RENDERING ENGINES OF A TEXTUAL WORLD

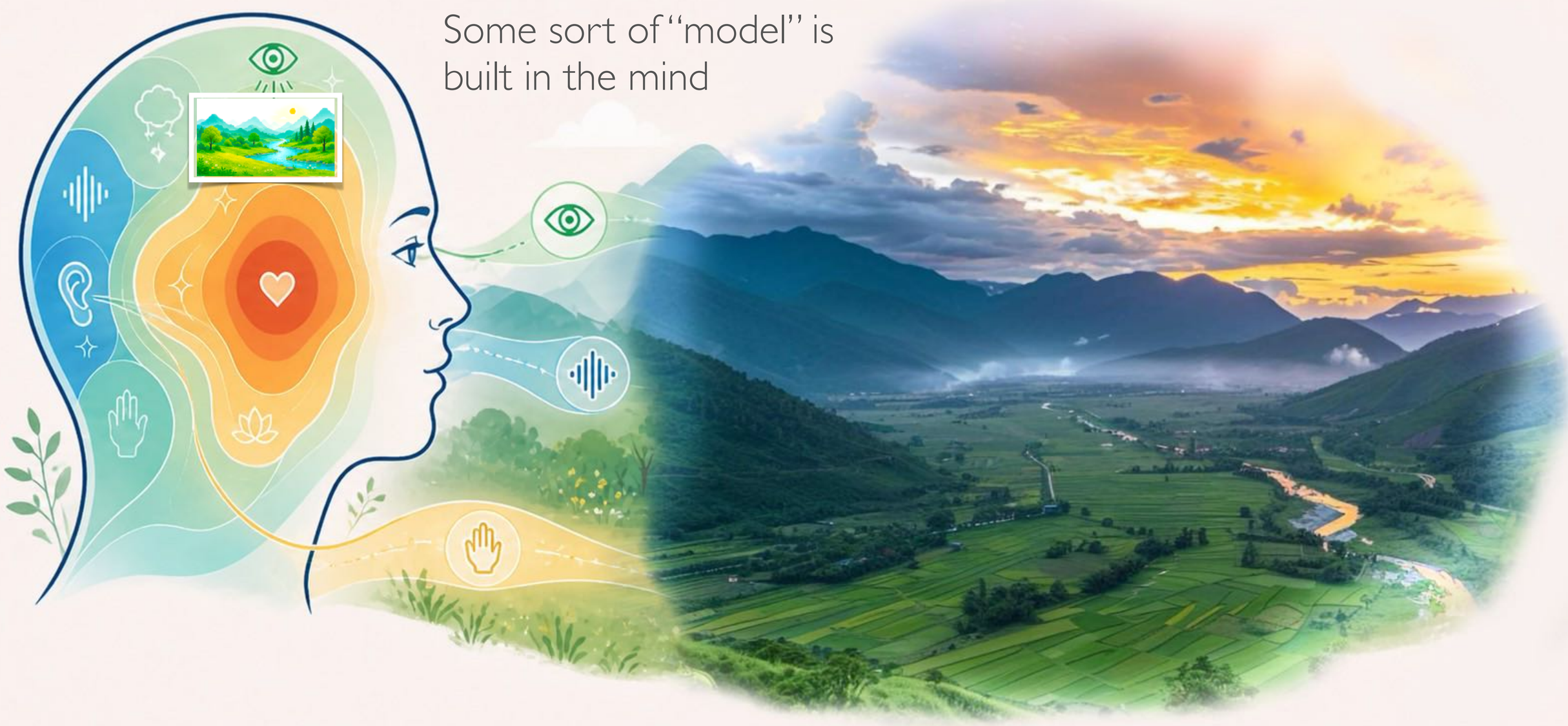
What LLMs Reconstruct (A Thought in Progress)

