

# An Explainable Diagnostic Framework for Neurodegenerative Dementias

Nathanaël Fijalkow

université  
de BORDEAUX

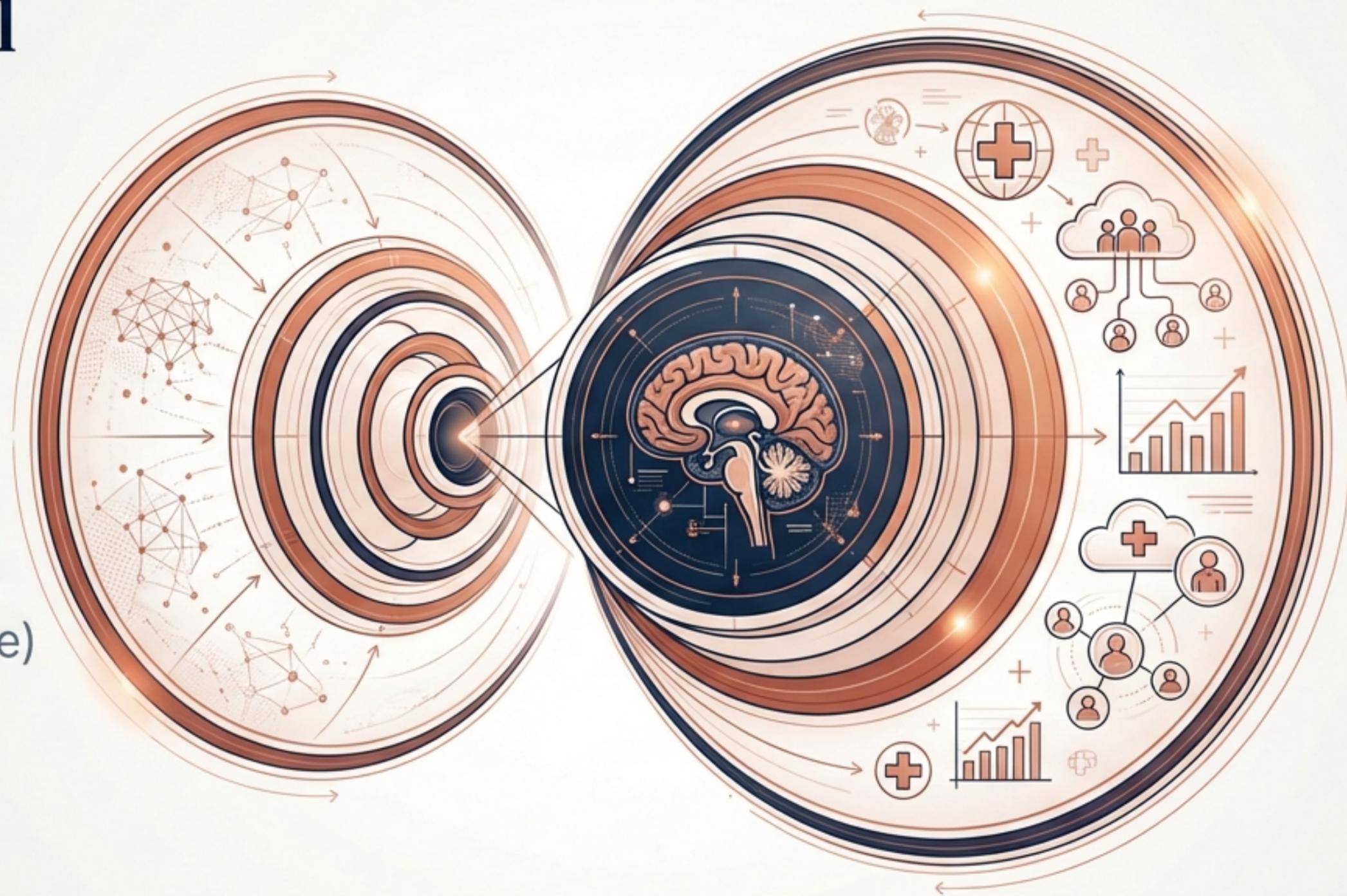


**LaBRI**

**Ni** N u k k A I

# The journey from general intelligence to clinical application

1. The state of Generative AI in 2026 (The Macro)
2. How GenAI helps analyse brain scans (The Micro)
3. A discussion on AI-assisted diagnosis (The Macro Future)



