


Data Science and AI for Medicine Training School

TRAINING: Introduction to Large Language Models

SPEAKER: Lars Hilgers



SACHSEN Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.



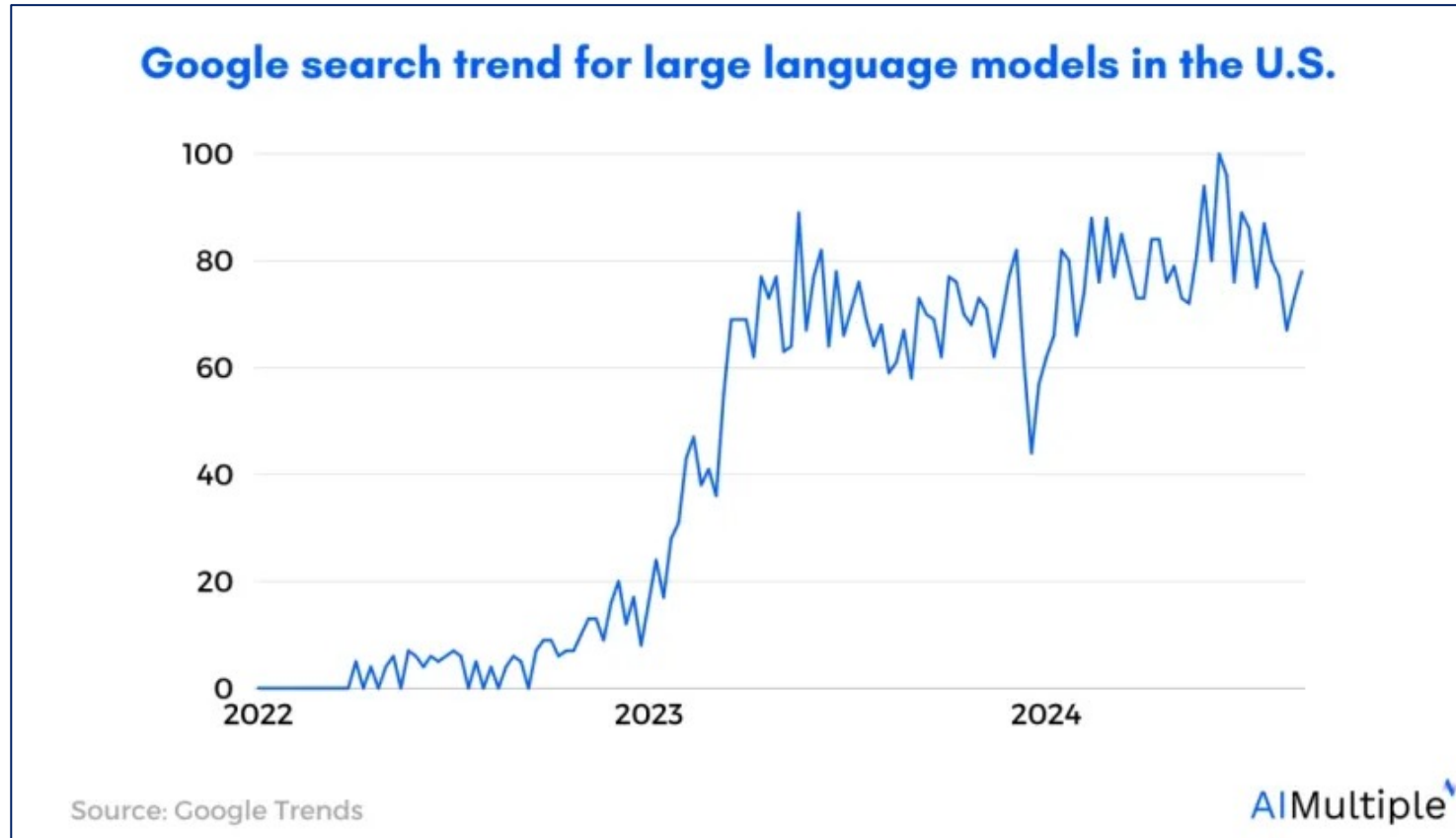
Come2Data
Kompetenzzentrum für
interdisziplinäre Datenwissenschaften

Data Science and AI for Medical Training School
Training: Introduction to Large Language Models

