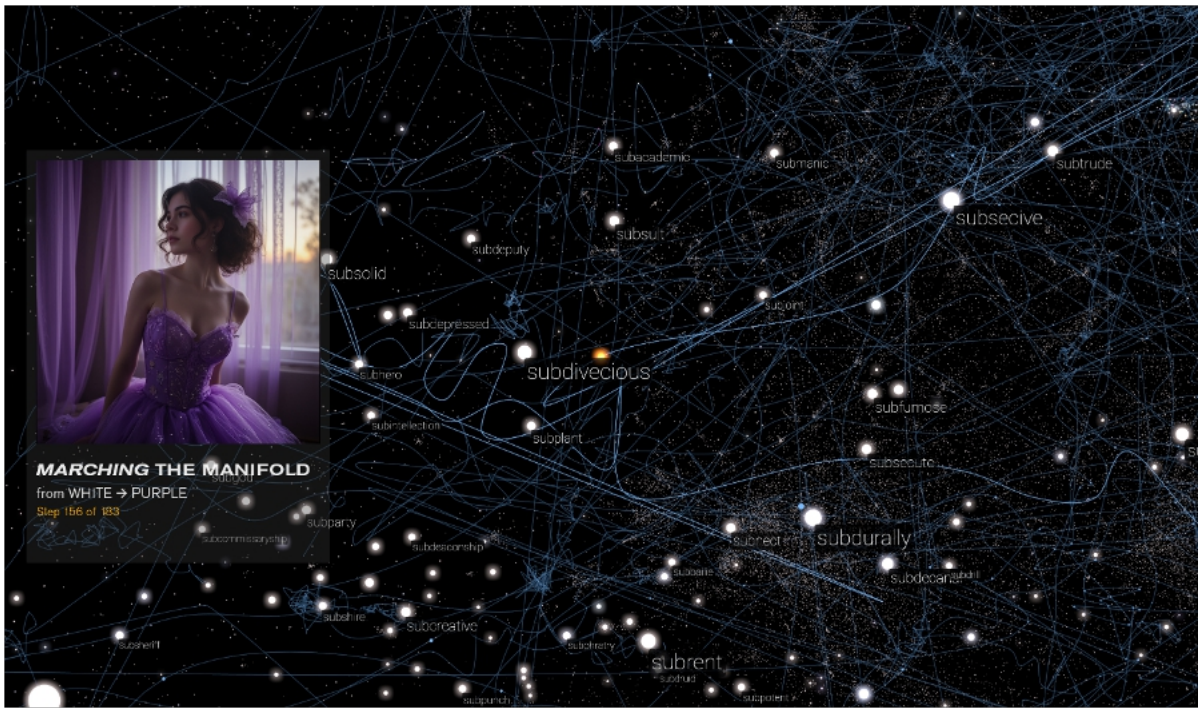
Marching The Manifold – Visualizing Semantic Relationships in CLIP's Latent Space

Tobias Bongartz[1,†]

[1] *Hochschule Bielefeld, Lampingstraße 3, 33649 Bielefeld, Germany*



Abstract

Latent spaces form the mathematical foundation of modern AI image generation systems, yet their high-dimensional complexity remains inaccessible to design practitioners who could benefit most from understanding them. The Bachelor Thesis "Marching The Manifold" transforms CLIP's (Contrastive Language-Image Pre-training)[1] abstract 768-dimensional embedding space into an interactive, galaxy-like 3D visualization where 235,886 embedded words are semantically clustered. The Web-Interface is architected with Vite, leveraging JavaScript and HTML alongside React and Three.js for real time application. A custom-developed Latent Navigator tool, programmed in Python, employs beam search algorithms with FAISS[2]-optimized nearest-neighbor calculations to identify meaningful pathways through this space, which users explore via a purpose-built physical controller. As users traverse these pathways, pre generated Image sets by the diffusion models FLUX.1[dev][3] and PixelWave Flux.1-dev 03[4] show corresponding imagery at each point, revealing the continuous semantic transitions between concepts at each point of the pathway.

The resulting system visualizes semantic relationships as navigable paths through the latent space. Transitions between concepts manifest as mostly coherent visual transformations in the generated images, while the celestial metaphor provides users with an intuitive frame of reference.

This visualization approach bridges the gap between technical AI implementation and creative practice by enabling designers to develop intuitive mental models of latent spaces through direct exploration. The project demonstrates how thoughtful interaction design can transform abstract mathematical structures into accessible tools for creative exploration.

