eccenca
mastering complexity

Edgard Marx | Principal Data Scientist

LEVERAGING AGENTIC DATA FLOWS WITH LARGE LANGUAGE MODELS

AIRBUS

SIEMENS

AstraZeneca

ERICSSON

TOTAL

NOKIA

NAVAL GROUP

BOSCH

DBpedia

SW//M

eccenca
mastering complexity

Outlook

Introduction

- Day-to-**Day**ta tasks

Motivation

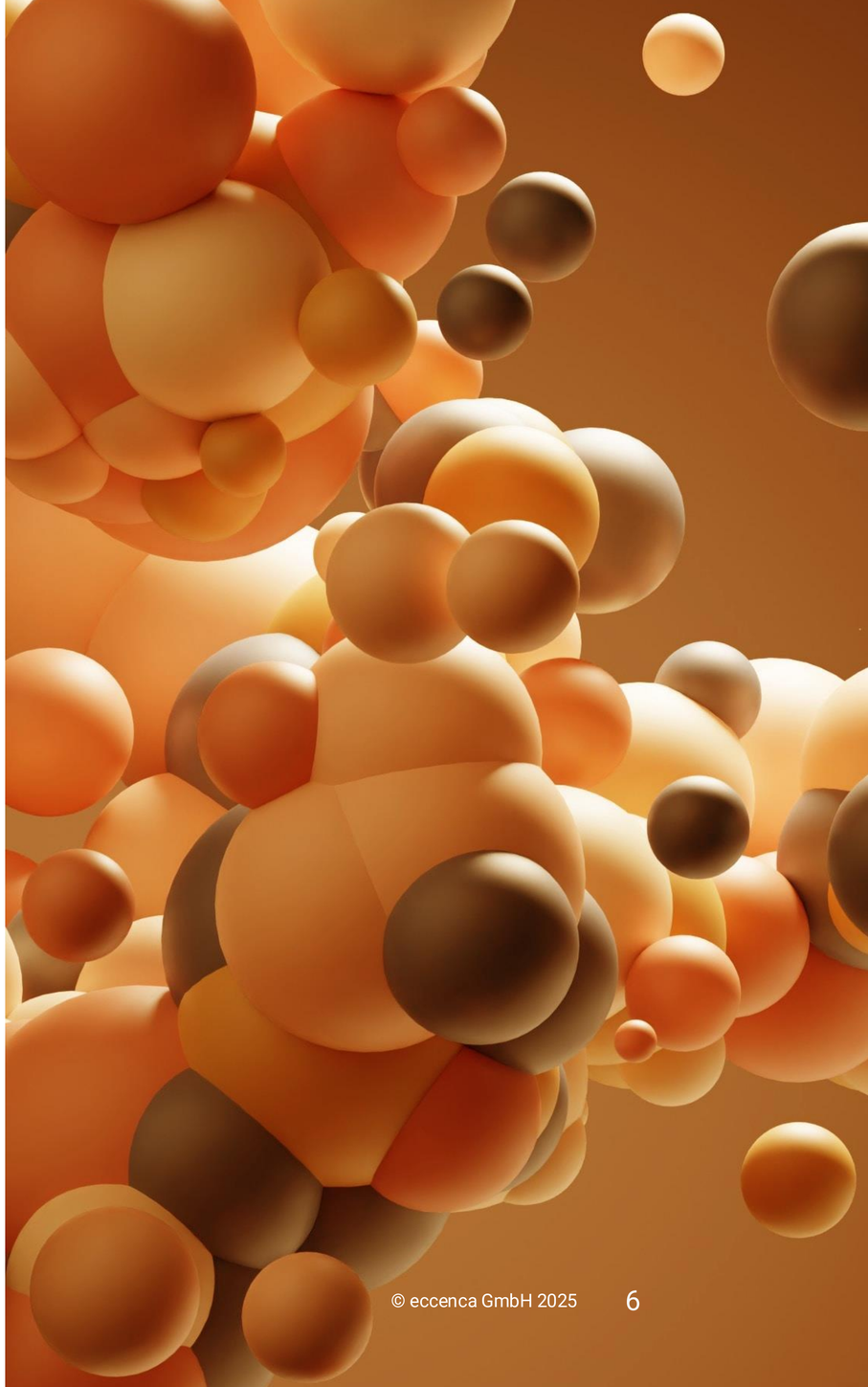
- Diving into Large language Models (LLMS)
- A shared Vision