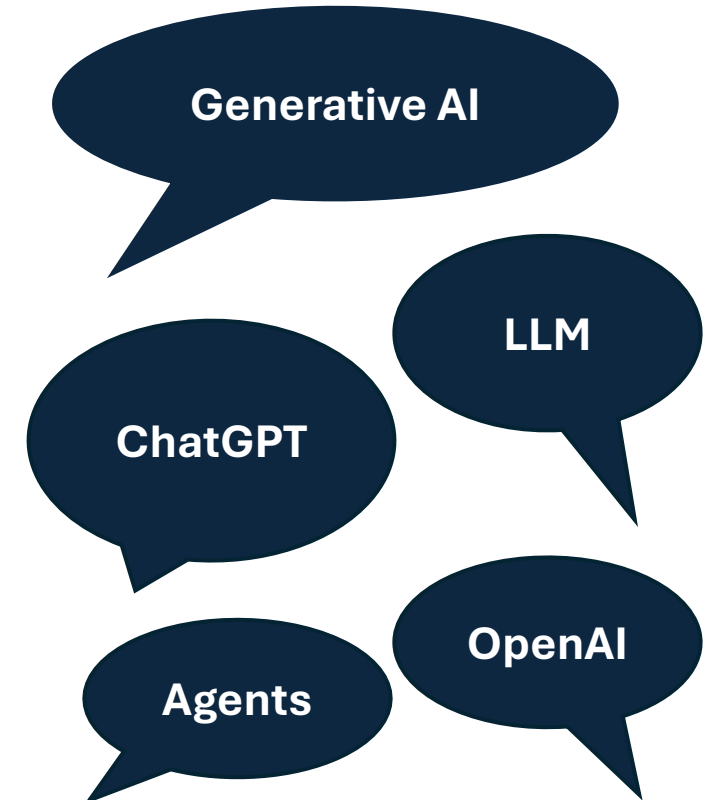
Slide 1



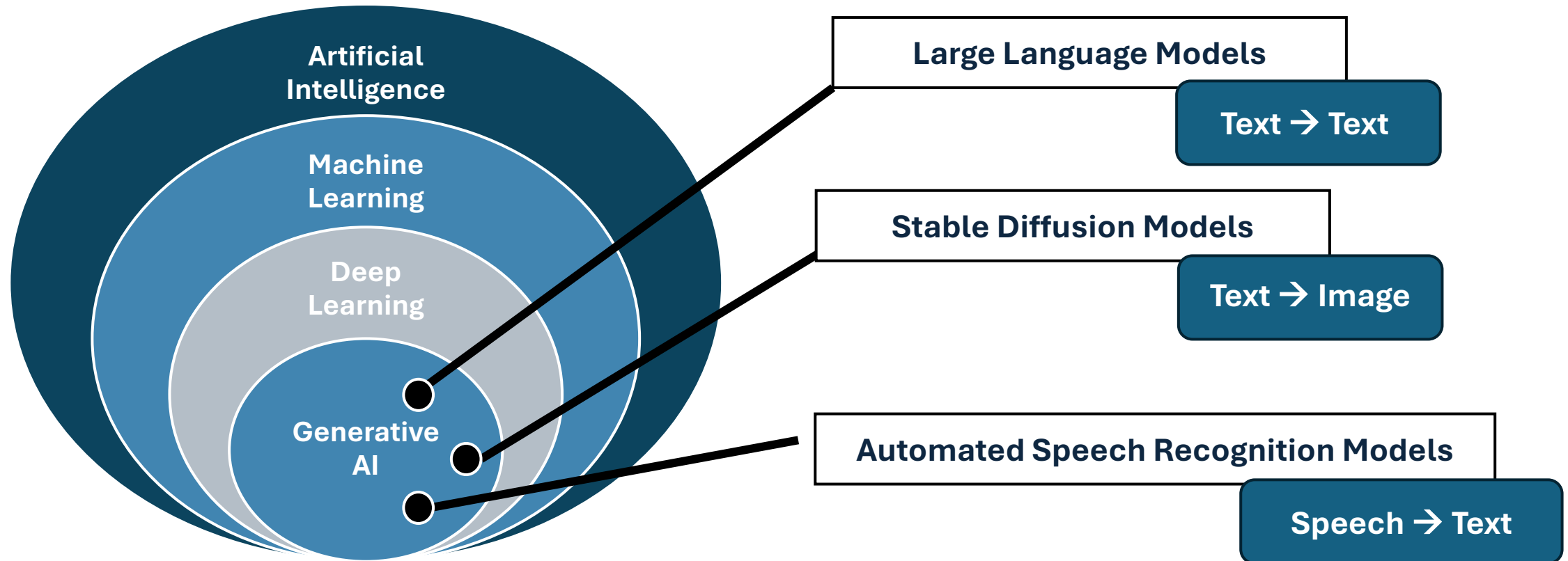
The Rise of Large Language Models



