

FINLLM TECHNICAL REPORT – APRIL 2025

# Why FinLLM Matters for AI Leaders in Financial Services

# Why FinLLM Matters for AI Leaders in Financial Services

At Aveni, our mission is to transform how financial institutions apply AI – responsibly, transparently, and with domain precision. That’s why we’re developing FinLLM, **the UK’s first large language model built specifically for the financial services sector.**

We’re not doing this alone. FinLLM is being developed in partnership with two of the UK’s leading financial institutions – Lloyds Banking Group and Nationwide – as well as the University of Edinburgh, a global centre of excellence in AI research. This unique collaboration combines deep financial domain expertise with cutting-edge language modelling and responsible AI design.

Where general-purpose models fall short, lacking context, control, and transparency, FinLLM is set to solve. It’s built from the ground up for the workflows, regulation, and language of financial services. Aveni Labs, Aveni’s Research and Development department has spent the last 6 months developing and rigorously testing FinLLM – **the results speak for themselves.**



*What follows are the key technical findings that demonstrate how FinLLM delivers superior, real-world performance across financial services tasks, unlocking innovation while managing the operational and regulatory demands of enterprise AI.*

## 1. Designed for Your Regulatory Reality

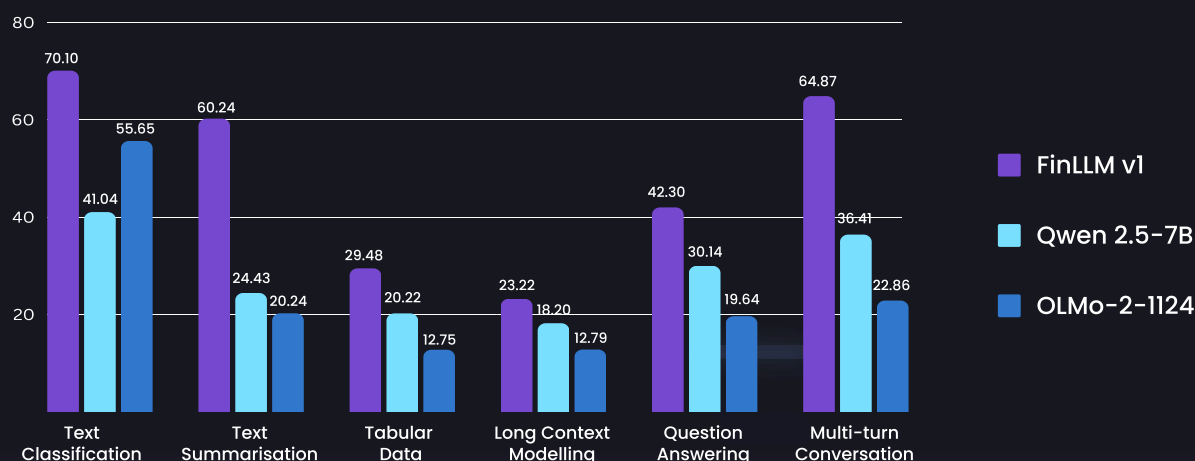
Generic LLMs aren't built for FS. FinLLM was. It was developed specifically to align with the [FCA](#), [PRA](#) and [EU AI Act](#) guidelines, addressing the sector's highest standards around data privacy, auditability, and responsible AI deployment. This means you can build with confidence knowing the model's design principles are rooted in regulatory compliance from day one.

## 2. Superior Performance on Financial Tasks

FinLLM consistently outperforms leading general-purpose and open-source LLMs (specifically Gemini 1.5 Flash and GPT-4o mini, LLaMA, and Mistral) across the kinds of tasks financial firms care most about:

- **Text classification**, such as flagging conduct risk in adviser calls
- **Long-context reasoning**, used in reviewing policy documents or investment reports
- **Tabular data analysis**, like interpreting balance sheets or onboarding forms
- **Multi-turn dialogue modelling**, essential for AI agents that hold coherent, compliant customer conversations

