



Swiss Institute of
Bioinformatics

Agentic AI for biodata exploration

Panayiotis Smeros

12 October 2026, ELIXIR-IMPACT Webinar

Problem

- Researchers want to access biological data, but **don't know which resource to target**

"How many methicillin-resistant strains of S. aureus are there?"

resistance genotype -> NCBI Pathogen Detection

the actual genomes -> BV-BRC, ENA

what changed under treatment -> GEO

what the protein does -> UniProt, STRING

what's been published -> PubMed

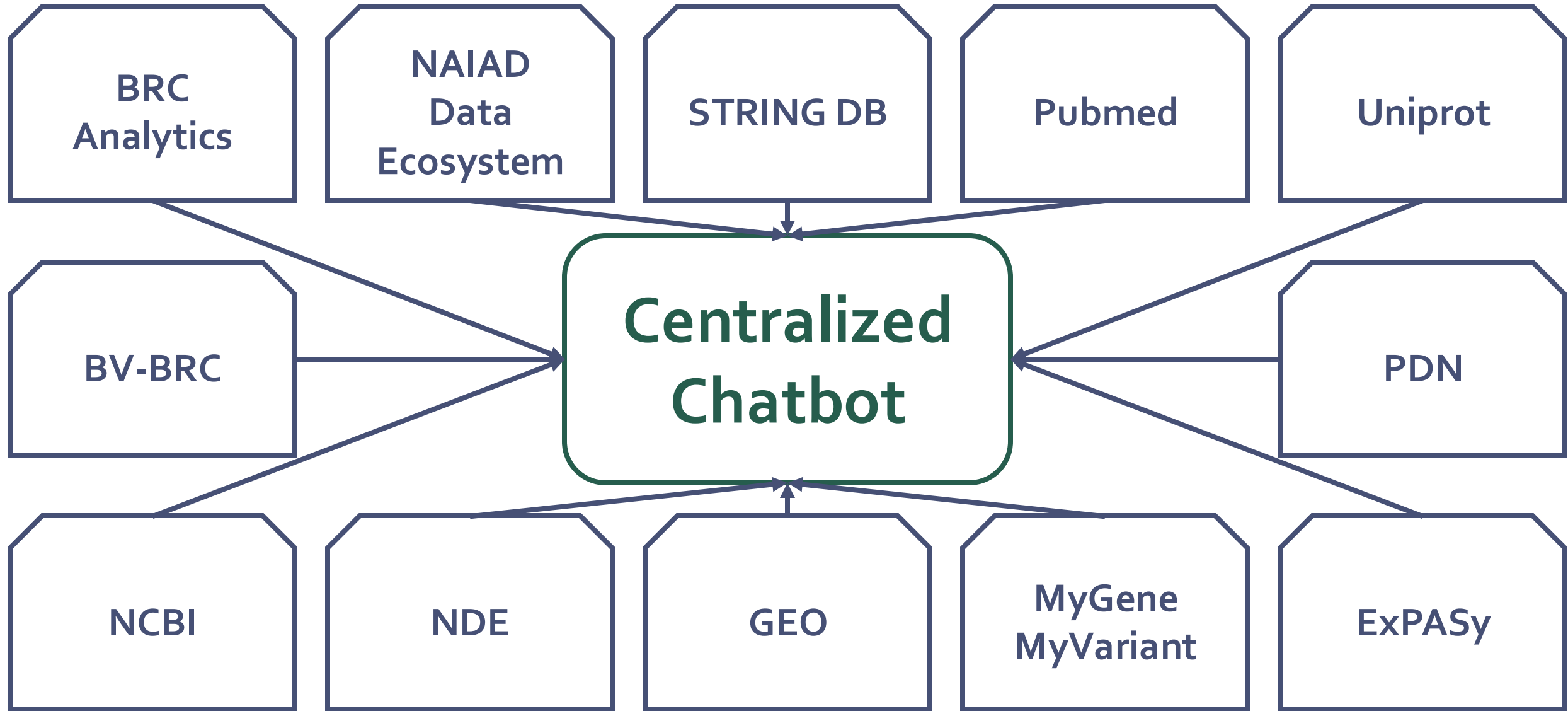
- **Many data sources, different data models, different access methods**

One centralized place to ask; all relevant data at hand

A group of six people (three women and three men) standing in a room with a blue wall. They are positioned around a whiteboard that contains handwritten notes. To the right, a large screen displays a colorful image of a brain with the text "SOL240-1501a" and some technical details. The people are wearing casual attire and name tags.