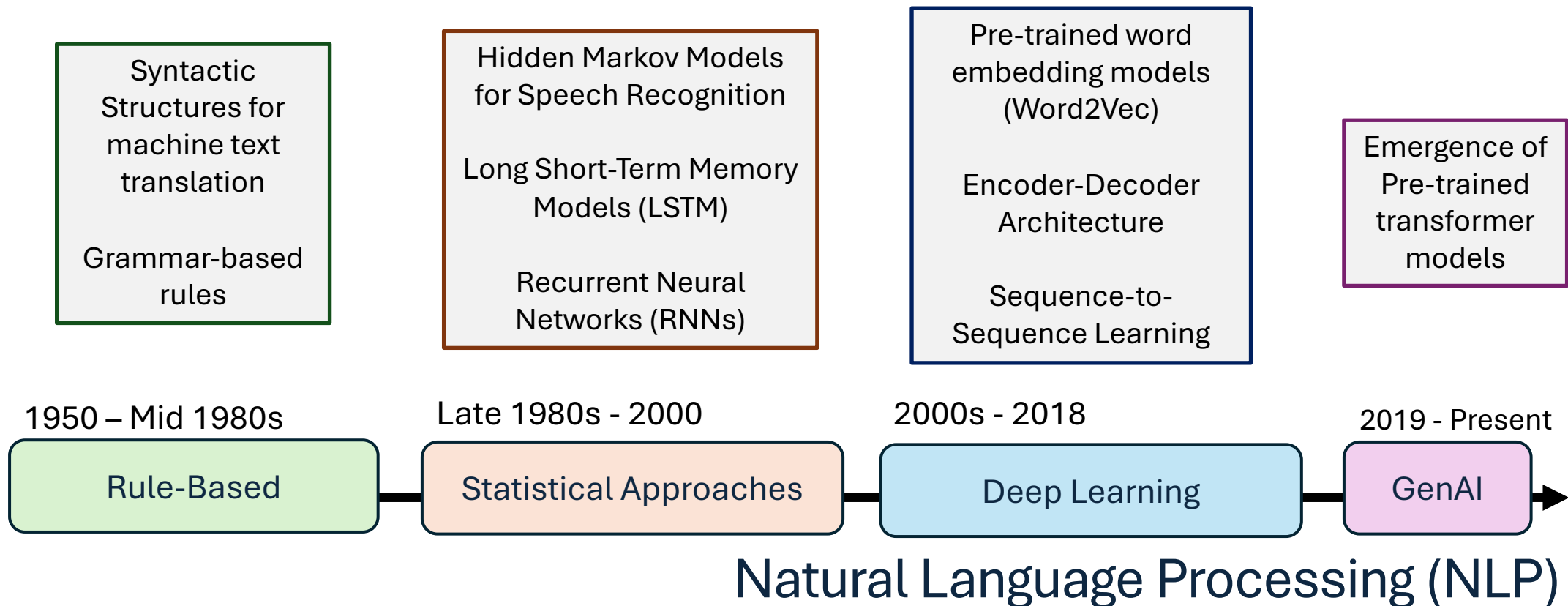
Source: <https://research.aimultiple.com/future-of-large-language-models/>



