Docket #: S26-140

Method and System for Constructing Auditable White-Box Reasoning Models by Introspective Symbolic Distilation and Fidelity-Gap Refinement of Language-Model Chain-of-Thought

Scientists have developed a novel framework that delivers frontier AI accuracy with full transparency and auditability.

While advanced language models deliver strong accuracy on scientific tasks like protein and toxicity prediction using minimal data, high computing costs prevent their deployment across million-candidate screening pipelines. Furthermore, their "black-box" nature lacks the auditable transparency required by regulatory bodies. Previous attempts to combine language models with decision trees have failed, producing either overfit rules when writing trees directly or double-digit accuracy drops when routing queries through tree nodes. Consequently, the industry faces an unaddressed gap for a scientific screening solution that is simultaneously accurate, cost-effective, and fully interpretable. The Lundberg lab at Stanford has developed a classification framework and platform which addresses these issues by distilling the reasoning of a large AI model into a transparent, structured decision framework that a smaller, more economical model can execute while maintaining high predictive accuracy at a fraction of the cost. The result is an AI system that is scalable, cost-effective, and fully interpretable, making it well suited for regulated industries where every prediction must be accompanied by a clear and auditable rationale.

Stage of Development: *Prototype*

Applications

- Pharmaceutical ADMET and safety screening
- Protein and antibody engineering
- Drug discovery

Advantages

- Inference costs are lower since only the small model runs at deployment
- Every decision follows a clear, human-readable path that domain experts can verify
- Matches the performance of the large model it was derived from

Innovators

- Noorsher Ahmed
- Zoe Wefers
- Mohini Misra
- Emma Lundberg

Licensing Contact

David Mallin

Licensing Manager, Physical Sciences

Email