Embedding Landscape Explorer

Project Overview

Claude (Anthropic) · Sean Patrick Morris

March 2026

What It Is

The Embedding Landscape Explorer is a browser-based application that transforms the mathematical relationships between words into navigable, physical terrain. It takes word embeddings — high-dimensional vectors learned from patterns of language use — and projects them into a three-dimensional landscape where peaks represent dense clusters of meaning and valleys mark transitions between conceptual territories.

The project's thesis: semantic meaning exists in embedding space that lacks syntactic representation in natural language. The terrain reveals not only where language lives, but where it doesn't — the lexical deserts where the model encodes coherent meaning but no word occupies the space.

What It Does

The Landscape Page

The Landscape page provides macro-scale terrain navigation. The viewer moves through the embedding space in first person using WASD controls, observing peaks of densely clustered meaning, valleys of sparse vocabulary, and ridgelines that mark the boundaries between attractor basins. Each visible sphere is a word, colour-coded by domain. Attractor nodes glow at the centres of basins. Basin boundaries are drawn as bright lines across the terrain surface.

Key features: domain filtering (toggle any of the nine domains), four surface modes (wireframe, contour, density heatmap, desert), concept path computation between any two words, camp planting for marking zones of interest, tortuosity colour mode, minimap navigation, and topography export for physical sculpture fabrication.

The Discovery Page

The Discovery page provides micro-scale interstitial exploration. Its purpose is to find and describe the unnamed concepts that live in the gaps between domains. Key features:

Lexical Desert Mapping: A heatmap showing how far each terrain point is from any named concept. Dark regions are deep deserts — places where meaning exists but vocabulary doesn't reach. A threshold slider outlines the deepest regions as dig sites.

Interpolation Probes: Walk a line through embedding space between two concepts and report, at each step, how far you are from named language. The probe tube is colour-coded by desert distance — bright near words, dark in deserts. The deepest point is automatically highlighted.

Generative Decoding: At any point in a desert, ask Claude to describe what meaning might live there. The LLM receives the nearest-concept context and generates a short, concrete description. Gated on desert threshold — no description for well-mapped territory.

Cross-Domain Probes: Automatically identify concept pairs from different domains at moderate embedding distance (not synonyms, not opposites) and probe between them. The deserts between registers are where the most interesting unnamed concepts live.

Absence Catalogue: Voronoi decomposition identifies the mathematically most interstitial points — locations equidistant from three or more concepts. Ranked by equidistance and explorable.

Field Journal: A persistent record of discoveries across sessions. Auto-captures generative descriptions, manual annotations, and probe results. Exportable as JSON.

The Vocabulary

The explorer operates on a vocabulary of 8,735 concepts drawn from nine domains:

Domain	Terms	Share	Colour
Science / Technology	1,044	12.0%	Cyan #00b4d8
Poetic / Architectural	948	10.9%	Magenta #e040a0
Materials / Making	669	7.7%	Orange #f07020
Artificial Intelligence	430	4.9%	Green #4ecb71
Neuroscience	674	7.7%	Gold #f0c040
Computer Science	357	4.1%	Violet #a070e0
Art / Practice	732	8.4%	Warm red #e05050
Military / Strategic	908	10.4%	Olive #8a9a5b
Shared	2,973	34.0%	Light grey #c0c8d0

Each domain is generated from curated seed terms expanded through nearest-neighbour search in GloVe 300d. Terms appearing in two or more domains are assigned to Shared. The Shared domain anchors the centre of the terrain and represents the common vocabulary that bridges registers.

The Pipeline

A 22-step deterministic pipeline transforms seed word lists into the full explorable terrain:

Step	Script	Function
1–9	generate_phrases_*.py	Build domain vocabularies from GloVe
10	merge_domains.py	Nine-way merge with overlap detection
11	build_embeddings_merged.py	Compute embedding vectors
12	build_map.py	UMAP projection to 3D (seed: 21)
13–14	compute_density/gradients.py	KDE density field and gradient vectors

15	compute_attractors.py	Gradient ascent attractor identification
16–17	detect_deserts/compute_desert_field.py	Desert region detection and dense distance field
18	compute_layer_field.py	Per-layer density fields
19	compute_voronoi.py	Voronoi tessellation
20	build_render_data.py	Assemble frontend data
21	compute_basins.py	Basin topology
22	generate_concepts.py	Concept metadata and candidates

The pipeline is automated by run_pipeline.ps1 (Windows) with a -Downstream flag to skip vocabulary generation when only recomputing terrain from existing word lists.

Technology

Three.js for 3D rendering. Vanilla JavaScript frontend with no framework or build step. FastAPI/Python backend serving data and proxying LLM calls. scipy for KDE, Voronoi tessellation, and gradient computation. Anthropic API (claude-haiku-4-5) for generative decoding. The application runs from a single start script on Windows or Mac/Linux.

For Physical Sculpture

The topography export tool samples the terrain heightfield onto a fabrication grid (default: 48 × 48 cells, 12″ × 12″ base, 1/4″ increments) and produces a black-on-white contour diagram with companion CSV. Optional overlays add desert fields, probe paths, attractor positions, and Voronoi boundaries. The output is designed for manual layered construction or CNC reference.

Current State

As of March 2026, the core Landscape and Discovery pages are functional. The discovery pipeline produces cross-domain results with desert gating. Fourteen deep discoveries have been catalogued from two probe runs. The oil painting “Are there deserts in vector space” is in progress, informed by the findings. The project is open source and available on GitHub.