Fraunhofer FIT

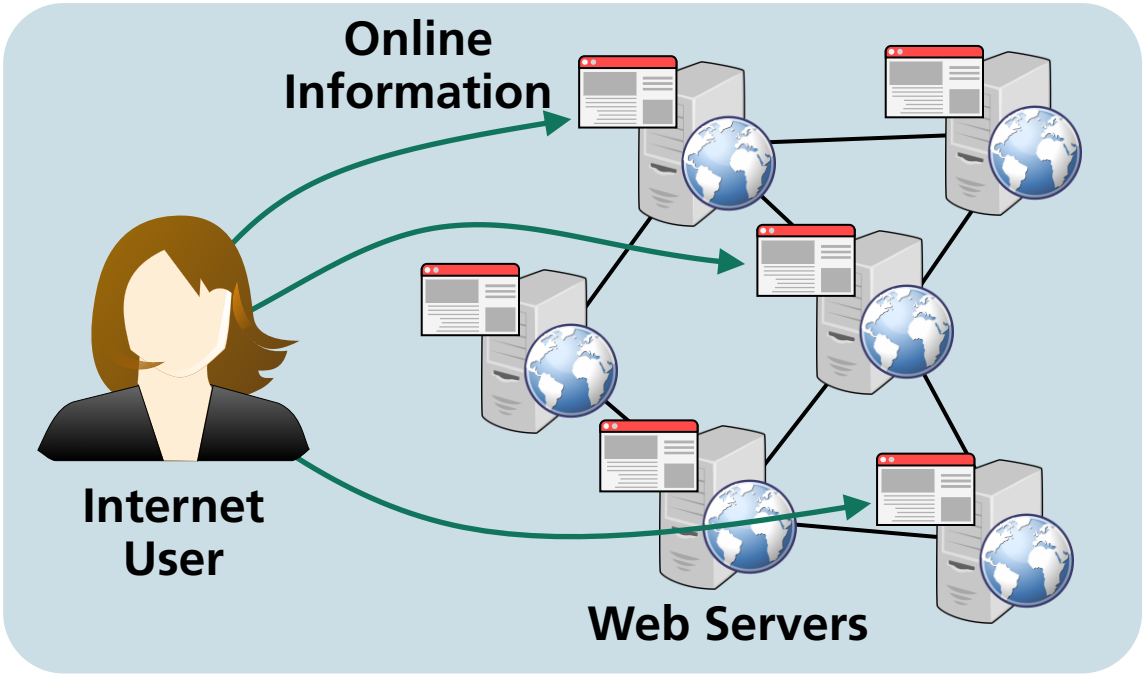
Fraunhofer Institute for Applied Information Technology FIT