Beyond the Imitation Game Conference
10th of May 2026

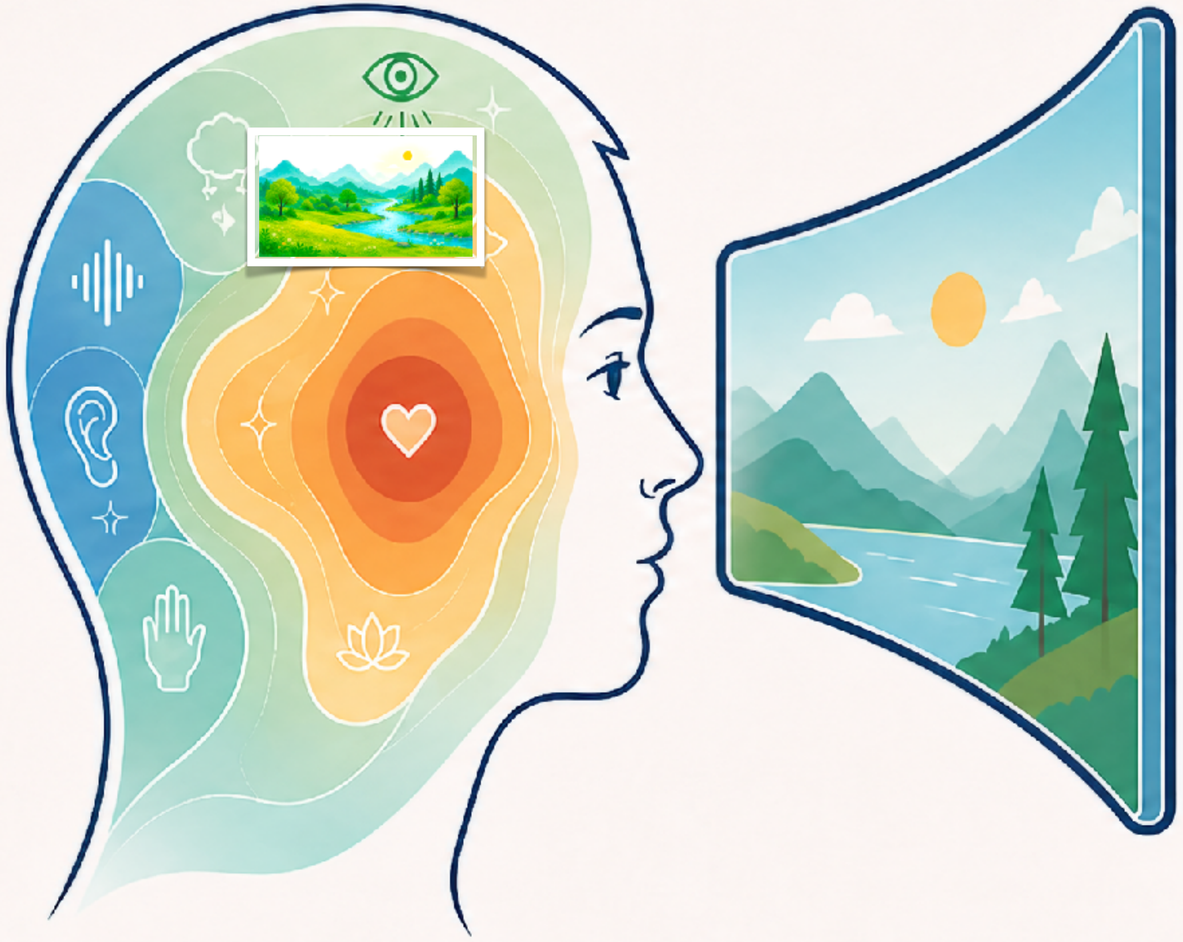

THE WORLD

TAKING IN THE WORLD

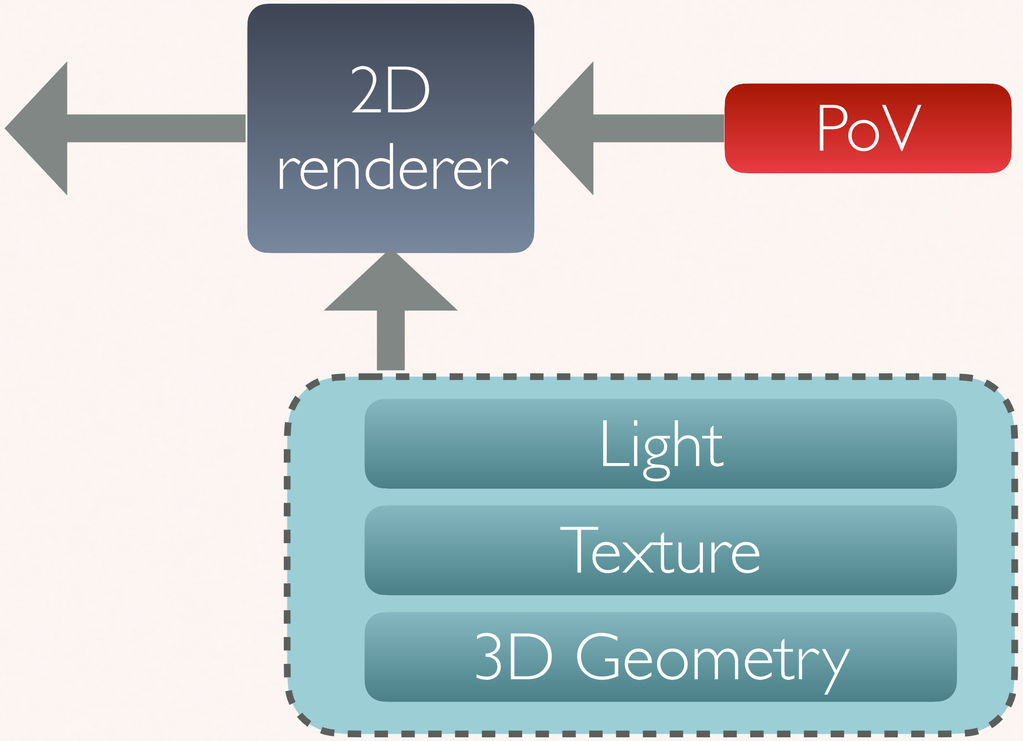
Some sort of "model" is built in the mind



RENDERING THE WORLD IN VR



How does this work?

