Irving Fisher Committee on
Central Bank Statistics

BIS

IFC-Bank of Italy Workshop on "Data science in central banking"

18-20 February 2025

Innovation project: large language model adaptation at Latvijas Banka[1]

Mārtiņš Liberts and Rihards Ostrovskis,
Bank of Latvia

[1] This contribution was prepared for the workshop. The views expressed in this publication are those of the authors and do not necessarily represent the official views of the Committee, its members, or the BIS.

Innovation Project: Large Language Model Adaptation at Latvijas Banka

Mārtiņš Liberts and Rihards Ostrovskis[1]

Abstract

Latvijas Banka has conducted an innovation project "Large Language Model Adaptation at Latvijas Banka (LLM-Lab)" according to the *iLab* framework. Innovation laboratory (*iLab*) framework is a method for targeted and rapid evaluation of new ideas, methods, or technologies. The project had two main objectives: identifying bank functions where large language models (LLMs) could be utilized, and developing a prototype for at least one of these identified functions. The project spanned four months and was executed by a team of nine employees from various departments, all working part-time for the project.

Keywords: Innovation framework, large language models, AI infrastructure

JEL classification: O31, O32, O33

[1] Chief Data Scientist (Martins.Liberts@bank.lv) and Data Scientist (Rihards.Ostrovskis@bank.lv), Data and Statistics Department at Latvijas Banka

1. iLab Framework

The Need for Innovation

Innovation is crucial for improving efficiency, optimising processes, and addressing organisational challenges. By embracing new ideas, institutions can streamline operations, freeing up valuable time for more strategic tasks. This approach not only enhances productivity but also helps reduce work overload, creating a more effective and sustainable work environment.

What is iLab?

iLab[2], or the "Innovation Laboratory," is not a physical space or organisational unit but a structured collaboration platform designed to turn ideas into real solutions. It provides a dynamic environment where professionals from different departments can work together, integrating technological advancements with business processes. iLab is a key tool for executing organisational strategies and fostering innovation within Latvijas Banka.

At iLab, employees develop ideas that align with the bank's strategic goals. The platform allows teams to test new concepts before they become official projects, encouraging cross-departmental collaboration and knowledge sharing. By bringing together expertise from various fields, iLab ensures that innovative ideas are explored and refined effectively.

How iLab Works

Each year, approximately three iLab projects are implemented, with participation open to employees from different departments. Participation is entirely voluntary, supported by direct managerial backing. To encourage a healthy competitive spirit, multiple teams may work on the same idea.

iLab follows a structured framework to ensure efficient idea development. The process begins with employees submitting their ideas for iLab projects and voting for ideas. The top voted ideas are then reviewed and prioritised by the Innovation Management Committee. Once selected, teams are formed with clearly defined roles, fostering a collaborative working environment.

During the research and development phase, teams explore potential technological solutions, define business processes, and create prototypes. Projects follow a three-month sprint cycle, using agile methods to experiment and refine their solutions. Teams provide monthly progress updates, culminating in a final review and presentation to the Innovation Management Committee. Finally a seminar is organised about lessons learned and results achieved during the iLab project. This is

[2] Artis Šmits (9 May 2023) "iLab – veids kā inovēt"[presentation slides in Latvian]. Latvijas Banka (internal publication).

the last deliverable of the iLab project. The seminar is open to all employees of Latvijas Banka.

Lessons Learned and Participant Insights

iLab projects produce several valuable outcomes, including the identification of business processes that can benefit from new technologies, a market analysis of available solutions, and an evaluation of the costs and benefits of implementing these innovations. Additionally, teams develop prototypes to demonstrate the feasibility of their ideas.

Participants have found iLab to be an inspiring and rewarding experience. It provides a structured way to explore new ideas, experiment with emerging technologies, and implement creative solutions. Employees appreciate the hands-on learning approach, which allows them to gain practical experience in an innovative environment.

Several key lessons have emerged from iLab's projects. Managerial support is crucial for allowing team members to engage with their projects. Flexibility is also essential, as enabling employees to join or leave teams easily helps maintain enthusiasm and productivity. Additionally, the initiative encourages a mindset where failure is seen as a learning opportunity rather than a setback. Even unsuccessful projects provide valuable insights and contribute to future innovation efforts. Finally, ensuring that iLab projects align with the organisation's strategic goals is critical for long-term success.

Participating in iLab is an exciting opportunity to contribute to organisational development while gaining new skills. It fosters a culture of innovation, encourages collaboration, and allows employees to challenge conventional ways of working. Moreover, iLab participation enhances professional growth and serves as a valuable addition to an employee's CV.

