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

CHIIR 2023

Niklas Deckers

Maik Fröbe

Johannes Kiesel

Gianluca Pandolfo

Christopher Schröder

Benno Stein

Martin Potthast

UNIVERSITÄT LEIPZIG

ScaDS.AI DRESDEN LEIPZIG

FRIEDRICH-SCHILLER- UNIVERSITÄT JENA

Bauhaus- Universität Weimar

Recent Advancements in Image Generation

Generate images, but conditioned on a user-defined text, the prompt

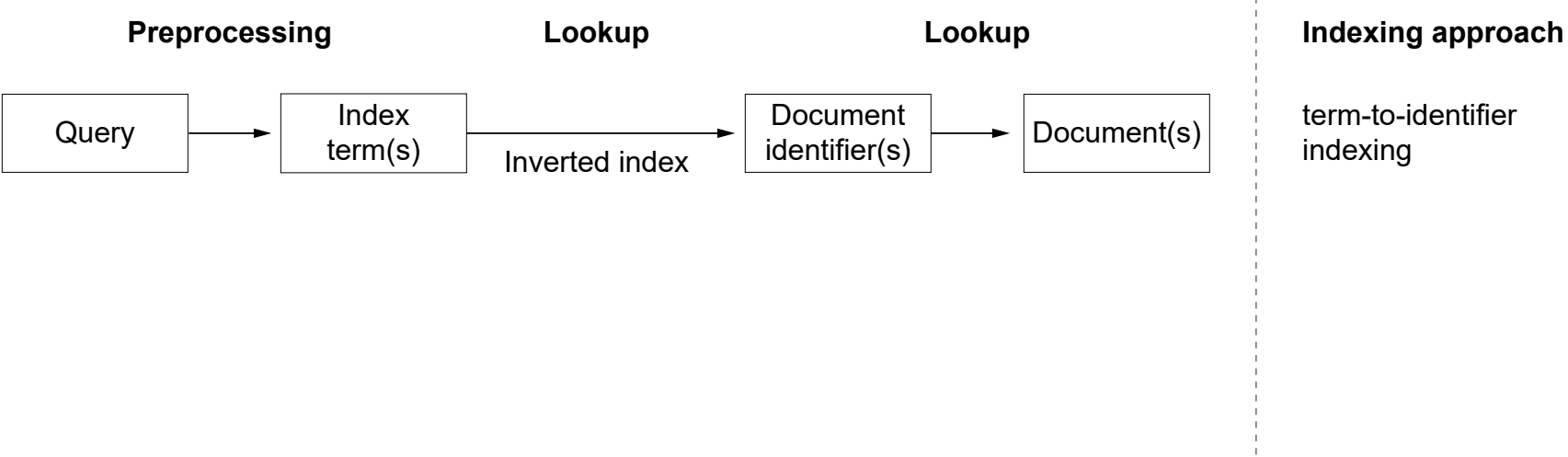`a beautiful sunset over the texan landscape`

Recent Advancements in Image Generation

Generate images, but conditioned on a user-defined text, the prompt


```
a beautiful sunset over
the texan landscape
```