# Part 1: Generative Artificial Intelligence

# Turning language into mathematics

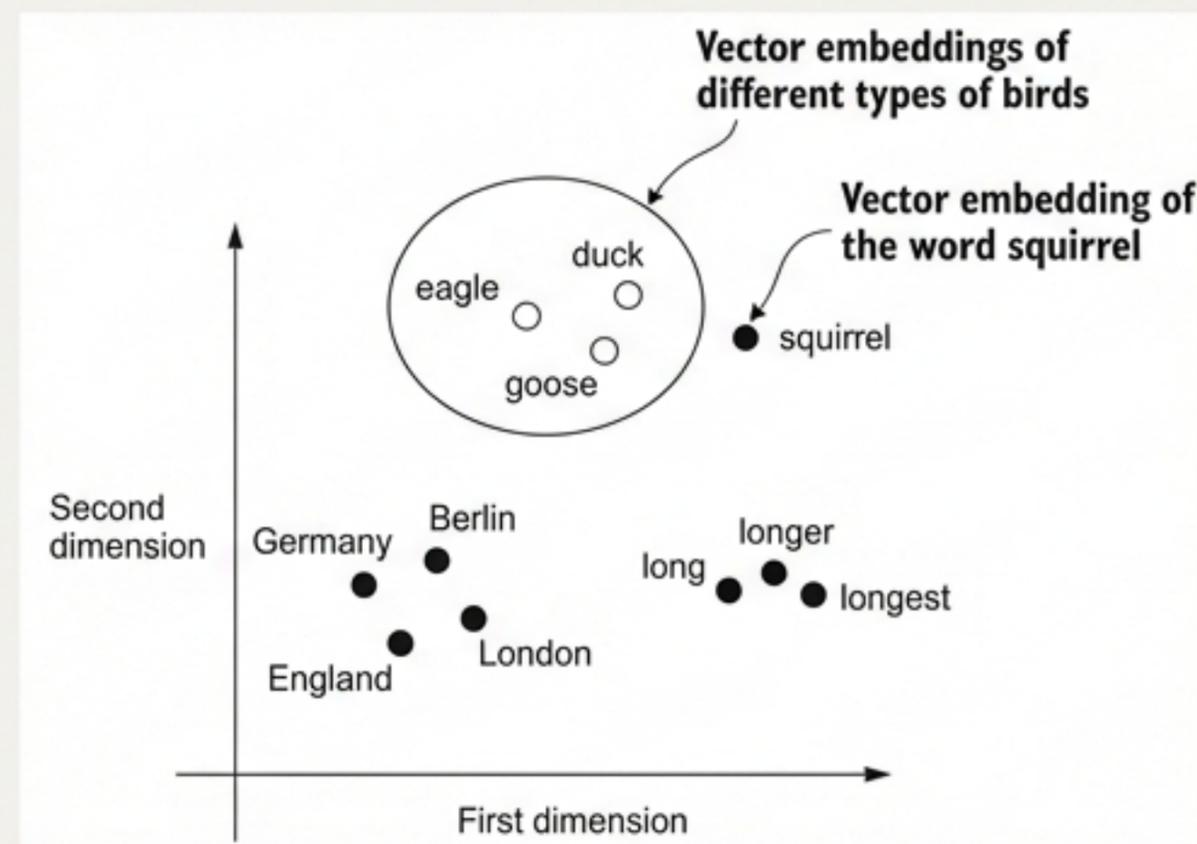
At its core, a language model simply takes an input text and predicts the next word—a concept fantasised about since the 1960s. The silent breakthrough arrived in 2003: capturing the way words are combined in sentences by turning language into numerical vectors called embeddings.

## A Neural Probabilistic Language Model

Yoshua Bengio  
R jean Ducharme  
Pascal Vincent  
Christian Jauvin

D partement d'Informatique et Recherche Op rationnelle  
Centre de Recherche Math matiques  
Universit  de Montr al, Montr al, Qu bec, Canada

BENGIOY@IRO.UMONTREAL.CA  
DUCHARME@IRO.UMONTREAL.CA  
VINCENTP@IRO.UMONTREAL.CA  
JAUVINC@IRO.UMONTREAL.CA



# The attention mechanism solves the context bottleneck

Between 2003 and 2017, the growing success of deep learning hit a wall: models struggled to understand longer text.

The 2017 breakthrough didn't do anything radically new. It proved that the attention mechanism alone was enough to enable long contexts, and its key property was its absolute simplicity.

## Attention Is All You Need

|   |   |  |  |
|---|---|--|--|
| <b>Ashish Vaswani*</b><br>Google Brain<br>avaswani@google.com | <b>Noam Shazeer*</b><br>Google Brain<br>noam@google.com | <b>Niki Parmar*</b><br>Google Research<br>nikip@google.com | <b>Jakob Uszkoreit*</b><br>Google Research<br>usz@google.com |
|---|---|--|--|