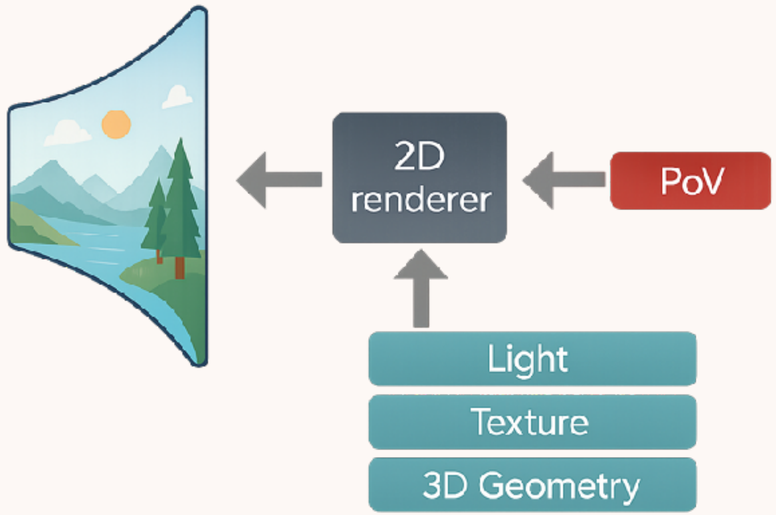


Internal Machine Model

Again, a "model" is still built in the mind

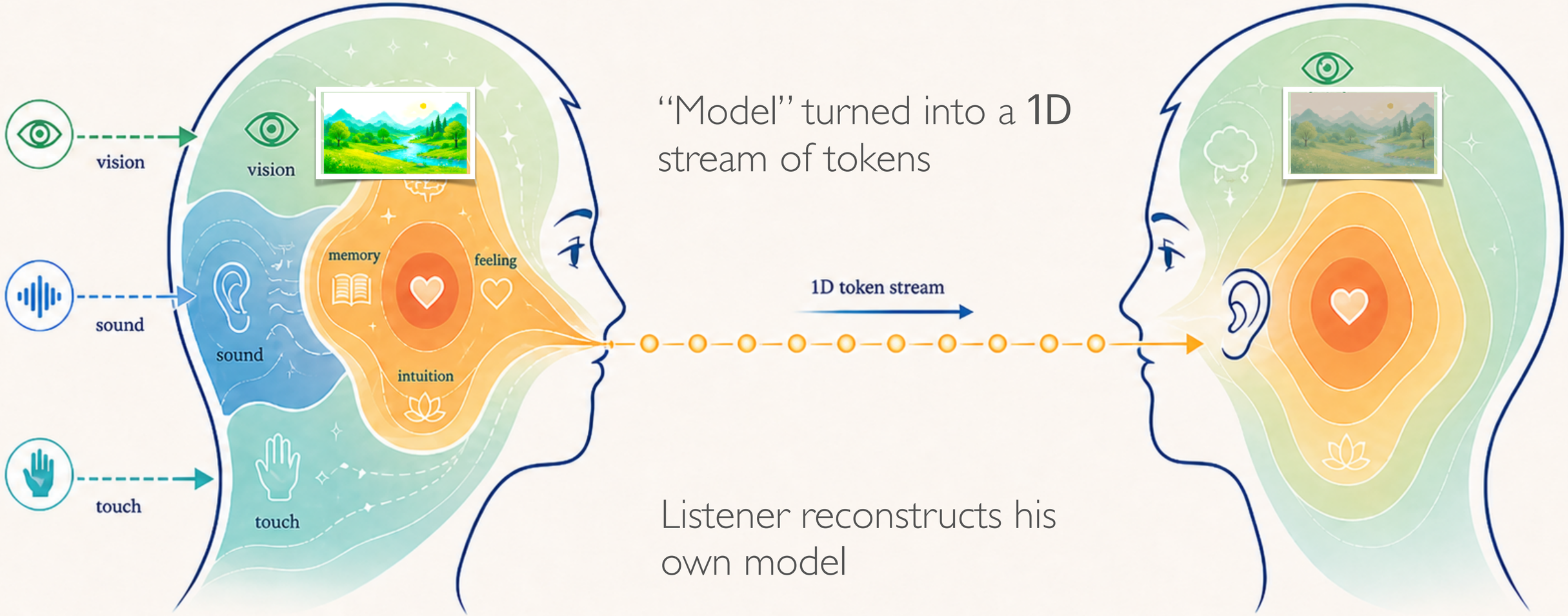
POV INSIDE A VR ENGINE?

- **3D+ model**
  - view-independent scene-structure: geometry, materials, textures, lights, relations
- **PoV**
  - camera state: position, direction, zoom / field of view
- **Rendering**
  - computing a 2D image from the model and the PoV



- Same model, many renders.
- **Matrix multiplication** is essential for rendering.