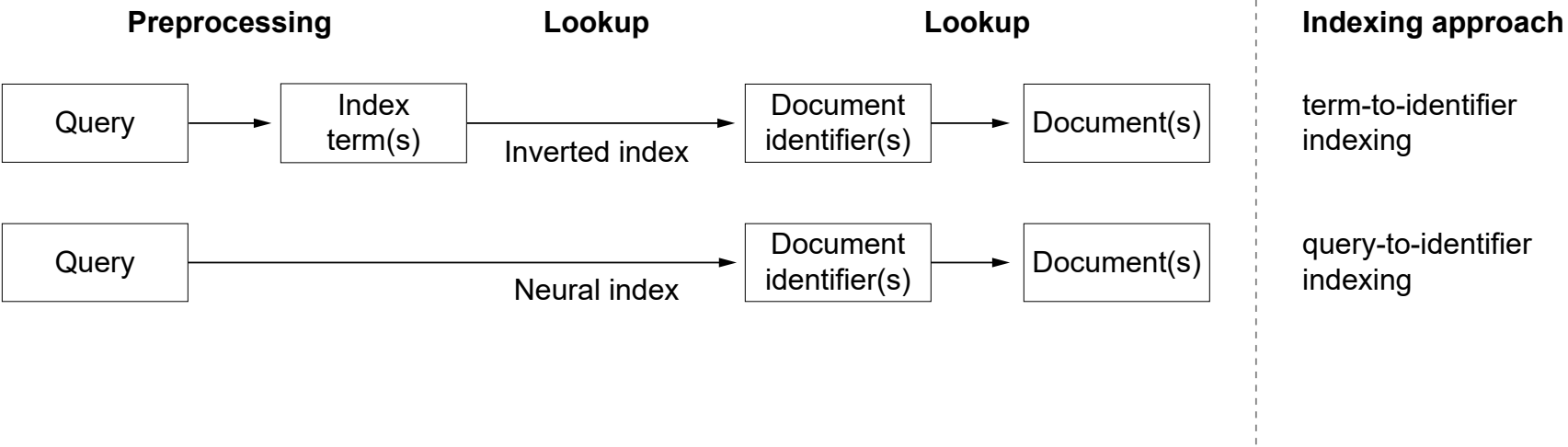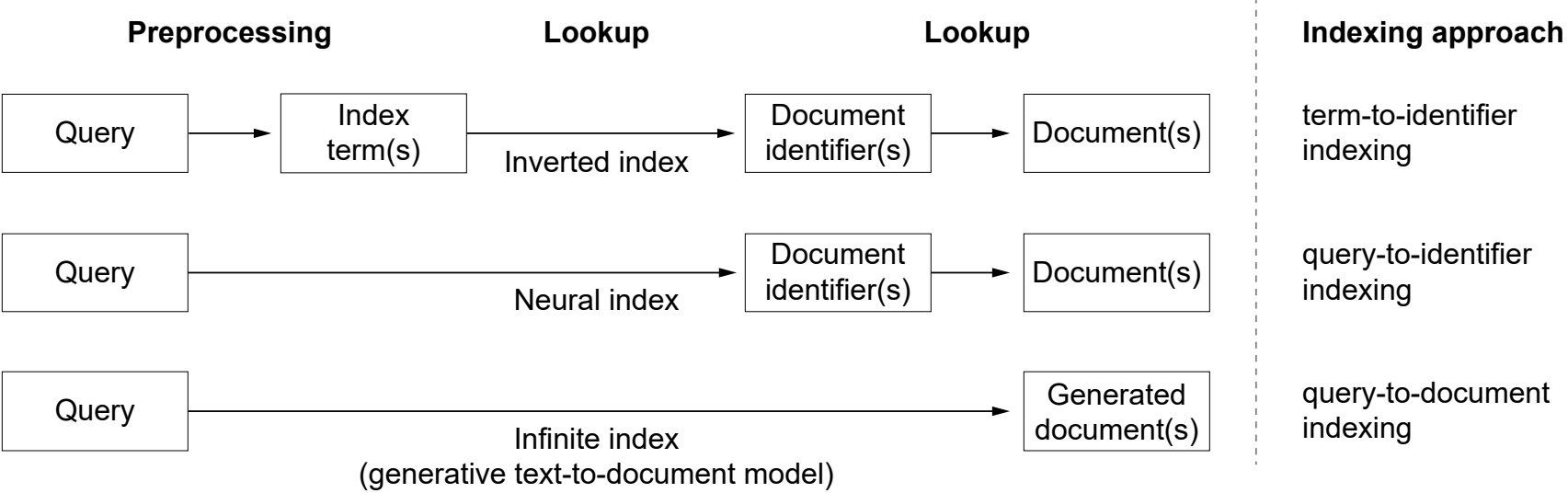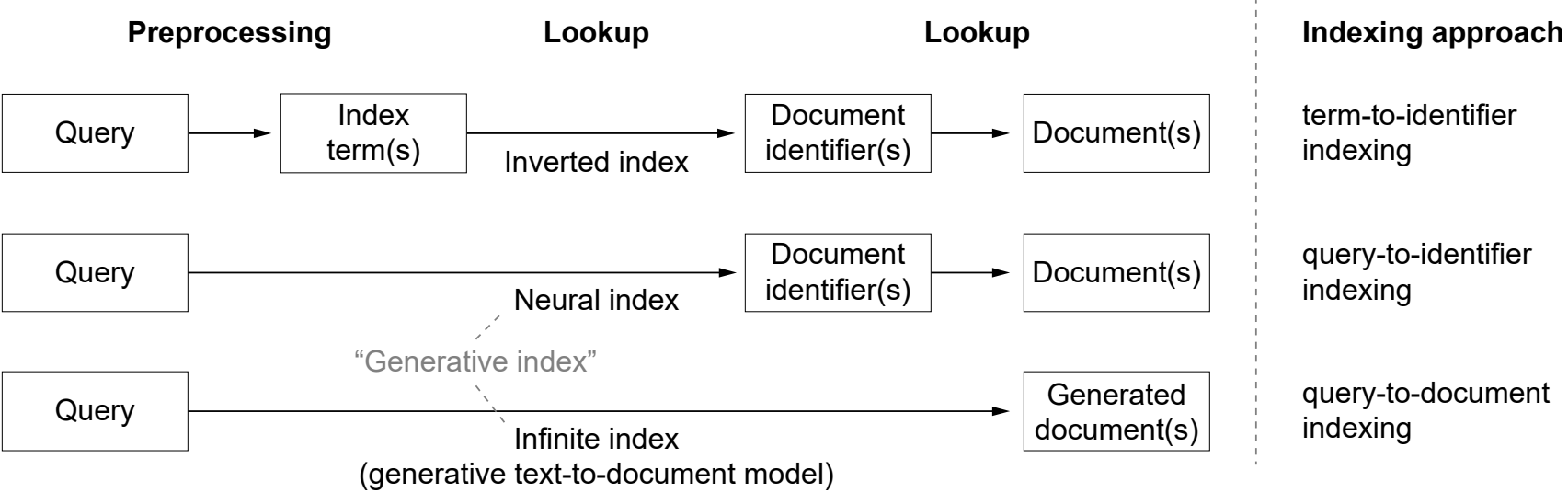| Name	Parameters	Month / Year
DALL·E (OpenAI)	12 B	01 / 2021
DALL·E 2 (OpenAI)	3.5 B	04 / 2022
Imagen (Google)	4.6 B	05 / 2022
Midjourney	n/a	07 / 2022
Stable Diffusion (Stability AI)	0.9 B	08 / 2022