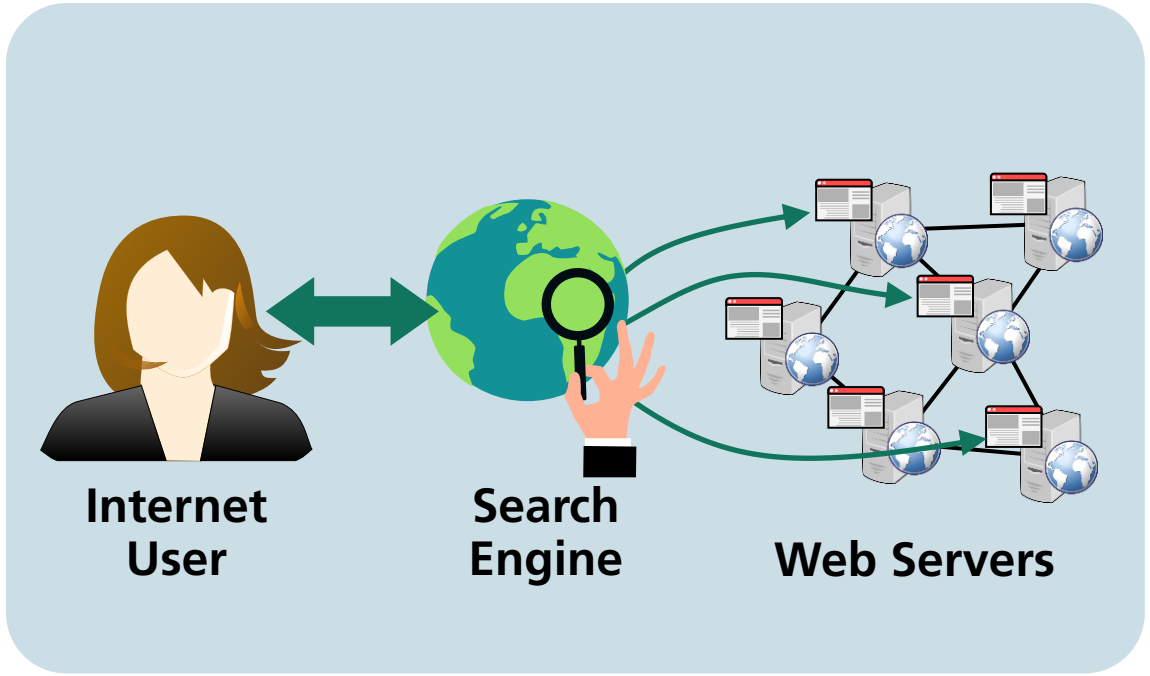
Roman Matzutt, Avikarsha Mandal

Towards a Reliable Web of Knowledge

Revolution of the Early Internet

From a User's Perspective





Early Internet

- Fragmented, direct access to information
- User has to know where to look

Search Engines

- Maintain central search index / interface
- User can query information

Recent Developments with LLMs

Why the Previous History Lesson?

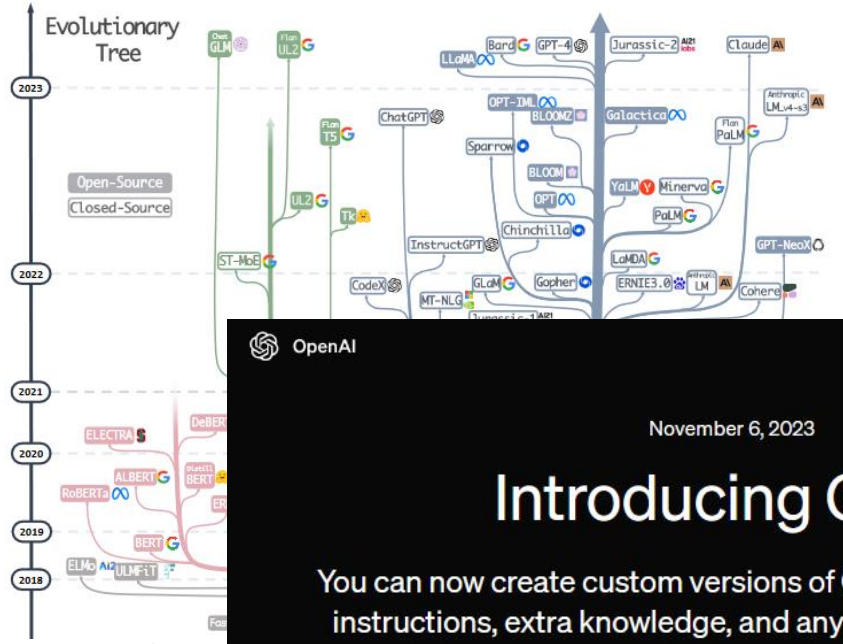
We already have a serverely fragmented LLM market!

- Both regarding the available models…
- …and services!

