Newsletter Data & AI

20/03/2025

THE SHORT OF IT ⏱️

- **Distillation Laws Guide Model Efficiency:** A new framework predicts when distillation outperforms supervised learning based on compute allocation.
- **V-JEPA Rethinks Video Learning:** Meta's model predicts abstract features instead of pixels, offering a more efficient alternative to generative AI with better physical understanding.

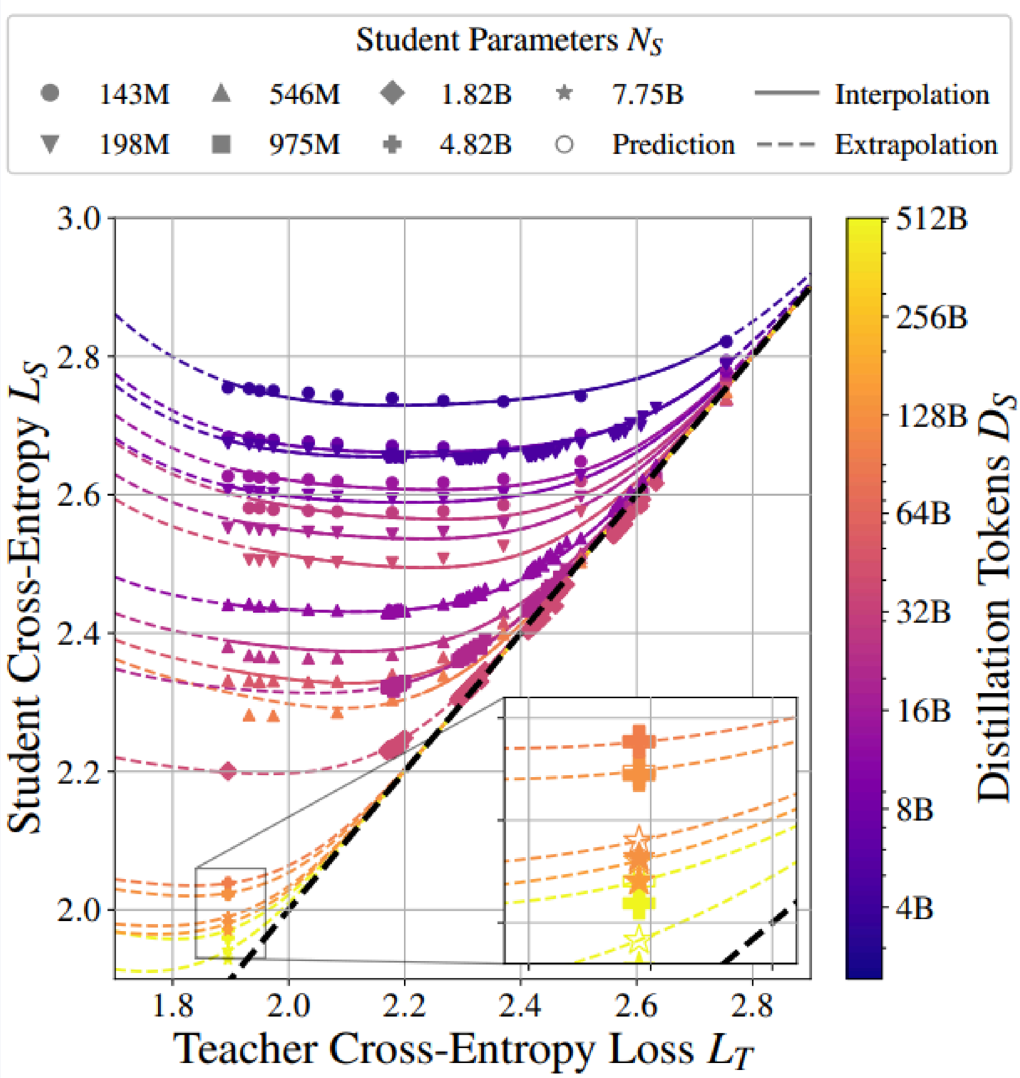
Glossary 📖

- **Distillation:** Training a smaller model using knowledge from a larger one.
- **Feature Prediction:** Learning by predicting missing parts of data instead of using predefined labels.

Trends

- [Paper] Distillation Scaling Laws

  Apple and Oxford researchers introduce a *scaling law* to predict student model performance based on compute allocation. Their findings show distillation is only efficient if a pre-trained teacher exists or multiple students are distilled; otherwise, supervised learning is preferable. The study provides a framework to optimize resource allocation and inference costs in large-scale model training.



