LLM DNA:

Tracing Model Evolution via Functional Representations

ICLR 2026 | Oral Presentation

Presenter: Zhaomin Wu

Authors: Zhaomin Wu, Haodong Zhao, Ziyang Wang, Jizhou Guo, Qian Wang, Bingsheng He

The LLM Landscape & Required DNA

Problem

2.5M+ models exist on HuggingFace; lineage information (e.g., fine-tuning origins) remains opaque and untracked.

Why It Matters

- **Model Routing**
- **Safety & Compliance Tracking**
- **Licensing & IP Auditing**

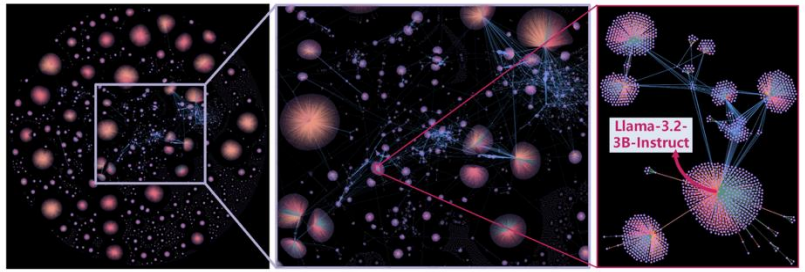


Figure 1. Visualization of Model Dependencies on HF. Purple nodes represent models, and edges represent model dependencies: Finetune, Adapter, Quantization, and Merge. **Left**: Global view, showing the widespread practices of model reuse. **Middle**: Zoomed-in view, where blue lines between models indicate they are merged into new models. **Right**: Subgraph view, illustrating the 3-hop neighbor models and their dependencies related to meta-llama/Llama-3.2-3B-Instruct.

Duan, et al. Position: Current Model Licensing Practices are Dragging Us intoa Quagmire of Legal Noncompliance. ICML'26

Formalizing LLM DNA

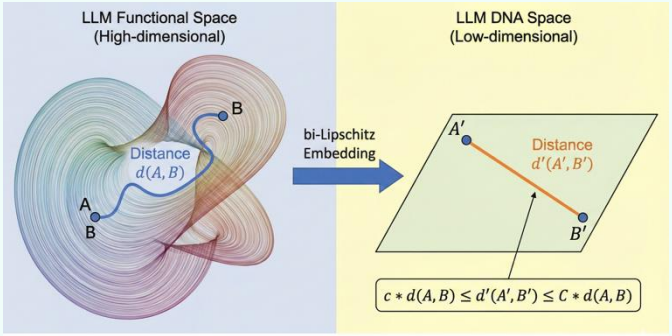
LLM as a Function: Treat a model as a **black-box function** (input → output).

DNA Definition: A low-dimensional embedding of functional space that preserves distances (**bi-Lipschitz**). Proven to exist for any LLMs.

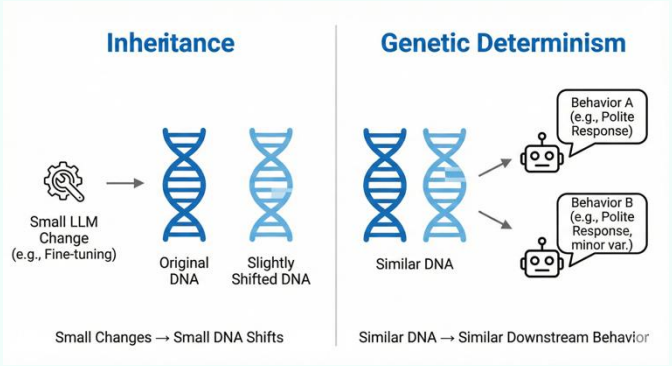
Key Properties

Inheritance: Small changes (e.g., fine-tuning) → small DNA shifts.

Determinism: Similar DNA → similar behavior.