Fig. 1: FinLLM Performance Comparison to Generic LLMs



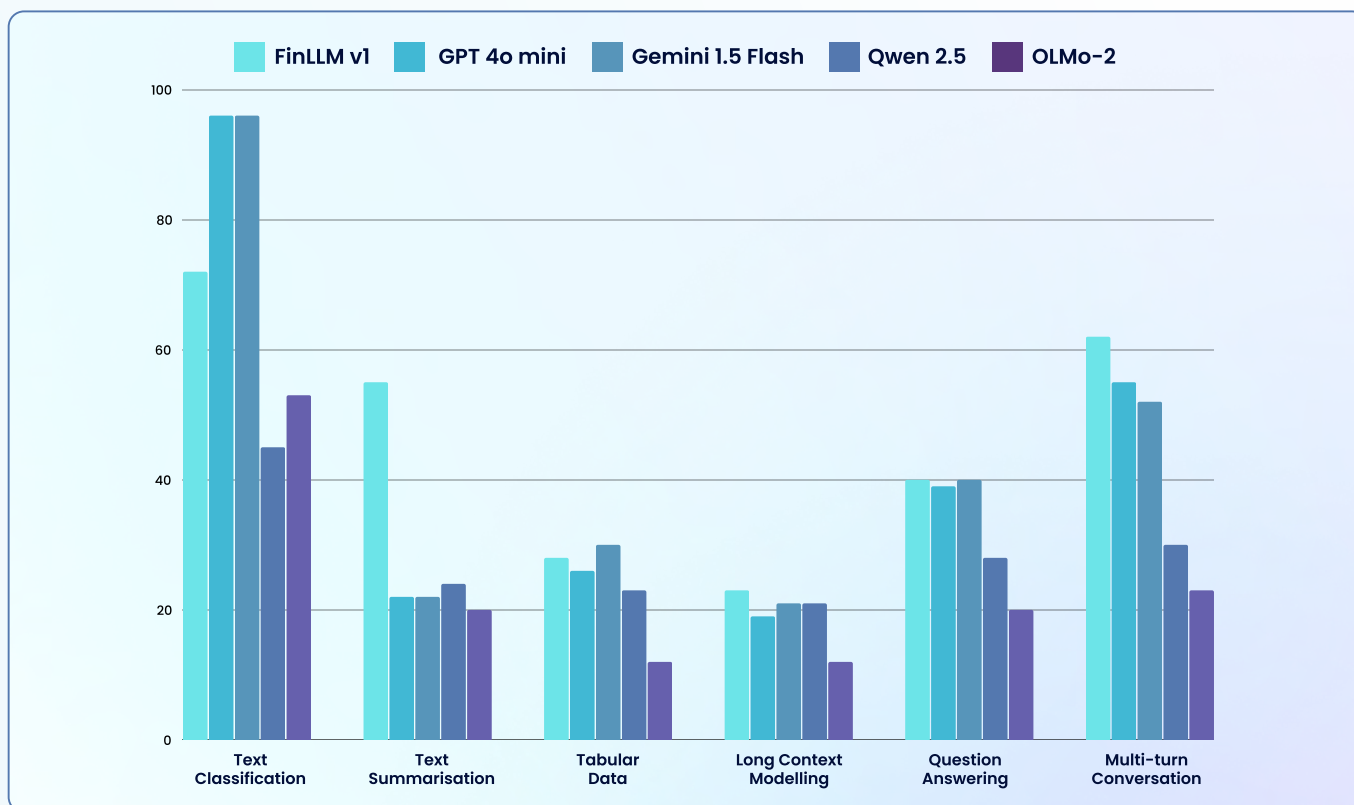
These capabilities were tested using AveniBench – FinLLM's in-house evaluation suite.



*AveniBench is a benchmark dataset built to cover real financial services tasks, such as compliance monitoring, customer vulnerability detection, and KYC document review. It ensures FinLLM is not just accurate in lab conditions, but relevant in production.*



## Comparison of Aveni's fine-tuned FinLLM 7b model against frontier general purpose models



The FinLLM 7b model demonstrates highly competitive performance across a broad range of tasks when benchmarked against frontier general-purpose models, many of which are significantly larger. In particular, **FinLLM excels in areas such as multi-turn conversation, question answering, and text summarisation**, often outperforming or matching models like GPT-4o mini and Gemini 1.5 Flash.

This strong performance, despite its relatively small size, highlights the effectiveness of domain-specific fine-tuning and model optimisation.

