aveni

3 December 2025

Alexandra Birch

Chair of Multilingual Natural Language Processing, School of Informatics
Chief Scientist and Co-Founder at Aveni

FinLLM

Achievements

Training and Evaluation

We have developed the infrastructure to be able to perform pre-training (CPT) and supervised fine-tuning (SFT) experiments. We have created a finance-specific evaluation suite (AveniBench) to quickly assess checkpoints.

Data

We have created a finance-specific CPT dataset (AveniPile) by applying a custom classifier to general purpose data, and by focused crawl. We have also produced a specialised finance-specific instruction tuning set (AveniBlocks)

Models

We released a set of models of different sizes, produced using a transparent and documented process and optimised for finance. These are deployed and available via internal endpoints.

Use-Cases

Targeting use-cases in Detect and Assist, we have demonstrated performance which exceeds that of OpenAI models with much smaller, locally deployed models. This included developing evaluation sets for Assist.

Data

- **AveniPile:** Financial CPT dataset; 91B+ tokens; deliberate tilt toward regulatory and educational sources.

- **AveniBlocks:** Ever expanding SFT dataset collection aligned with our use-cases.

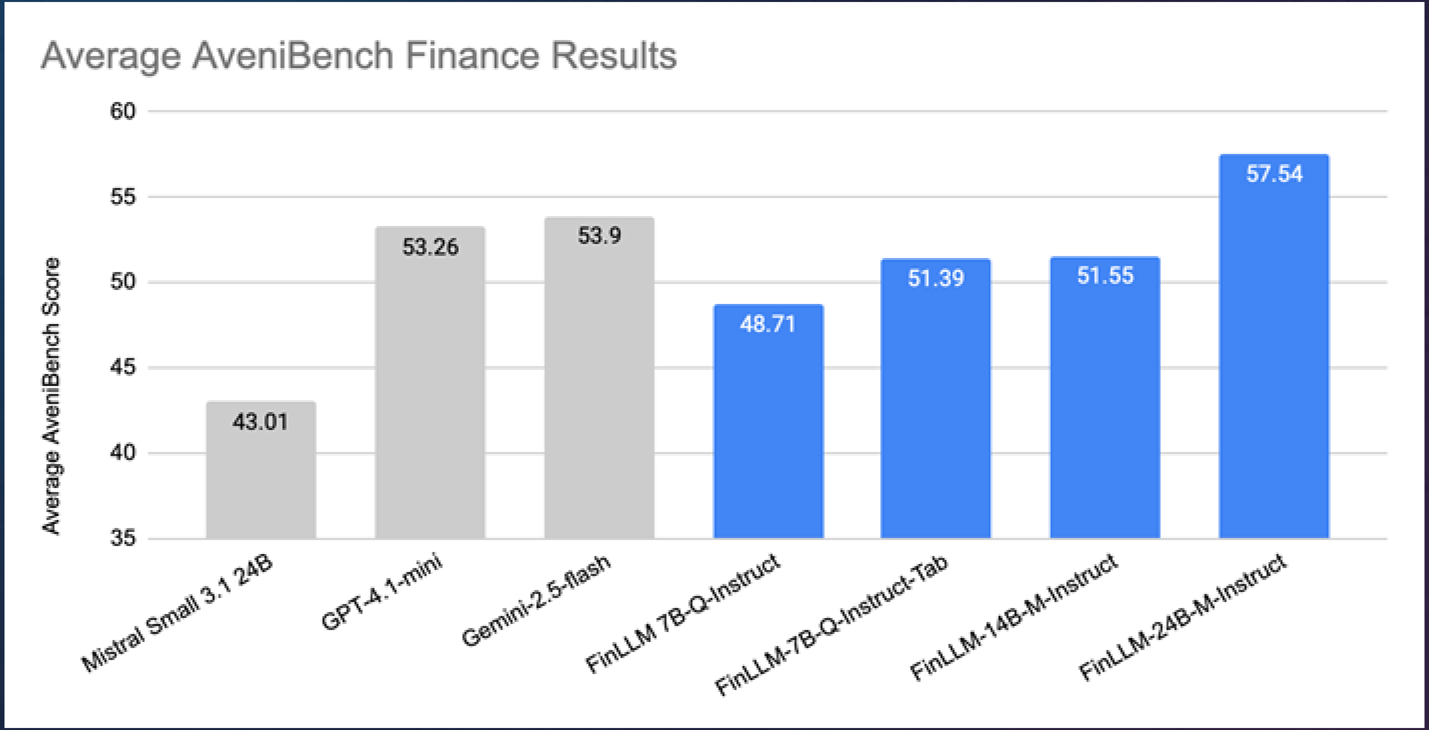
- **Financial classifier:** (teacher Llama-3.3-70B, student ModernBERT) reaches Gold F1 = 0.893 to surface finance-relevant text.