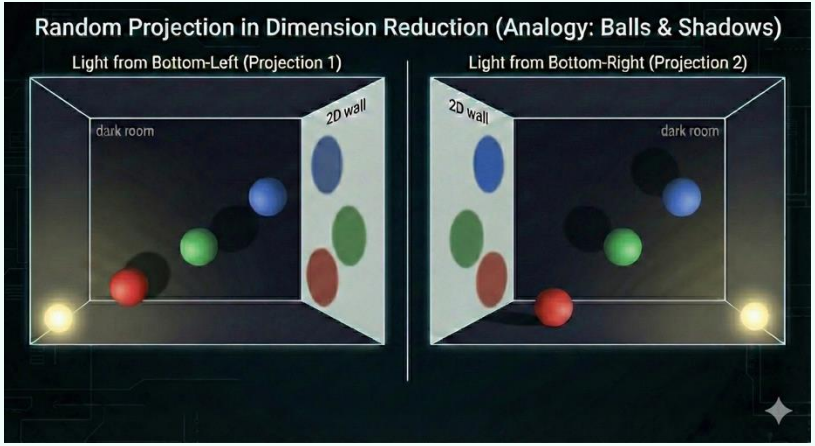
Bi-Lipschitz: Distances don't collapse or explode (preserved within constant factors c and C).



RepTrace: DNA Extraction

Direct distance calculation is intractable — RepTrace uses sampling and semantic mapping.

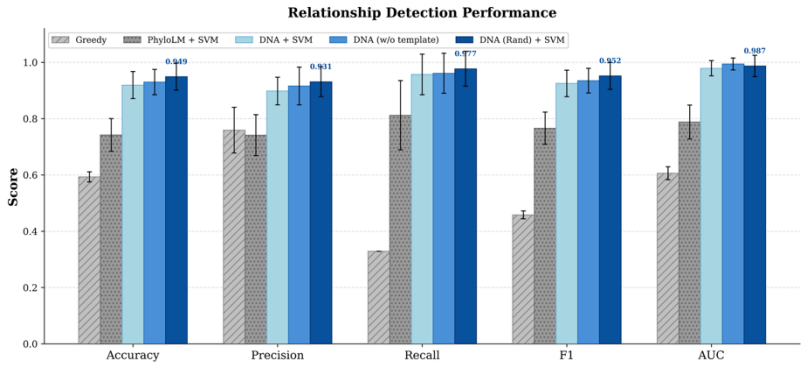
1. **Input:** Feed **random** word sequences (probe dataset) into LLM to get response.
2. **Embed:** Map responses to vectors via a frozen sentence-encoder.
3. **Concatenate:** Stack embeddings into a high-dimensional vector.
4. **Project:** Reduce dimensionality with a **fixed random** Gaussian matrix.
5. **Result:** A low-dimensional, training-free DNA fingerprint (τ_f).



Validation: Lineage Detection & Model Routing

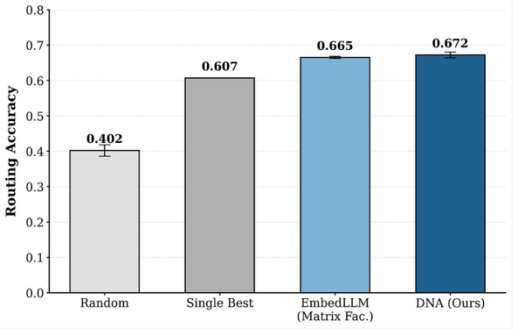
Relationship Detection

DNA (Rand) + SVM achieves **highest AUC (~0.99)** across 305 LLMs, outperforming both PhyloLM [1] and standard DNA.
Recall > Precision: uncovers previously undocumented model dependencies.



Model Routing

DNA routing surpasses EmbedLLM [2] in model routing without task-specific training.

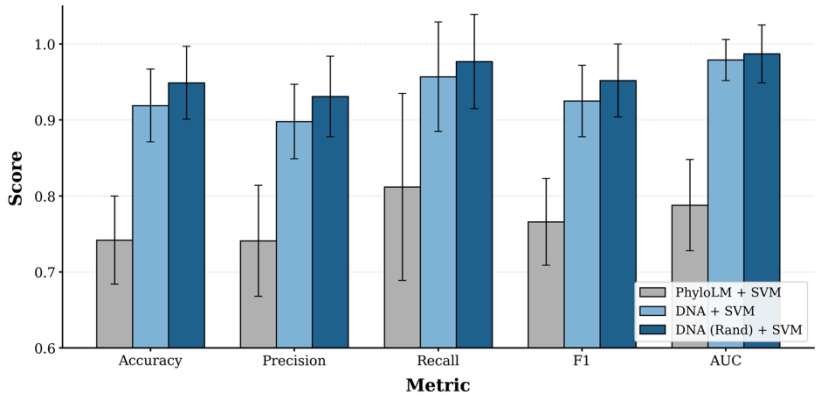


References:
[1] Yax et al. PhyloLM: Inferring the phylogeny of large language models and predicting their performances in benchmarks., ICLR 2025.
[2] Zhuang et al. EmbedLLM: Learning compact representations of large language models., ICLR 2024.