Generative Models as an Index

See image generation with a prompt as
image retrieval with a query, but on an infinite index

Generative Models as an Index

See image generation with a prompt as
image retrieval with a query, but on an infinite index

| **Preprocessing** | **Lookup** | **Lookup** | **Indexing approach** |

```
┌─────────┐      ┌─────────┐                    ┌─────────────┐    ┌──────────────┐
│  Query  │ ───► │  Index  │ ───────────────►   │  Document   │──► │ Document(s)  │
│         │      │ term(s) │    Inverted index  │ identifier(s)│   │              │
└─────────┘      └─────────┘                    └─────────────┘    └──────────────┘
```

term-to-identifier
indexing

Generative Models as an Index

See image generation with a prompt as
image retrieval with a query, but on an infinite index

Preprocessing	Lookup	Lookup		Indexing approach

Preprocessing   **Lookup**   **Lookup**   **Indexing approach**

Query → Index term(s) → Document identifier(s) → Document(s)

Inverted index

term-to-identifier indexing

Query → Document identifier(s) → Document(s)

Neural index

query-to-identifier indexing

Generative Models as an Index

See image generation with a prompt as
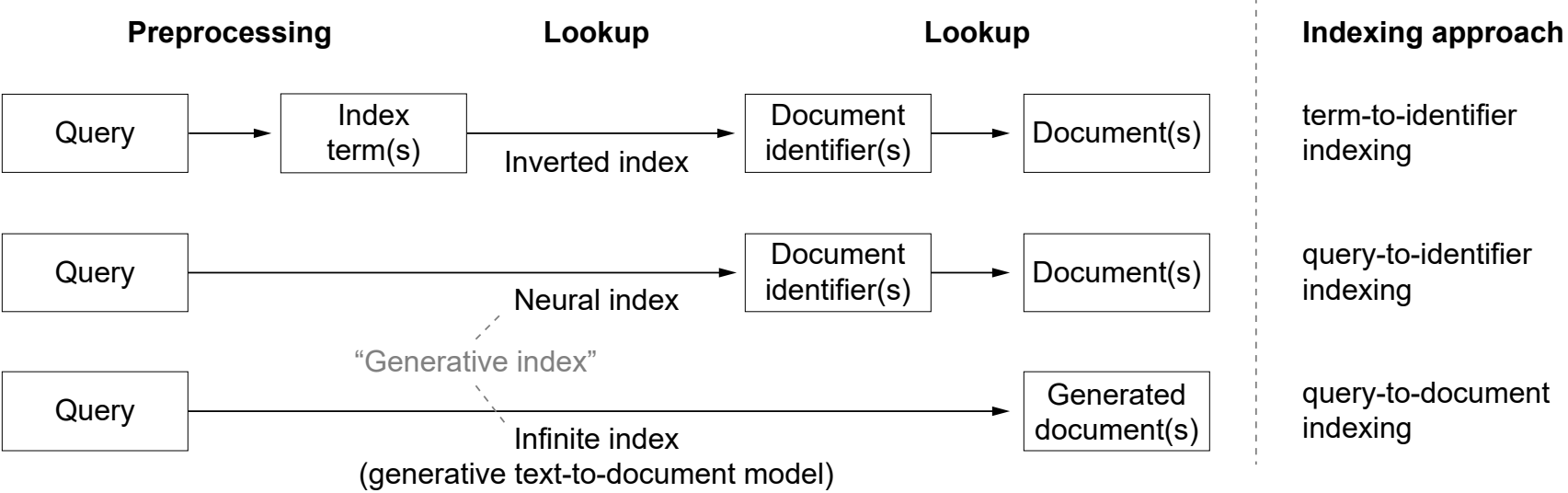image retrieval with a query, but on an infinite index