Keywords

Latent Space Navigation, Computational Visualization, CLIP Embeddings, Physical Computing, AI Transparency

WCDCC 2025: First Workshop on Computational Design and Computer-aided Creativity 2025, 23 June, 2025, Campinas, BR

[1,*] Corresponding author.

tb.bongartz@gmail.com (T. Bongartz);

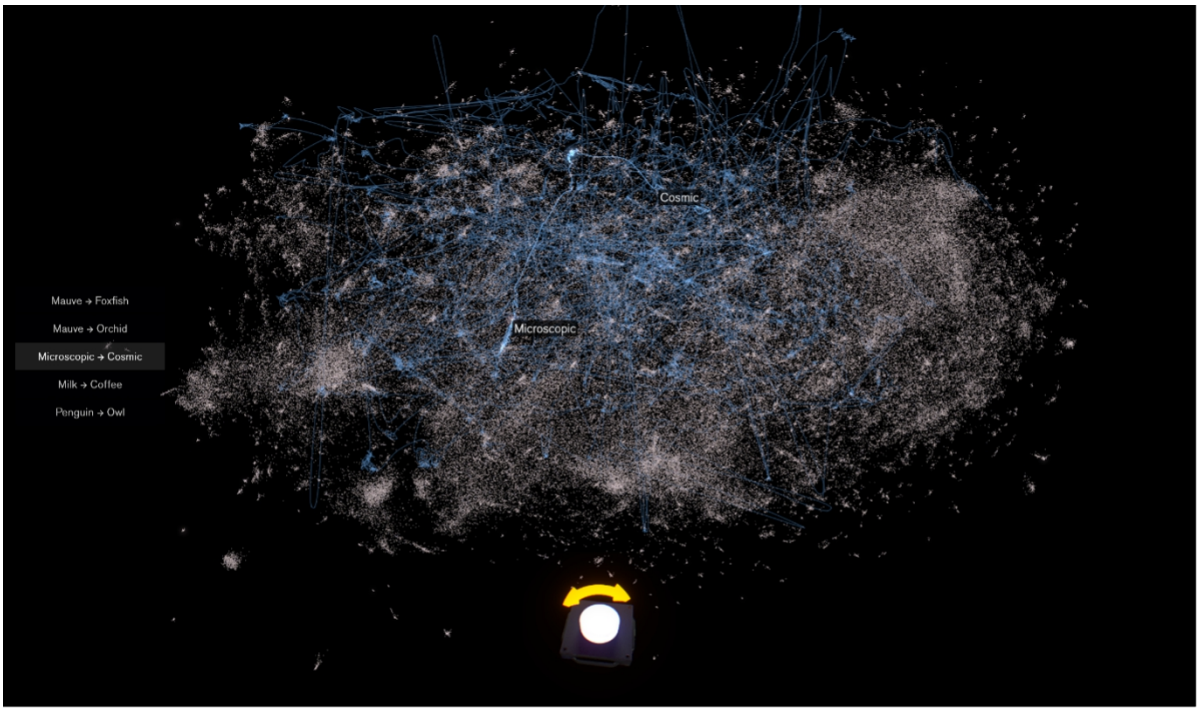


Figure 1: The main interface in idle mode: On the left side resides the menu for scrolling through the predefined pathways. Every white point represents a word-embedding in latent space. The Blue lines represent selectable and traversable pathways.

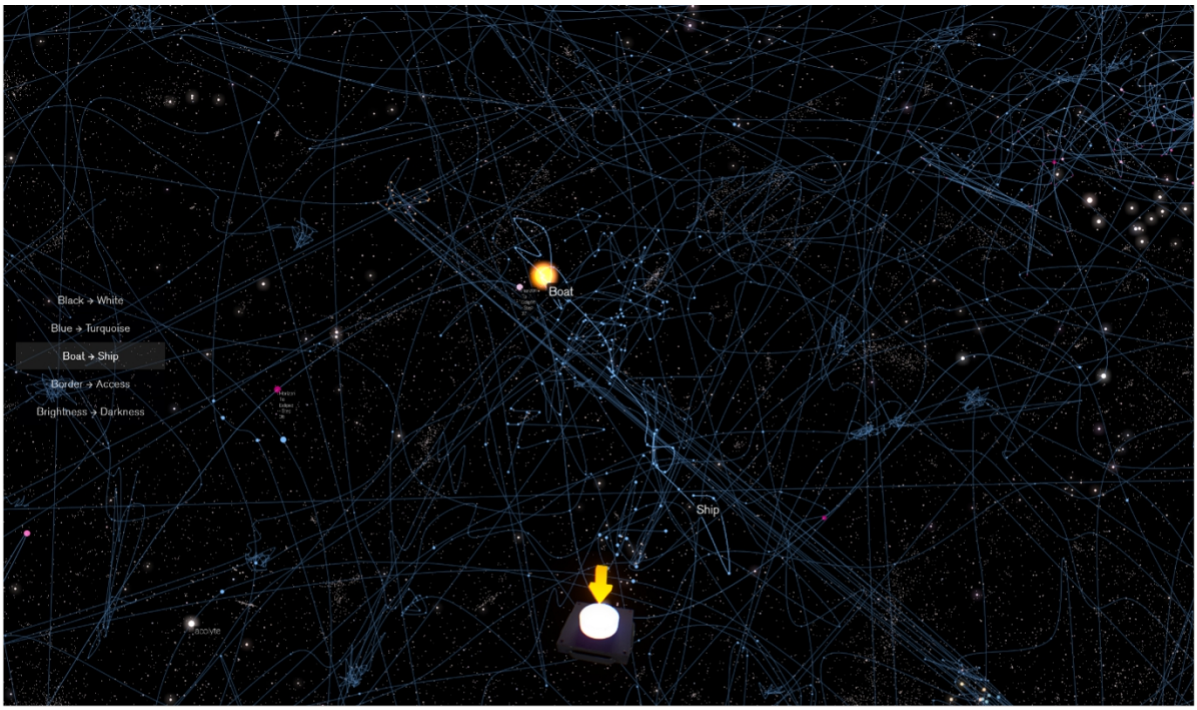


Figure 2: Interacting with the physical controller: Scrolling through the pathways shows them in whole.