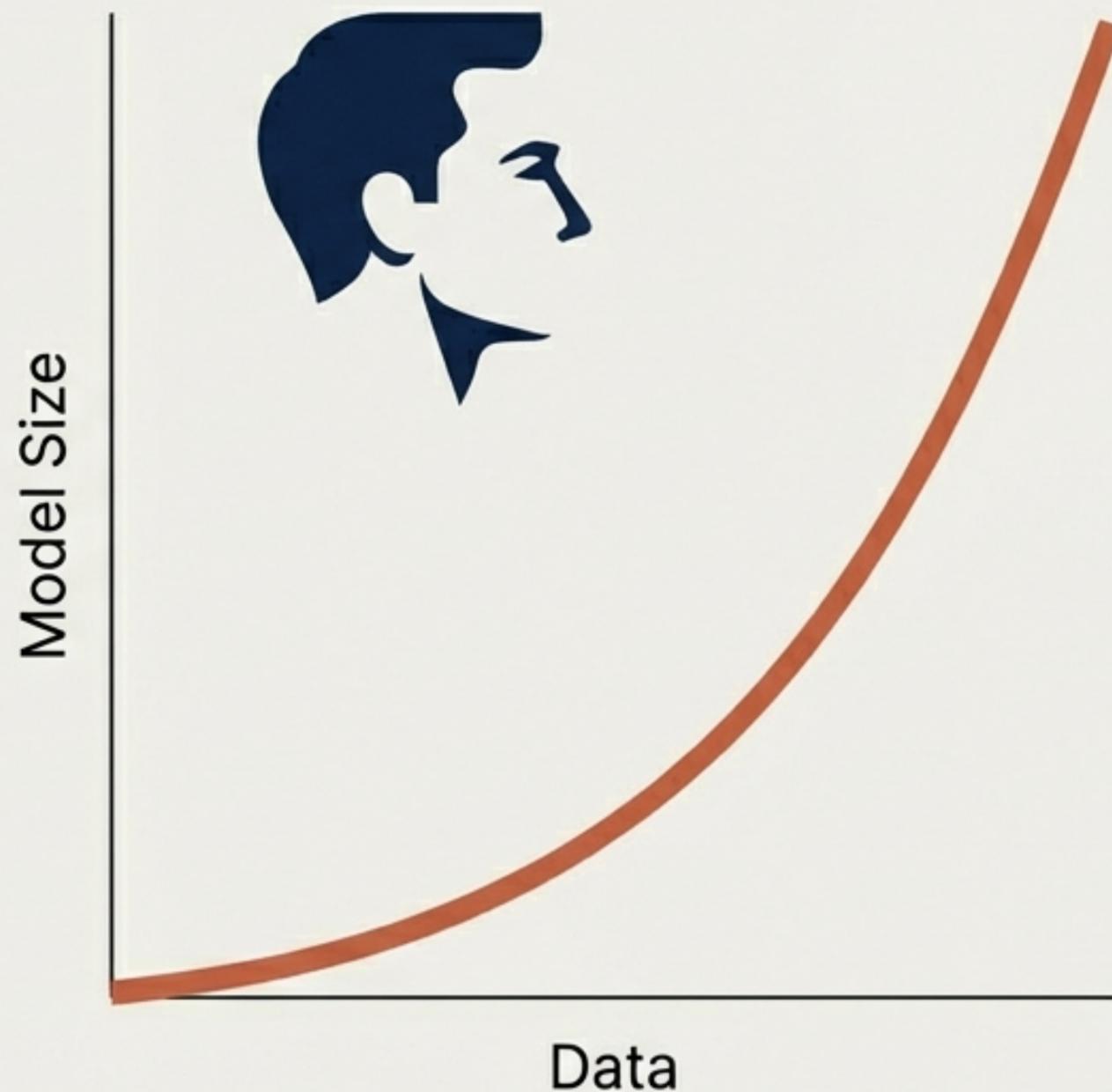
|  |   |  |
|--|---|--|
| <b>Llion Jones*</b><br>Google Research<br>llion@google.com | <b>Aidan N. Gomez*</b><br>University of Toronto<br>aidan@cs.toronto.edu | <b>Łukasz Kaiser*</b><br>Google Brain<br>lukaszkaizer@google.com |
|--|---|--|