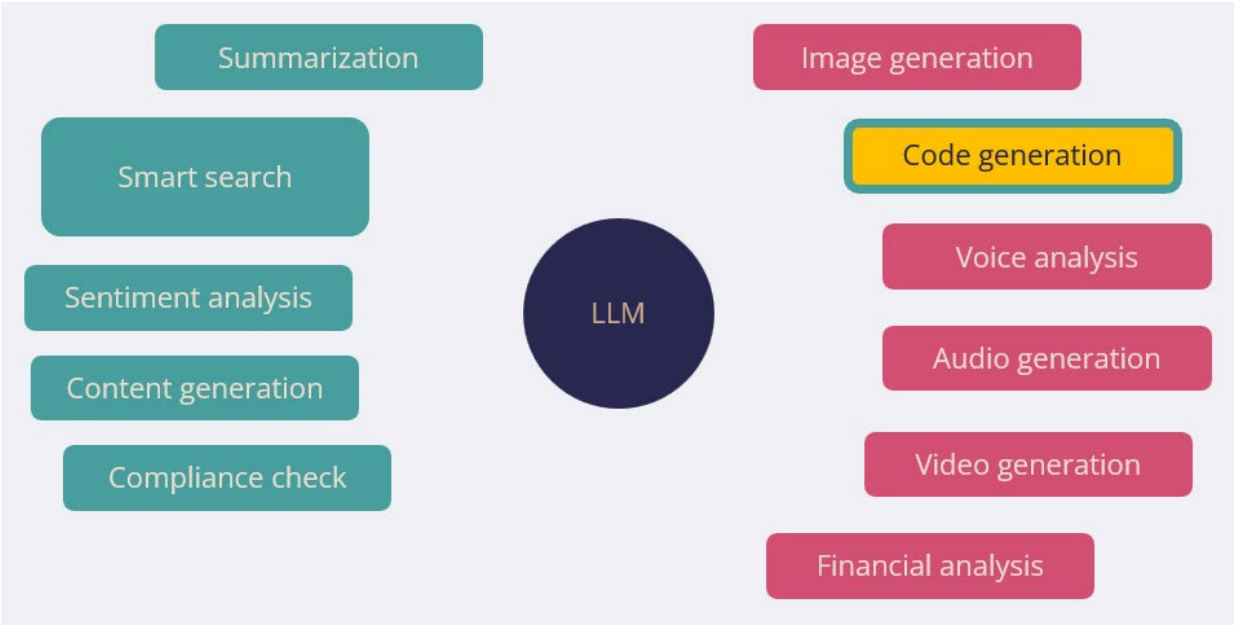
By fostering a culture of continuous learning and innovation, iLab has become an important driver of technological advancement and strategic growth within Latvijas Banka. It empowers employees to experiment, collaborate, and develop creative solutions.

2. LLM-Lab

One of the iLab projects in 2024 at Latvijas Banka was LLM-Lab. Name of this iLab is combination from the name of the iLab framework and technology researched in it – LLM. The goal of this project was to build at least one prototype that is useful for addressing real-world daily challenges. As LLM is general-use technology it was required to narrow down the scope, so it was decided that we will focus on projects that involve only text and not any multi-modal models. As shown in Graph 1, summarization, smart search, sentiment analysis, content generation and compliance appear to be the most promising categories at Latvijas Banka. The existing market products for code generation are effective, while tools for image generation, voice analysis, audio generation, and video generation do not align with the organization's primary objectives. For financial analysis, specialized tools are already in use. Therefore, the ones on the left were chosen as the primary focus.

Scope of LLM-Lab: Key Use-Case Categories

Graph 1



Source: Latvijas Banka LLM-Lab report to Innovation Management Committee

Course of the project

The project team was highly cross-functional including nine colleagues with technical (data scientists, IT architect, SupTech developers and experts) and business (experts from cash operations, law and strategy) backgrounds from several different departments. As it is typical for the iLab projects, also for LLM-Lab there were no members whom this was sole job responsibility.

We remained quite flexible throughout the project by having regular weekly meetings to monitor progress. Overall length of the project was set to around four months.

The overseeing body of this project is Innovation Management Committee involving board members, heads of departments and seasoned experts, ensuring that progress remains aligned with the organization's overall objectives.

Objectives

In order to achieve the goal of the project, first objective was to identify functions at Latvijas Banka where this technology could be effectively applied and prioritize them. Although this type of task differs from typical projects, which usually arise from a clearly identified problem, the technology and experience were still at an early stage of maturity, making it necessary to step back and identify problems to be solved by the LLM technology. After surveys and communication with every department at the organization 39 use cases were identified, two of which had already been initiated by

a few members of LLM-lab before. It is worth noting that all of the use cases widely differ by their complexity, scale and gains.