Where do Large Language Models fit into the AI Space

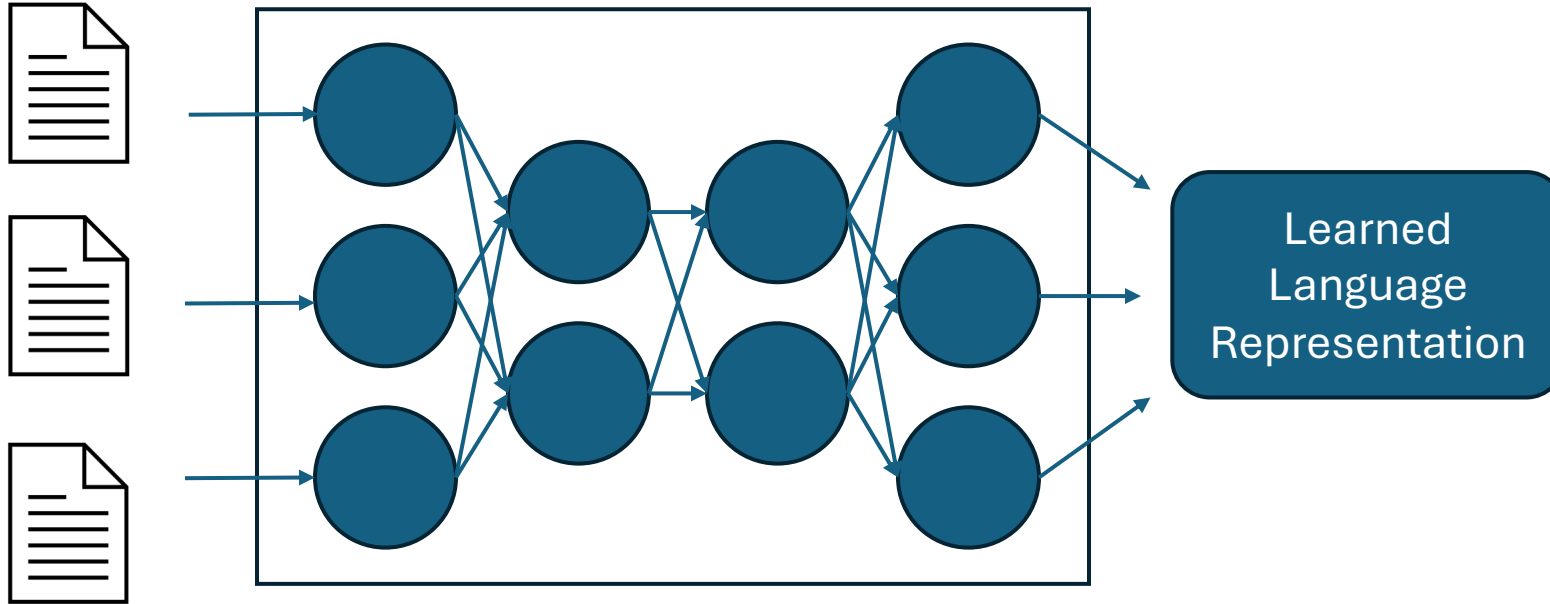


The progression towards Large Language Models

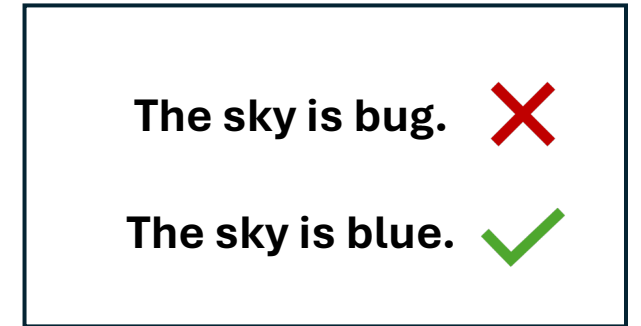


<https://blog.dataiku.com/nlp-metamorphosis>

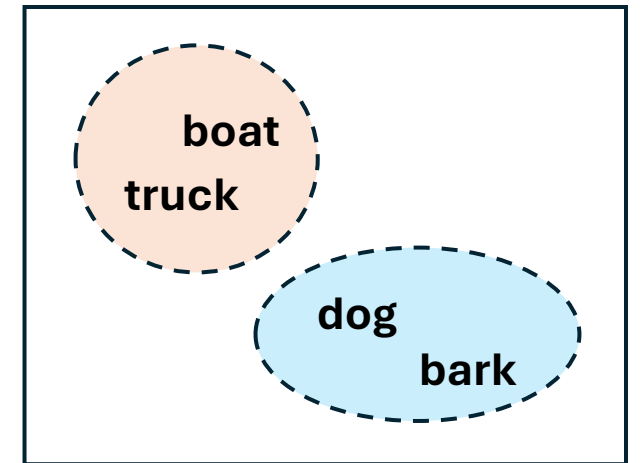
How Natural Language is Modelled within LLMs