RepTrace: Robustness Validation

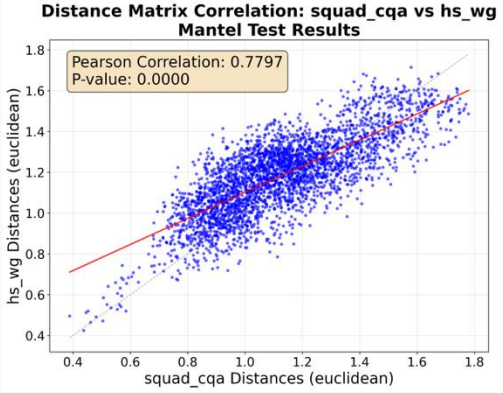
Random Input Robustness

Replace real prompts with random 100-word sequences — relationship prediction remains strong



Dataset Stability

Mantel test: DNAs from disjoint mixes (SQuAD+CQA vs HellaSwag+Winogrande) are highly correlated



LLM DNA is robust to probe dataset

DNA Distribution of 305 Models

t-SNE Visualization

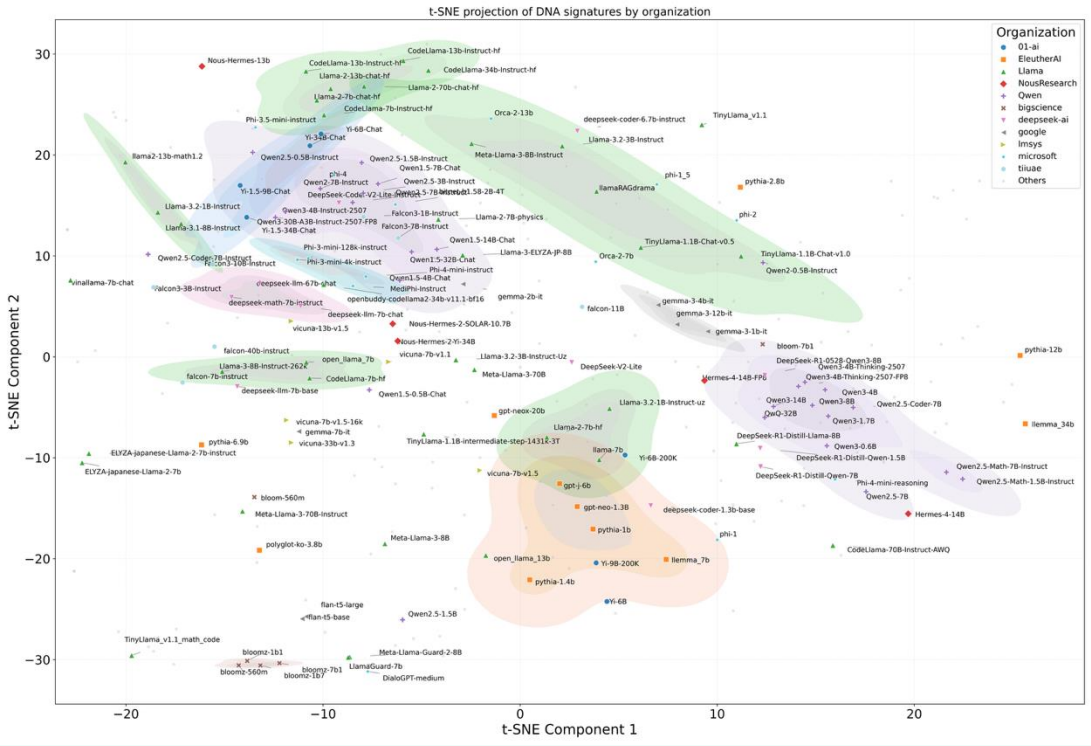
305 LLMs projected into 2D space — models cluster by organization and model family