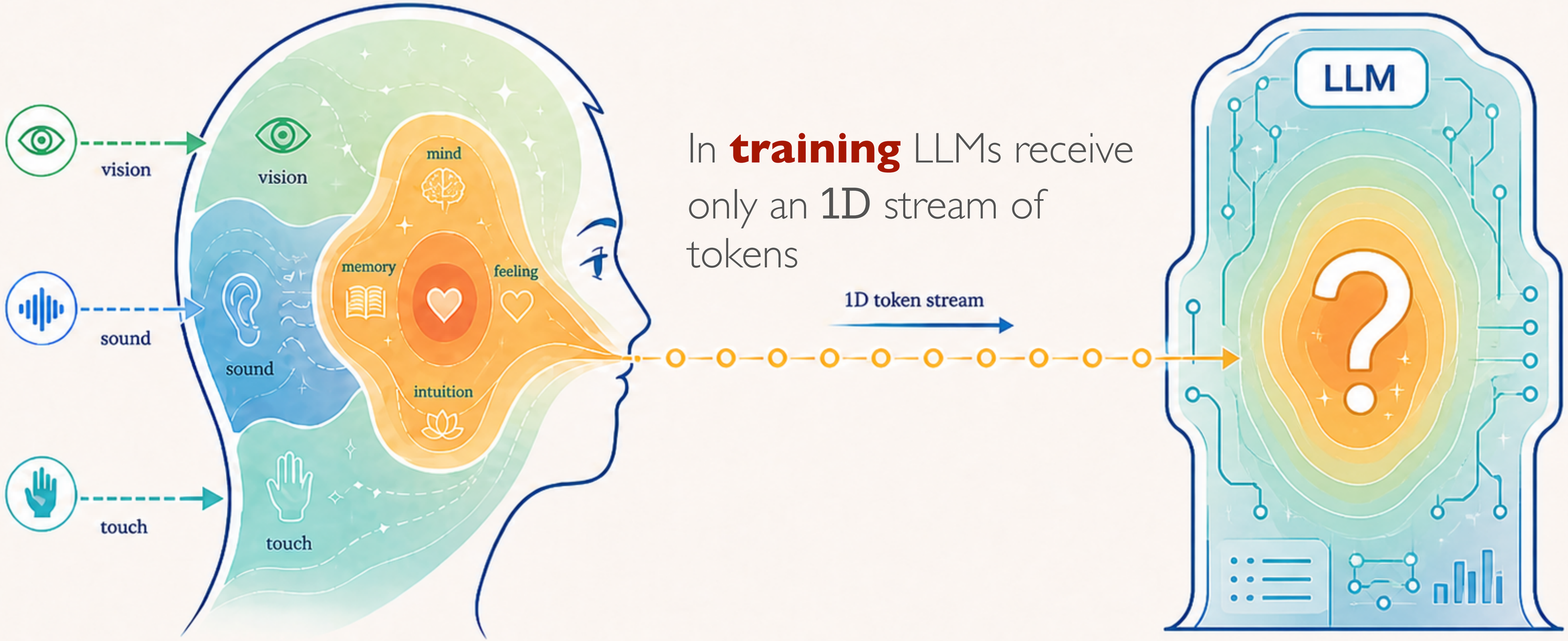
TELLING ABOUT THE WORLD



"Model" turned into a 1D stream of tokens

Listener reconstructs his own model

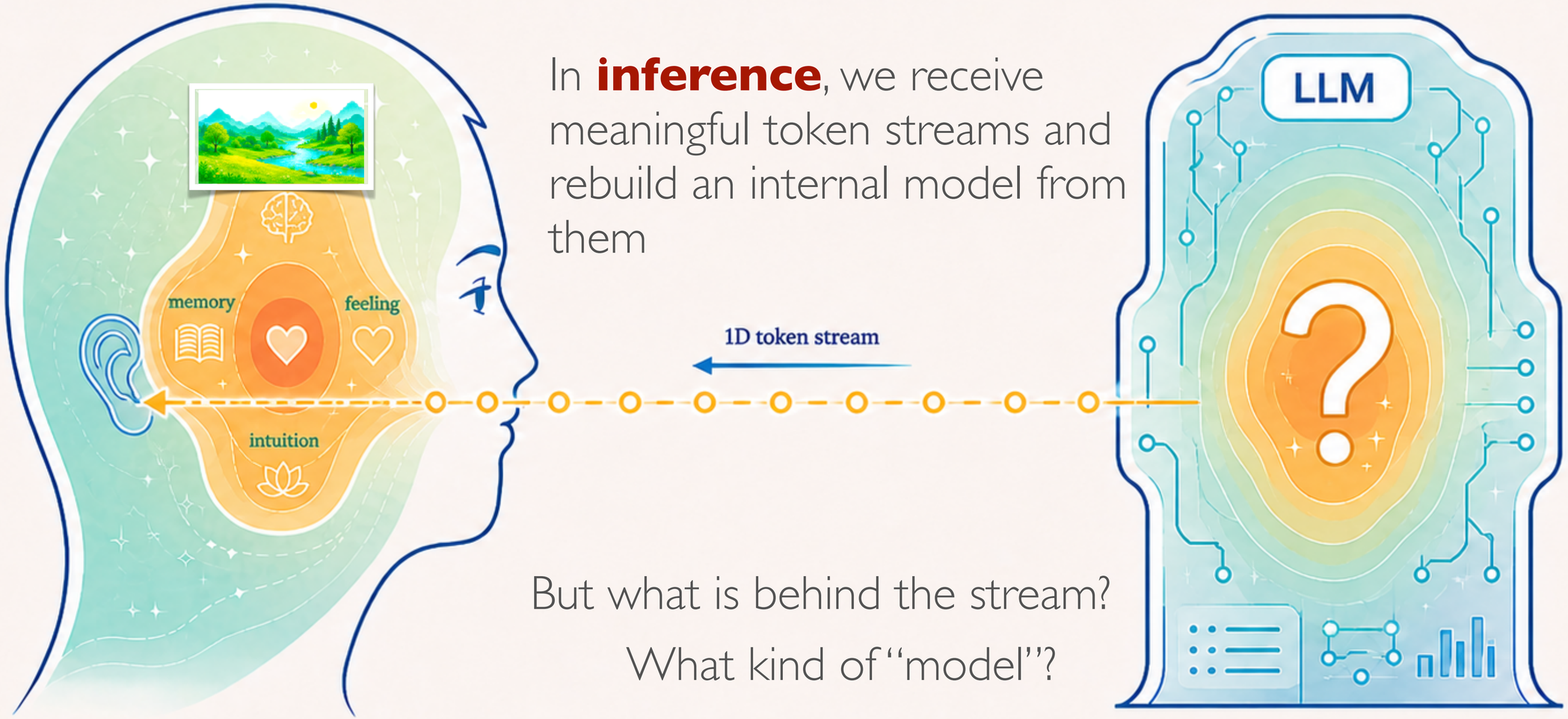
WHAT ABOUT LLMS?



