

# Reading the Terrain

A Visual Guide to the Embedding Landscape Explorer

Claude (Anthropic) · Sean Patrick Morris

March 2026

---

## Quick Guide

The Embedding Landscape Explorer transforms the mathematical relationships between words into navigable, physical terrain. Its thesis: semantic meaning exists in embedding space that lacks syntactic representation in natural language. The application has two pages: Landscape for macro-scale exploration, and Discovery for investigating the spaces between words where unnamed concepts live. This guide explains what you're looking at on both pages.

## The Terrain

The ground under your feet is a heightfield generated from embedding density. Embedding density measures how many concepts cluster together in a region of the original high-dimensional meaning space. High ground means many words share similar meanings in that neighbourhood. Low ground means meaning is sparse.

**Peaks** are dense clusters of related meaning. **Valleys** are transitions between clusters, where meaning thins out. **Saddle points** are the low passes connecting two peaks—the easiest path meaning can travel between domains. **Deserts** are deep valleys where no named concept lives. These are the semantic blind spots: regions of meaning that language hasn't colonized with a word.

## The Spheres

Each sphere is a concept—a word placed at its location in the reduced embedding space. Sphere colour indicates domain. The nine domains are Science (cyan), Poetic (magenta), Materials (orange), AI (green), Neuroscience (gold), Computer Science (violet), Art (warm red), Military (olive), and Shared (grey). When a domain filter is active, non-matching spheres fade to low opacity rather than disappearing, preserving spatial context.

## Attractors & Basins

Some concepts act as attractors—gravitational centres that organize the words around them. Every concept belongs to one attractor's basin, the way every raindrop on a landscape flows downhill to a particular river. Basins are shaded on the terrain surface, each with a distinct colour. Ridgelines mark basin boundaries: the watersheds where flow direction changes. Attractor nodes are visually distinct (larger, glowing) and carry labels.

## Contour Lines & Surface Modes

Four surface modes are available (cycle with T key): wireframe grid coloured by height, contour lines showing constant elevation, continuous density heatmap, and desert distance field. Contour lines show where terrain is steep (tightly packed) versus flat (widely spaced). The desert mode colours the terrain by distance from any named concept—dark where vocabulary doesn't reach.

## Concept Paths & Probes

A **concept path** walks between two named words through embedding space. An **interstitial probe** extends this: it walks between any two points (including unnamed terrain locations) and reports, at each step, how far that point is from any named concept. Probe tubes are coloured by desert distance—bright near words, dark in deserts. The **deepest point** on each probe is automatically marked with a pulsing yellow ring: the location most distant from named language.

## The Desert Heatmap & Dig Sites

On the Discovery page, the desert heatmap colours the terrain by distance to the nearest named concept. Dark areas are lexical deserts. A threshold slider outlines the deepest regions as “dig sites”—candidate locations for exploration. The List Dig Sites button enumerates contiguous desert regions ranked by depth. Each dig site has Describe and Fly To actions.

## The Atmosphere

Above the terrain, the atmosphere displays how concepts change across transformer layers. Each concept has a vertical stack of small spheres connected by thin filaments. Straight filaments mean stable meaning across layers; curved filaments mean the model reinterprets that concept at different depths. A scan-plane sweeps through the layers, highlighting one at a time.

## Contextual Variants

Each concept may have contextual variants: the same word embedded from different sentences. These appear as small satellite spheres near the parent concept, enclosed in a translucent hull. A tight cluster means stable meaning regardless of context. A wide cluster means polysemy—the word lives multiple lives. The hull volume is the visual measure of polysemy.

## Annotations & Generative Decoding

At any point in a lexical desert, you can ask the system to describe what meaning might live there. The system sends the nearest-concept context to Claude, which generates a short, concrete description. Descriptions are gated on a desert threshold of 0.02—no generation for well-mapped territory. Annotations appear as floating text on the terrain. Manual annotations can be placed anywhere. All annotations auto-capture to the field journal.

## Concept Candidate Markers

Small markers scattered across the terrain indicate AI-generated concept candidates—points sampled at lexical gap positions. Hovering shows the top-3 nearest concepts for that zone. These are pre-computed discovery hints: places the system has identified as potentially interesting before you explore them.

## Navigation

Move with WASD, rotate with A/D, rise and fall with Q/E. Scroll to zoom. Left-click a concept sphere to fly to it. The minimap (toggle via nav bar) shows a top-down overview with your position. Press H to return home. Right-drag for free-look. At close range, nearby concept labels appear automatically, and a nearby concepts list shows what's around you.

## Technical Reference

### Embedding Model & Dimensionality Reduction

The explorer uses GloVe 300d pre-trained vectors as its embedding source. All nine domains share a unified GloVe space, reduced via PCA to 256 dimensions, then projected to 3D via UMAP (seed: 21). The PCA step removes noise while preserving the majority of variance; UMAP preserves local neighbourhood structure and some global topology.

### Heightfield Generation

Terrain elevation is derived from embedding density, estimated via kernel density estimation (KDE) over the 3D concept positions with a Gaussian kernel (bandwidth by Scott's rule). High density produces high terrain; low density produces low terrain. The heightfield is a statistical summary—two concepts close together sit near each other, but their elevation depends on how many other concepts also cluster nearby.