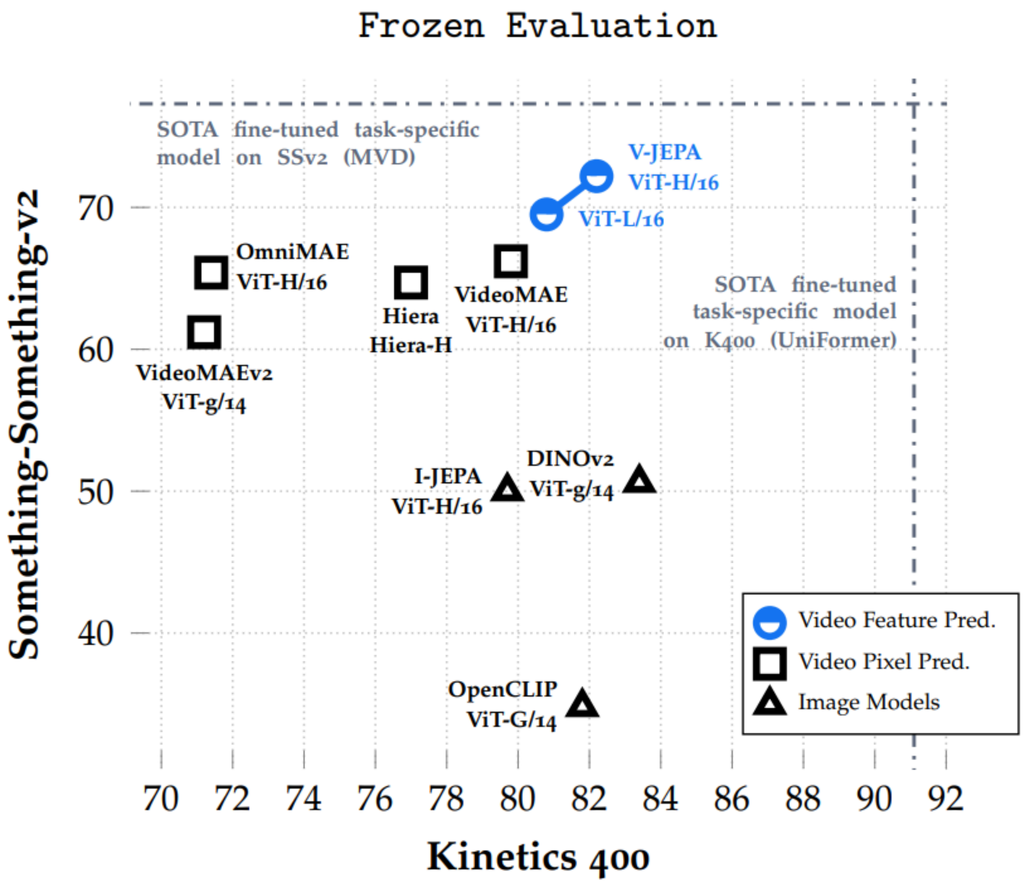
- [Paper] Imputation for Prediction: Beware of Diminishing Returns.

  The paper finds that *better imputations offer minimal predictive gains*, especially with expressive models. Across 19 datasets, simple imputations perform competitively, particularly when missingness indicators are used. While imputation quality matters for linear outcomes, its impact in real-world settings is limited, suggesting that *investing in complex imputations often yields little benefit*.

State Of The Art

- [Paper] Revisiting Feature Prediction for Learning Visual Representations from Video

  V-JEPA from Meta, INRIA, and NYU is a vision model trained on video using *feature prediction* without pretrained encoders or text. It outperforms prior methods on motion and appearance tasks with a frozen backbone, showing *feature prediction enables efficient, versatile visual representations* for self-supervised learning.



- [Paper] The Belief State Transformer

  The *Belief State Transformer (BST)* is a new *goal-conditioned next-token predictor* that processes both prefix and suffix information to predict the next and previous tokens. This approach improves planning-heavy tasks like structured text generation, outperforming existing methods in *story writing* and *graph-based reasoning*, leading to better *goal-conditioned decoding and text coherence*.

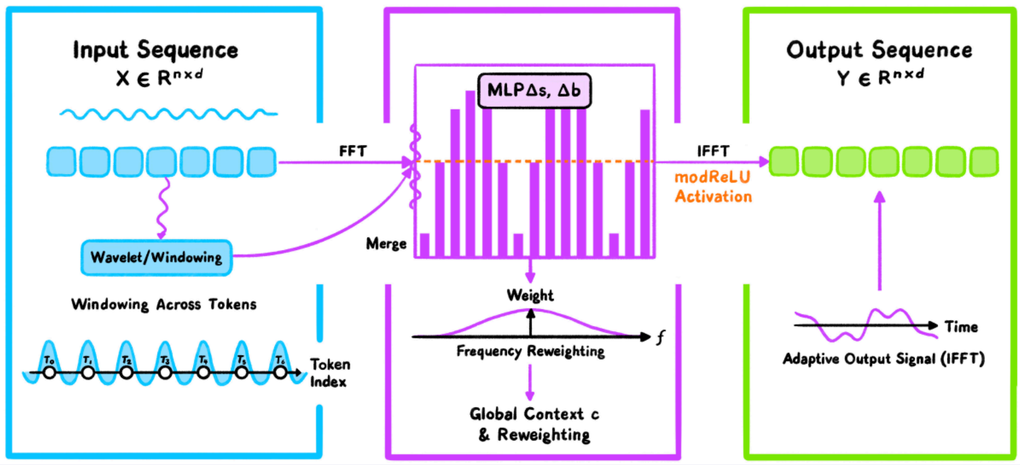
Miscellaneous

- [Paper] Causality Can Systematically Address the Monsters Under the Bench(marks)

  Causal inference offers a systematic way to improve LLM evaluation, addressing biases, artifacts, and unreliable failure analyses in benchmarks. The authors introduce *Common Abstract Topologies (CATs)* (causal graph templates) to formalize reasoning structures and refine assessments. Case studies show how this approach enhances transparency, clarifies model reasoning, and strengthens evaluation frameworks in machine learning.