To delve a bit further into the list of use cases it is worth adding that the list of use cases is not final and remains open to revision; modifications and additions are only welcome. The larger the list, the higher the probability to find the opportunities with the highest value. We defined the top use cases by:

- categorization by categories mentioned in Graph 1;
- information classification levels at Latvijas Banka[3];
- alternatives in other institutions;
- complexity of implementation;
- initiative from stakeholders;
- impact of finished product.

Our first pilot prototypes were chosen based on initiative from stakeholders, impact and information confidentiality level appropriate for Microsoft Azure for easier experimentation and implementation so we can focus more on the solution engineering itself, less on supporting technology. As developing entirely new solutions extra effort from stakeholders is mandatory, and for most of the use cases it was by far the most important aspect for prioritization.

Second objective was to develop at least one generative AI solution within Latvijas Banka. As the participation rate and effort from the team was high, it was decided to work on more than one potential solution and continue working and building prototypes for four initiatives:

- Emission Prospectus compliance checks against Commission Delegated Regulation (EU) 2019/980[4] for licensing and sanctions purposes;
- Smart search in large documents related with cash instructions and procedures;
- Data extraction from letters received from State Police regarding cash issues;
- Assistant for drafting specification for environmentally friendly public procurements.

Additionally, another very relevant aspect – as we are a relatively small central bank, making the correct decisions and spending our resources wisely is very important for us.

A key supporting objective was to gain knowledge about the technology, the methodology for applying it, and the typical real-world scenarios in which it is used. Notably, the project's objectives were not pursued strictly in sequence but were approached iteratively, particularly regarding the learning process. Some theoretical

[3] More in detail in *Information Classification* chapter.

[4] Some of the Annexes of Delegated regulation - 2019/980 - EN - EUR-Lex are being used to find information inside the Prospectus document. If all necessary information is included in a Prospectus document then it complies to the rules and is correct. Available: https://eur-lex.europa.eu/eli/reg_del/2019/980/oj/eng

groundwork was completed in advance, while additional insights emerged during hands-on activities.

LLM-Lab LLM System Prototypes

LLM System Prototypes

Table 1

	Prospectus	Cash Instruction Documents	Data Extraction from State Police Letters	Environmentally Friendly Public Procurement Assistant
Key Use-Case Categories	Smart search + compliance check	Smart search + summarization	Smart search	Smart search + content generation
Complexity (relative)	Medium	Medium	Easiest	Hardest
Knowledge base	Prompt	RAG	Prompt	RAG
Knowledge size	80-100 A4 p.	500 A4 p.	1-2 A4 p.	1000 A4 p.
Language	English	Latvian	Latvian	Latvian
Base model	gpt-4/gpt-4o	gpt-4/gpt-4o	gpt-4/gpt-4o	gpt-4/gpt-4o
Information classification level	Restricted access	Restricted access + specially protected (excluded from prototype)	Restricted access	Public + restricted access
Result	50 documents per year, multiple hours saved per document	Harder to measure in time	200 documents per year, 20 minutes saved per document	Harder to measure in time
Interface	Background service integrated in internal platform	Chat	Background service	Chat
Types of prompts	Pre-defined	Flexible	Pre-defined	Flexible

Table 1 showcases four prototype systems that demonstrate the multifaceted nature of smart search when integrated with large language models. Although each prototype presents a slightly different combination of functionalities – from compliance checking and summarization to content generation – they all rely, at their core, on some form of information extraction. In an LLM context, smart search is understood as the capability to interpret user queries and retrieve, extract, or synthesize relevant content from large text corpus. Complexity is relative among the chosen use-cases – it helped us to decide which prototypes to focus more on at the beginning.

The size of the knowledge base varies for each of our business cases. We began with Prospectus and data extraction from State Police letters because they rely primarily on what we consider the simpler aspect of an LLM system – smart search – and involve a much smaller knowledge base. In particular, data extraction from State Police letters has very short inputs, so it was clear from the outset that a vector database was unnecessary. This simplified the overall LLM system and avoided any risk of inaccuracies when extracting chunks of text.

Language is not an obstacle anymore if the most recent and powerful LLMs are used. According to OpenAI already gpt-4 was performing well for Latvian[5] – far above average. Smaller models for smaller languages could bring some obstacles.

We later concluded that Prospectus smart search also did not require Retrieval Augmented Generation (RAG). It produced better results without it, especially after the release of the GPT-4o model, which features an expanded input context window of 128,000 tokens[6] – compared to the 8,000 or 32,000 tokens[7], of the GPT-4 model. Since 100 pages amount to roughly 80,000 tokens, they comfortably fit within a context window, and we did not observe "lost in the middle"[8] issues.