**Illia Polosukhin\***  
illia.polosukhin@gmail.com

# The triumph of scale and general computation

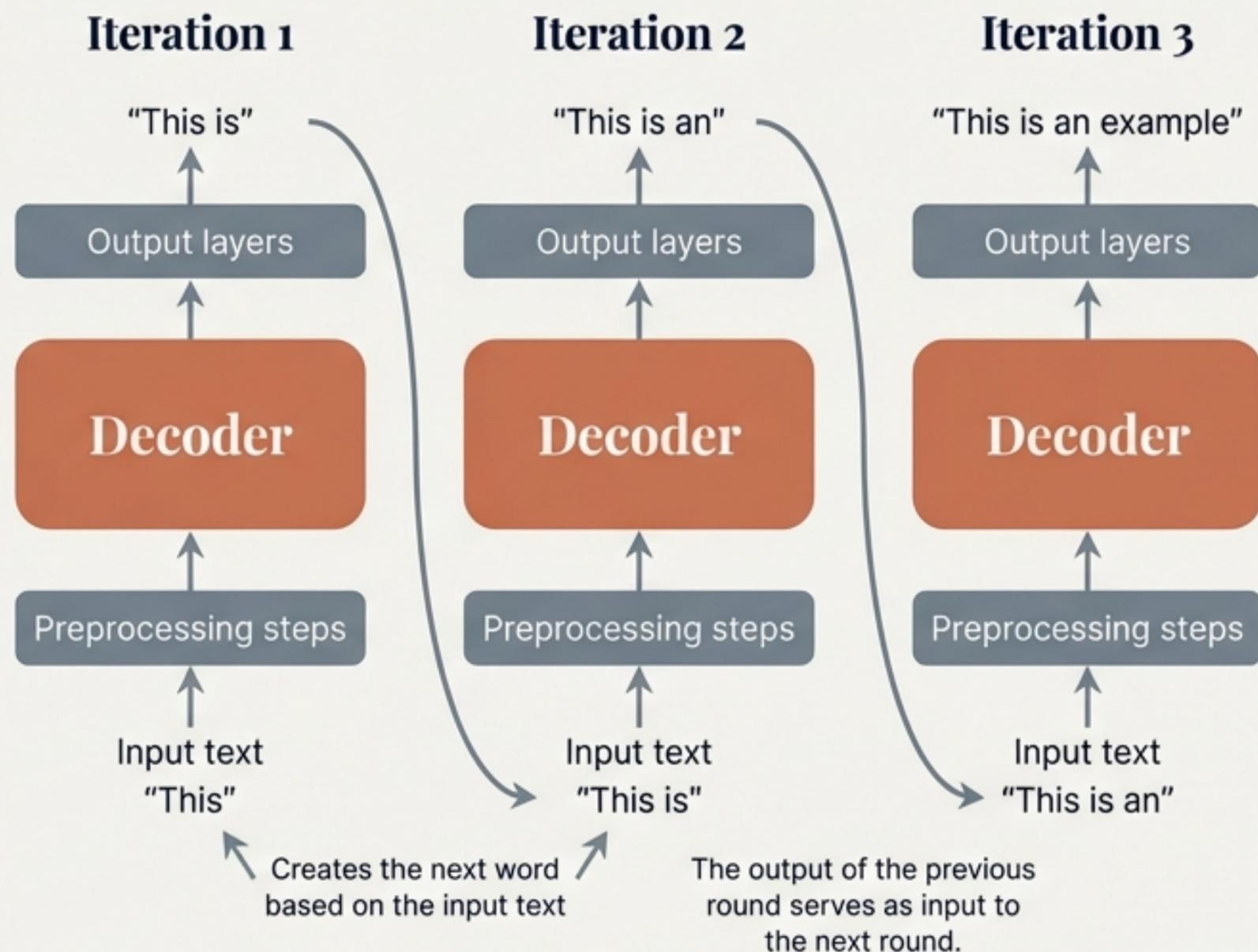
OpenAI's Noam Brown noted in 2022 that incredible progress in AI boils down to scale across two dimensions: model size and training datasets.

This echoes Richard Sutton's Bitter Lesson from 2019: 70 years of AI research proves that general methods leveraging vast computation ultimately defeat tailored, rule-based methods by a large margin.



# Modern models read and predict iteratively

Almost all LLMs today are variants of the Transformers introduced in 2017. They are autoregressive—to produce an answer, they read the prompt and their own generated text over and over again.





# The three phases of a model's lifecycle

- 1. Pre-training: The model acquires natural language and world knowledge by reading the internet.
- 2. Post-training: The model learns to solve concrete, specific tasks.
- 3. Inference: The trained model is deployed for use and is typically no longer modified.