Preprocessing	Lookup	Lookup	Indexing approach
Query → Index term(s)	—Inverted index→ Document identifier(s)	→ Document(s)	term-to-identifier indexing
Query	—Neural index→ Document identifier(s)	→ Document(s)	query-to-identifier indexing
Query	Infinite index (generative text-to-document model)→	Generated document(s)	query-to-document indexing

Generative Models as an Index

See image generation with a prompt as
image retrieval with a query, but on an infinite index

Preprocessing	Lookup	Lookup	Indexing approach

Query → Index term(s) → (Inverted index) → Document identifier(s) → Document(s) — term-to-identifier indexing

Query → (Neural index) → Document identifier(s) → Document(s) — query-to-identifier indexing

"Generative index"

Query → Infinite index (generative text-to-document model) → Generated document(s) — query-to-document indexing

Generative Models as an Index

See image generation with a prompt as
image retrieval with a query, but on an infinite index

Preprocessing	**Lookup**	**Lookup**	**Indexing approach**

Query → Index term(s) — Inverted index → Document identifier(s) → Document(s)

term-to-identifier indexing

Query → Document identifier(s) → Document(s)

Neural index

query-to-identifier indexing

"Generative index"

Query → Generated document(s)

Infinite index
(generative text-to-document model)

query-to-document indexing

Recent examples: Bing, YouChat, NeevaAI

The Infinite Index

- Every query yields (some) results
- For each query, there exist infinitely many possible results
  with different random seeds

The Infinite Index

- ❑ Every query yields (some) results
- ❑ For each query, there exist infinitely many possible results
  with different random seeds

Golden treehouse in lush forest, better homes and hardens magazine, big glass windows, intricate woodworking, polaroid

ethiopian landscape, highly detailed, digital painting, concept art, sharp focus, cinematic lighting, diffuse lighting, fantasy, intricate, elegant, lifelike, photorealistic, illustration, smooth

a cybernetic samoyed and beagle, concept art, detailed face and body, detailled decor, fantasy, highly detailed, cinematic lighting, digital art painting, winter, nature, running

The Infinite Index

❑ Prompt embedding space is continuous

and allows for approaches like gradient descent

The Infinite Index

❑ **Prompt embedding space is continuous**
and allows for approaches like gradient descent

| Prompt A Seed 1 | | Prompt B Seed 1 | | Prompt C Seed 1 | | Prompt D Seed 1 | | Prompt E Seed 1 | | Prompt E (modified) Seed 1 |
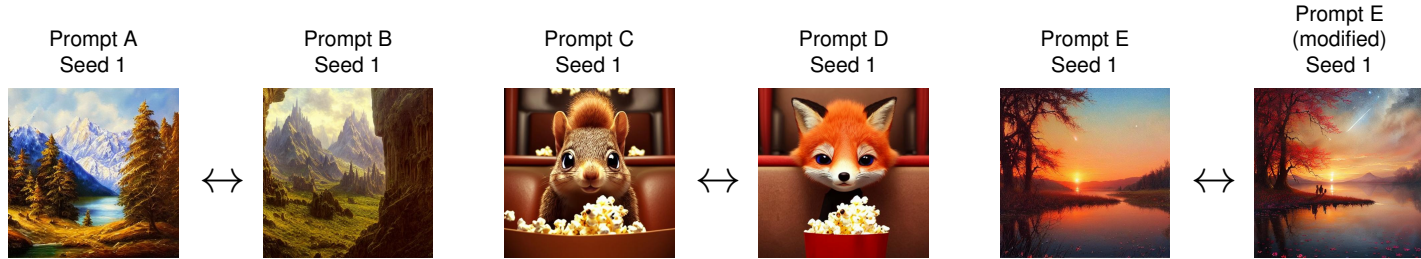

```
beautiful mountain landscape,
lake, snow, oil painting 8 k hd
```
interpolated to
```
a beautiful and highly detailed
matte painting of the epic
mountains of avalon, intricate
details, epic scale, insanely
complex, 8 k, sharp focus,
hyperrealism, very realistic,
by caspar friedrich, albert
bierstadt, james gurney, brian
froud,
```

```
Cute small squirrel sitting in
a movie theater eating popcorn
watching a movie ,unreal
engine, cozy indoor lighting,
artstation, detailed, digital
painting,cinematic,character
design by mark ryden and pixar
and hayao miyazaki, unreal 5,
daz, hyperrealistic, octane
render
```
interpolated to
```
Cute small fox sitting [...]
```

```
a beautiful painting of a
peaceful lake in the Land
of the Dreams, full of
grass, sunset, red horizon,
starry-night!!!!!!!!!!!!!!!!!!!!,
Greg Rutkowski, Moebius,
Mohrbacher, peaceful, colorful
```
and a gradient ascent on an aesthetic score