However, for our other two prototypes, the knowledge base is too large to fit into a context window, making a vector database and RAG necessary to handle the data effectively. The information classification level for all prototypes permits the use of cloud-based LLM APIs, except for some cash-instruction documents, which were excluded due to their relatively small quantity. Nevertheless, we designed the systems to be as modular as possible so that we can switch to locally hosted models at any time, for any reason.

For solutions integrated into established processes – such as Prospectus or data extraction from State Police letters – measuring results is relatively straightforward, given the clear input, output, and time savings compared to previous methods. However, for chatbots, which are used for a broader range of tasks by more colleagues, quantifying benefits is more challenging. This also applies to the user interface: while background services can be seamlessly integrated into existing workflows, adopting a chat interface requires changes in both user behaviour and organizational culture.

The types of prompts significantly affect the complexity of evaluating the system. When using a fixed set of predefined prompts – especially with closed-ended questions, such as extracting data from state police letters – it becomes easier to build a corresponding evaluation dataset. However, developing chat systems with flexible, open-ended questions requires a larger and more diverse dataset, ensuring that the distribution of questions aligns with the underlying knowledge base.

Current status, as of writing this paper (March 2025), is that Prospectus and data extraction from State Police letters have been in production for several months and is being used on regular basis. The rest of the solutions are at the last cycles of evaluation and soon to be implemented in production.

[5] OpenAI "GPT-4". Available: https://openai.com/index/gpt-4-research/

[6] Azure OpenAI Documentation. Available: https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/models?tabs=global-standard%2Cstandard-chat-completions

[7] Except gpt-4 turbo had 128k input context window length, but gpt-4o is a superior model

[8] N. F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang, "Lost in the Middle: How Language Models Use Long Contexts," *arXiv preprint arXiv:2307.03172*, 2023. [Online]. Available: https://arxiv.org/pdf/2307.03172

Conclusion and results of LLM-Lab

To begin with, not only was the initial goal achieved, but four prototypes, instead of one, were developed, and future projects and priorities were clearly outlined. Beyond the direct impact, the initiative laid a flexible yet rigorous foundation by involving participants from across the organization, thus fostering a much broader understanding of applied GenAI among both technical and business colleagues. As of March 2025, this groundwork has already contributed to numerous projects and will continue to support the development of future initiatives.

Overall, LLM-Lab provided a highly impactful and well-structured approach to exploring this new and immensely broad topic.

3. Way forward

Democratising LLM Technology

To fully unlock the potential of large language models (LLMs), we aim to democratise their use across the organisation, ensuring that every employee – not just technical specialists – can create and customise their own AI models. This includes providing a no-code interface, allowing users to build and fine-tune models effortlessly, regardless of their technical expertise. By making AI model development accessible to all, we foster innovation, improve efficiency, and enable employees to tailor AI solutions to their specific needs.

Achieving this vision requires a robust infrastructure that is functional, intuitive, and secure. Employees need a seamless and reliable environment where they can experiment with LLMs, refine their models, and integrate them into their workflows. Security is paramount, ensuring that sensitive data remains protected while maintaining ease of use.

To ensure broad usability, each custom model must be accessible through two key interfaces. First, a web-based chat interface will allow employees to interact with their models in a user-friendly manner, making AI-driven assistance readily available for various tasks. Second, an API integration will enable these models to be embedded into existing information systems, streamlining operations and enhancing automation.

By providing these tools and infrastructure, we create an ecosystem where AI is not just a specialised technology but a widely available resource that empowers every employee to innovate and improve their daily work.

Information Classification

The current information classification framework used at Latvijas Banka distinguishes between three levels of information:

- publicly available information;
- restricted access information;

- specially protected restricted access information.

Publicly available information can be stored and processed across all environments without restrictions. However, restricted access information may only be stored and processed on on-premises infrastructure or Microsoft Azure Cloud, while general cloud platforms are not permitted for this type of data storage. The most sensitive category, specially protected restricted access information, must be stored exclusively on on-premises infrastructure to ensure the highest level of control.

This is a crucial requirement we should be taking into account when building our LLM infrastructure. This is a reason we have chosen to build two environments, namely:

- Environment with cloud models hosted in Microsoft Azure – usable for restricted access information;
- Environment with local models hosted on premises – usable for restricted access information and also specially protected restricted access information.

This dual-environment approach ensures that AI capabilities can be leveraged securely while adhering to strict data protection standards.

Technology Stack

We have chosen Open WebUI[9] as a core technology for building the LLM infrastructure. Open WebUI is an open-source project delivering extensible, feature-rich, and user-friendly self-hosted AI platform. Open WebUI serves as the primary interface for user interactions, whether through direct usage or system integrations.