Transformer Architecture with Attention Mechanism

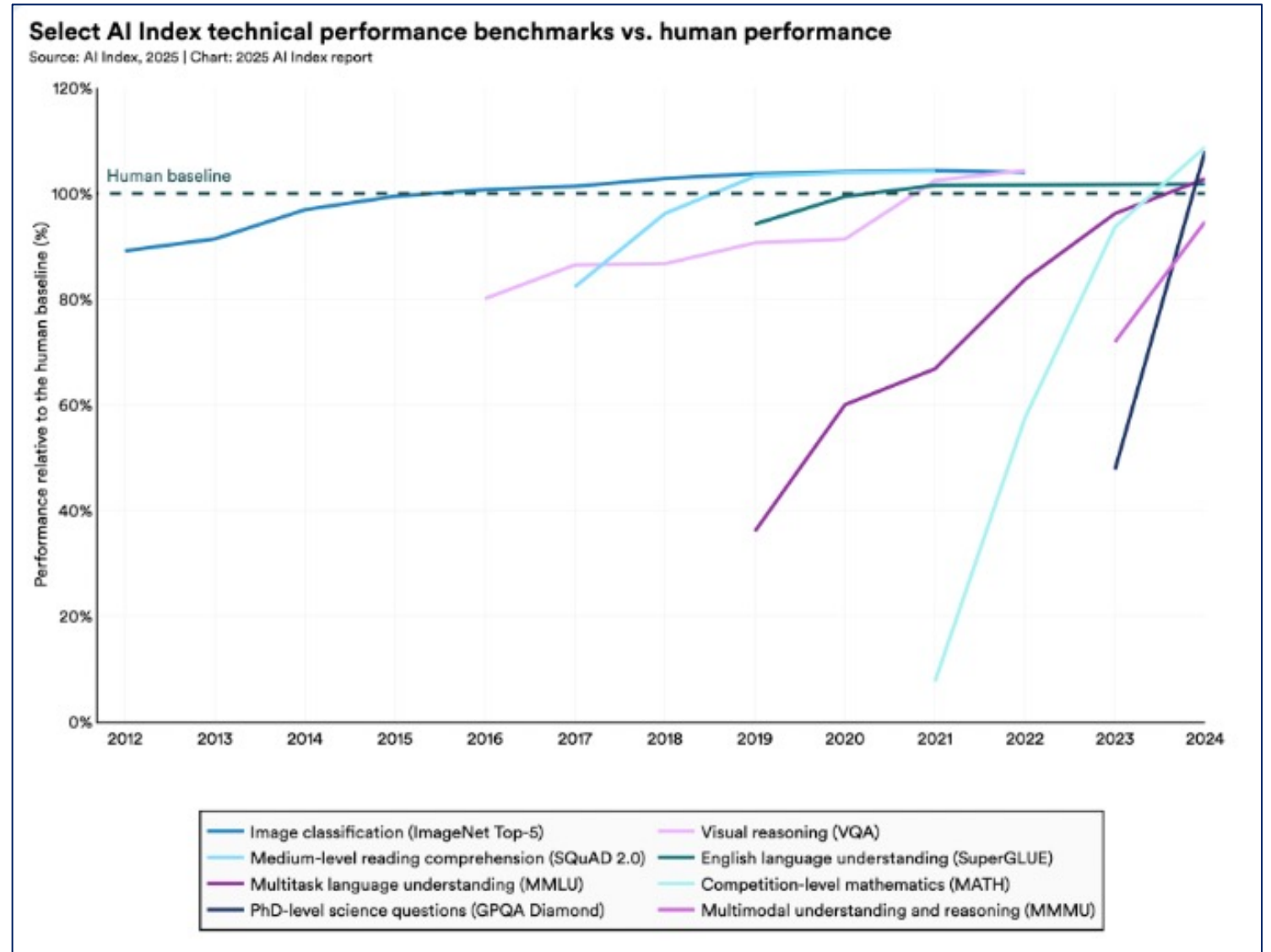


Textual Patterns



Semantic Patterns

Zero-Shot LLMs are highly performative



<https://hai.stanford.edu/ai-index/2025-ai-index-report>

LLMs are well suited for Medical Tasks

Properties of Medical Applications

Unstructured Data
Multimodal Data
Personalized Solutions
Large amounts of Knowledge
Complex Reasoning

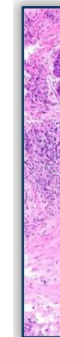
Electronic Health Records (EHR)

