Some thoughts

Michael W. Mahoney

ICSI, LBNL, & Dept of Statistics, UC Berkeley

Sept 2026

Some context

Foundations:

- “Free Decompression with Algebraic Spectral Curves (arXiv:2605.03634)
- “Learning to Discover Iterative Spectral Algorithms (arXiv:2602.09530)
- “Understanding the Implicit Biases of Design Choices for Time Series Foundation Models” (arXiv:2510.19236, ICLR26)

Implementations:

- “On Neural Scaling Laws for Weather Emulation through Continual Training (arXiv:2603.25687)
- “Reward Under Attack: Analyzing the Robustness and Hackability of Process Reward Models (arXiv:2603.06621, ICML26)
- “Zero-shot Forecasting by Simulation Alone (arXiv:2601.00970, ICLR26)

Applications:

- “Zatom-1: A Multimodal Flow Foundation Model for 3D Molecules and Materials (arXiv:2602.22251)
- “A 3D, Structure-Based, Deep Learning Approach for Predicting the Regioselectivity of Transition-Metal Catalysis (chemrxiv.10001648)
- MatterChat: A Multi-Modal LLM for Material Science (arXiv:2502.13107, NatureMI)

Some context

Foundations:

- “Free Decompression with Algebraic Spectral Curves (arXiv:2605.03634)
- “Learning to Discover Iterative Spectral Algorithms (arXiv:2602.09530)
- “Understanding the Implicit Biases of Design Choices for Time Series Foundation Models” (arXiv:2510.19236, ICLR26)

Implementations:

- “On Neural Scaling Laws for Weather Emulation through Continual Training (arXiv:2603.25687)
- “Reward Under Attack: Analyzing the Robustness and Hackability of Process Reward Models (arXiv:2603.06621, ICML26)
- “Zero-shot Forecasting by Simulation Alone (arXiv:2601.00970, ICLR26)

Applications:

- “Zatom-1: A Multimodal Flow Foundation Model for 3D Molecules and Materials (arXiv:2602.22251)
- “A 3D, Structure-Based, Deep Learning Approach for Predicting the Regioselectivity of Transition-Metal Catalysis (chemrxiv.10001648)
- MatterChat: A Multi-Modal LLM for Material Science (arXiv:2502.13107, NatureMI)

Question: How can I use your problem to do my ML?

Question: How ignorant can I be and still solve your problem?

Question for the panel

Question: What distinguishes a scientific foundation model from a general purpose LLM fine-tuned on papers?

Question for the panel

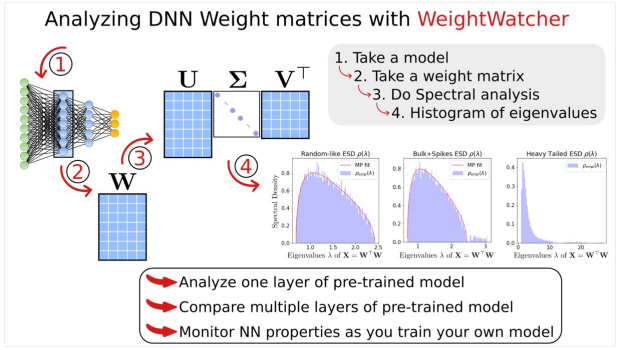
Question: What distinguishes a scientific foundation model from a general purpose LLM fine-tuned on papers?

That's an odd question:

- Are these equations "scientific":

$$\Delta s^2 = c^2\Delta t^2 - \Delta x^2 \quad \text{(Invariant Spacetime Interval)}$$
$$\Delta x' = \gamma(\Delta x - v\Delta t) \quad \text{(Spatial Transformation)}$$
$$\Delta t' = \gamma(\Delta t - \tfrac{v\Delta x}{c^2}) \quad \text{(Temporal Transformation)}$$
$$\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \quad \text{(Lorentz Factor)}$$

- We can ask scientific questions about LLMs:



What is a foundation model?

The people who coined the word (arXiv:2108.07258) spent 150 pages to say:

- It is a foundation. Upon which one can build things. Period. Nothing more. The rest are details.

What is a foundation model?

The people who coined the word (arXiv:2108.07258) spent 150 pages to say:

- It is a foundation. Upon which one can build things. Period. Nothing more. The rest are “details.”

If you use the word ”foundation” in a different sense:

- That’s fine.
- But you mean something completely different.
- There is no reason to think that any of the remarkable stuff you see in AI will hold for you.

Who has digested “the bitter lesson” of AI/ML?