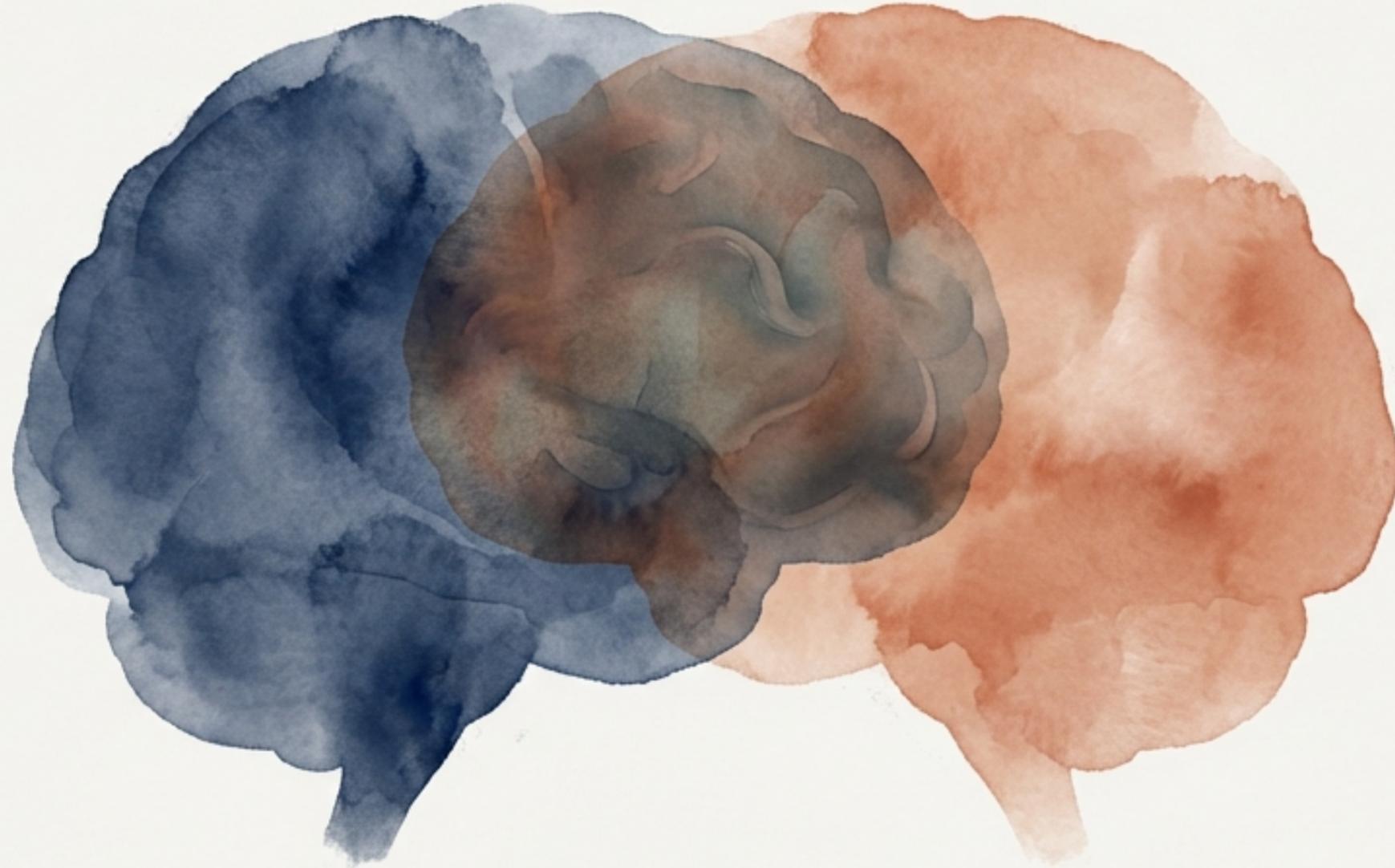


# Part 2: GenAI for Analysing Brain Scans

Joint work with Andrew Zamai, Pierrick Coupé, Vincent Planche,  
Boris Mansencal, and Laurent Simon

# The open clinical challenge of differential diagnosis

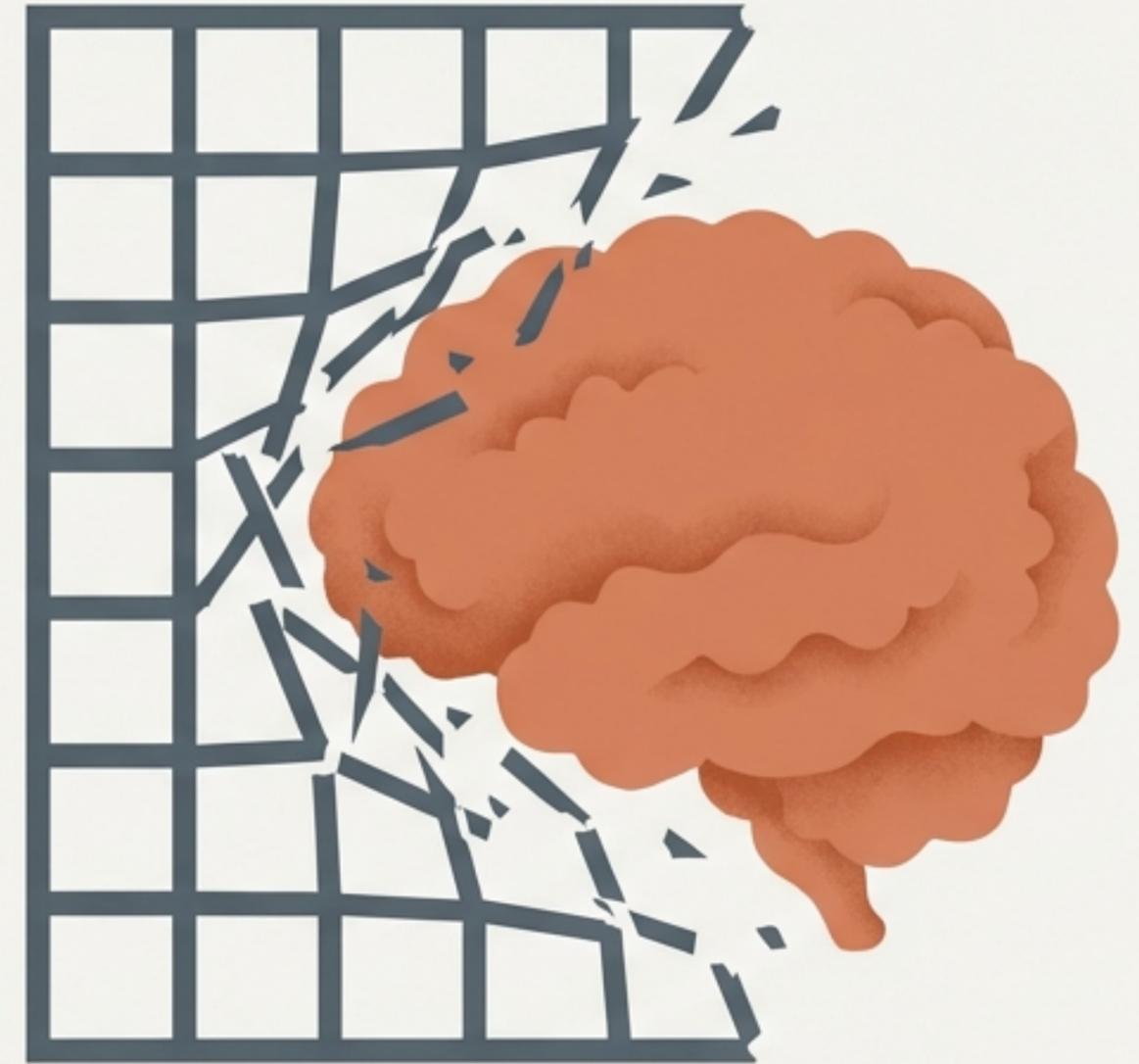
Diagnosing neurodegenerative diseases remains incredibly difficult due to the complex overlap in symptom presentation and the striking similarity of patterns observed in neuroimaging.