In the cloud environment, LLM models are hosted on Azure AI Services[10], while in the on-premises environment, they are managed by Ollama[11]. The backend processes, including request handling and data storage, are managed by supporting components such as LiteLLM[12] (an open-source gateway for cloud hosted models) and PostgreSQL[13] (database engine), but these remain transparent to users. Graph 2 explains several usage scenarios. These usage scenarios illustrate how different users interact with the AI system according to their specific roles and security requirements.

[9] Timothy Jaeryang Baek et. al. "Open WebUI" [Online]. Available: https://docs.openwebui.com/

[10] Microsoft "Azure AI services documentation" [Online]. Available: https://learn.microsoft.com/en-us/azure/ai-services/

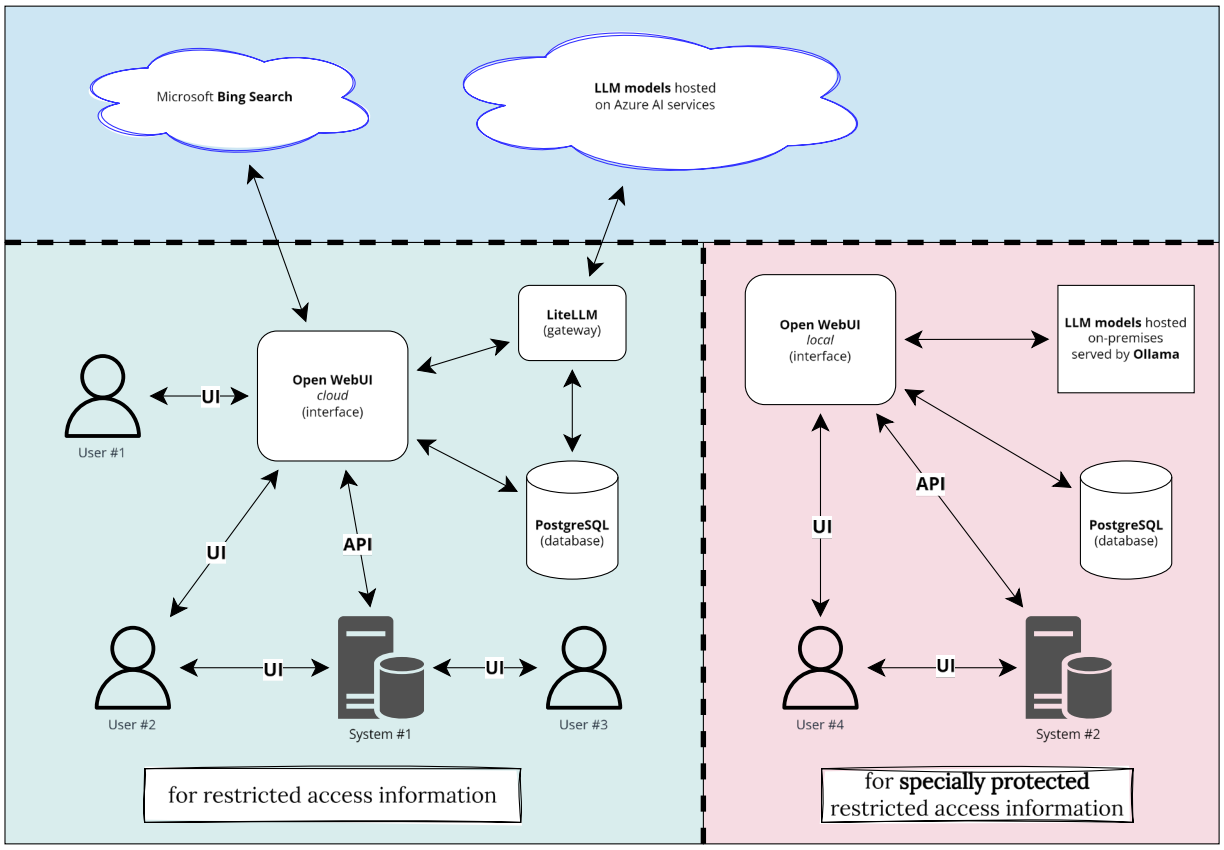
[11] Jeffrey Morgan et. al. "Ollama Documentation" [Online]. Available: https://github.com/ollama/ollama/blob/main/docs/README.md

[12] Ishaan Jaffer and Krrish Dholakia "Lite LLM" [Online]. Available: https://docs.litellm.ai/

[13] The PostgreSQL Global Development Group "PostgreSQL 17.4 Documentation" [Online]. Available: https://www.postgresql.org/docs/current/

Architecture sketch

The sketch explains two environments for cloud based and on-premises hosted models Graph 2



[1] A sketch for an Open WebUI architecture for restricted and specially protected restricted information

Sources: Authors own drawing

User #1 represents researchers using cloud-based AI. A researcher requires AI-powered assistance for analysing publicly available or restricted access information. They interact with Open WebUI (*cloud*), which provides access to LLM models hosted on Azure AI Services. If needed, the system can retrieve additional data from Microsoft Bing Search.