Showcase

- Practical Application

Conclusion

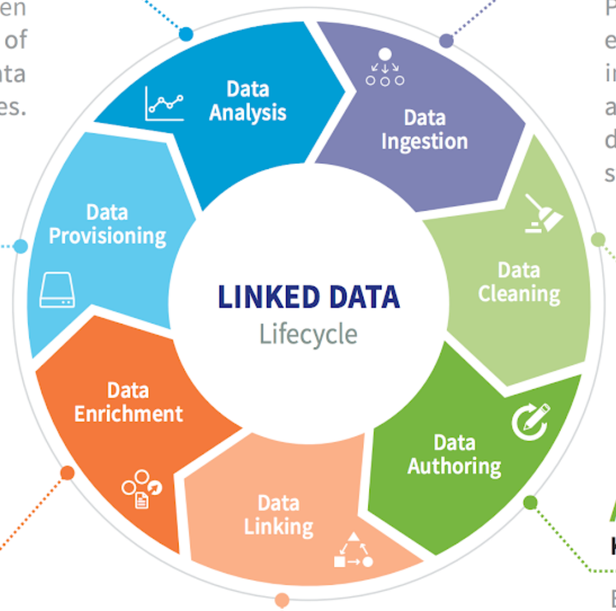
- Challenges
- Future Directions
- Questions



eccenca
mastering complexity

Day-to-Dayta tasks

The (linked) data life cycle



ANALYSIS
Insights from linked data

Discover and analyze complex - and often hidden - relations in your data by means of customized data portals and state-of- the-art data visualization based on Semantic Data Lakes.

INGESTION
Avoid data migration

PoolParty is a semantic middleware that enriches and contextualizes your data with relevant information. Link and integrate different datasets by applying semantic knowledge models. Enterprise data gets smarter while remaining in the original storage system.

PROVISIONING
Data-driven personalization

Use semantic technologies to make your content and data management workflows smarter and more dynamic. Create personalized digital experiences. Feed your bots and machine learning algorithms with higher-quality data.

CLEANING
Data quality management

PoolParty allows organizations to ensure that their data is consistent, reliable and usable for analysis. Benefit from highly automated 'Data Repair' features.

ENRICHMENT
Actionable data

Provide automatic tagging functions based on controlled vocabularies to ensure precise content classification. Semantic models help to enrich all content and data consistently and based on standards.

AUTHORING
Knowledge graph capabilities

Enable collaboration and develop synergies between your departments by establishing defined knowledge graphs and controlled vocabularies

LINKING
The power of connected data

Link internal and external data resources and provide your organization with a customized and relevant knowledge platform.

eccenca
mastering complexity

Day-to-Dayta tasks

“Data practitioners spend 80% of their valuable time finding, cleaning, and organizing the data”

