5th Workshop of the Cross-Language Evaluation Forum, CLEF 2004, Bath, UK, September 15-17, 2004, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 3491)*, Carol Peters, Paul D. Clough, Julio Gonzalo, Gareth J. F. Jones, Michael Kluck, and Bernardo Magnini (Eds.). Springer, 664–675. https://doi.org/10.1007/11519645_65

[53] Vivian Liu and Lydia B. Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 384, 23 pages. https://doi.org/10.1145/3491102.3501825

[54] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. *arXiv preprint arXiv:2204.09007* (2022).

[55] Yang Liu, Alan Medlar, and Dorota Glowacka. 2022. ROGUE: A System for Exploratory Search of GANs. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*, Enrique Amigó, Pablo Castells, Julio Gonzalo, Ben Carterette, J. Shane Culpepper, and Gabriella Kazai (Eds.). ACM, 3278–3282. https://doi.org/10.1145/3477495.3531675

[56] Xiaolu Lu, Alistair Moffat, and J. Shane Culpepper. 2016. The effect of pooling and evaluation depth on IR metrics. *Inf. Retr. J.* 19, 4 (2016), 416–445.

[57] Piotr Mirowski, Kory W. Mathewson, Jaylen Pittman, and Richard Evans. 2022. Co-Writing Screenplays and Theatre Scripts with Language Models: An Evaluation by Industry Professionals. *arXiv preprint arXiv:2209.14958* (2022).

[58] Bhaskar Mitra and Nick Craswell. 2018. An Introduction to Neural Information Retrieval. *Found. Trends Inf. Retr.* 13, 1 (2018), 1–126. https://doi.org/10.1561/1500000061

[59] Alistair Moffat and Justin Zobel. 2008. Rank-biased precision for measurement of retrieval effectiveness. *ACM Trans. Inf. Syst.* 27, 1 (2008), 2:1–2:27.

[60] Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. 2016. MS MARCO: A Human Generated MAchine Reading COmprehension Dataset. In *Proceedings of the Workshop on Cognitive Computation: Integrating neural and symbolic approaches 2016 co-located with the 30th Annual Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, December 9, 2016 (CEUR Workshop Proceedings, Vol. 1773)*, Tarek Richard Besold, Antoine Bordes, Artur S. d'Avila Garcez, and Greg Wayne (Eds.). CEUR-WS.org. http://ceur-ws.org/Vol-1773/CoCoNIPS_2016_paper9.pdf

[61] Rodrigo Frassetto Nogueira, Wei Yang, Jimmy Lin, and Kyunghyun Cho. 2019. Document Expansion by Query Prediction. *CoRR* abs/1904.08375 (2019). arXiv:1904.08375 http://arxiv.org/abs/1904.08375

[62] OpenAI. 2022. DALL·E: Creating Images from Text. https://openai.com/blog/dall-e/.

[63] Jonas Oppenlaender. 2022. Prompt Engineering for Text-Based Generative Art. *arXiv preprint arXiv:2204.13988* (2022).

[64] Srishti Palani, Zijian Ding, Stephen MacNeil, and Steven P. Dow. 2021. The "Active Search" Hypothesis: How Search Strategies Relate to Creative Learning. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia). ACM, New York, NY, USA, 325–329. https://doi.org/10.1145/3406522.3446046

[65] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. 2022. DreamFusion: Text-to-3D using 2D Diffusion. *CoRR* abs/2209.14988 (2022). https://doi.org/10.48550/arXiv.2209.14988 arXiv:2209.14988

[66] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 8748–8763. http://proceedings.mlr.press/v139/radford21a.html

[67] 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).

[68] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-Shot Text-to-Image Generation. In *Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 8821–8831. https://proceedings.mlr.press/v139/ramesh21a.html

[69] Ali Razavi, Aaron Van den Oord, and Oriol Vinyals. 2019. Generating diverse high-fidelity images with vq-vae-2. *Advances in neural information processing systems* 32 (2019).

[70] Navid Rekabsaz, Oleg Lesota, Markus Schedl, Jon Brassey, and Carsten Eickhoff. 2021. TripClick: The Log Files of a Large Health Web Search Engine. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, Fernando Diaz, Chirag Shah, Torsten Suel, Pablo Castells, Rosie Jones, and Tetsuya Sakai (Eds.). ACM, 2507–2513. https://doi.org/10.1145/3404835.3463242

[71] Laria Reynolds and Kyle McDonell. 2021. Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. In *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama Japan, May 8-13, 2021, Extended Abstracts*, Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, and Takeo Igarashi (Eds.). ACM, 314:1–314:7. https://doi.org/10.1145/3411763.3451760

[72] 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*. 10684–10695.

[73] Ian Ruthven. 2008. Interactive information retrieval. *Annu. Rev. Inf. Sci. Technol.* 42, 1 (2008), 43–91. https://doi.org/10.1002/aris.2008.1440420109

[74] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho, David J Fleet, and Mohammad Norouzi. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. *arXiv preprint arXiv:2205.11487* (2022).