Radiology images

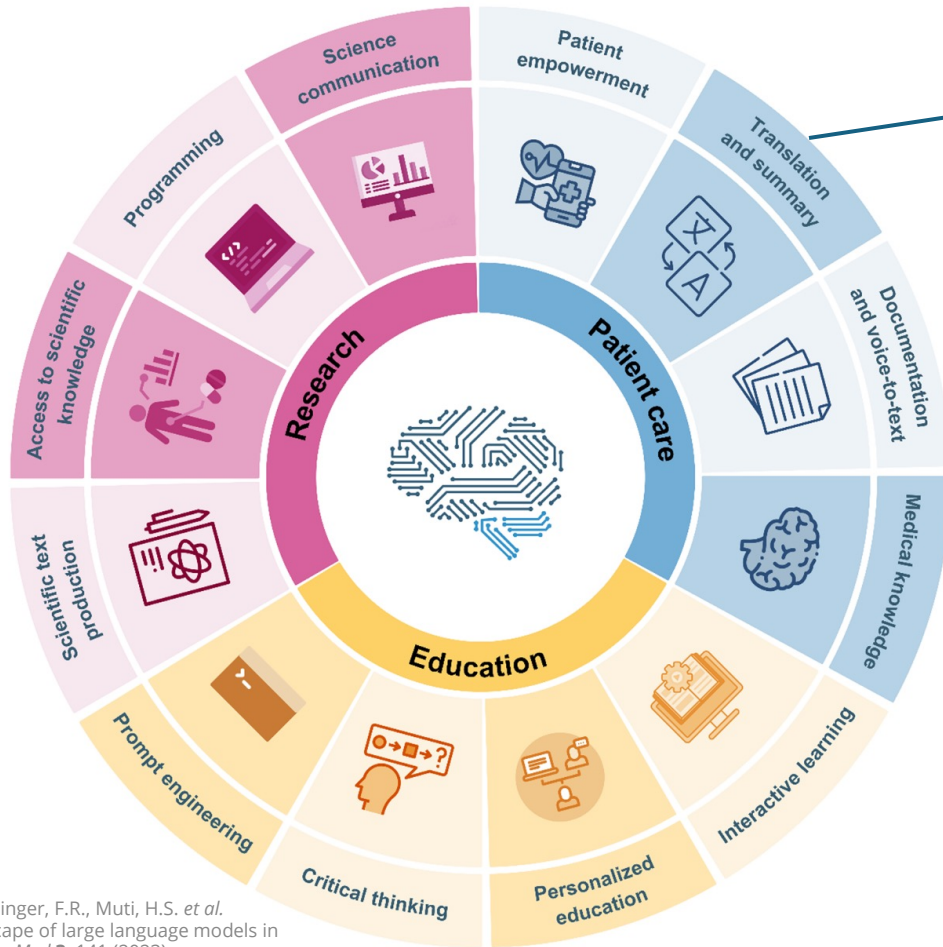
Intervention videos

Histology images

Discharge letters



LLMs show potential for diverse clinical applications



nature medicine

Article <https://doi.org/10.1038/s41591-024-02855-5>

Adapted large language models can outperform medical experts in clinical text summarization

Received: 23 October 2023
 Accepted: 2 February 2024
 Published online: 27 February 2024

Check for updates

Dave Van Veen^{1,2}, Cara Van Uden^{2,3}, Louis Blankemeier^{1,2}, Jean-Benoit Delbrouck², Asad Aali⁴, Christian Bluethgen^{2,5}, Anuj Pareek^{2,6}, Malgorzata Polacin⁵, Eduardo Pontes Reis^{2,7}, Anna Seehofnerova^{8,9}, Nidhi Rohatgi^{8,10}, Poonam Hosamani⁵, William Collins⁵, Neera Ahuja⁵, Curtis P. Langlotz^{2,8,9,11}, Jason Hom⁵, Sergios Gatidis^{2,9}, John Paul¹ & Akshay S. Chaudhari^{2,9,11,12}

npj | digital medicine

Privacy-preserving large language models for structured medical information retrieval

[Isabella Catharina Wiest](#), [Dyke Ferber](#), [Jiefu Zhu](#), [Marko van Treeck](#), [Sonja K. Meyer](#), [Radhika Juglan](#), [Zunamys I. Carrero](#), [Daniel Paech](#), [Jens Kleesiek](#), [Matthias P. Ebert](#), [Daniel Truhn](#) & [Jakob Nikolas Kather](#) ✉

Clusmann, J., Kolbinger, F.R., Muti, H.S. *et al.*
 The future landscape of large language models in medicine. *Commun Med* 3, 141 (2023).
<https://doi.org/10.1038/s43856-023-00370-1>