In **training** LLMs receive only an 1D stream of tokens

From many such streams, they build a high-dimensional representational space: 10.000+ D

“DESCRIBE ME A NICE MEADOW”



In **inference**, we receive meaningful token streams and rebuild an internal model from them

But what is behind the stream?

What kind of “model”?

A TEXT RENDERING ENGINE



Some "text landscape" is selected
around the "PoV", which is the prompt

TWO WAYS TO LOOK AT A PROMPT

- A prompt can be represented as:
  - **a point** in a sentence embedding space
  - **a collection of points** in a token embedding space
- A sentence embedding space is better to build a visualization:
  - the initial prompt is one point
  - each completed answer is one point
  - sampling many answers shows the region of the embedding space the prompt licenses

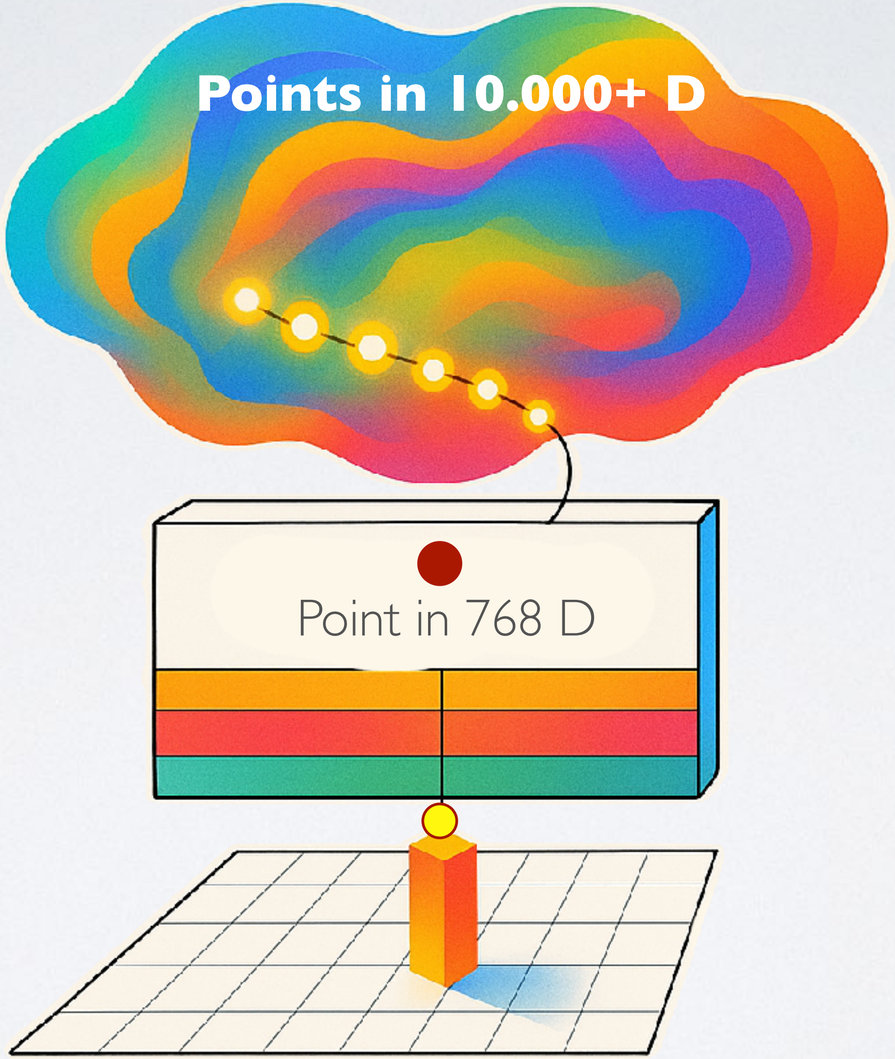
SEEING THE MANIFOLD

- The prompt we'll use: "Describe me a nice meadow."
- We give the prompt to a local LLM (qwen3:14b) and sample 100 different completions at temperature 1.5.
- Each completion is taken as a single full text, then passed to a sentence embedding model (nomic-embed-text), which turns the text into a single 768-dimensional vector.
- Multidimensional scaling projects all of these vectors down to 3D, preserving distances as well as possible.