3

Our Solution



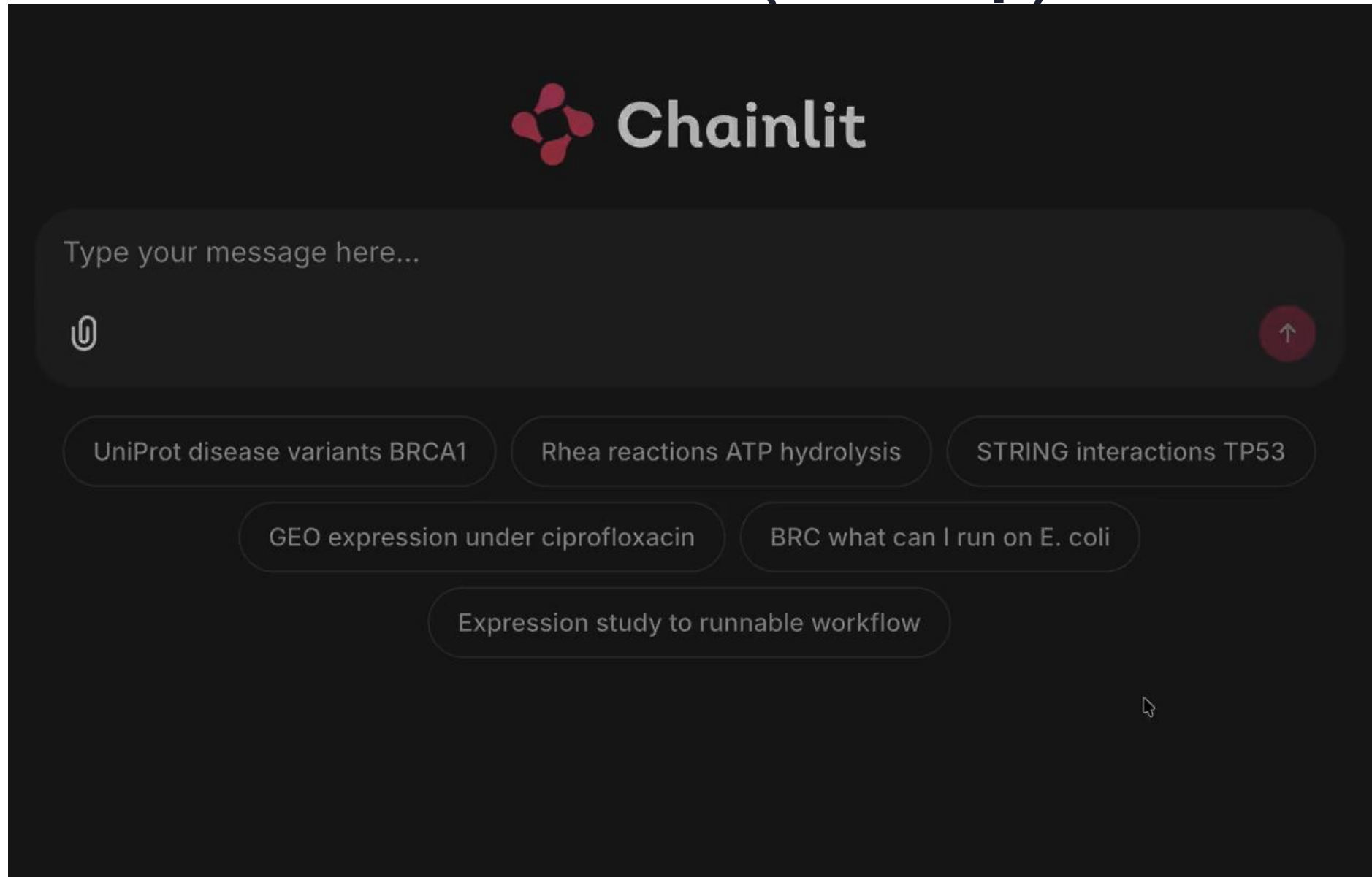
13 MCP Servers – 114 MCP Tools

Research Questions

- How many methicillin-resistant strains of *S. aureus* are there? Then: read the genome queries; which groups study these; tell me about strain X; structural implications.
- What is the latest in AMR for a named organism or drug?
- Question to test bias towards repositories: What tools and datasets can I use to compare gut microbiome of human subjects with *C. difficile*
- Question to trick system into providing inaccurate response: what resources (tools/data/pipelines) could I use to analyze airborne *Vibrio cholerae* transmission
- Can you show me a collection of influenza genomes that I could use to study the evolution of the virus over time