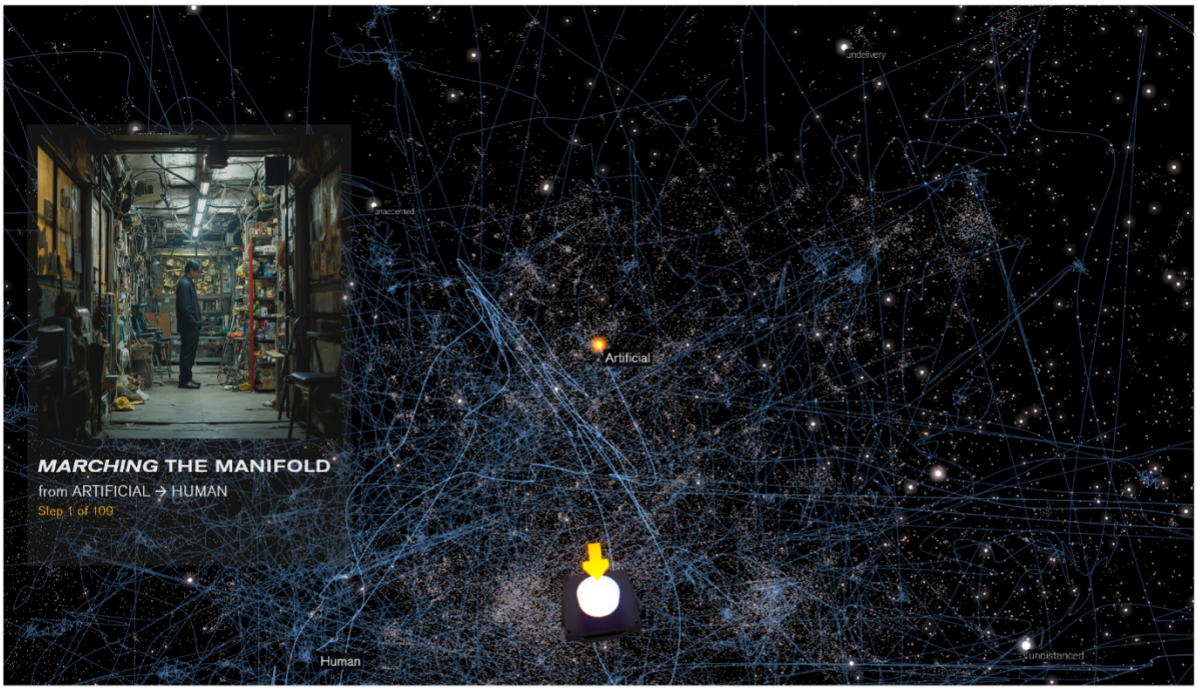


Figure 3: Entering a pathway by pressing on the physical controller highlights the first point of the pathway with the corresponding image. Turning the dial, enables the user to select the next point of the pathway, the view follows, the corresponding image is shown.

Figure 4: Each white point represents one embedded word concept. The corresponding name of the point is shown when the camera gets close enough.

Figure 5: The 3D-printed physical controller on a podium. The final design incorporates a single, push-enabled turning knob, driven by an RP2040-based development board. The projection of the web interface is situated in the background on the wall.

1. Video Documentation

A video documentation can be seen at: https://youtu.be/L29iY3emqsA

2. The Custom Controller

A 3D rendering of the physical controller can be seen at: https://youtu.be/Vok3NN-SeLs

Declaration on Generative AI

During the preparation of this work, the author used Perplexity and Claude 3.7 Sonnet Thinking in order to: Improve writing style and for grammar and spelling check. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] Alec Radford, Ilya Sutskever, Jong Wook Kim, Gretchen Krueger, Sandhini Agarwal, "CLIP: Connecting Text and Images," CLIP: Connecting Text and Images (blog), January 5, 2021, https://openai.com/index/clip/.

[2] Matthijs Douze et al., "The Faiss Library" (arXiv, September 6, 2024), URL: https://doi.org/10.48550/arXiv.2401.08281

[3] "Black-Forest-Labs/FLUX.1-Dev · Hugging Face," accessed January 1, 2025, URL: https://huggingface.co/black-forest-labs/FLUX.1-dev.

[4] "PixelWave - FLUX.1-Dev 03 | Flux Checkpoint | Civitai," December 14, 2024, URL: https://civitai.com/models/141592/pixelwave.