Inheritance

Fine-tuned models locate near their base models, confirming the DNA inheritance property

HF-Undocumented Connection

Hidden lineage links visible — e.g., `vicuna-7b-v1.1` and `orca-2-13b` share unexpected proximity



Phylogenetic Analysis

Family Grouping

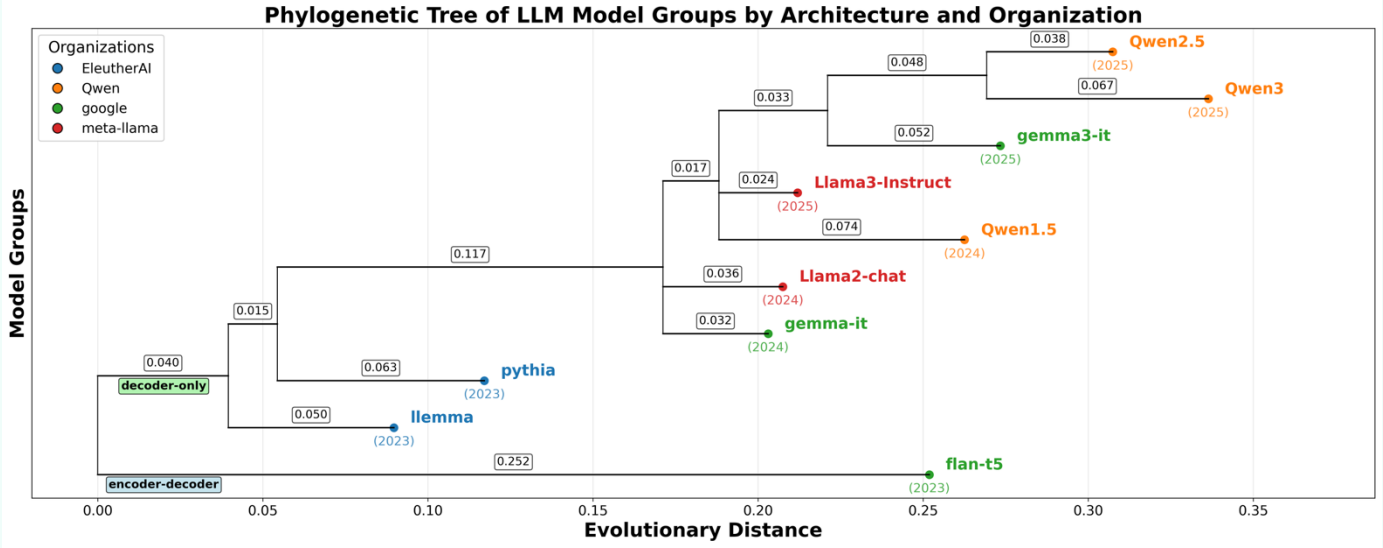
Correctly groups models by family.

General Trend

Captures generational evolution (encoder-decoder → decoder-only).

Evolutionary Rates

Reveals varying evolutionary rates across families.



Fine-Tuning Effect on DNA

Setup

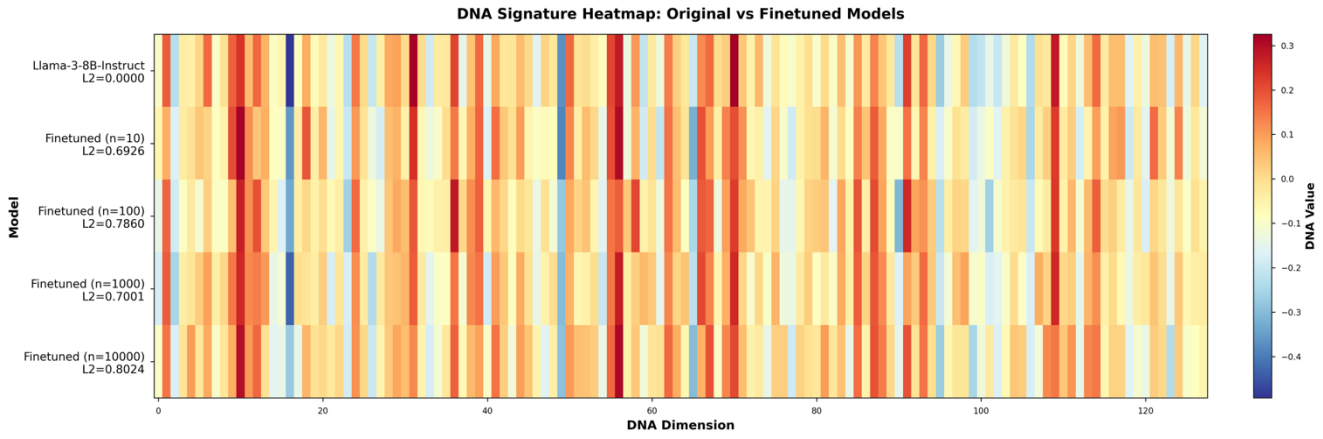
Full-parameter fine-tuning of Llama3-8B-Instruct on OpenMathInstruct-2 with increasing data