- Recommend tools and workflows to identify proteins in virus or bacteria families that are candidates for small molecules
- From the GH: launch an analysis of Tuberculosis genomes to identify antibiotic resistance profiles among subjects from different countries. Are there differences in patterns of drug resistance by country? Over time?

Demo Time!

Demo Time! (Backup)



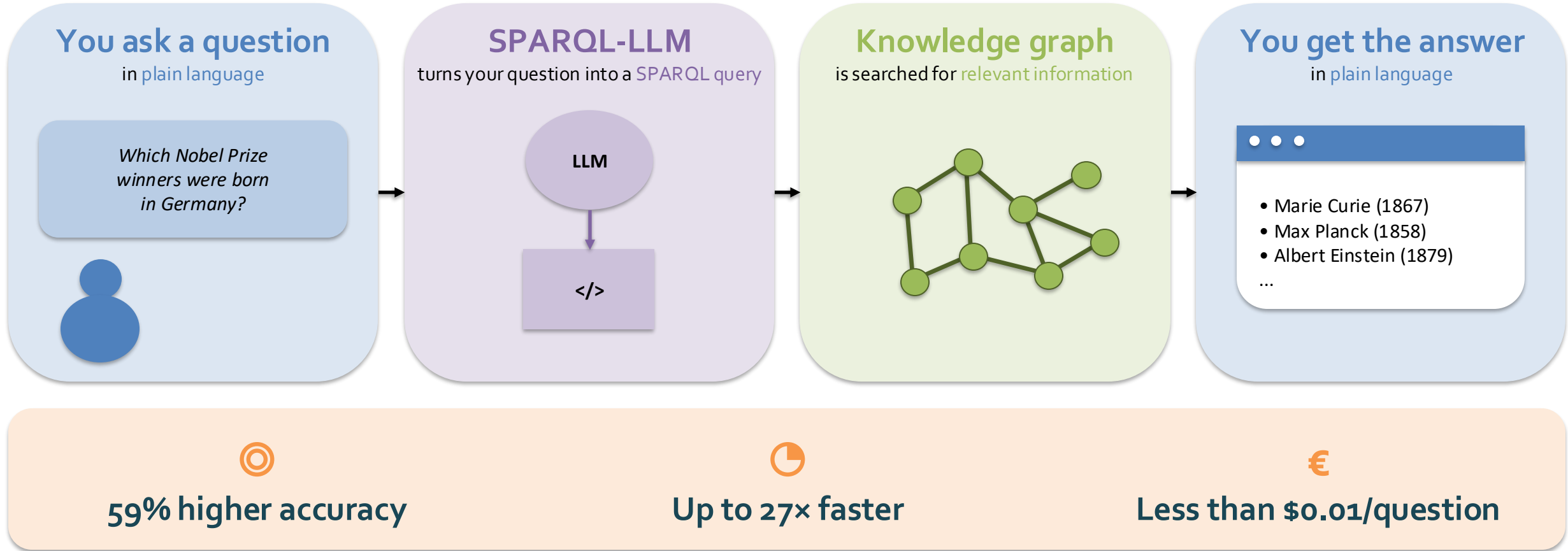
Open Questions

- **Evaluation:** LLM answers with vs. without MCPs
- **Selection Bias:** servers with more tools or better docstrings win the routing
- **Competing Sources:** NCBI and ENA disagreed in one question, which do we keep?
- **Interoperability:** cross-source joins are delegated to the LLM; No shared identifier model
- **Authentication:** per-source credentials, protection from agent scraping
- **Deployment:** 13 lightweight MCP servers (soon public!)