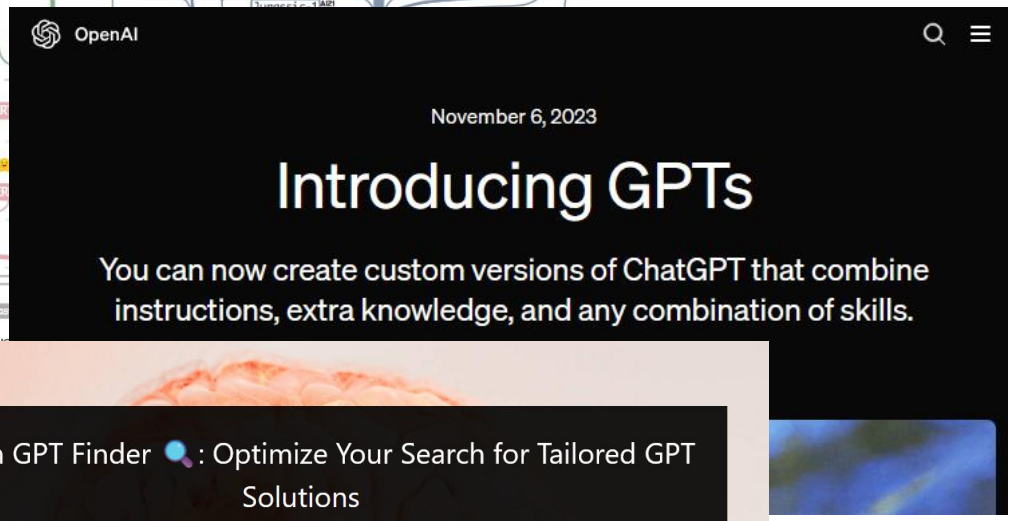
This leads to undesirable customer experience:

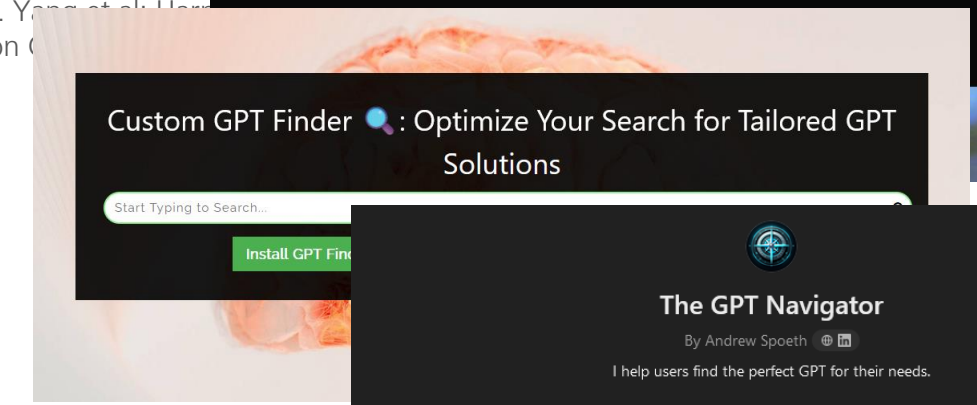
- It becomes increasingly hard to traverse the „LLM Jungle“
- Do users now have to know the best LLM for their use case again?