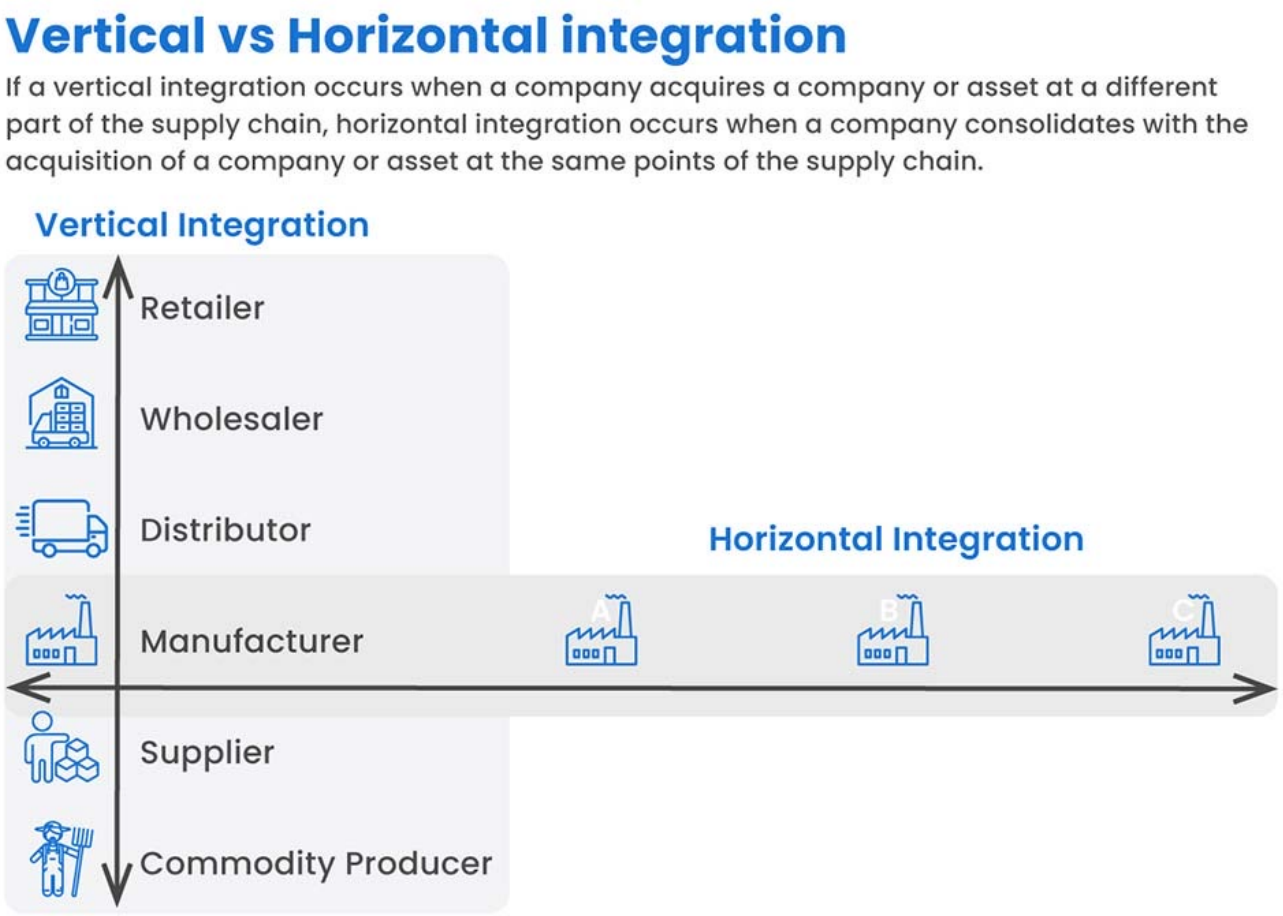
I gave a talk two years ago at SciFM and asked the audience this.

- No one had digested it.
- Two or three people in the audience had even heard of it.

"general methods that leverage computation are ultimately the most effective, and by a large margin”

- Build general models that are as domain ignorant as possible.
- Put in your domain insight as late as possible.

How to view Scientific Machine Learning



What might be possible with a meaningfully *Scientific* FM?

Train on data from:

- Atmosphere: Climate and Weather processes
- Land: Water and Ecosystem processes
- Subsurface: Heterogeneous flows and seismicity
- Language models: e.g., if you want to learn 1/f noise

Transfer learn on data from:

- Astronomy: to discover habitable exoplanets
- Materials Science: to learn physics across scales in an end-to-end way
- Chemistry: to learn interatomic potentials for MD simulation
- Fire/Floods/Etc.: to learn distributions of extreme events well enough to create insurance markets
- Nuclear Physics: to learn classified data from public data

How is this even possible? My data are special/unique?

- No; NOT so.
- You are NOT so unique/special: ML algorithms predict movies you watch better than you do

Some failure modes …

- **PINNS**:
  - "Characterizing possible failure modes in physics-informed neural networks," Krishnapriyan, Gholami, Zhe, Kirby, and Mahoney, arXiv:2109.01050, NeurIPS2021.
- **Neural ODEs**:
  - "Learning continuous models for continuous physics," Krishnapriyan, Queiruga, Erichson, and Mahoney, arXiv:2202.08494, Communications Physics 2023.
- **Neural Operators**:
  - "The False Promise of Zero-Shot Super-Resolution in Machine-Learned Operators," Sakarvadia, Hegazy, Totounferoush, Chard, Yang, Foster, and Mahoney, arXiv:2510.06646, ICLR2026.

Some directions

Illustrative recent proof-of-principle directions:

- **SuperBench**: A Super-Resolution Benchmark for SciML
- **Traditional vs Modern ML UQ**: Over- vs under-parameterized models
- **ContinuousNet**: “numerical” convergence tests
- **ProbConserve**: a posteriori correction for conservation constraints
- **Weight Diagnostics**: WeightWatcher analysis and HTSR
- **Time Series**: LEM, ConvLEM, Chronos
- **NeurDE**: long-term prediction of nonlinear conservation laws*

*ML horizontal: kinetic theory formulation to use ML to couple between different scales