- **Data Processing:**
  a. Multi-stage process (Content Extraction, Toxicity Detection, Deduplication, Language Identification).
  b. Risk-based pseudonymisation tied to a finance taxonomy: L0 48.4%, L1 45.6%, L2 6.0%—protecting PII while retaining factual utility.

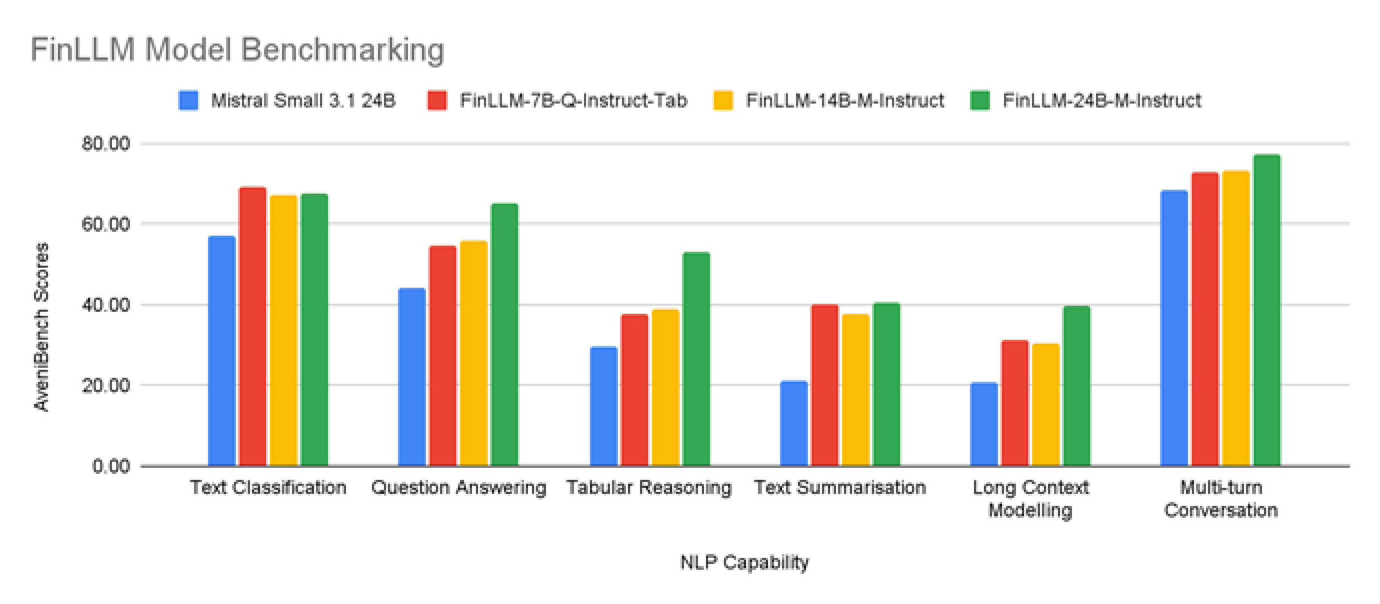
Training and Evaluation

- **Two-phase CPT:** warm-plateau → anneal + model merging (slerp)
- Targeted **synthetic tabular/math data** for model specialisation
- LLM → SLM methodology: pruning → distillation → annealing → merging → SFT
- **AveniBench:** finance, general, and safety tasks spanning classification, tabular QA/reasoning, long-context, and multi-turn dialogue; public subset released.