DNA Distance Grows

DNA distance increases with fine-tuning data volume — more training equals larger DNA shift

Global Structure is Stable

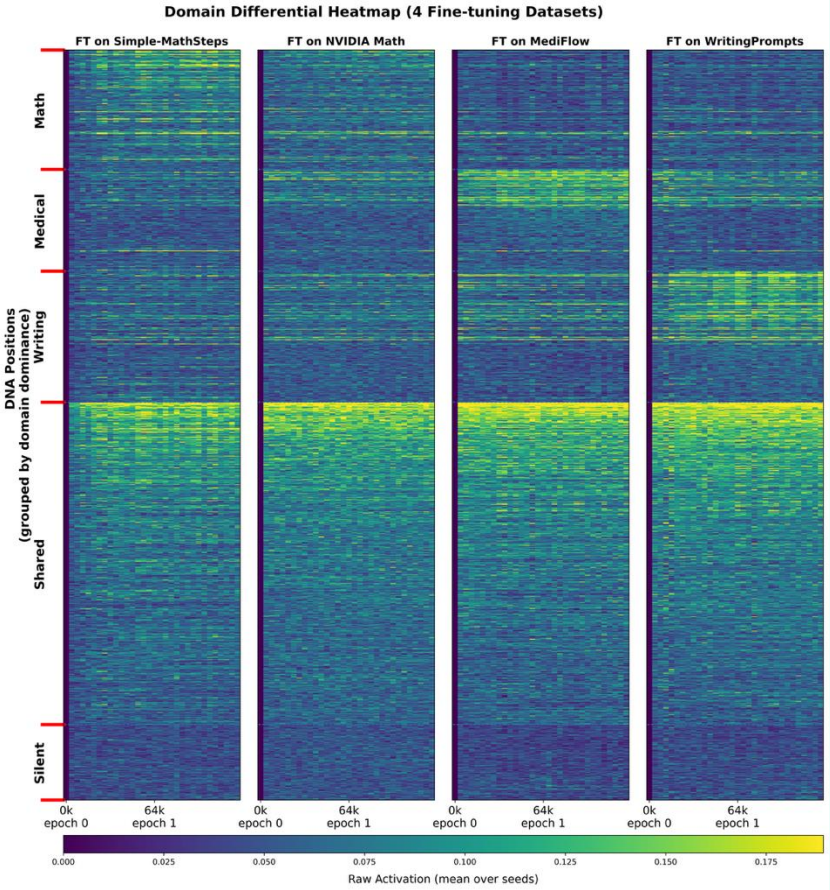
Global structure remains stable — inherited traits persist even through extensive fine-tuning



DNA Sensitivity to Fine-Tuning Domain

Domain-Specific Changes

Some DNA dimensions shift significantly when fine-tuning on math (OpenMathInstruct, NVIDIA Math)



Conclusion & Takeaways

Questions?

LLM DNA: A functional fingerprint of model behaviors and relationships.

Hidden Lineages: Successfully uncovers previously undocumented relationships.

Broad Applications: Enables routing without task training, licensing audits, API verification, and model crossbreeding.

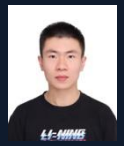
```
$ pip install llm-dna
```

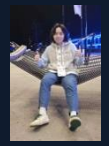
Website: https://dna.xtra.science/
GitHub: https://github.com/Xtra-Computing/LLM-DNA


Zhaomin Wu


Haodong Zhao


Ziyang Wang


Jizhou Guo


Qian Wang


Bingsheng He