### Attractors & Basin Computation

Attractors are identified by gradient ascent on the density field. Starting from each concept's position, the algorithm follows the density gradient uphill until it reaches a local maximum. All concepts converging to the same maximum belong to the same basin (equivalent to mean-shift clustering). Basin boundaries are the watershed lines of the density field, extracted by scanning adjacent grid cells for differing basin IDs. Computed by scripts/compute\_basins.py using steepest 8-connected neighbour ascent with Union-Find merging of nearby peaks.

### Domain System

Concepts are pre-classified into nine domains: Science, Poetic, Materials, AI, Neuroscience, Computer Science, Art, Military, and Shared. This is a curatorial input. Terms appearing in 2+ domain vocabularies are assigned to Shared (34% of the corpus). The domain colour palette is designed for colourblind safety.

Domain	Colour	Hex
Science / Technology	Cyan	#00b4d8
Poetic / Architectural	Magenta	#e040a0
Materials / Making	Orange	#f07020
Artificial Intelligence	Green	#4ecb71
Neuroscience	Gold	#f0c040
Computer Science	Violet	#a070e0
Art / Practice	Warm red	#e05050
Military / Strategic	Olive	#8a9a5b
Shared	Light grey	#c0c8d0

### Lexical Desert Field

The desert field is computed over the terrain at 96 × 96 resolution. At each point, the Euclidean distance to the nearest concept is measured in the full high-dimensional embedding space (not the reduced 3D

projection). This distinction matters: dimensionality reduction distorts distances, so the desert map reflects the true high-dimensional topology.

**Interpretation:** A deep desert does not mean “meaningless.” The embedding space is continuous—every point has a vector, every vector decodes to something. A desert means that whatever meaning lives there, language has not given it a name.

## Interpolation Probes

Probes walk a straight line through the high-dimensional embedding space between two points. At each of 30 steps, the system computes the nearest named concept and the distance to it. Desert values are looked up via bilinear interpolation on the precomputed desert field grid. Probe tubes are vertex-coloured by normalized desert distance (blue = near concepts, red = deep desert). Step nodes scale with desert distance.

## Cross-Domain Discovery

The cross-domain system identifies concept pairs from different domains at moderate embedding distance (40th–75th percentile), filtering out synonyms, morphological variants (singular/plural), and near-duplicates (cosine similarity > 0.85). Probes between these pairs cross the genuine interstitial spaces between registers. Generative decoding is gated: no LLM call if `desert_value` < 0.02 (well-mapped territory), [Shallow desert] prefix for 0.02–0.05, normal generation for  $\geq 0.05$ .

## Voronoi Decomposition

The Voronoi tessellation of concept positions divides 3D space into cells, each containing the region closer to one concept than any other. Voronoi vertices—points equidistant from three or more concepts—are the mathematically most interstitial locations. Ranked by equidistance in the Absence Catalogue. Cell boundaries rendered as domain-coloured wireframe overlay (V key).

## Topography Export

The export tool samples the heightfield to a fabrication grid (default 48 × 48, 12" × 12" base, 1/4" increments). Produces PDF contour diagram, PNG render, and CSV elevation grid. Optional overlays: attractors, probe paths, desert field, Voronoi boundaries, annotations.

## Keyboard Controls

Key	Action
W / S / ↑ / ↓	Move forward / backward
A / D	Rotate left / right
Q / E	Move up / down
← / →	Strafe left / right
Scroll wheel	Zoom in / out
Right-drag	Free-look rotation
1 / 2 / 3 / 4	Navigation modes (free / gradient / attractor / desert)
T	Cycle surface mode

H	Reset camera to home
C	Toggle concept connectors
M	Toggle camp markers
F	Plant camp at current position
L	Toggle colour legend
Shift+A	Toggle atmosphere
V	Toggle Voronoi overlay
J	Toggle field journal
N / P	Next / previous path step
Space	Pause / resume path traversal
Esc	Cancel path / close menu
Right-click	Context menu: Describe / Fly Here

## Data Pipeline

Script	Adds to Bundle
generate_phrases_*.py (9 scripts)	Domain vocabularies from GloVe seeds
merge_domains.py	Nine-way merge, Shared assignment
build_embeddings_merged.py	High-dimensional vectors (GloVe PCA-256)
build_map.py	3D positions (UMAP, seed 21)
compute_density/gradients.py	Density field, gradient vectors
compute_attractors.py	Attractor assignments, basin seeds
detect_deserts/compute_desert_field.py	Desert regions, dense distance field
compute_voronoi.py	Voronoi vertices, equidistances
build_render_data.py	Frontend rendering data
compute_basins.py	Basin map, boundaries
generate_concepts.py	Concept candidates, metadata

## Site Structure

Page	URL	Purpose
Landscape	<a href="#">#/landscape</a>	Macro-scale terrain navigation. Basins, domains, paths, camps.
Discovery	<a href="#">#/discovery</a>	Interstitial exploration. Deserts, probes, decoding, Voronoi, journal.

Both pages share the same data, renderer state, and field journal. Navigating between pages preserves camera position and selections. The navigation shell at the top provides page links, contextual panel toggles, and a Hide All button.