FinLLM Instruct models (Overall Finance)

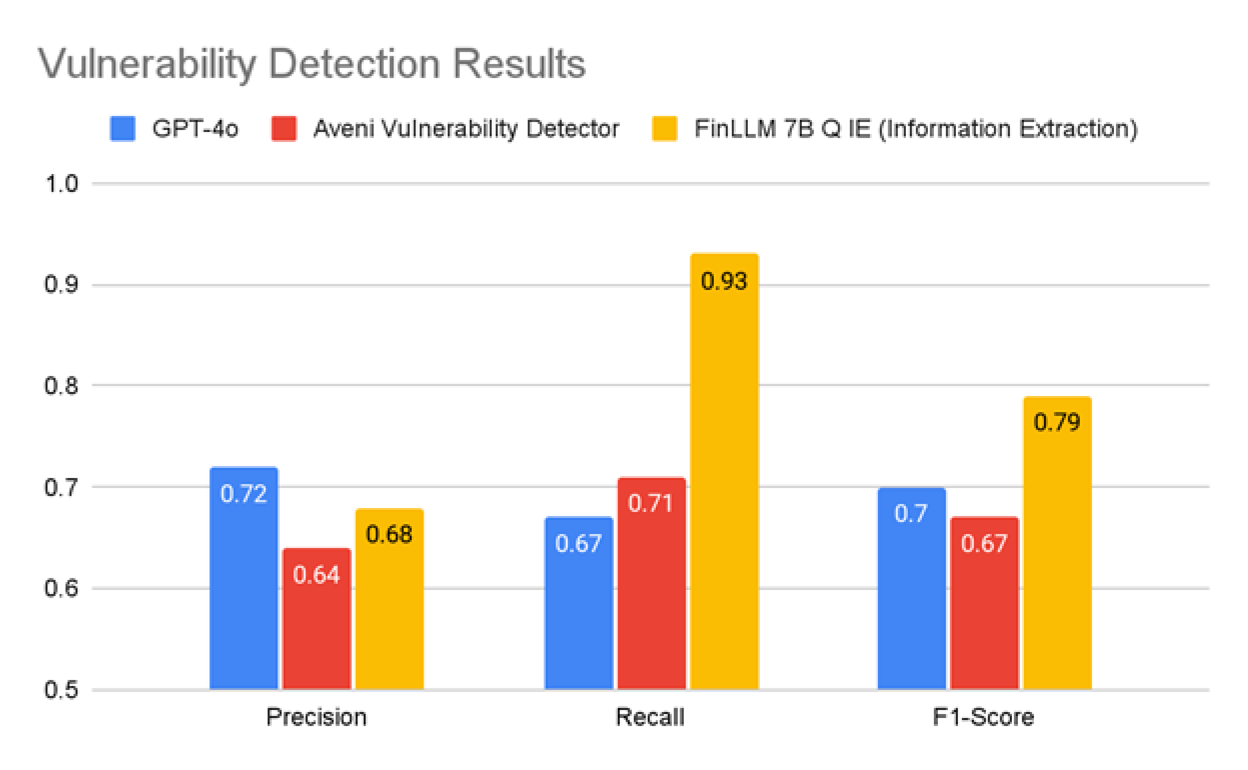


FinLLM Instruct models (Capabilities)



Aveni Detect (vulnerability in calls)

FinLLM-7B-Q-Instruct-IE attains best **F1 = 0.79**, competitive hit rate vs larger commercial models, simplifying a multi-step pipeline.