10.000+D Geometry

768 D Geometry

3D Geometry



Point in 3 D

Series of tokens

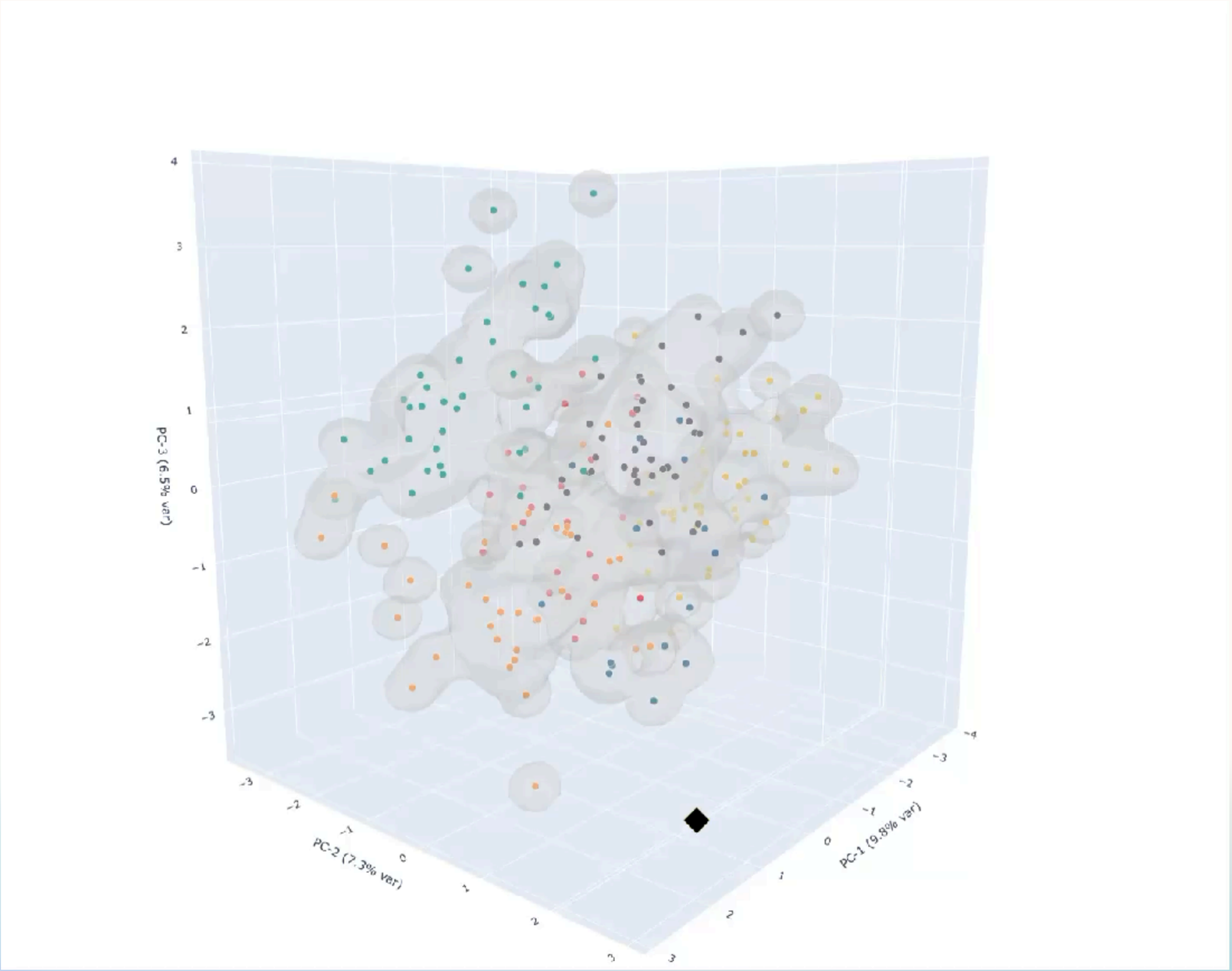
Sentence

Point in 3D

“DESCRIBE ME A MEADOW AT SUNRISE”

