distil labs

IBM Power Systems x ROCKETGRAPH

Private AI on IBM Power

Small Language Models Transform Graph Analytics
Complete Data Sovereignty • 10x Faster • 100x More Efficient

Paul Chapman

PaulChapman@uk.ibm.com

IBM Global Power, AI & Modernisation Solutions Leader

The Enterprise AI Dilemma

The Challenge

Organizations want AI-powered natural language querying for graph databases, but cloud-based LLMs create insurmountable barriers.

Critical Issues

- Sensitive data must be sent to external servers
- Query patterns and schemas exposed to third parties
- 2-5 second response times per query
- High financial and environmental costs
- Compliance and regulatory roadblocks

The Breakthrough Solution

Small Language Models (SLMs)

A specialized Granite 3.3 8B model fine-tuned for Rocketgraph's OpenCypher queries, running entirely on IBM Power hardware.

Unlike massive cloud LLMs with over 1 trillion parameters, our SLM delivers targeted performance with just 8 billion parameters.

Key Principle

Specialization beats generalization when you know exactly what you need to accomplish.

Training Approach: Public documentation and synthetic data only - no customer information ever used.

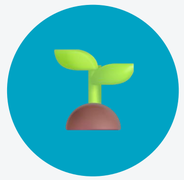
Four Pillars of Excellence

Complete Data Sovereignty

Runs entirely on your IBM Power infrastructure behind your firewall

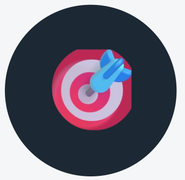
10x Faster

Sub-200ms query translation vs 2-5 seconds for cloud LLMs



100x More Efficient

Dramatically reduced energy consumption for sustainable AI



85% of Claude 4 Performance

Purpose-built for Rocketgraph OpenCypher queries

How It Works

Training Process

1. **Public Documentation** - Rocketgraph OpenCypher specifications
2. **Synthetic Data** - 900+ schemas, 15,000+ validated examples
3. **Fine-Tuning** - IBM Granite 3.3 8B specialized for OpenCypher
4. **On-Premise Deployment** - Your hardware, your control

Real-World Example

Natural Language Query:

"Show me all transactions over $10,000 involving accounts opened in the last 30 days"

Translated in <200ms

OpenCypher Query:

```
MATCH (a:Account)-[t:TRANSACTION]-()
WHERE t.amount > 10000
```

Performance Comparison

The Numbers Speak

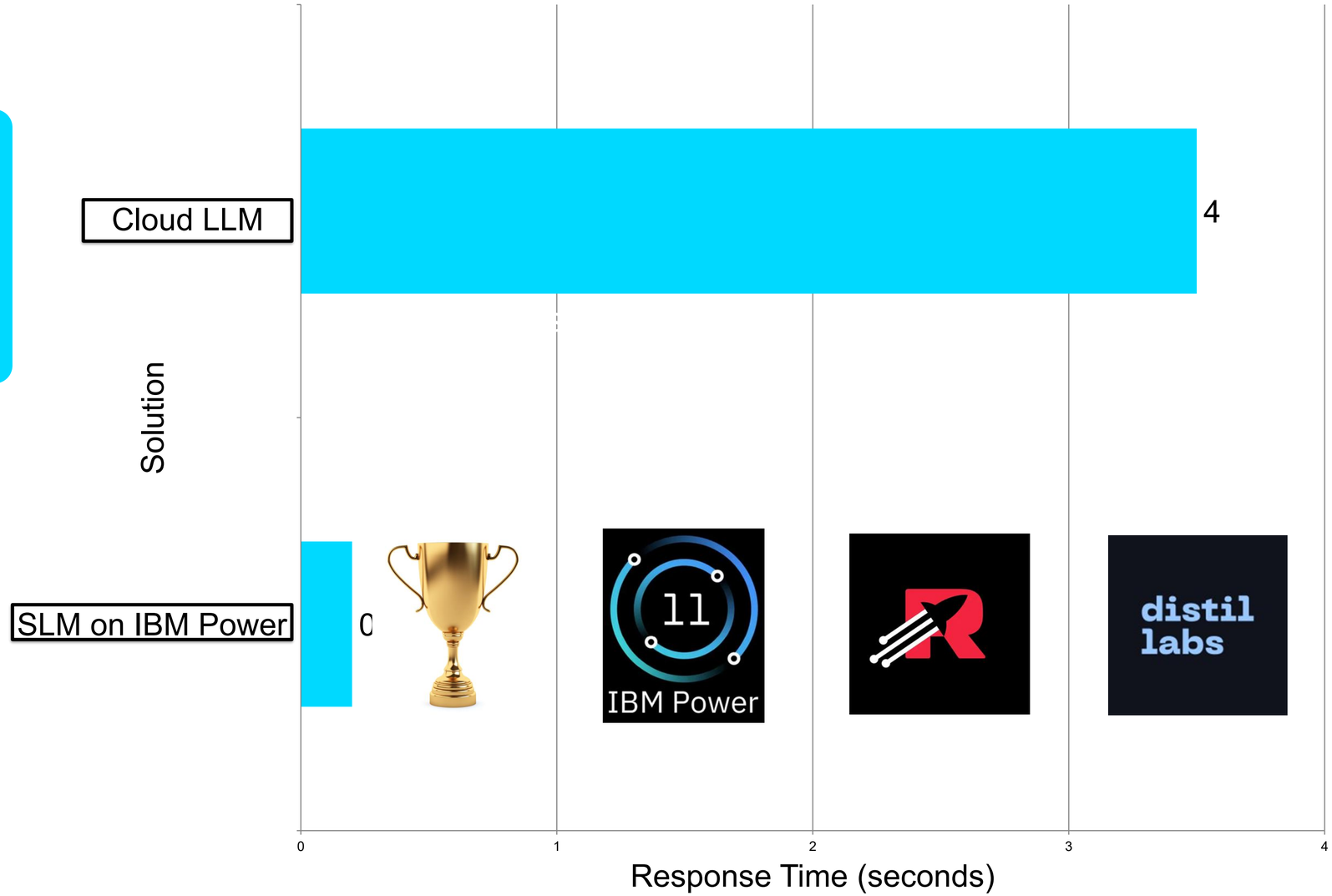
2-5s

Cloud Response Time

200ms

SLM Response Time on IBM Power

For fraud detection teams running hundreds of queries per hour, this speed difference is transformational.