In the worst case, we may revert to Early Internet-style browsing

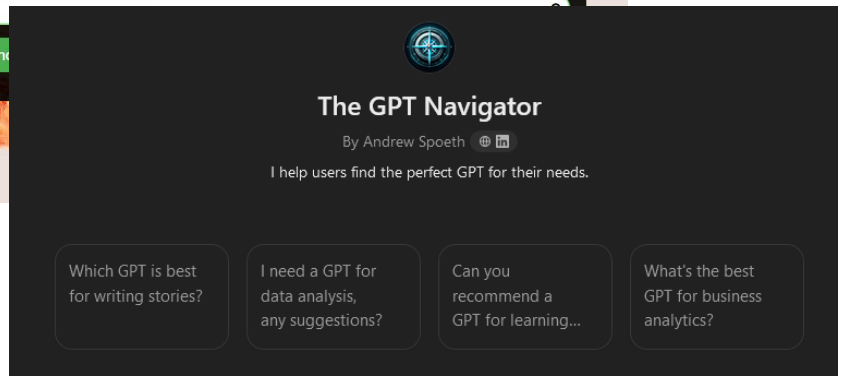


J. Yang et al. Harn… on …





https://gptfinder.co/



https://chatgpt.com/g/g-RizxV7UDv-the-gpt-navigator

Fraunhofer FIT

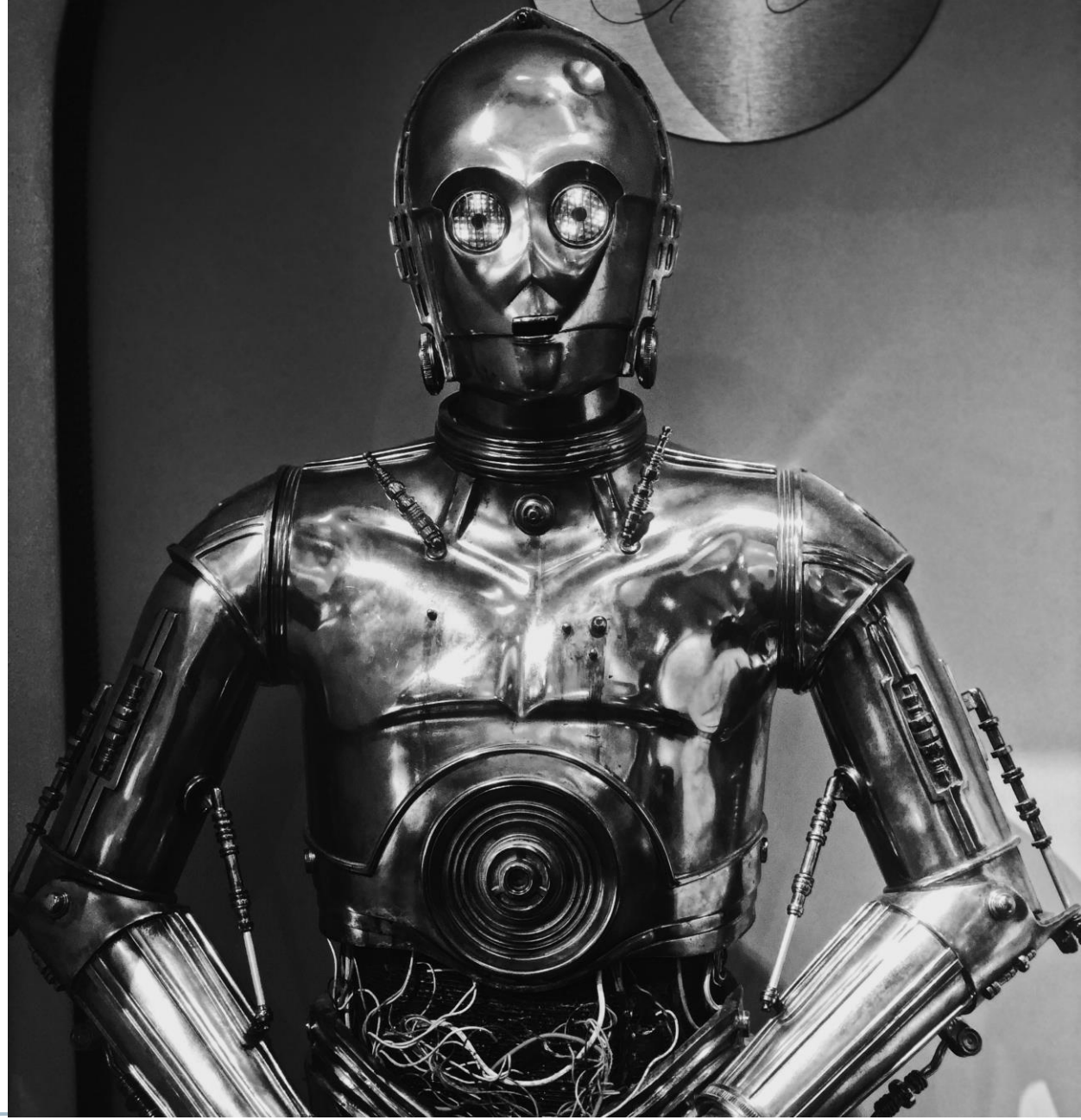
Toward a Web of Knowledge

Desirable User Experience

Instead, offer LLMs as integral part of the user interface for information retrieval

- Already offered by search engine / tech providers
  - Microsoft Copilot, Google Gemini, Apple Intelligence
- Also, development toward on-the-fly information retrieval
  - E.g., OpenAI ChatGPT-4o

