FORTISSIMO
PLUS

FFPLUS SUCCESS STORY:
INNOVATION STUDY

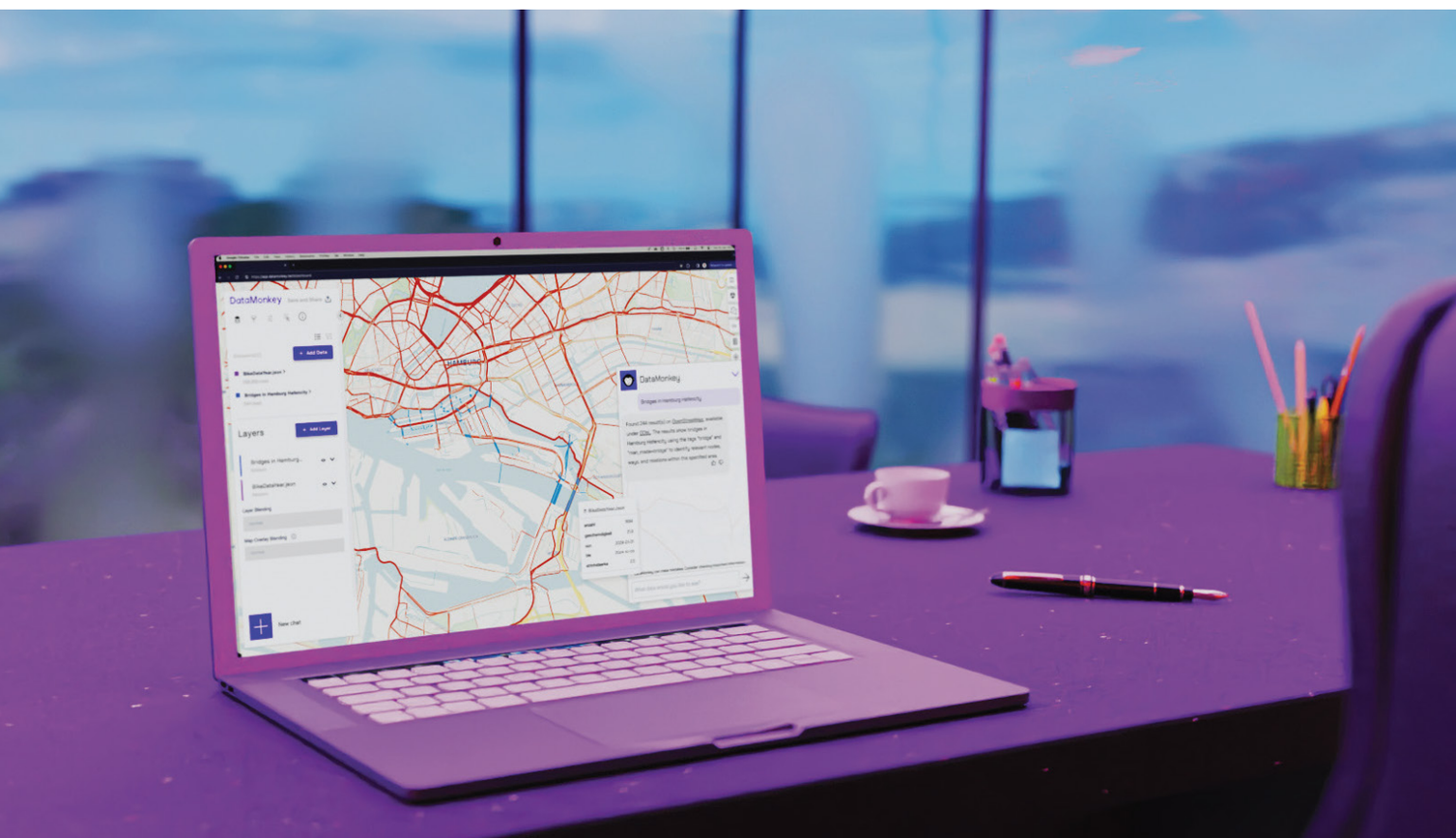
GEO-LLAMA – LARGE LANGUAGE MODEL FOR GEOGRAPHIC DATA

ORGANIZATIONS

DataMonkey, a brand under Urban Monkeys GmbH, is a German technology SME specialising in AI-driven data access solutions. The company focuses on lowering barriers to advanced analytics for business experts, data analysts and IT teams by combining large language models with domain-specific data pipelines. For this project, DataMonkey led the design, training and validation of a geospatially-aware LLM, working with EuroHPC infrastructure to deliver a scalable, European alternative for geospatial intelligence.

THE CHALLENGE

Geospatial data is critical e.g. for mobility planning, climate-risk analysis and real-estate management, yet remains difficult to exploit. Although most business data has a spatial component, many SMEs and public organisations lack GIS expertise, suitable tools and computing capacity. Open data such as OpenStreetMap is valuable but complex, requiring specialised queries and handling highly inconsistent tagging.



Technology used: HPC, AI, GenAI, Geospatial Technologies
Industry Sector: Built Environment

THE SOLUTION

DataMonkey developed Geo-Llama, a geospatially aware large language model that enables users to query and combine open geographic data using natural language. The solution integrates Generative AI, Retrieval-Augmented Generation pipelines and HPC-accelerated fine-tuning. Access to EuroHPC GPU resources enabled efficient training of multi-billion-parameter models, delivering a scalable, high-precision geospatial analytics platform for non-experts.

THE IMPACT

Geo-Llama fundamentally improves access to geospatial intelligence by reducing the need for specialised GIS skills. Through natural-language interaction, organisations can analyse and combine open geographic data while reducing data-preparation effort by up to 95%, accelerating decision-making e.g. in mobility, utilities, real estate and sustainability.

Access to large-scale HPC resources was critical to train and optimise multi-billionparameter models, ensuring the accuracy, robustness and scalability required for business-grade deployment. For DataMonkey, this enabled faster product development, reduced technical risk and a stronger competitive position in the European AI market. More broadly, Geo-Llama delivers an EU-based, compliant geospatial AI solution that supports digital sovereignty, encourages reuse of open data and enables more sustainable planning, infrastructure management and environmental decision-making across Europe.

BENEFITS

- +35% higher query accuracy compared to baseline models.
- Training dataset of over 135,000 high-quality input–output pairs covering 91% of OpenStreetMap key usage.
- Up to 95% reduction in data-preparation time for geospatial analyses.
- Enables non-technical users to perform complex geographical queries via natural language.
- Strengthens DataMonkey's position as a leading EU-based geospatial AI provider.