Related Information Retrieval Concepts

- Image retrieval including reverse image search
  - Query-to-image as image generation
  - Image-to-image including inpainting

- Retrieval for creative tasks

- Interactive retrieval
  - Query reformulation
  - Query log analysis
  - Query suggestion

Interaction Methods

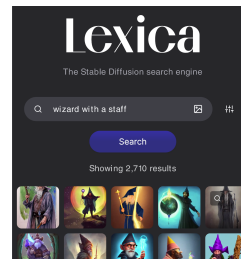(a) Prompting in a community-maintained Stable Diffusion web interface
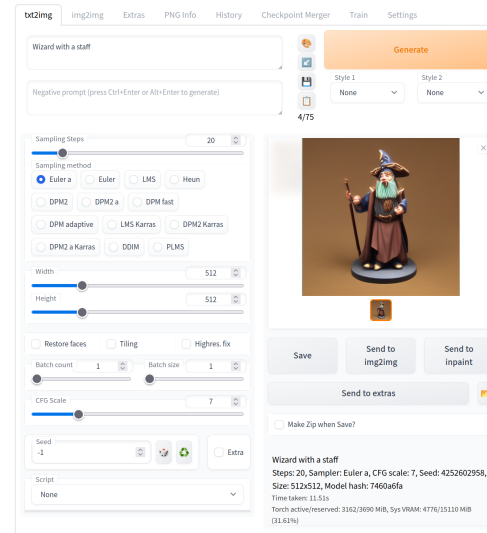
(a)

Interaction Methods

(a) Prompting in a community-maintained Stable Diffusion web interface

(b) Lexica search engine for generated images along their prompts
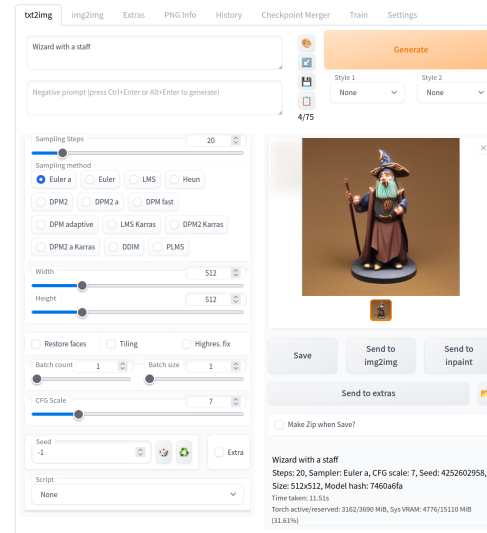
(a)

(b)

# Interaction Methods

(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
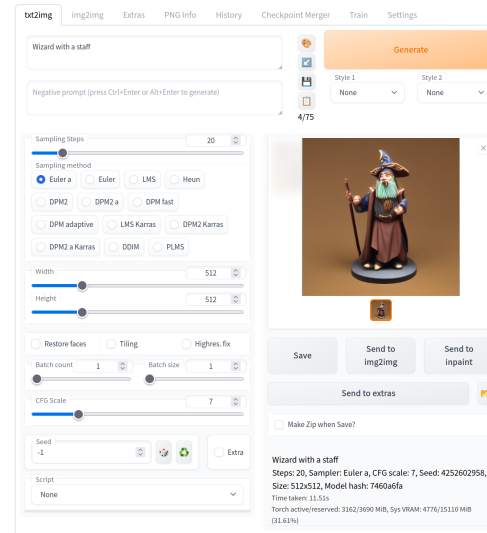
(a)
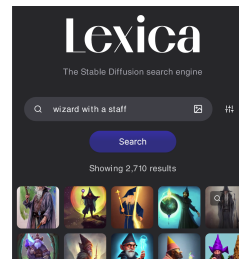
(b)