Beyond performance, FinLLM offers compelling advantages in terms of efficiency, cost, and environmental impact. Smaller models are faster to deploy, require less compute, and are more accessible for real-time applications, making them **highly suitable for regulated industries such as financial services**. This release marks the foundation of a broader roadmap, larger versions such as FinLLM 30b are in development, and as model capacity increases, we expect to see continued gains in accuracy and performance.



### 3. Tailored, Transparent, and Tunable

FinLLM gives your teams direct control over the model:

- **Open model variants** (1B and 7B parameters) that can be securely deployed and fine-tuned on-prem or in Virtual Private Cloud (VPC) environments.
- **Training tuned for financial communication**, so the model understands nuance in situations such as product disclosures, customer complaints, and suitability discussions.
- **Model transparency built in**, so outputs can be interrogated, audited, and explained to regulators or internal risk teams.

This is made possible by innovations like Finance Classifier 2.0 – a proprietary component developed by **Aveni** that classifies whether content is truly ‘financial’



*Finance Classifier 2.0 is a filtering system used to identify high quality, in domain, and highly relevant data that can be used for model training. It ensures that the data the model learns from is genuinely financial. It weeds out noisy or irrelevant text (e.g. marketing jargon, outdated forum chatter) and prioritises high-quality sources like regulatory guidance, financial advice transcripts, and product literature.*





## 4. Safe, Ethical, and Future-Proof

FinLLM's architecture supports safer, more accountable GenAI, essential in high-stakes domains like finance as well as in use within agentic frameworks where autonomous action is involved versus just information generation.

Its training process incorporates safeguards, for example all data included in the training of FinLLM undergoes filtering to flag sensitive, toxic and bias data before the model even sees a single token, and future releases will include input/output guardrails and response justifications to meet growing regulatory expectations.

As well as the ethical aspect, one of the most exciting developments is FinLLM's integration into agentic RAG systems.



**Agentic Retrieval-Augmented Generation (RAG)** means combining FinLLM's language capabilities with retrieval engines that fetch facts and context from real-time document sources, like policies, case logs, or regulatory guidance. This allows the AI to cite sources, answer accurately, and operate with traceability, giving teams confidence in outputs that are both informed and auditable.

## 5. Built with the Industry, for the Industry

FinLLM wasn't built in a vacuum. It was shaped in collaboration with Lloyds Banking Group, Nationwide, and the University of Edinburgh, ensuring every design decision was informed by live, *high-priority* problems in financial services, with use cases from scaling call QA to reducing the burden of regulatory reporting.





# Turning FinLLM into Business Value

FinLLM is more than a model, it's a critical driver of AI infrastructure, designed for the complexity, compliance, and scale of UK financial services. Whether you're modernising operations, building next-gen products, or accelerating regulatory transformation, FinLLM provides the foundation you need.

Here's how forward-thinking institutions could leverage it:

## **AI-Powered Customer Interaction Analysis**

Automate the review of adviser-client conversations to identify vulnerability, conduct risk, and service gaps, at scale and with confidence.

## **Next-Gen Digital Assistants**

Deploy domain-specialist AI agents to support customers, advisers, and ops teams with compliant, context-aware guidance across financial products and policies.

## **Intelligent Document Understanding**

Extract insights from lengthy, complex documents like suitability letters, mortgage offers, or KYC forms, without human review bottlenecks.

## **Faster, Safer Complaint Resolution**

Classify, triage and summarise complaints with high accuracy—reducing response times while improving consistency, defensibility and oversight.

## **Risk and Compliance Monitoring at Scale**

Continuously scan communications, documents and decisions for regulatory red flags—proactively identifying issues before they escalate.

## **Accelerated Product Development**

Equip product and engineering teams with a fine-tuned LLM that understands financial workflows, speeding up prototyping and safe deployment of AI-powered services.

## **Policy-Aware Automation**

Implement AI systems that operate within the boundaries of your internal policies and risk appetite, with explainability and auditability built in.

FinLLM marks a turning point for financial AI. It's not just a model, it's a catalyst for change. Purpose-built for the language, regulation, and complexity of the industry, it empowers financial institutions to reimagine what's possible with GenAI. For leaders ready to shape the future, FinLLM is the critical foundation.



**Get in touch with our team to  
explore how FinLLM can be  
deployed in your organisation**

Reach us at [hello@aveni.ai](mailto:hello@aveni.ai) to start  
the conversation.

