THOUGHT EXPERIMENT

Beyond Words: Could an LLM Learn Directly from the Electromagnetic Spectrum?

What the geometry inside language models suggests about training AI on frequency, phase and amplitude instead of word tokens

By Andrew Pritchard, Gemini & MS-CoPilot · AI · Mechanistic interpretability · Thought experiment

The shapes inside the machine

Mechanistic interpretability research suggests that large language models organise what they learn into recognisable geometric structures within their activation space. Cyclic concepts such as months and days form rings, years trace helices, grammatical categories pack into polytopes, scales such as sentiment form lines, and hierarchies branch like trees.

The infographic below summarises these five families of shape and the kinds of concept each one represents.

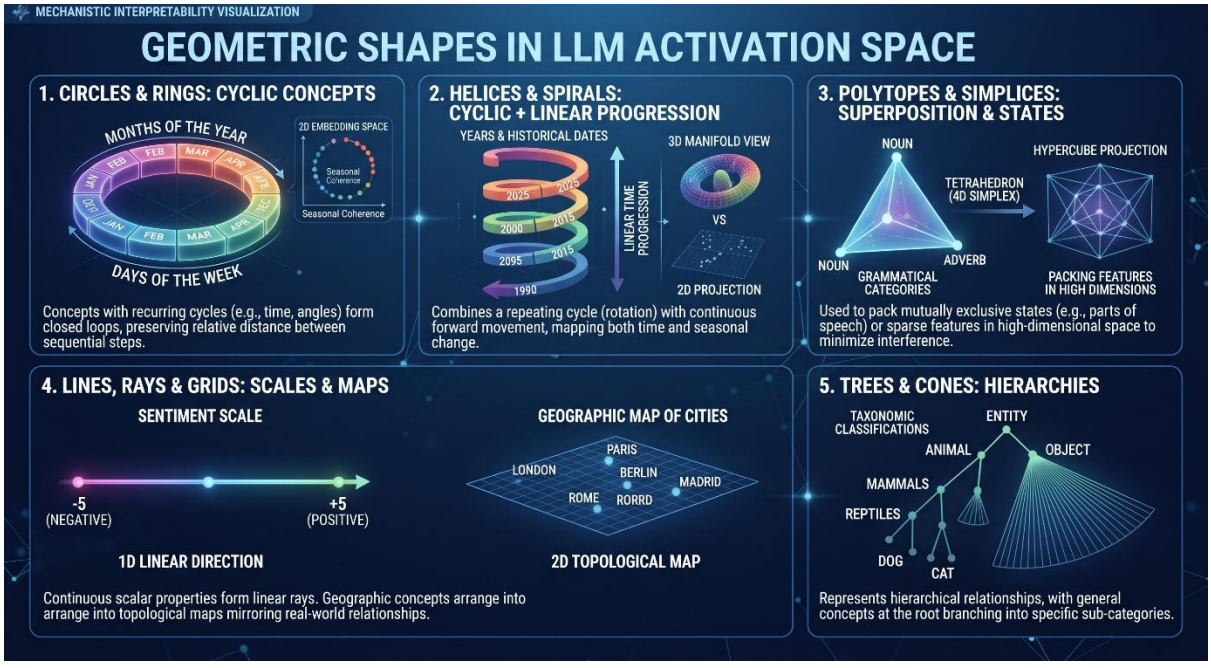

Figure 1: Geometric shapes in LLM activation space: rings, helices, polytopes, lines and grids, and trees.

The thought experiment: swap words for waves

If geometry is how a model organises structure, does it matter that the input was language? This thought experiment asks what would happen if the tokens were not

words at all, but raw electromagnetic signals described by frequency, phase and amplitude. A multi-band EM feed, drawn from Earth's ionosphere and from space, would be converted by FFT and spectral tokenisation and passed through a deep network, which would learn a "physical manifold".

On this manifold, the same families of shape would take on physical meaning: circles and rings for phase coherence and periodic cycles, helices for frequency sweeps and Doppler drift, polytopes for mode superposition and polarisation states, and lines and grids for parameter scales and intensity maps.

Using the geometry as a consistency checker

The infographic below visualises the core of the thought experiment: how a neural network trained on raw electromagnetic signals builds a physical manifold, and how that geometry can be used to evaluate new, incoming signals.