But: Challenges of LLMs get aggravated in this scenario



Recap: Challenges of LLMs

Aggravated in Single-Point-of-Failure Deployment

Response Accuracy

- Statistical responses without reasoning
- Hallucination of “likely” information
- Identification of tasks that require calling external functionality (e.g., computations)

Data Availability

- Vast amounts of training data required
- Privacy and copyright issues
- Tendency cover knowledge in training data (fine tuning)

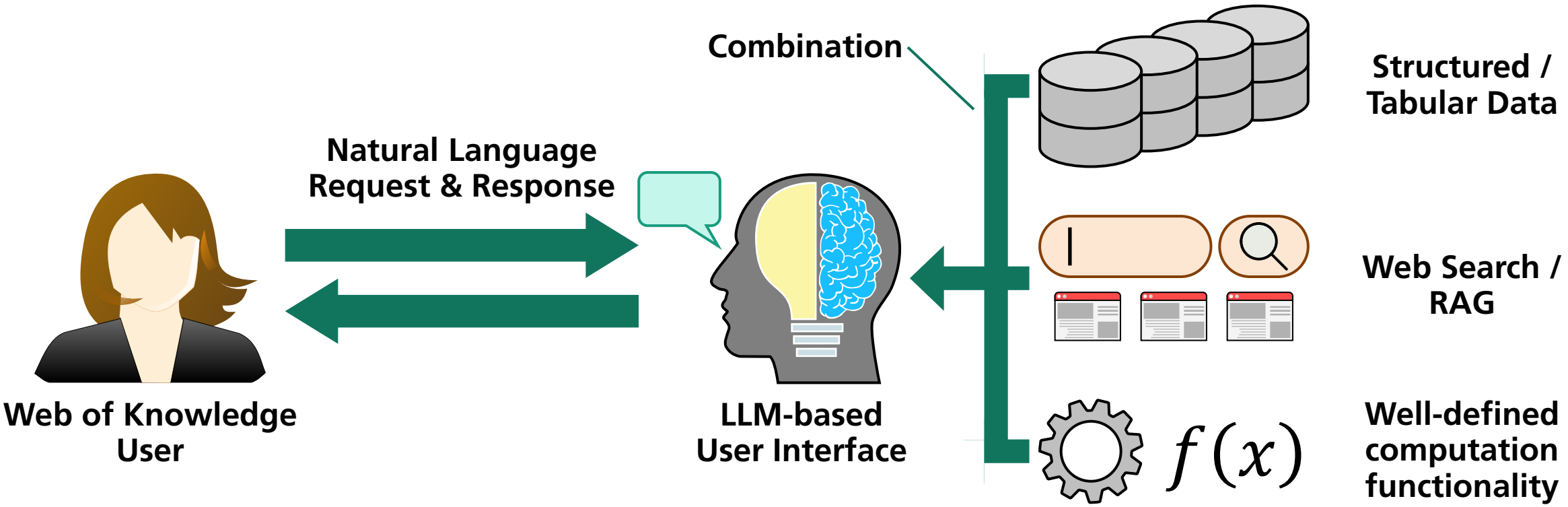
Why do LLMs have to encode knowledge?
→ Focus on the true strengths of LLMs!

Get inspired by RAG to rethink knowledge retrieval

LLMs in the Center of the Web of Knowledge

Return to Strengths

Main idea: Focus on what LLMs provide at the core: Natural-language user interfaces

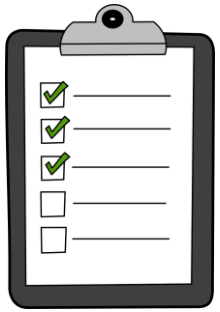


Challenges

Or: Research Opportunities

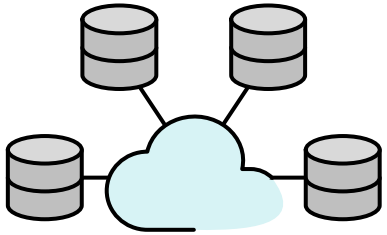
Standardization

- Requirement for APIs
- Data ontologies and derivation of structured data
- Access structures and permission management



Linking Data Sources

- Find and combine data from multiple sources
- Data heterogeneity
- Different data owners
- Example: “Should I go by car or train?”