User #2 represents analysts integrating AI with internal systems. They work with restricted access information and requires AI support within an internal system. They interact with Open WebUI (*cloud*) and integrate AI functionalities via an API, allowing automation and enhanced workflow efficiency. The infrastructure ensures secure handling of restricted access information while providing easy access to AI capabilities.

User #3 represents users utilising LLM technology exclusively through system integration rather than direct interaction with Open WebUI. Their internal system connects via an API, enabling seamless access to LLM models hosted in Azure AI Services. This integration allows AI-powered automation while maintaining security and compliance with information classification policies.

User #4 represents analysts working with specially protected restricted access information. Their interaction mirrors that of User #2, except they use Open WebUI

(*local*) to access LLM models hosted on-premises, managed by Ollama. This setup ensures that all data remains within a highly secure infrastructure while enabling AI-powered analysis and decision-making.

The integration of these technologies offers several benefits. A unified interface provided by Open WebUI simplifies interactions with various LLMs through a single, cohesive platform. The framework supports custom model development through techniques such as prompting and retrieval-augmented generation (RAG), enabling employees to customise AI models according to their specific needs. Finally the framework allows usage and sharing of created custom models through the web based chat interface or API integration in existing information systems.

Conclusions

The iLab framework proved to be an effective platform for initiating a structured and goal-oriented exploration of LLM technology at Latvijas Banka. Through this process, we identified multiple potential use cases for LLM integration and successfully developed four functional prototypes. The positive outcomes of the LLM-Lab project have reinforced our commitment to further implementing LLM technology across various banking processes.

A key priority moving forward is the democratisation of LLM technology, ensuring that AI-powered tools are accessible to all employees. We recognise this as a crucial step toward modernising and optimising our institution's operations. To achieve this, we are actively developing a functional, intuitive, and secure environment that enables seamless interaction with both generic and custom LLMs, fostering innovation and efficiency across all departments.

Innovation project: large language model adaptation at Latvijas Banka

Mārtiņš Liberts and Rihards Ostrovskis
Latvijas Banka

February 18, 2025

Table of contents

iLab framework

iLab framework

- *iLab* = **Innovation Laboratory**
- A method for targeted and rapid testing of new ideas, solutions, and technologies
- Benefits of *iLab* – **structured innovation**:
  - Work smarter in the long term, not harder
  - Achieve the bank's strategic goals
  - Enhance professional capabilities
  - Foster constructive collaboration with colleagues from other departments
  - Purposefully go the *extra mile*
  - Diversify work life with interesting and exciting projects

Key principles of *iLab*

- **Permission to Fail**
  - Mistakes are allowed, but inaction is not tolerated
  - “Fail fast” is considered a positive outcome
- **Structured and Purposeful Work**
  - Focused, structured work without unnecessary bureaucracy
- **Deliverables**
  - Ends with a deliverable: feasibility of the idea, prototype, potential costs, benefits, or a well-evaluated decision to abandon the idea

The process of *iLab*

- Submission of ideas
- Voting for submitted ideas
- Approval of *iLab* projects by the Innovation Management Committee
- Project work for three months
- Monthly interim reports
- Final deliverable

LLM-Lab project

LLM-Lab objectives

- **Identify Bank Functions**
  - Determine the bank functions where implementing an LLM-based solution would be most beneficial
- **Develop AI Solution**
  - Create a generative AI solution within Latvijas Banka

LLM-Lab innovation steps

- Vision
- Technical capacity
- Business problem
- User identification
- Experimenting and prototyping

Vision for solutions

- Summary creation
- Smart search
- Sentiment analysis
- Text generation
- Compliance assessment

Business scenarios

- 39 possible business scenarios were identified, some examples:
  - Processing email correspondence with market participants
  - Evaluating EU regulatory projects for compliance with Latvian interests
  - Processing EU questionnaires and preparing responses
  - Generating flowcharts from text to evaluate procedure quality
  - Obtaining information on large monetary transactions that could affect the balance of payments

Prototypes

- Emission Prospectus compliance checks against the Commission Delegated Regulation (EU) 2019/980 for licensing and sanctions purposes
- Smart search in large documents related with cash instructions and procedures
- Information extraction from letters received from State Police regarding cash issues (**in production**)
- Assistant for drafting specification for green public procurements

Way forward

Information classification

	On-premises	Microsoft Azure Cloud	Cloud
Publicly available information	✓	✓	✓
Restricted access information	✓	✓	✗
Specially protected restricted access information	✓	✗	✗