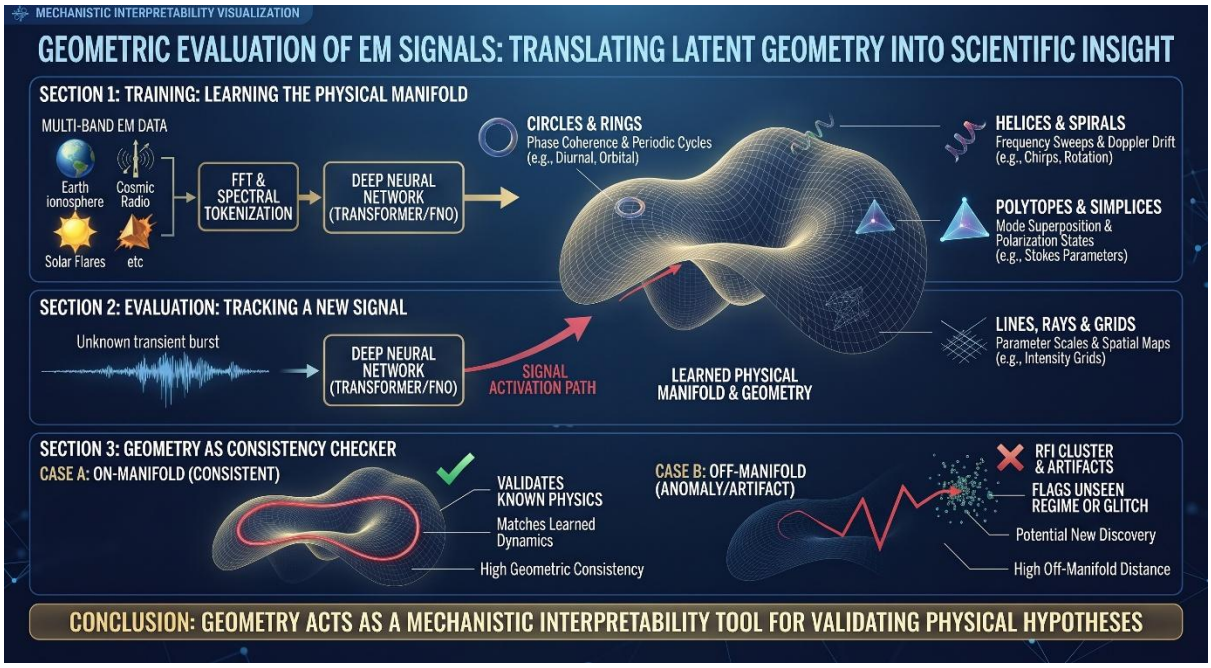

Figure 2: Training a physical manifold from multi-band EM data, then tracking a new signal's activation path against it.

Figure 3 complements Figure 2 by showing the same pipeline in a second layout, with the on-manifold and off-manifold outcomes spelled out, including the RFI artefact cluster and the high off-manifold distance that flags an anomaly.

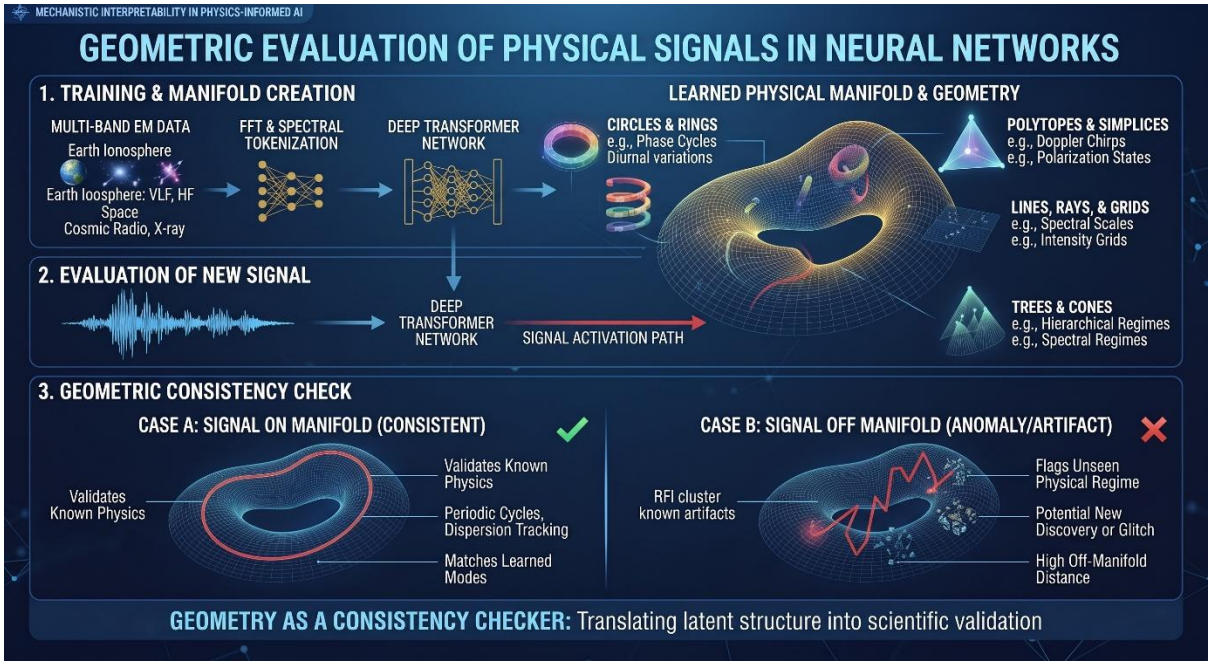

Figure 3: A companion view of the same process, Geometric Evaluation of Physical Signals in Neural Networks, showing the on-manifold and off-manifold cases.

1. Training and manifold creation

The model ingests multi-band EM data from Earth and space, processes it via spectral tokenisation, and learns a structured geometric landscape. The shapes discussed above are physically realised on this manifold, representing known physical behaviours such as phase coherence, dispersion and polarisation states.

2. Evaluation of a new signal

A previously unseen signal is fed into the network, and its "activation path" (the red line in the figure) is tracked as it moves through the high-dimensional activation space.

3. Geometric consistency check

The path is then compared against the learned manifold, with two possible outcomes.

- **Case A, on-manifold:** the signal's path smoothly follows the existing geometry, for example tracing a clean loop. This validates that the signal is consistent with known physics, confirming its modes and coherence.
- **Case B, off-manifold:** the path deviates sharply, makes erratic jumps, or enters a cluster of known artefacts such as RFI. This flags an anomaly, indicating an unseen physical regime, a potential new discovery, or an instrumental glitch.

Closing thought

This "inverse" approach turns the neural network's latent geometry into a powerful, data-driven consistency checker for validating physical signals against the laws of the universe captured during training.