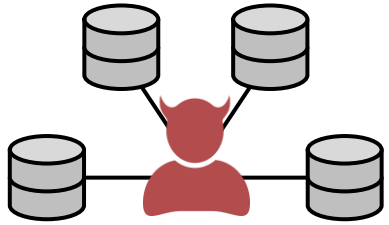
Data Provenance

- Provide means to validate received information
- Trade off: response simplicity and verifiability
- Enable dynamic investigation with optional auxiliary data (quick info vs. research)

[citation needed]

Over-Centralization

- Service providers not guaranteed to stay neutral
- May become person in the middle
- Withhold crucial information
- Alter retrieved data

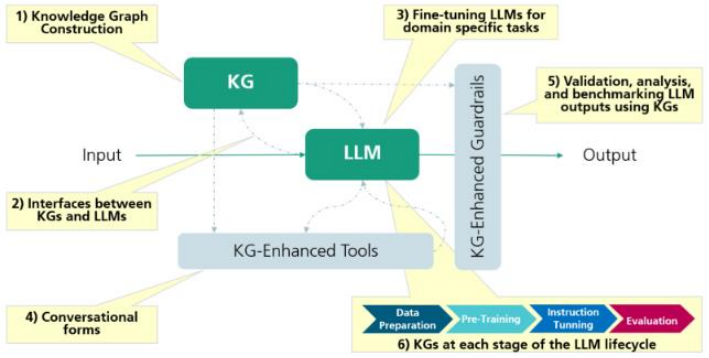


Next Steps

Related Activities at FIT

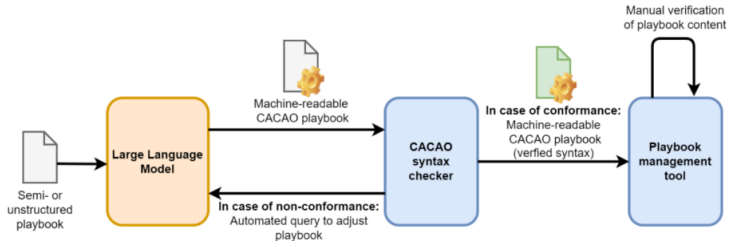
Knowledge-Enhanced LLMs

- Enhance reasoning of LLMs by introducing a structured knowledge store
- Realized by constructing knowledge graphs that is queried by the LLM
- Translate natural language queries via Text-to-SPARQL technology



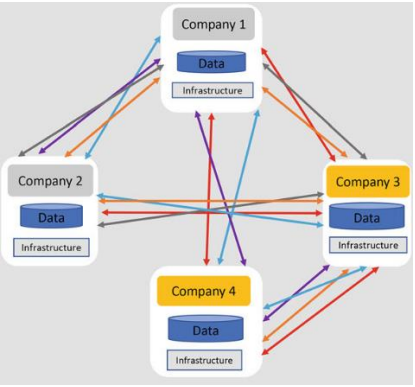
Structured Process Descriptions

- Translate natural-language descriptions of workflows into machine-readable, standardized form for LLM compatibility
- Current focus: Shareable cybersecurity playbooks: Bidirectional translation
- Envisioned application to general process descriptions



Data Spaces

- Federated architectures to facilitate self-sovereign data exchanges
- Provide infrastructure for sharing industrial data that is hosted at the respective origin companies
- Example initiatives: IDS and GAIA-X



H. Tardieu: Role of Gaia-X in the European Data Space Ecosystem

Conclusion

Summary

Web of Knowledge (WoK):

- Understand LLMs as excellent interfaces based on human natural language instead of knowledge encoders