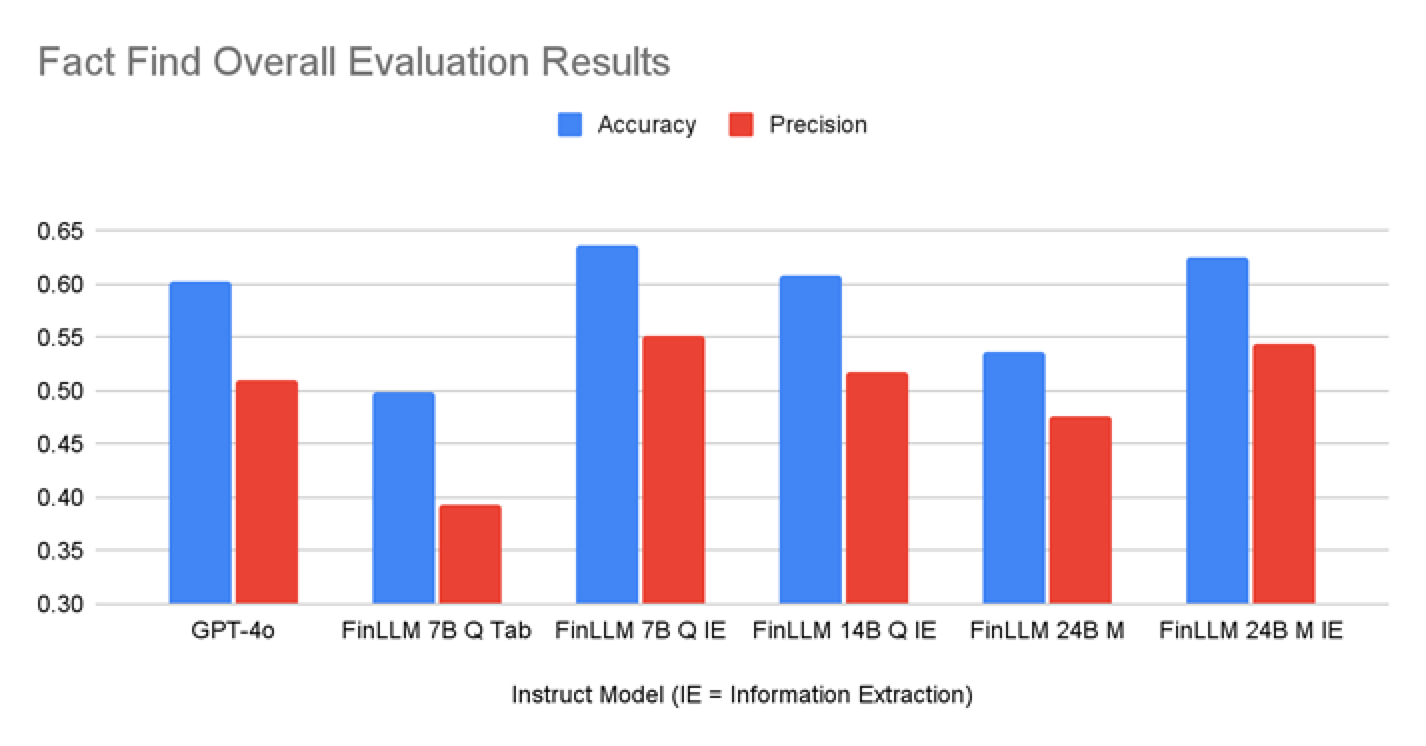


Use Cases

Aveni Assist (FactFind extraction)

FinLLM variants outperform GPT-4o on accuracy (0.64 Acc / 0.71 F1 for 7B-Q-Instruct-IE).

Extensive efforts in robust evaluation



Product

- Tools for crawling, filtering and cleaning **data**
- Tools for training and evaluating **models**
- Our own products will be **powered** by FinLLM

- FinLLM will finetuned to create **custom, small models SLMs**
- FinLLM will not stand still, the **core model** must be improved as new requirements come up - new models, techniques, and datasets become available

- The Aveni platform will generate **data** which can be used to improve FinLLM – learning from experience

More Information

FinLLM Year 1 Overview:
https://aveni.ai/resources/finllm-year-in-review/

FinLLM Year 1 Technical report:
https://labs.aveni.ai/finllm-year-one-building-a-model-for-uk-financial-services/