

Multimodal Machine Learning Lab

Winter Semester 2024/2025

Niklas Deckers and Martin Potthast

Agenda

- ❑ Feedback for User Stories and Storyboarding
- ❑ About the Reviews
- ❑ Discussion: Are Text Prompts Really Helpful?
- ❑ Prototyping a Probability-Based System
- ❑ Outlook

Feedback for User Stories and Storyboarding

- ❑ Two kinds of ideas: Giving some directions to choose from, or giving the user control to choose their own direction (but with less support)
- ❑ Many of the proposed methods had the interaction added *on top* by allowing for modifications via prompt engineering (as the core framework for the iteration)
- ❑ How can the introduced approaches be generalized to the general prompt engineering problem (given a mental target function and a start idea, find the optimal image)?

About the Reviews

- ❑ Very nice, encouraging and constructive feedback
- ❑ Sometimes, technical backbones of the systems were identified as contributions (instead of core ideas for the user stories or story boards)
- ❑ Reviews complement each other nicely
- ❑ Will hand out all reviews to everyone, please read at least reviews for your own system plus the meta reviews for all other systems

Discussion: Are Text Prompts Really Helpful?

- ❑ What is the objective of the interactive approaches?
- ❑ Working towards a specific (pre-defined) target image? Then it would need to be shown that the iteration steps really work towards this target.
- ❑ Or exploring a space of options? Then text prompting seems counter-intuitive since the user still has to come up with the idea.

Prototyping a Probability-Based System

- ❑ Target of this stage: Understanding the problems with creative generation
- ❑ Given a subject (i.e. start prompt) and the probability space of images (defined by Stable Diffusion)
- ❑ How can the probability space defined by user surprise be modeled?
- ❑ How can this problem be boiled down?
- ❑ Good example tasks (esp. for evaluation)?
- ❑ Building blocks of such a system?

Prototyping a Probability-Based System

- ❑ From the discussion:
- ❑ How can surprise be defined? Does it mean that the user has never seen a given image in their life or never seen it generated by the machine (e.g., users would be surprised by an image of themselves)
- ❑ Some distance measure needs to be defined to avoid generating very small-step modifications of previously generated (or known) images
- ❑ Related problems: Searching for documents on the web (in a text domain, allowing processing on a token level); diversity is important (achieving maximum diversity within 10 blue links)
- ❑ Recommender systems (Spotify etc.): Based on other user data, on a simple discrete space
- ❑ Subsymbolic (images, wavefiles) vs. symbolic (text, MIDI, SVGs etc., where components have a meaning) might also make things easier
- ❑ Simplification by making experiments on a small discrete space (100 pre-selected images)

Prototyping a Probability-Based System

- ❑ How can the system be adapted to individual users?
- ❑ User profiles (similarly to Spotify). Improves with more users
- ❑ Could also use linear coefficients of pre-defined prompt-based vectors to represent users (This image is surprising to old people, This image is surprising to tech-savvy people, ...)
- ❑ Simplest form: List of all images shown to the user so far
- ❑ Forming a user profile from feedback (what surprises the user, what does not?)
- ❑ How many images should be shown? 5? 100? Feedback with a small exponent base should be able to yield results quickly, too (cf. Akinator)
- ❑ Interpolation on the space of surprising images
- ❑ Space of surprising images might not be static but change when images are shown

Prototyping a Probability-Based System

- ❑ Work packages:
- ❑ Web UI (Henry): Main loop for the iteration and interactive web UI to select preferences. Should include webis demo template. May be based on some suitable framework (gradio, streamlit, nicegui, pywebio)
- ❑ Recommender (Klara): From the start prompt and the current user config (which can be mapped back to the CLIP space), derive recommended samples for the next iteration
- ❑ Optimizer (Aleyna and Paul): Based on the feedback, adapt the user profile (e.g. by gradient-based optimization)
- ❑ (Recommender and Optimizer resemble components of an active learning system)
- ❑ Generator (Moritz): Generate images quickly
- ❑ User Profile Host (Aleyna and Paul): Defined the mapping from the CLIP space to the, e.g., 10-dimensional user profile space (using prompts etc.), and back

Prototyping a Probability-Based System

- Tasks:
- Define interfaces, e.g., as a graph for which module calls which including arguments (until 21.11.24)
- Create a mock-up of all components and put them together for next week's lab (Web UI: Loop and simple commandline-based interaction; Recommender: random perturbations from the current position; Optimizer: Some small step adjustments through e.g. interpolation; Generator: Slow image generation baseline; User Profile Host: Some simple initial mapping)
- It might be that it is easier to first model user preference instead of user surprise

Outlook

- ❑ For the optimization problem (given a mental target function and a start idea, find the optimal image): We need to give a sense of direction. Idea: Extracting and providing dimensional axes (e.g., moods or object properties). Optimizing one axis after the other. Depending on the prompt, what could be these axes (cf. the game Guess Who)?
- ❑ Next week: Towards evaluation and lab experiments