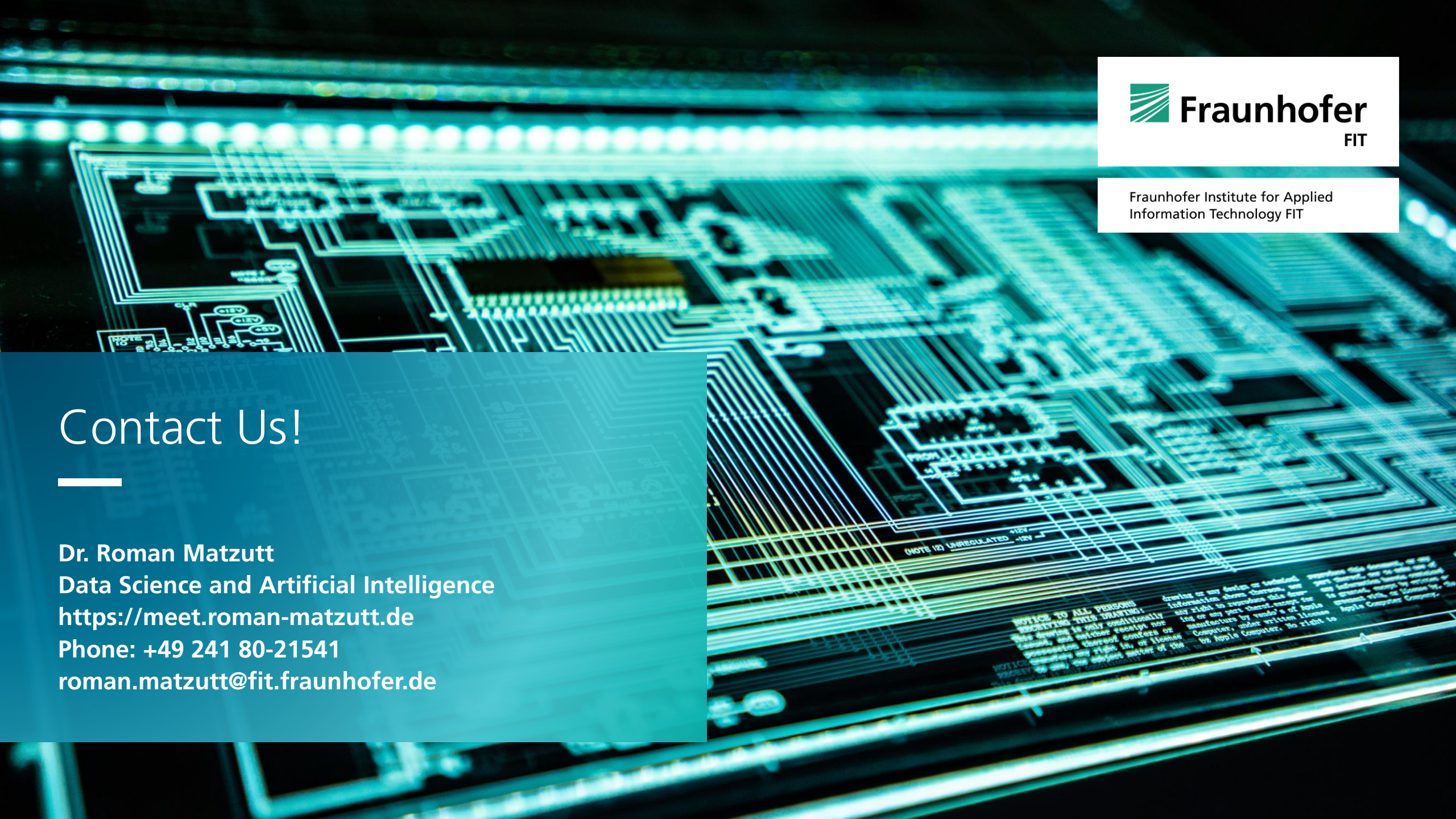
Exchange old for new challenges

- Potential to reduce hallucination, data ownership issues
- But: Challenges regarding standardization, linkability, verifiability

Ongoing initial efforts

- Connecting knowledge graphs and standardized processes to LLMs

Interested? Let's cook together!



Fraunhofer
FIT

Fraunhofer Institute for Applied Information Technology FIT

Contact Us!

Dr. Roman Matzutt
Data Science and Artificial Intelligence
https://meet.roman-matzutt.de
Phone: +49 241 80-21541
roman.matzutt@fit.fraunhofer.de