Technology stack

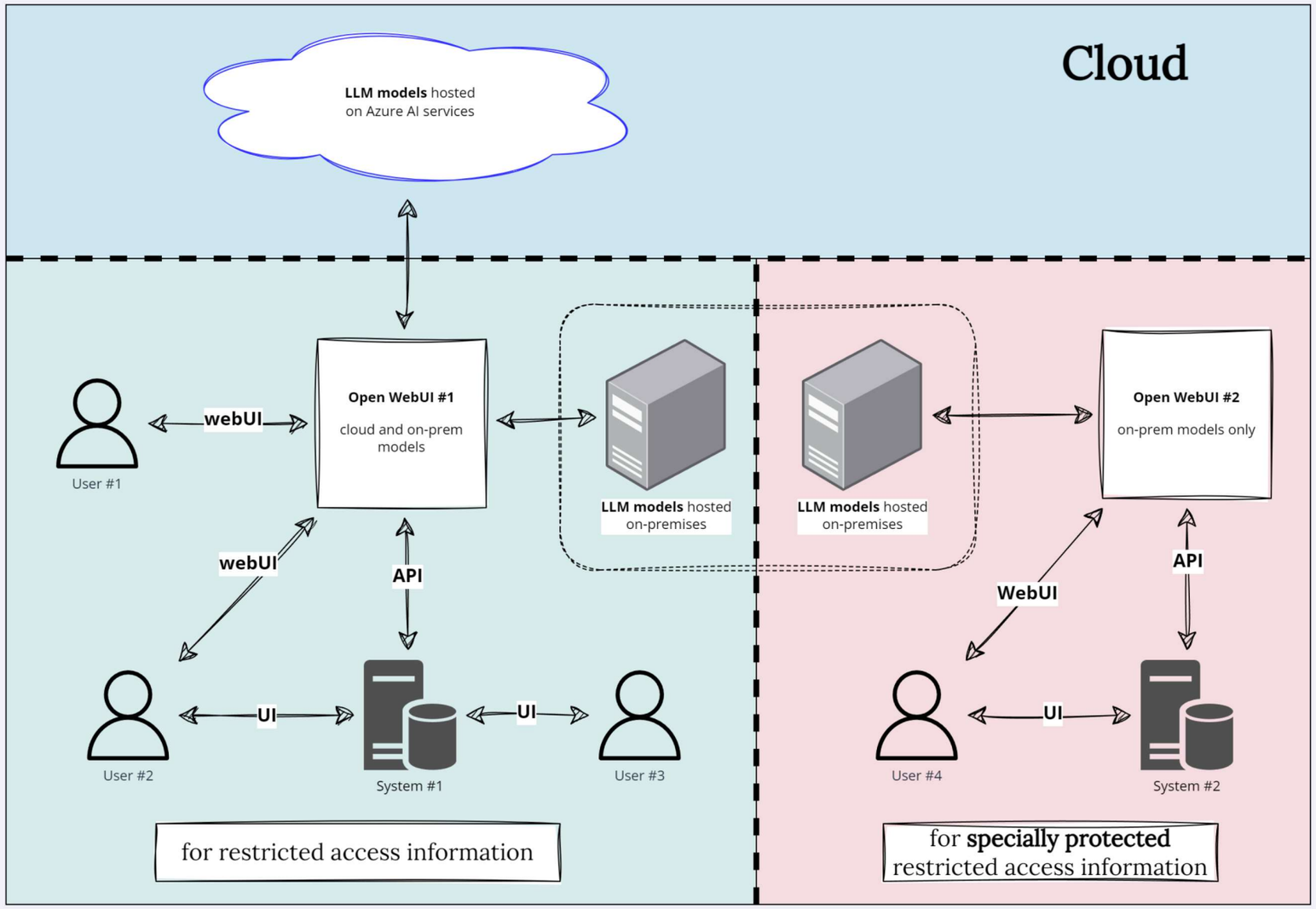
Azure AI Services for hosting OpenAI GPT models in cloud