# Rigid systems fail in a fuzzy reality

Expert systems failed because they **lacked space for uncertainty**. Medical reasoning requires **navigating arguments**, mitigations, and potential measurement errors—there is rarely a single ground truth.

Existing Machine Learning and CNN approaches also fall short. They suffer from **data scarcity** in rare diseases, **exploit biases**, and **lack explainability**. They can point to a relevant brain region, but they **cannot explain** why it matters.



# Translating numbers into natural language

LLMs have read all medical research, but they inherently struggle with numerical values.

The solution leverages trusted, non-AI segmentation from the **Volbrain platform** and translates that quantitative data into qualitative text.

Instead of feeding the AI an atrophy 4 times over standard deviation, we provide the phrasing **a severe atrophy**.

| Volbrain Segmentation ID | Region        | Volumetric Score | Z-Score | Volume (mm3) |
|--------------------------|---------------|------------------|---------|--------------|
| 00123-AB                 | Hippocampus_L | 3.45             | -2.89   | 2450         |
| 00123-AB                 | Hippocampus_R | 3.68             | -2.65   | 2510         |
| 00123-AB                 | Amygdala_L    | 1.12             | -1.34   | 1120         |
| 00123-AB                 | Amygdala_R    | 1.20             | -1.28   | 1180         |
| 00123-AB                 | Ventricle_L   | 14.50            | +4.10   | 8900         |
| 00123-AB                 | Ventricle_R   | 15.10            | +4.35   | 9150         |

TRANSLATION

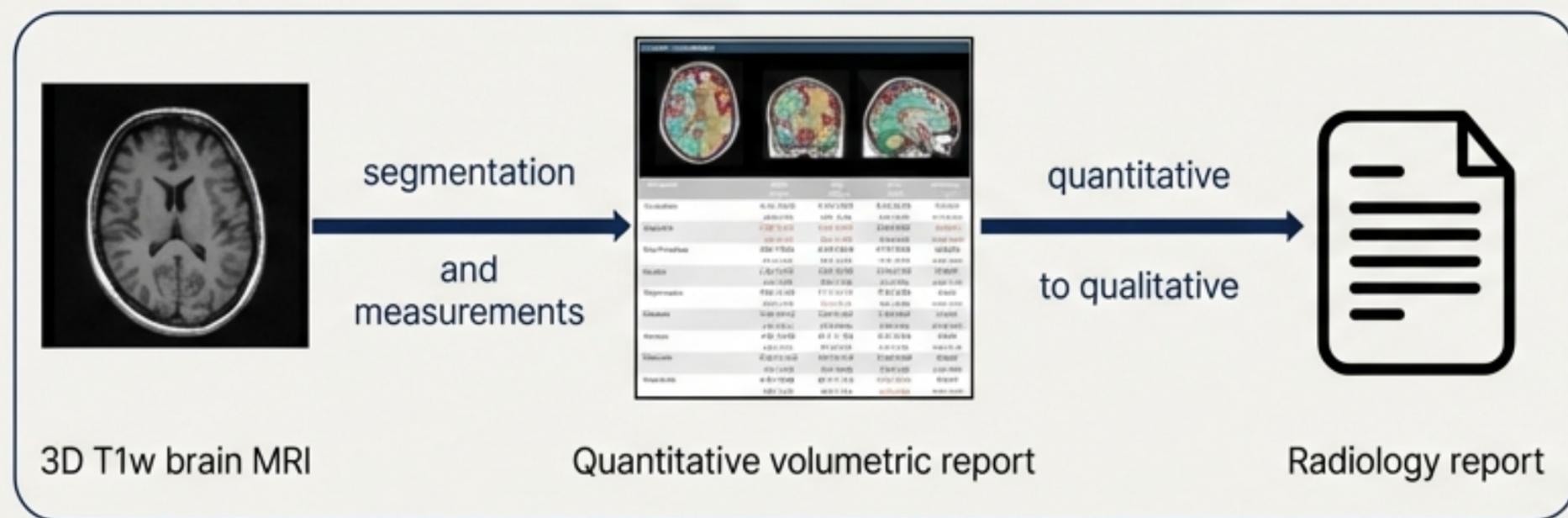


The patient exhibits marked bilateral **hippocampal atrophy**, more pronounced in the left hemisphere. There is moderate volume loss in the left amygdala, while the right amygdala remains relatively preserved.