(c)

Interaction Methods

(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
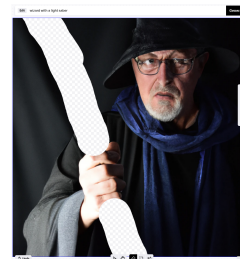
(d) Upscaling and variation generation in Midjourney
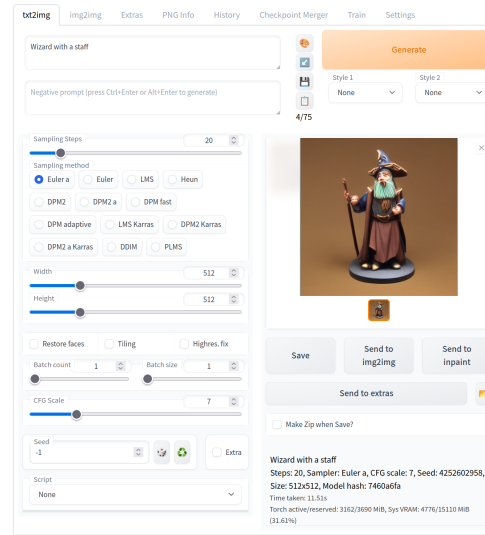
(a)
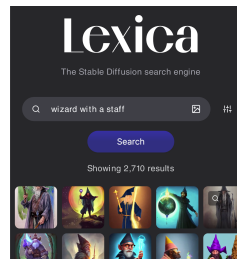
(b)

(c)

(d)

Interaction Methods

(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

(d) Upscaling and variation generation in Midjourney

**Equivalent to
search interfaces**
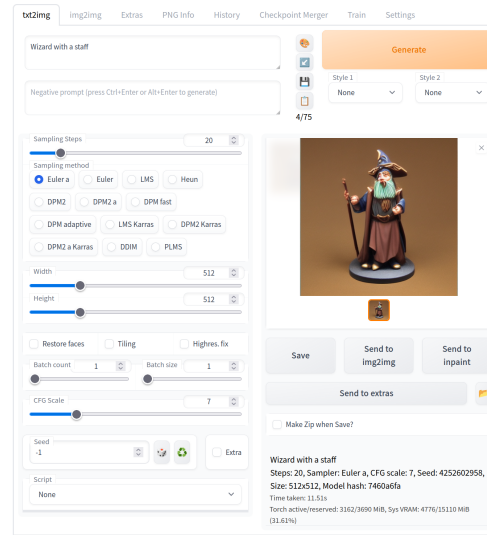
(a)
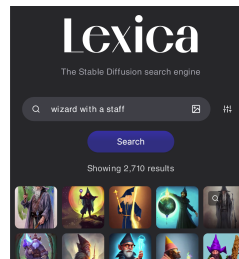
(b)

(c)

(d)

User Behavior

- ❑ Result relevance depends on the information needs of users
  as with regular image retrieval

- ❑ Descriptive approach:

- ❑ Creative approach:

User Behavior

❑ Result relevance depends on the information needs of users
  as with regular image retrieval

❑ Descriptive approach:
  – The user has an idea of a fixed target image
  – Generates image that approximates their ideas

❑ Creative approach:

User Behavior

❑ Result relevance depends on the information needs of users
   as with regular image retrieval

❑ Descriptive approach:
   – The user has an idea of a fixed target image
   – Generates image that approximates their ideas

❑ Creative approach:
   – The user has no clear vision or goal, but a set of constraints
   – Iteratively refines their prompt in a feedback-loop with random elements introduced by the system

User Behavior

❑ Result relevance depends on the information needs of users
as with regular image retrieval

❑ Descriptive approach:
  – The user has an idea of a fixed target image
  – Generates image that approximates their ideas

❑ Creative approach:
  – The user has no clear vision or goal, but a set of constraints
  – Iteratively refines their prompt in a feedback-loop with random elements introduced by the system

❑ Similar behavior in terms of query log analysis
a general prompt is extended with details to become more specific

Relevance in Text-To-Image Generation

❑ Current approach to measure relevance:
Users click a like button for good images
but that does not generalize well for unseen prompts

❑ Model user intent just by looking at the prompt?
Initiative is entirely with the user