Embedding space: **768 D**

Intrinsic: **26 D**



100 different ending answers

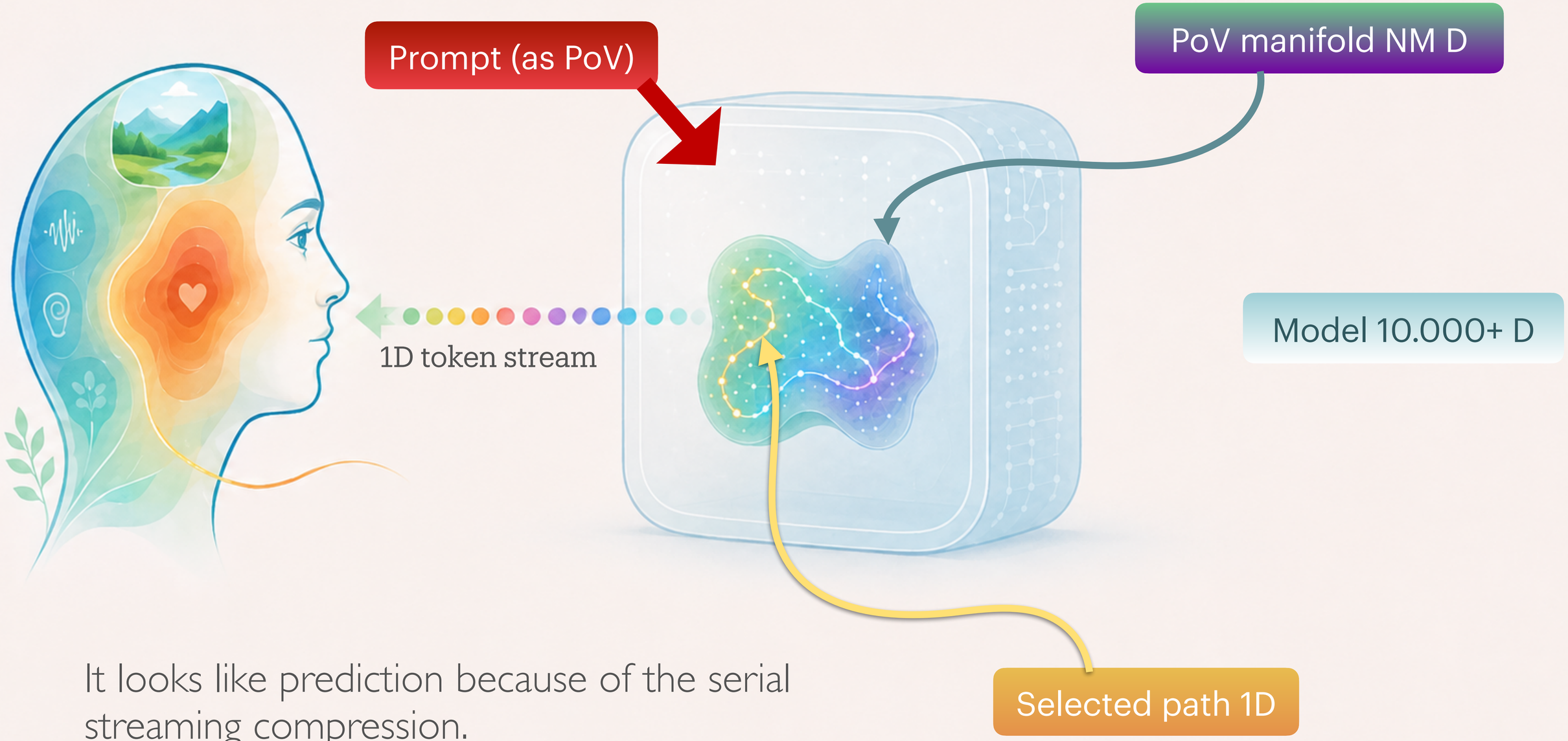
WHAT DOES IT (REALLY) MEAN?

- The responses lie on a roughly 26-dimensional surface curving through 768-dimensional space.

WHAT IS THIS MANIFOLD?

- All possible responses lie on a manifold (a convoluted surface) in a much lower-dimensional subspace inside the embedding space.
- The prompt appears to induce a structured manifold in the representation space.
- It is like the LLM selects a "region" of its embedding space based on the prompt, and then its answers are simply anything encountered in this region (sentences).
- This is the same architecture as a VR engine: a stored high-dimensional model, a viewpoint that selects which region is active, and a rendering process that compresses it back to the output channel: a flat 2D image.

THIS IS NOT JUST “NEXT TOKEN PREDICTION”



It looks like prediction because of the serial streaming compression.