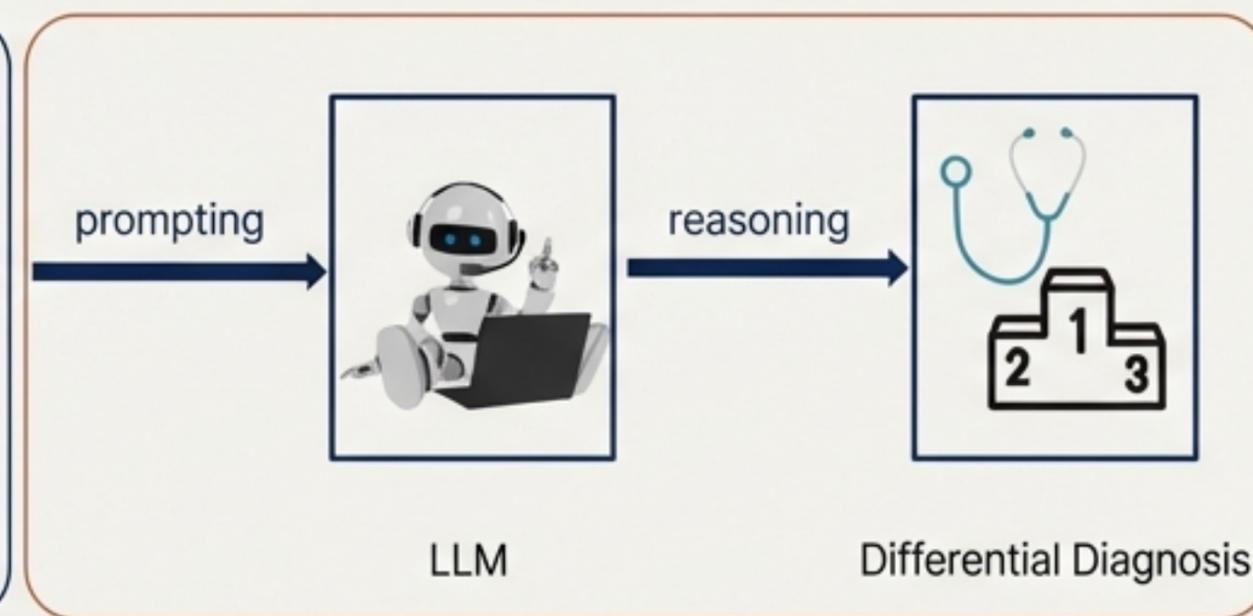
**Ventricular enlargement** is significant, indicative of substantial global cerebral atrophy, with ventricular volume exceeding four standard deviations above the mean, suggesting a **severe atrophy**.

# The explainable diagnostic pipeline

## Synthetic report generation from neuroimaging



## Reinforcement learning-optimized reasoning



# Transparent reasoning grounded in trusted facts

The final diagnosis itself is not the most vital output. The true value lies in the natural language reasoning that thoroughly assesses the reports.

Crucially, all facts debated by the AI are grounded in **trusted, traditional segmentation techniques**—no AI is involved in generating the base measurements.

**DeepSeek-R1-Distill-Llama-8B GRPO fine-tuned**

Okay, so I'm trying to figure out the possible diagnosis based on this T1-weighted brain MRI report.

Let me think about each possible diagnosis: [...]

So, considering all this, the most likely is bvFTD because of the extensive frontal and limbic atrophy, especially the left side.

**Wait, but the report also shows** significant atrophy in the caudate and thalamus, which are more typical of nfvPPA. **So maybe that's a contender. Also,** the hippocampi are atrophied, which is a point for AD, but not as strong as the others.

Putting it all together, the most likely is bvFTD, followed by svPPA and nfvPPA. **AD is possible but less likely** because of the distribution. Normal aging is less likely due to the severity.

# The Bitter Lesson, applied to medicine

Rule-based systems could never achieve this level of nuance.

Leveraging the vast internal knowledge of Large Language Models provides the necessary flexibility. It creates a space for argumentation, discussion, and measured conclusions that rigid systems simply cannot replicate.



# Part 3: AI-Assisted Diagnosis

# Defining the future role of the clinician

AI is undeniably useful in healthcare, but what exact role will it play?

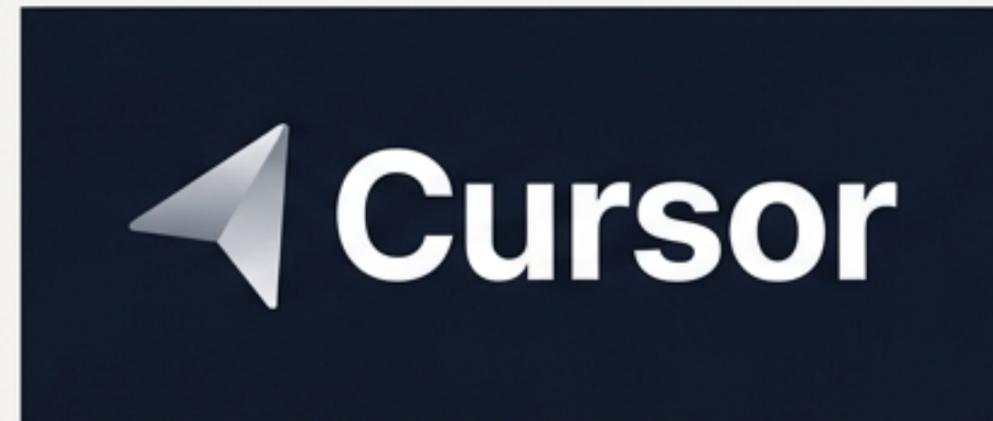
To understand the future of diagnosis, we look to the 70-year evolution of software engineering. Every revolution in coding made programming more accessible and reliable, vastly enlarging the pyramid of users at the base.