❑ Retrieval methods such as query suggestions can be transferred from
information retrieval to text-to-image model generation
and retrieval evaluation measures can be adopted

Evaluating Retrieval on Text-To-Image Models

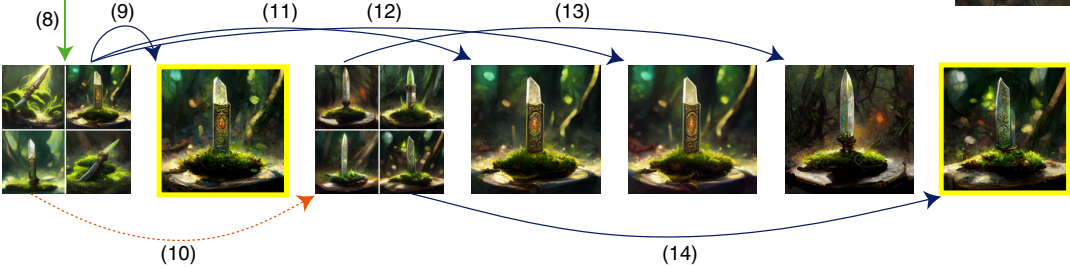Problems with the Infinite Index:

- ❑ Unending stream of alternative generated images leads to problems with recall-oriented evaluation measures
  and requires sampling

- ❑ Near duplicates with high relevance can be generated
  and thus evaluation requires deduplication

Case Study: Game Artwork Search

Initial prompt: an ancient golden dagger lying on moss, illuminated by godrays, close up, digital painting, matte painting, midjourney, concept art, detailed art, scifiart cinematic painting, magic the Gathering, volumetric light, masterpiece, volumetric realistic render, epic scene, 8k, post-production detailed art, scifiart cinematic painting --q 2
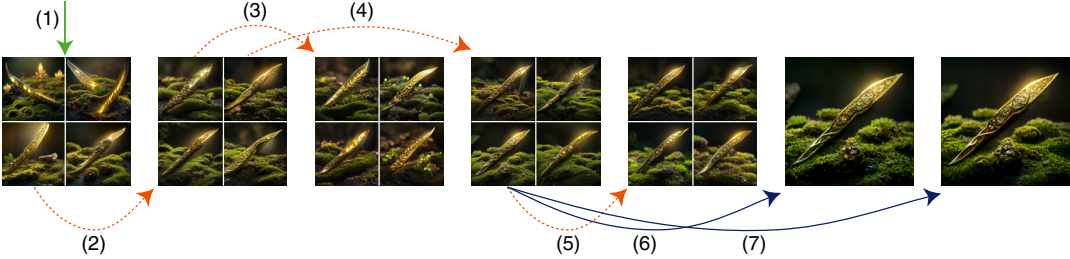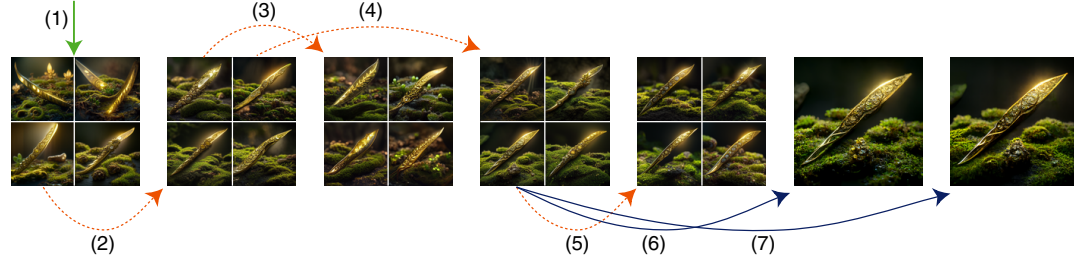

Reformulated prompt: a medieval dagger lying on moss, lit by god rays, art by Adrian Smith + Paul bonner, magic gathering style, warcraft, blizzard style, hearthstone, fantasy concept art, medieval, masterpiece, mystical, witchcraft
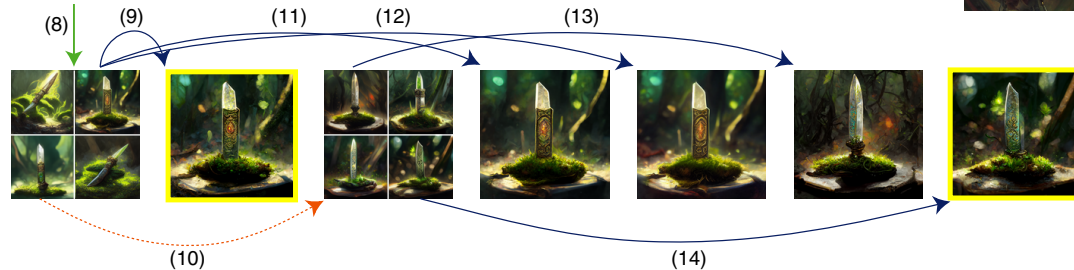
Case Study: Game Artwork Search