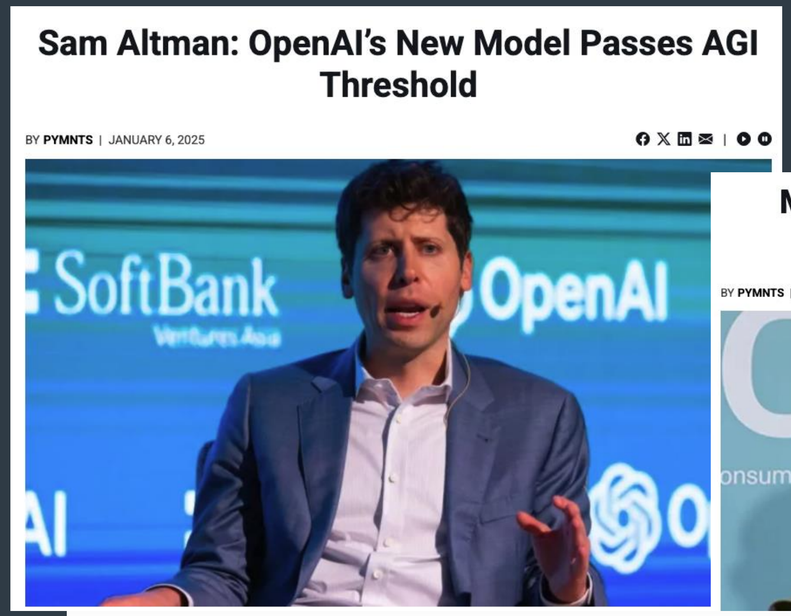
Challenges in Data Tasks

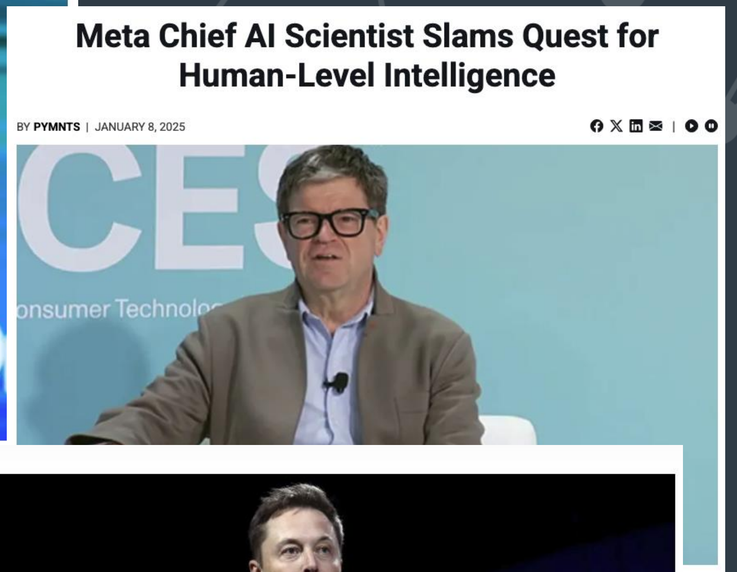
- Handling Missing Data
- Removing Duplicate Entries
- Correcting Data Types
- Standardizing Data Formats
- Dealing with Outliers
- Data Cleaning Tools and Techniques
- Addressing Inconsistent Data

Diving into LLMs

Clonflicting Visions

“In fact, we’re not even close to matching the understanding of the physical world of any animal, cat or dog.”
Lecun