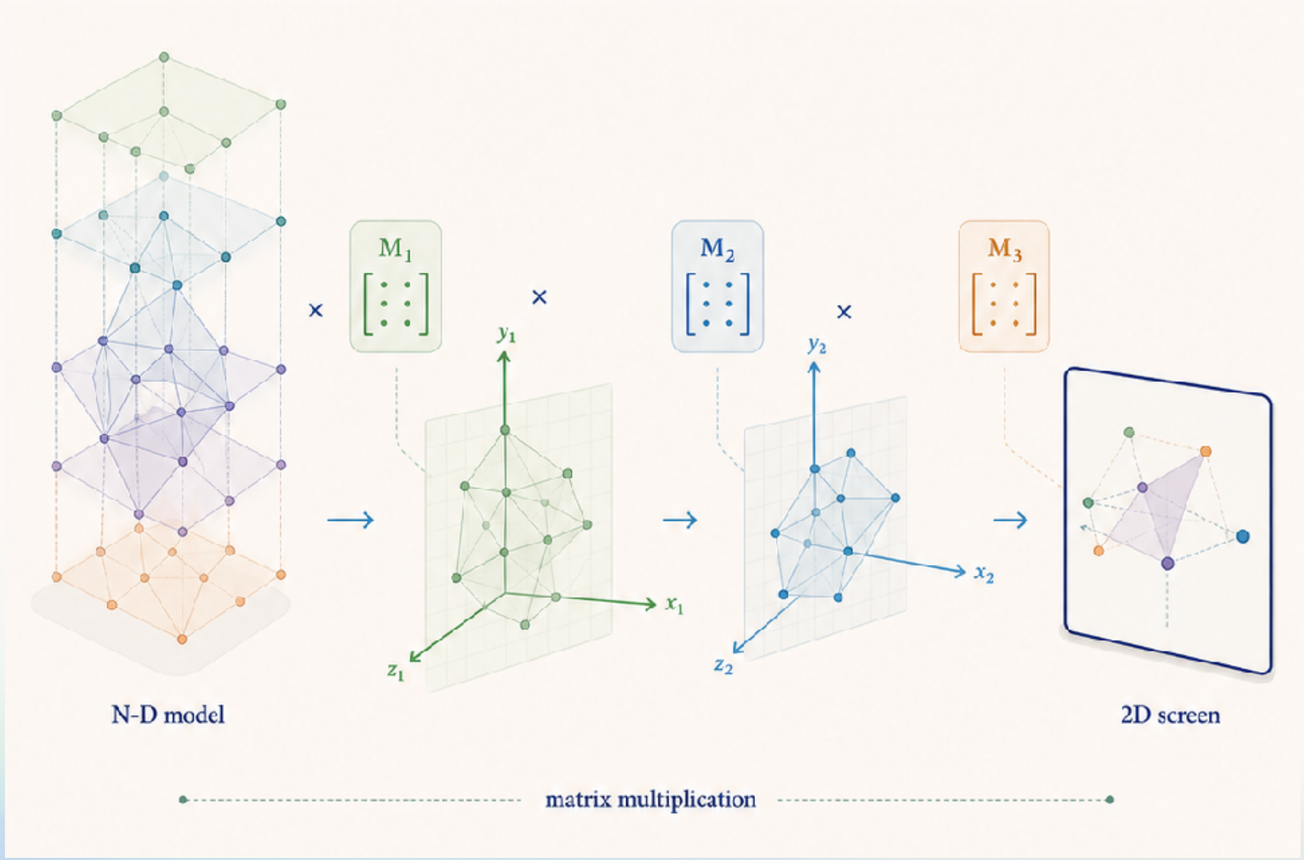
Prompt / PoV: “What is the capital of France?”



1D token stream

LOGITS IN NEXT TOKEN PREDICTION

MATRIX MULTIPLICATION IN RENDERING



Graphic cards are implementing the fundamental rendering operation: Matrix Multiplication (matmul).

It is by matmul that moving through different dimension representations is possible: Internal model + PoV -> 2D screen

MATMUL COINCIDENCES

- **Both LLMs and VR rely on matrix multiplication** as their core primitive, even if they are completely different animals.
- This is not a coincidence. Both transform structured representations through intermediate coordinate systems before producing an output constrained by the output channel.
  - VR: model space → camera space → screen space
  - Transformer: token embeddings → contextualized hidden states → next-token scores → 1D token stream

CLOSING THE ANALOGY

- **Rendering** = computing an output from a stored model and a selector.
  - PoV + world's geometry in VR.
  - Prompt + embedded language model in LLMs.
- Works like this:
  - Selector **activates a region** of the model's high-dimensional internal geometry.
  - Transformation **compresses that region** into the low-dimensional output channel.

WHY THIS MATTERS PHILOSOPHICALLY

- "Next-token prediction", as a description, is describing the output structure. It is silent about the middle-layer structure.
- An LLM has no default standpoint. Like a VR model without a camera, the geometry is fully there but nothing renders until a prompt supplies the perspective. That means, it has no "ego", it is just a renderer without opinions.
- Like a multiplayer VR engine rendering many cameras at once, an LLM can render many views from the same stored geometry — including mutually contradictory ones, since the geometry contains every position its training data licensed.

MORE (PHILOSOPHICAL) THOUGHTS

- LLMs are exceptional tools to understand language and our minds. They work much as the telescope or microscope in Early Modern Science: we can see hidden things. Expect a new science to be born.

- Our minds seem to work the same way: when a question is asked, we instantly know "the manifold" on which our answer lies. We use context (audience, level, etc.) but we essentially move on the manifold. This is why talking can be "automatic", whatever we select it mostly makes sense.

- LLMs build a model of the world from some sort of back-engineering of our texts?

SITUATING THE CLAIM

- Park, Choe & Veitch (2023ff) — linear/geometric structure of concepts in activation space (hierarchy is orthogonality)
- Huh et al. (2024), The Platonic Representation Hypothesis — convergence as evidence of stored structure (image/text representations convergence)
- Gärdenfors, Conceptual Spaces — concepts as geometry in human minds


THANK YOU!
QUESTIONS?