Initial prompt: an ancient golden dagger lying on moss, illuminated by godrays, close up, digital painting, matte painting, midjourney, concept art, detailed art, scifiart cinematic painting, magic the Gathering, volumetric light, masterpiece, volumetric realistic render, epic scene, 8k, post-production detailed art, scifiart cinematic painting --q 2

Reformulated prompt: a medieval dagger lying on moss, lit by god rays, art by Adrian Smith + Paul bonner, magic gathering style, warcraft, blizzard style, hearthstone, fantasy concept art, medieval, masterpiece, mystical, witchcraft
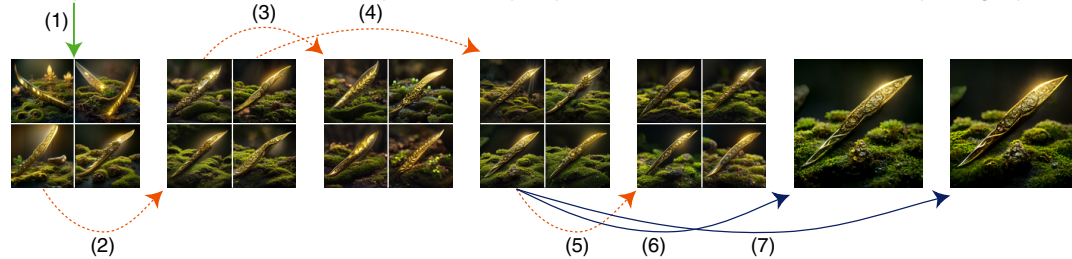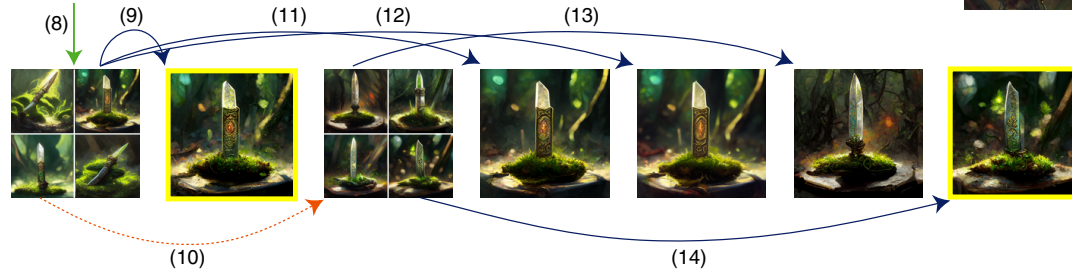
Central findings:

❑ Uncertainty never fully ceases

❑ Sense of direction, but lack of control

Case Study: Game Artwork Search

Initial prompt: an ancient golden dagger lying on moss, illuminated by godrays, close up, digital painting, matte painting, midjourney, concept art, detailed art, scifiart cinematic painting, magic the Gathering, volumetric light, masterpiece, volumetric realistic render, epic scene, 8k, post-production detailed art, scifiart cinematic painting --q 2

Reformulated prompt: a medieval dagger lying on moss, lit by god rays, art by Adrian Smith + Paul bonner, magic gathering style, warcraft, blizzard style, hearthstone, fantasy concept art, medieval, masterpiece, mystical, witchcraft

Central findings:

❑ Uncertainty never fully ceases

❑ Sense of direction, but lack of control

Prompt engineering forms a central limitation in current approaches

Outlook: Supporting the User to Navigate the Infinite Index

- ❑ Given prompt templates as a source of inspiration

- ❑ More modalities (depth maps, segmentation maps) for more precise control in the image space

- ❑ Interactive gradient-based methods to modify the prompts
  using active learning

- ❑ Generalization of these concepts to text-to-text domains

Outlook: Supporting the User to Navigate the Infinite Index

- Given prompt templates as a source of inspiration

- More modalities (depth maps, segmentation maps) for more precise control in the image space

- Interactive gradient-based methods to modify the prompts
  using active learning

- Generalization of these concepts to text-to-text domains

Thank you!

Backup: Interpolating Seeds

Seed 1   Seed 2

Latents → Interpolated latents ← Latents