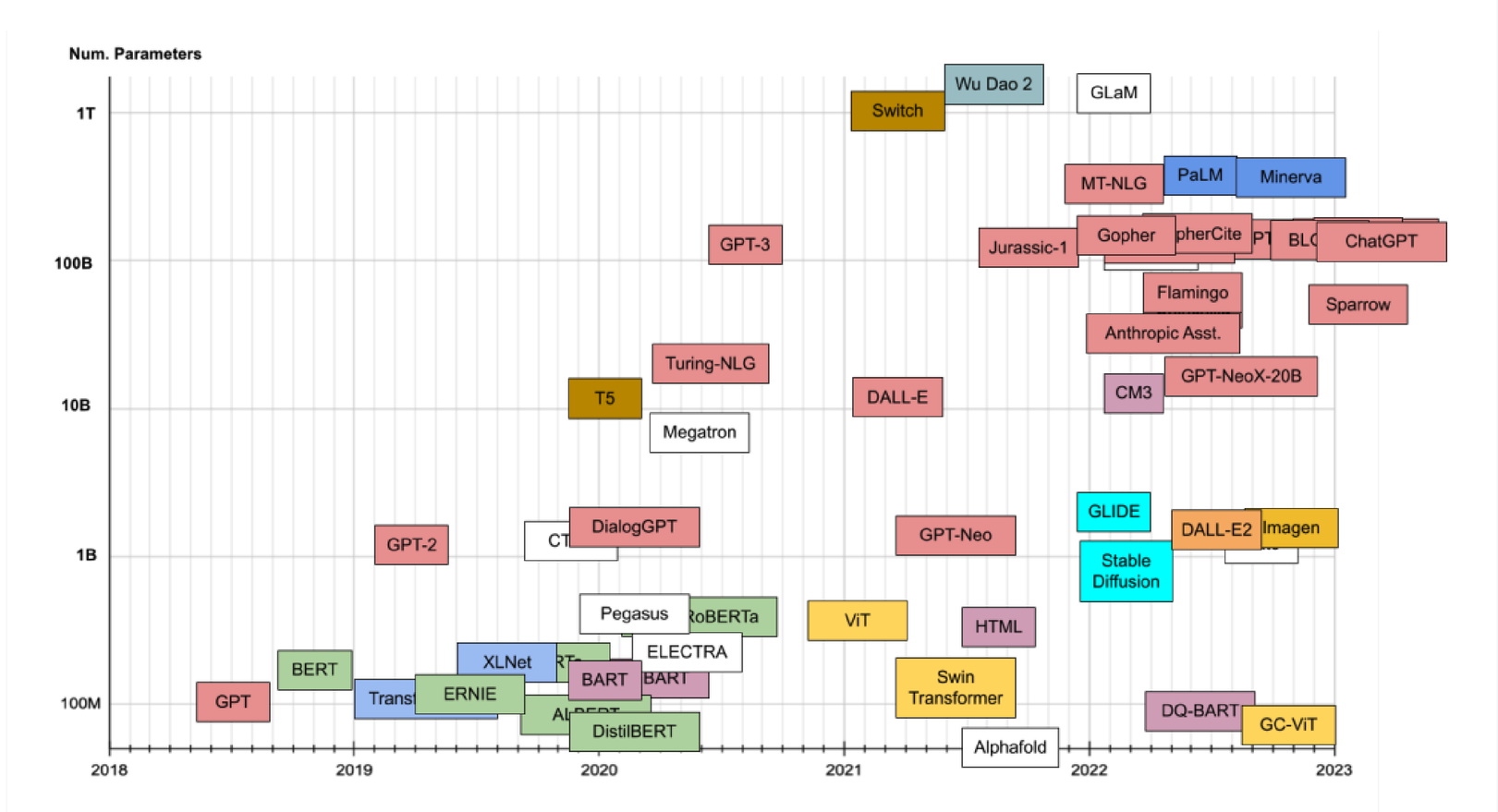


eccenca
mastering complexity

LLM Evolution

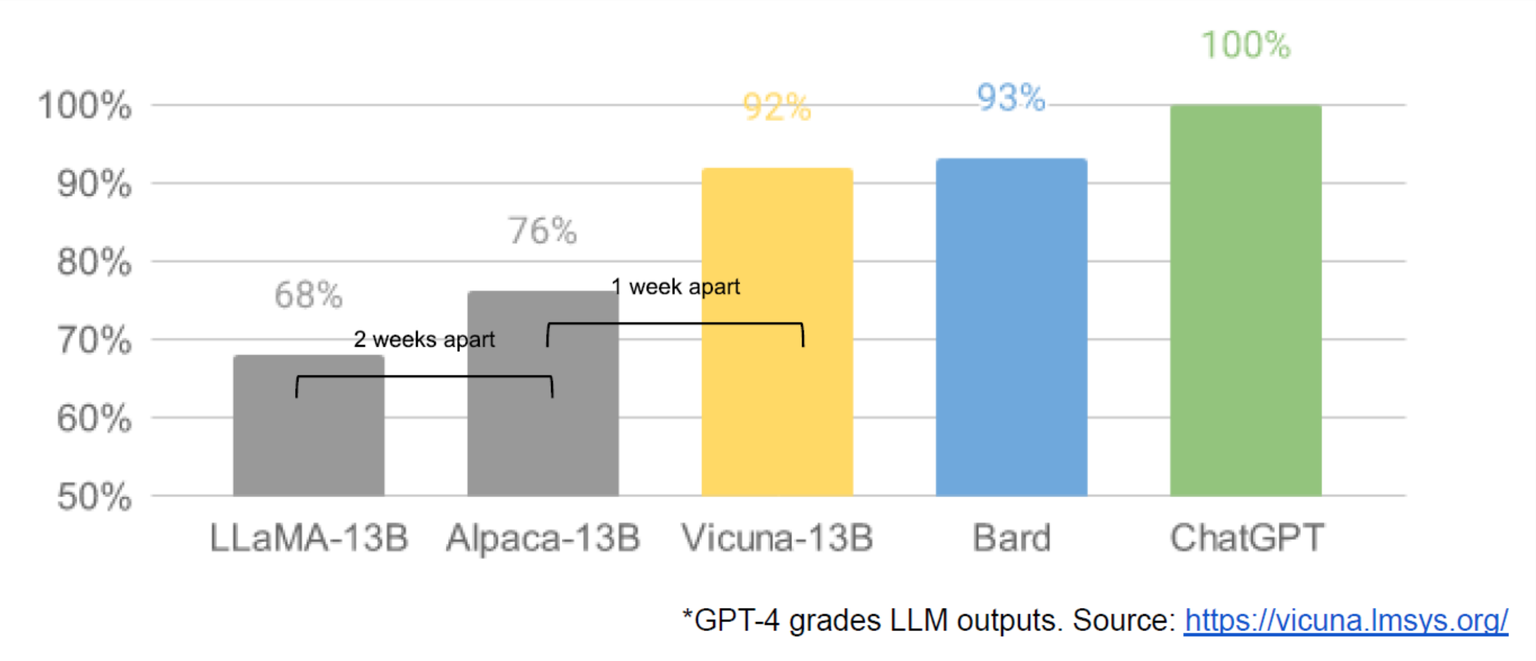
Open-Source LLMs

- MosaicML
- Falcon
- Open-Assistant
- Llama
- Qwen
- Mistral
- Phi
- DeepSeek



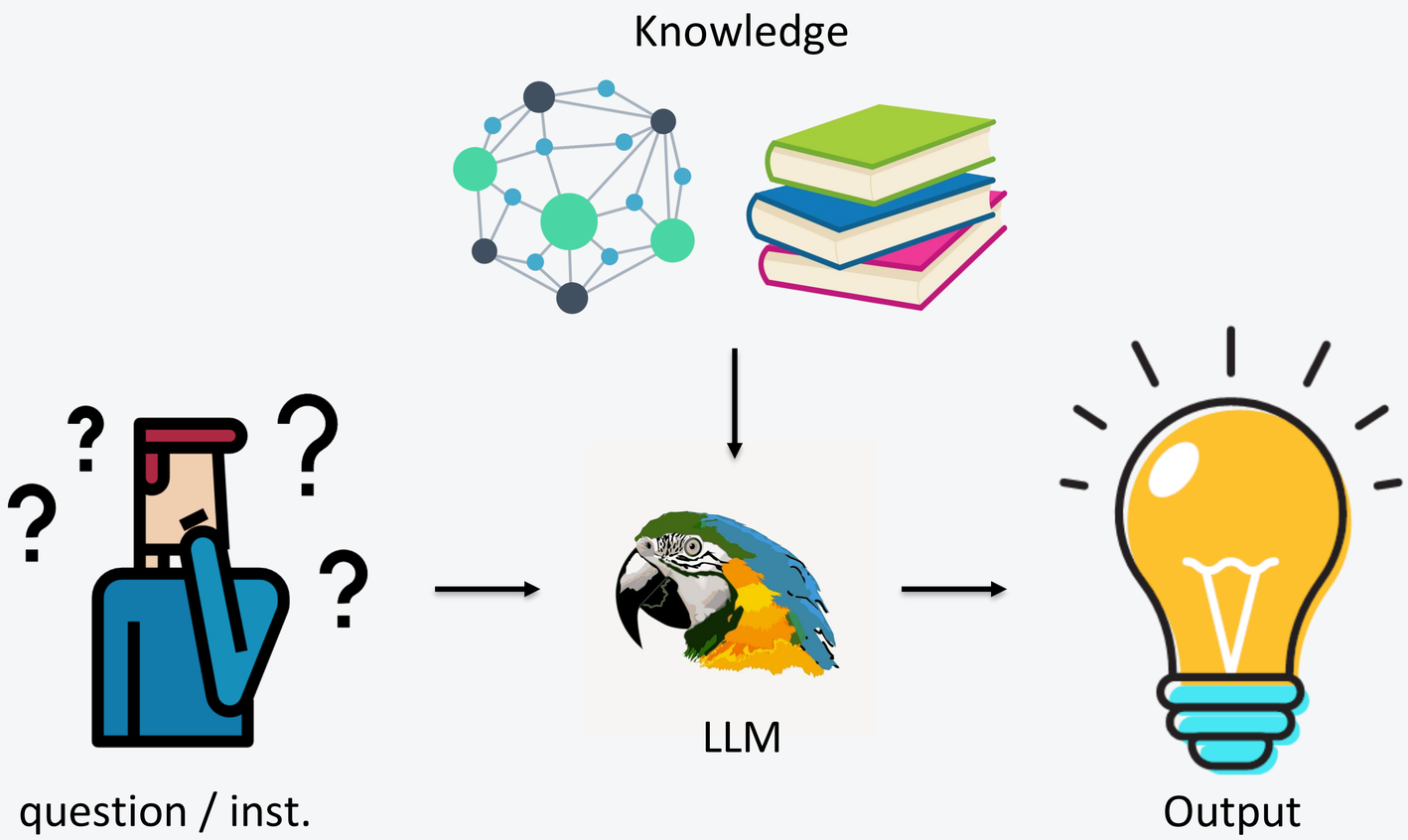
Fine-tuning

- **Deep Task Specialization**
  - **Adapts LLMs** to master niche domains (e.g., legal, medical, or brand-specific language).
  - **How?** By training on small, high-quality datasets → reshapes model weights for **domain expertise**.
    → *Results in unmatched accuracy for targeted use cases.*
- **Resource-Intensive but Powerful**
  - **Requires:**
    - Curated task-specific data (100s–1000s of examples).
    - Computational power (GPUs/TPUs) & technical expertise.
  - **Trade-off:** Higher upfront effort than RAG, but delivers **self-contained, optimized models** needing no live retrieval.
  - **Key Takeaway:** Fine-tuning creates purpose-built AI experts but demands significant investment.



Retrieval Augmented Generation (RAG)

- **Retrieval-First Architecture**
  - **Step 1:** Retrieve relevant data from trusted sources (e.g., databases, documents).
  - **Step 2:** Augment LLM prompts with this context.
    → *Grounds responses in real-time, domain-specific knowledge.*
- **Mitigates Hallucinations & Outdated Info**
  - Uses live external data instead of relying solely on pre-trained model knowledge.
  - Reduces factual errors and keeps outputs current/accurate.
- **Cost-Efficient & Adaptable**
  - No retraining needed for new data—update the knowledge base instantly.
  - Works with any LLM (e.g., GPT, LLaMA) for flexible deployment.