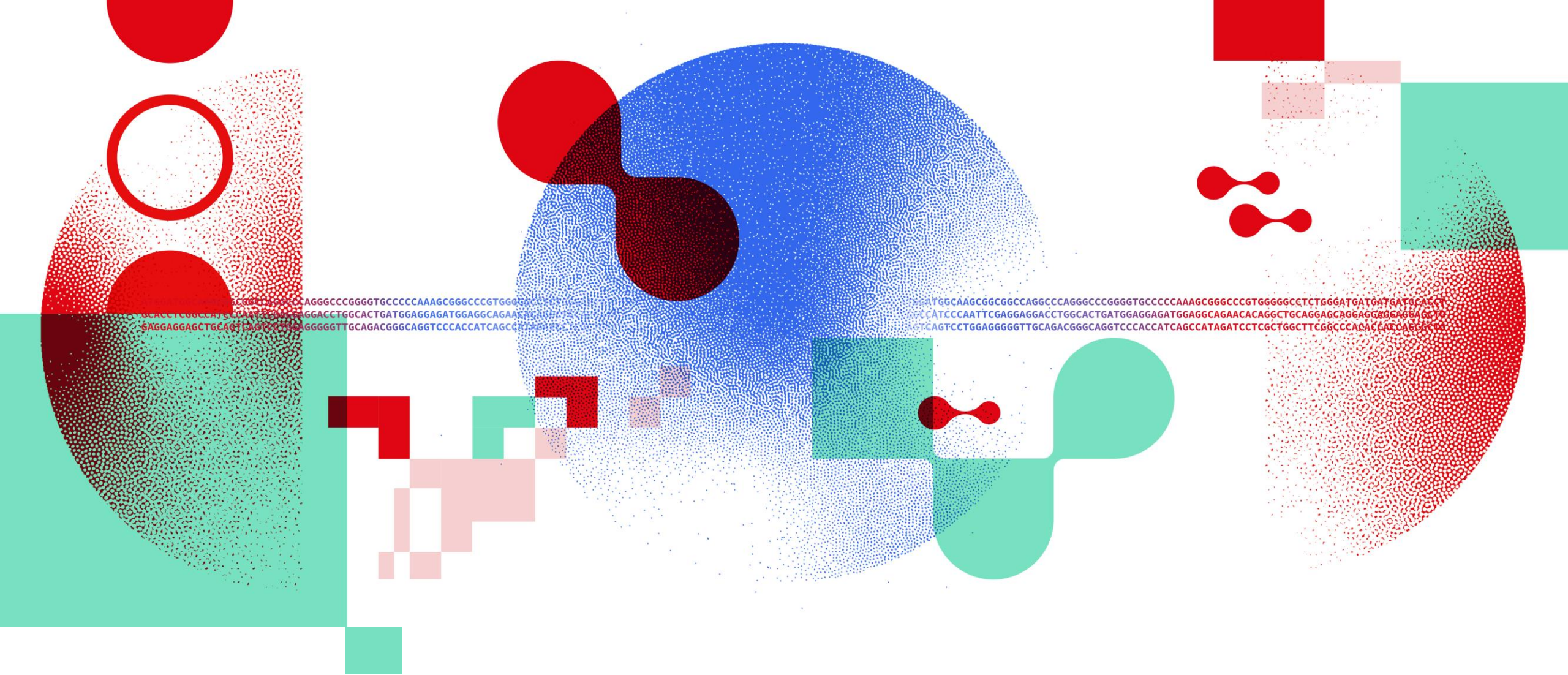
ELIXIR-IMPACT Course in January

- Building Agentic AI Applications for Biodata Exploration (**BAIBE**)
- **Learning outcomes**
 - **Understand** the basics of Large Language Models
 - **Explore** the interplay of LLMs and bioinformatics Knowledge Bases (KBs)
 - **Develop** tools to enable AI agents to interact with bioinformatics KBs
 - **Interact** programmatically with LLMs
 - **Build** a simple agentic application for answering questions over biodata

SPARQL-LLM in a nutshell



<https://dl.acm.org/doi/abs/10.1145/3847195>



Thank you!

KNOWLEDGE REPRESENTATION FOR LIFE

panayiotis.smeros@sib.swiss