- [Paper] The FFT Strikes Again: An Efficient Alternative to Self-Attention

  FFTNet replaces self-attention with a Fast Fourier Transform (FFT)-based approach, achieving *O(n log n)* efficiency for global token mixing. By operating in the frequency domain, it captures long-range dependencies while a *learnable spectral filter* dynamically emphasizes key frequency components. Tested on *Long Range Arena and ImageNet*, FFTNet outperforms both self-attention and fixed Fourier methods, offering a faster, scalable alternative for sequence modeling.

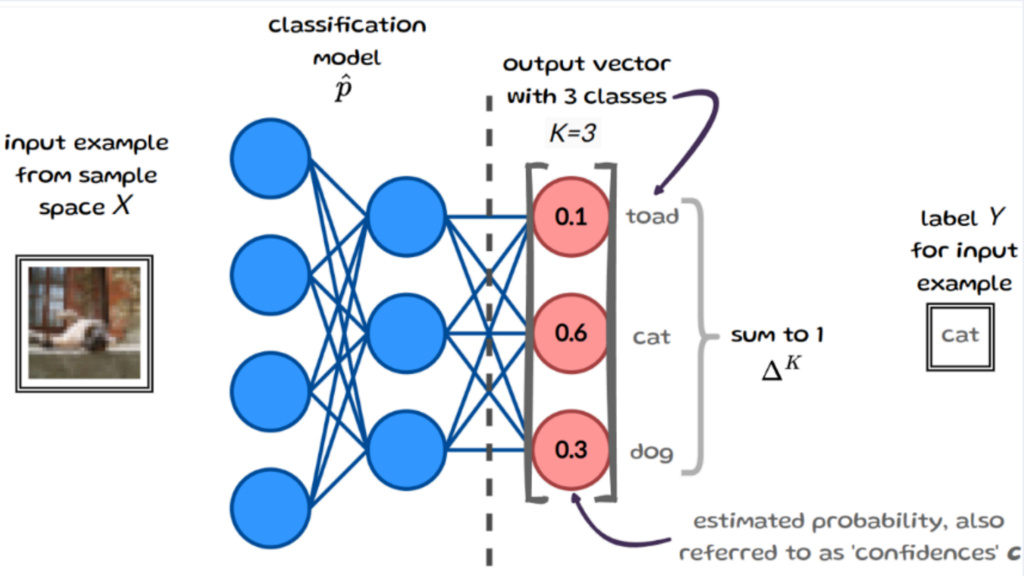


- [Blog] Understanding Reasoning LLMs

  Sebastian Raschka explores reasoning models, LLMs specialized for complex multi-step tasks like math and coding. He outlines four key approaches to improving them: inference-time scaling, pure RL (Reinforcement Learning), SFT (Supervised fine-tuning) with RL, and distillation. Examining DeepSeek R1, he highlights how reasoning can emerge from RL alone and why SFT with RL remains dominant. The article also covers cost-effective strategies like distillation and journey learning for developing reasoning models on a budget.

- [Blog] Understanding Model Calibration: A Gentle Introduction & Visual Exploration

  Model calibration ensures a model's confidence aligns with real-world outcomes, improving reliability in predictions. Maja Pavlovic explores *Expected Calibration Error (ECE)*, its limitations, and alternative approaches like *adaptive binning, multi-class calibration, and human uncertainty calibration*. The article highlights why ECE remains widely used despite its flaws and examines newer evaluation metrics for better model assessment.



- [Blog] The Data Validation Landscape in 2025

  Data validation ensures data quality by checking formats, missing values, and anomalies before analysis or reporting. The article reviews leading validation tools, including Great Expectations, Pointblank, Pandera, and Pydantic, highlighting their strengths and best use cases. It recommends Pandera for mixed teams, Great Expectations for production environments, and Pydantic for API and form validation, offering a practical guide for selecting the right tool.

Events

- [Conference] NVIDIA GTC 2025

  The NVIDIA GPU Technology Conference (GTC) 2025 is scheduled from March 17 to 21, 2025, at the San Jose McEnery Convention Center in San Jose, California. The event brings together researchers, developers, and industry leaders to explore advancements in AI, accelerated computing, and related technologies. For more details and highlights, visit the GTC 2025 website.

Thank you for your engagement. We eagerly anticipate sharing further advancements in AI with you.