[75] Tetsuya Sakai. 2007. Alternatives to Bpref. In *SIGIR 2007: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, The Netherlands, July 23-27, 2007*, Wessel Kraaij, Arjen P. de Vries, Charles L. A. Clarke, Norbert Fuhr, and Noriko Kando (Eds.). ACM, 71–78.

[76] Tetsuya Sakai. 2008. Comparing metrics across TREC and NTCIR: The robustness to system bias. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008, Napa Valley, California, USA, October 26-30, 2008*, James G. Shanahan, Sihem Amer-Yahia, Ioana Manolescu, Yi Zhang, David A. Evans, Aleksander Kolcz, Key-Sun Choi, and Abdur Chowdhury (Eds.). ACM, 581–590.

[77] Rob Salkowitz. 2022. Midjourney Founder David Holz On The Impact Of AI On Art, Imagination And The Creative Economy. *Forbes* (Sept. 2022). https://www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy/

[78] R. Keith Sawyer. 2012. *Explaining creativity: The science of human innovation.* Oxford University Press, New York, NY, US.

[79] Greg Schohn and David Cohn. 2000. Less is More: Active Learning with Support Vector Machines. In *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000), Stanford University, Stanford, CA, USA, June 29 - July 2, 2000*, Pat Langley (Ed.). Morgan Kaufmann, 839–846.

[80] Christopher Schröder, Andreas Niekler, and Martin Potthast. 2022. Revisiting Uncertainty-based Query Strategies for Active Learning with Transformers. In *Findings of the Association for Computational Linguistics: ACL 2022, Dublin, Ireland, May 22-27, 2022*, Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, 2194–2203. https://doi.org/10.18653/v1/2022.findings-acl.172

[81] Christoph Schuhmann, Richard Vencu, Romain Beaumont, Robert Kaczmarczyk, Clayton Mullis, Aarush Katta, Theo Coombes, Jenia Jitsev, and Aran Komatsuzaki. 2021. LAION-400M: Open Dataset of CLIP-Filtered 400 Million Image-Text Pairs. *CoRR* abs/2111.02114 (2021). arXiv:2111.02114 https://arxiv.org/abs/2111.02114

[82] Tal Schuster, Roei Schuster, Darsh J. Shah, and Regina Barzilay. 2020. The Limitations of Stylometry for Detecting Machine-Generated Fake News. *Comput. Linguist.* 46, 2 (June 2020), 499–510. https://doi.org/10.1162/coli_a_00380

[83] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*. PMLR, 2256–2265.

[84] Gowthami Somepalli, Vasu Singla, Micah Goldblum, Jonas Geiping, and Tom Goldstein. 2022. Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models. *CoRR* abs/2212.03860 (2022). https://doi.org/10.48550/arXiv.2212.03860 arXiv:2212.03860

[85] Statista Inc. 2022. Digital Media Report - Video Games. https://www.statista.com/study/39310/video-games/.

[86] Yi Tay, Vinh Q. Tran, Mostafa Dehghani, Jianmo Ni, Dara Bahri, Harsh Mehta, Zhen Qin, Kai Hui, Zhe Zhao, Jai Prakash Gupta, Tal Schuster, William W. Cohen, and Donald Metzler. 2022. Transformer Memory as a Differentiable Search Index. *CoRR* abs/2202.06991 (2022). arXiv:2202.06991

https://arxiv.org/abs/2202.06991

[87] Nicola Tonellotto. 2022. Lecture Notes on Neural Information Retrieval. *CoRR* abs/2207.13443 (2022). https://doi.org/10.48550/arXiv.2207.13443 arXiv:2207.13443

[88] Simon Tong and Edward Y. Chang. 2001. Support vector machine active learning for image retrieval. In *Proceedings of the 9th ACM International Conference on Multimedia 2001, Ottawa, Ontario, Canada, September 30 - October 5, 2001*, Nicolas D. Georganas and Radu Popescu-Zeletin (Eds.). ACM, 107–118. https://doi.org/10.1145/500141.500159

[89] Antti Ukkonen, Pyry Joona, and Tuukka Ruotsalo. 2020. Generating Images Instead of Retrieving Them: Relevance Feedback on Generative Adversarial Networks. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*, Jimmy X. Huang, Yi Chang, Xueqi Cheng, Jaap Kamps, Vanessa Murdock, Ji-Rong Wen, and Yiqun Liu (Eds.). ACM, 1329–1338. https://doi.org/10.1145/3397271.3401129

[90] Salahuddin Unar, Xingyuan Wang, Chuan Zhang, and Chunpeng Wang. 2019. Detected text-based image retrieval approach for textual images. *IET Image Process.* 13, 3 (2019), 515–521. https://doi.org/10.1049/iet-ipr.2018.5277

[91] Ellen M. Voorhees. 2001. The Philosophy of Information Retrieval Evaluation. In *Evaluation of Cross-Language Information Retrieval Systems, Second Workshop of the Cross-Language Evaluation Forum, CLEF 2001, Darmstadt, Germany, September 3-4, 2001, Revised Papers (Lecture Notes in Computer Science, Vol. 2406)*, Carol Peters, Martin Braschler, Julio Gonzalo, and Michael Kluck (Eds.). Springer, 355–370.