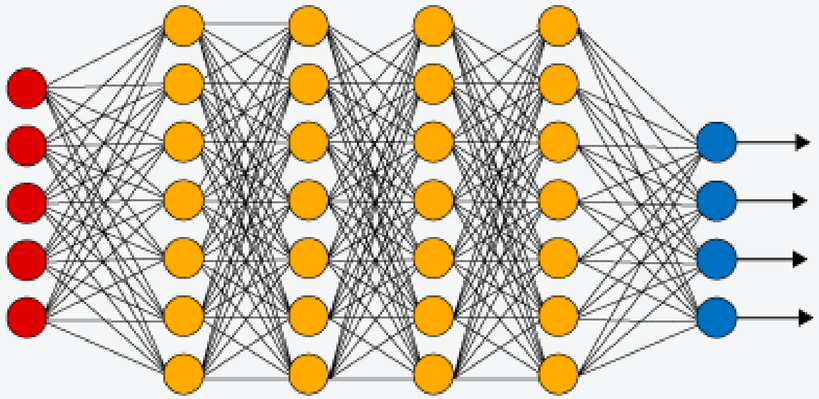


Shared Vision

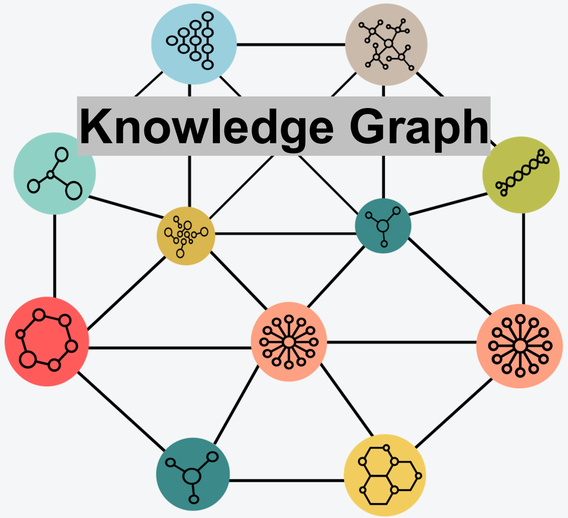
Best of both Worlds - Neuro-Symbolic AI

Integration of Knowledge Graphs with Machine Learning Models

Machine Intelligence



Background Knowledge



Human Intelligence

Knowledge authoring, curation, validation

eccenca
mastering complexity

Shared Vision

KG-LLM Integration Opportunities

Knowledge Graphs

- Represent a base of validated, trustworthy knowledge
- Help organize and integrate enterprise knowledge from various sources
- Provide input for enterprise and domain-specific training and fine-tuning of LLMs

Large Language Models

- Help curating knowledge in the KG by suggesting and recommending
- Create mappings, queries etc. for the KG
- Can become a frontend for human interaction with the KG

eccenca
mastering complexity

Pratical Application

eccenca
mastering complexity

Challenges

Navigating Trade-offs in LLM Customization

Fine-tuning

- **Resource Intensity**
  - Demands heavy computation (GPU/TPU), time, and technical expertise
  - Requires large *curated datasets* → costly/data-scarce domains struggle
- **Knowledge Rigidity**
  - Model "freezes" post-training → fails with new trends/updates
  - Retraining needed for adaptation → operational bottlenecks

RAG

- **Retrieval Dependency**
  - Output quality = f(knowledge base quality + retrieval accuracy)
  - Hallucinations persist if retrieval fails or sources are outdated
- **Latency & Complexity**
  - Real-time retrieval → slower responses than pure LLMs
  - Synchronizing vector DBs/APIs adds architectural overhead
  - Size of the Context

eccenca
mastering complexity

Any Questions?
Get in Touch!

eccenca
mastering complexity

eccenca GmbH
Hainstraße 8
D-04109 Leipzig
Germany

Try for free our Commnunity Edition Sandbox!

https://eccenca.my

+49 341 2650 8028

info@eccenca.com
https://eccenca.com