Performance statement source: https://www.distillabs.ai/blog/distil-labs-enables-rocketgraphs-private-ai-on-ibm-power-with-small-language-models

Real-World Impact: Financial Services

The Scenario

Financial institutions analyzing transaction patterns for fraud detection need to query complex relationship graphs containing highly sensitive data.

The Requirements

- Zero data exposure risk
- Sub-second query response
- Regulatory compliance
- High query volume capacity

The Results

100%
Data Privacy

<200ms
Query Speed

No Limits
API Throttling

100%
Compliant

Powered by Industry Leaders

Rocketgraph

Graph-native analytics platform with OpenCypher querying

Distil Labs

Knowledge distillation experts enabling specialized SLM deployment

IBM Power

Enterprise-grade infrastructure with Granite 3.3 8B model

A collaboration proving that enterprises can adopt AI without compromising security, speed, or sustainability

Ready to Get Started?

For Rocketgraph Customers

- Runs on your existing IBM Power infrastructure
- No external dependencies or API keys required
- Seamless integration with current installation
- Full support from combined teams

Your data. Your control. Your SLM.

Learn More

Contact us to explore how SLMs can transform your graph analytics:

Technical Documentation:

`https://docs.distillabs.ai`

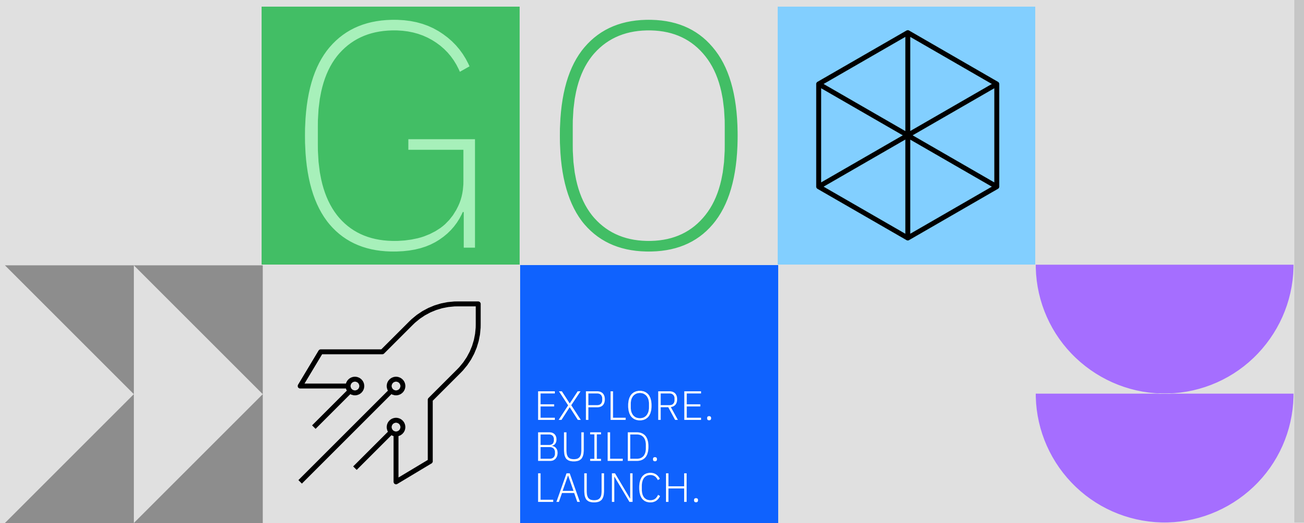
Full Blog Post:

`distillabs.ai/blog`

Schedule a demo to see private AI in action on IBM Power

Thank you

The Future of Enterprise AI is Here



Paul Chapman

PaulChapman@uk.ibm.com

IBM Global Power, AI & Modernisation Solutions Leader

Notices and disclaimers

This document is distributed “as is” without any warranty, either express or implied. In no event shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.

Case studies and client examples are presented as illustrations of how customers or IBM has used IBM products in production or test environments and the results they may have observed. Actual performance, cost, savings or other results in other operating environments may vary.

Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM.

Not all offerings are available in every country in which IBM operates.

Any statements regarding IBM’s future direction, intent or product plans are subject to change or withdrawal without notice.

Certain comments made in this presentation may be characterized as forward looking under the Private Securities Litigation Reform Act of 1995.

Forward-looking statements are based on the company’s current assumptions regarding future business and financial performance. Those statements by their nature address matters that are uncertain to different degrees and involve a number of factors that could cause actual results to differ materially. Additional information concerning these factors is contained in the Company’s filings with the SEC.

Copies are available from the SEC, from the IBM website, or from IBM Investor Relations.

Any forward-looking statement made during this presentation speaks only as of the date on which it is made. The company assumes no obligation to update or revise any forward-looking statements except as required by law; these charts and the associated remarks and comments are integrally related and are intended to be presented and understood together.