[92] Ellen M. Voorhees. 2019. The Evolution of Cranfield. In *Information Retrieval Evaluation in a Changing World - Lessons Learned from 20 Years of CLEF*, Nicola Ferro and Carol Peters (Eds.). The Information Retrieval Series, Vol. 41. Springer, 45–69.

[93] Ellen M. Voorhees, Ian Soboroff, and Jimmy Lin. 2022. Can Old TREC Collections Reliably Evaluate Modern Neural Retrieval Models? *CoRR* abs/2201.11086 (2022). arXiv:2201.11086

[94] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. 2021. Towards Real-World Blind Face Restoration With Generative Facial Prior. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021*. Computer Vision Foundation / IEEE, 9168–9178. https://doi.org/10.1109/CVPR46437.2021.00905

[95] Yujing Wang, Yingyan Hou, Haonan Wang, Ziming Miao, Shibin Wu, Hao Sun, Qi Chen, Yuqing Xia, Chengmin Chi, Guoshuai Zhao, Zheng Liu, Xing Xie, Hao Allen Sun, Weiwei Deng, Qi Zhang, and Mao Yang. 2022. A Neural Corpus Indexer for Document Retrieval. *CoRR* abs/2206.02743 (2022). https://doi.org/10.48550/arXiv.2206.02743 arXiv:2206.02743

[96] Ryen W. White, Mikhail Bilenko, and Silviu Cucerzan. 2007. Studying the use of popular destinations to enhance web search interaction. In *SIGIR 2007: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, The Netherlands, July 23-27, 2007*, Wessel Kraaij, Arjen P. de Vries, Charles L. A. Clarke, Norbert Fuhr, and Noriko Kando (Eds.). ACM, 159–166. https://doi.org/10.1145/1277741.1277771

[97] Ryen W. White, Ian Ruthven, and Joemon M. Jose. 2005. A study of factors affecting the utility of implicit relevance feedback. In *SIGIR 2005: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil, August 15-19, 2005*, Ricardo A. Baeza-Yates, Nivio Ziviani, Gary Marchionini, Alistair Moffat, and John Tait (Eds.). ACM, 35–42. https://doi.org/10.1145/1076034.1076044

[98] Wikipedia contributors. 2022. Infinite monkey theorem — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Infinite_monkey_theorem&oldid=1122059899 [Online; accessed 10-January-2023].

[99] Zuobing Xu, Ram Akella, and Yi Zhang. 2007. Incorporating Diversity and Density in Active Learning for Relevance Feedback. In *Advances in Information Retrieval, 29th European Conference on IR Research, ECIR 2007, Rome, Italy, April 2-5, 2007, Proceedings (Lecture Notes in Computer Science, Vol. 4425)*, Giambattista Amati, Claudio Carpineto, and Giovanni Romano (Eds.). Springer, 246–257. https://doi.org/10.1007/978-3-540-71496-5_24

[100] Christoph Zauner. 2010. *Implementation and benchmarking of perceptual image hash functions.* Master's thesis. Upper Austria University of Applied Sciences, Hagenberg Campus.

[101] Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending Against Neural Fake News. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2019/file/3e9f0fc9b2f89e043bc6233994dfcf76-Paper.pdf

[102] Cha Zhang and Tsuhan Chen. 2002. An active learning framework for content-based information retrieval. *IEEE Transactions on Multimedia* 4, 2 (2002), 260–268. https://doi.org/10.1109/TMM.2002.1017738

[103] Yinglong Zhang and Robert Capra. 2019. Understanding How People Use Search to Support Their Everyday Creative Tasks. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (Glasgow, Scotland UK). ACM, New York, NY, USA, 153–162. https://doi.org/10.1145/3295750.3298936

[104] Yinglong Zhang, Rob Capra, and Yuan Li. 2020. An In-Situ Study of Information Needs in Design-Related Creative Projects. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (Vancouver BC, Canada). ACM, New York, NY, USA, 113–123. https://doi.org/10.1145/3343413.3377973

[105] Shengyao Zhuang, Houxing Ren, Linjun Shou, Jian Pei, Ming Gong, Guido Zuccon, and Daxin Jiang. 2022. Bridging the Gap Between Indexing and Retrieval for Differentiable Search Index with Query Generation. *CoRR* abs/2206.10128 (2022). https://doi.org/10.48550/arXiv.2206.10128 arXiv:2206.10128