Ollama for hosting open-source models on-premises

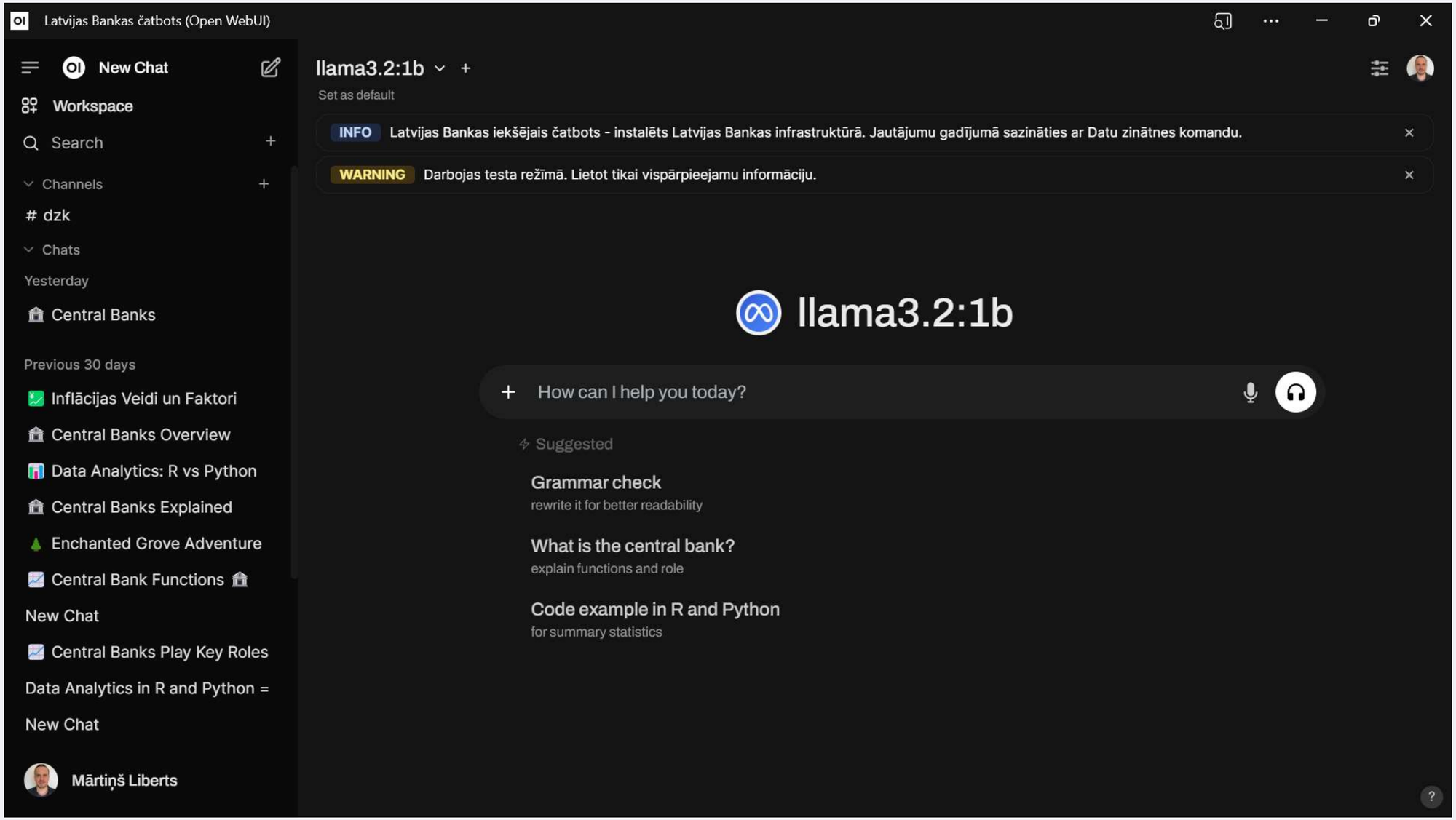
Open WebUI as user and application interface for LLM models

- Unified Interface: Simplify interactions with different LLMs through a single, integrated platform.
- Ease of Implementation: Quick start integration with comprehensive documentation and community support.
- Custom models built by prompting and retrieval-augmented generation (RAG)

Architecture sketch



UI example


Latvijas Bankas čatbots (Open WebUI)
New Chat
Workspace
Search
Channels
dzk
Chats
Yesterday
Central Banks
Previous 30 days
Inflācijas Veidi un Faktori
Central Banks Overview
Data Analytics: R vs Python
Central Banks Explained
Enchanted Grove Adventure
Central Bank Functions
New Chat
Central Banks Play Key Roles
Data Analytics in R and Python =
New Chat
Mārtiņš Liberts
llama3.2:1b
Set as default
INFO Latvijas Bankas iekšējais čatbots - instalēts Latvijas Bankas infrastruktūrā. Jautājumu gadījumā sazināties ar Datu zinātnes komandu.
WARNING Darbojas testa režīmā. Lietot tikai vispārpieejamu informāciju.
llama3.2:1b
How can I help you today?
Suggested
Grammar check
rewrite it for better readability
What is the central bank?
explain functions and role
Code example in R and Python
for summary statistics

Summary

Summary

- *iLab* framework – a method for targeted and rapid testing of new ideas, solutions, and technologies
- *LLM-Lab* project – the first step for implementing an LLM-based solutions
- Hybrid architecture with LLM models hosted on-premises and in cloud accessible through an unified interface provided by the Open WebUI

Thank you!