# From active assistant to peer collaborator

In 2021, GitHub Copilot introduced the AI as an assistant: the human remains the main programmer, typing requests while the AI speeds up the process.

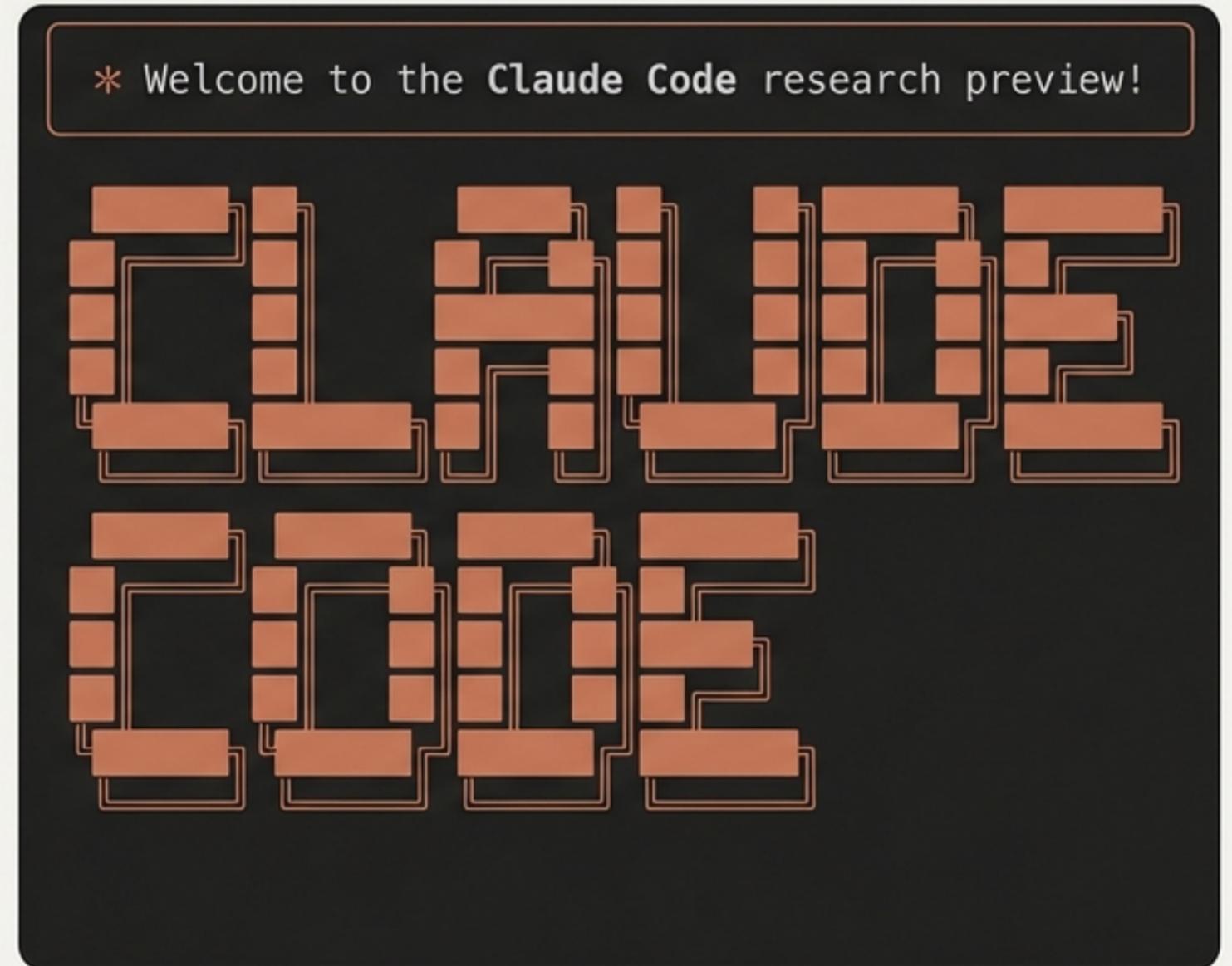
By 2023, Cursor elevated the AI to a peer: it gains autonomy to write code alongside the user, with both parties co-authoring and testing together.



# Natural language as the ultimate abstraction layer

In 2024, **Claude Code shifted the paradigm entirely**. The AI is no longer designed to help the developer write code; **it acts as the developer itself**.

You can express intent with zero programming knowledge. The **user becomes the architect**, delegating the execution entirely to the machine.



# Will doctors become the architects of patient care?

Just as software engineering transitioned from manual coding to high-level architectural design via natural language, clinical diagnosis faces a similar paradigm shift. The clinician transitions from executing the reasoning to architecting the strategy.