Come2Data
 Kompetenzzentrum für
 interdisziplinäre Datenwissenschaften

Data Science and AI for Medical Training School
 Training: Introduction to Large Language Models

Slide 8



Limitations of LLMs and their effect in Clinical Settings

HALLUCINATIONS OR
FACTUAL INACCURACIES



*Need for
Validated Outputs
for High-Risk
Applications*

DATA PRIVACY



*Patient
Privacy and
Regulatory
Concerns*

TRAINING CUTOFFS
LIMITING KNOWLEDGE
AVAILABILITY



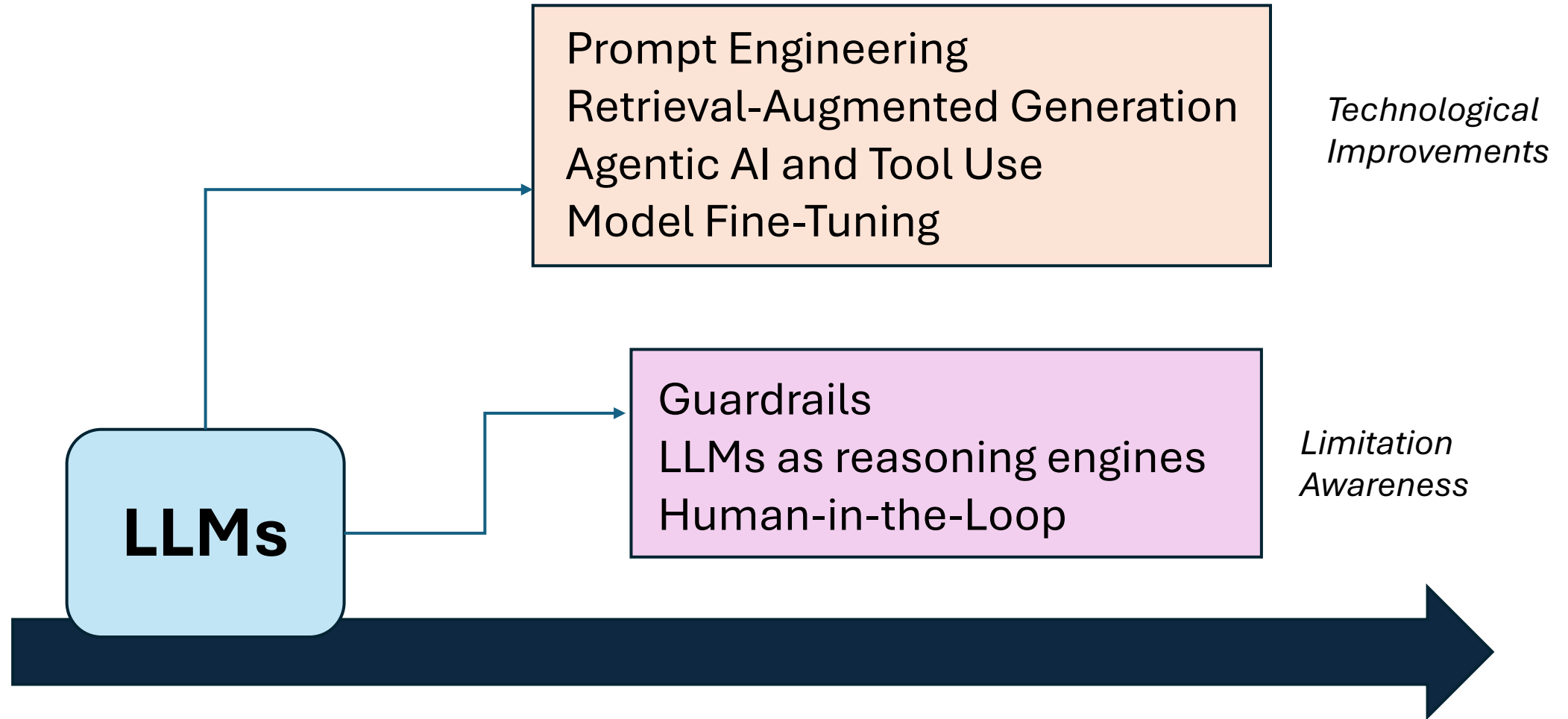
*Concordance
with updated
medical policies
and knowledge*

MODEL BIAS

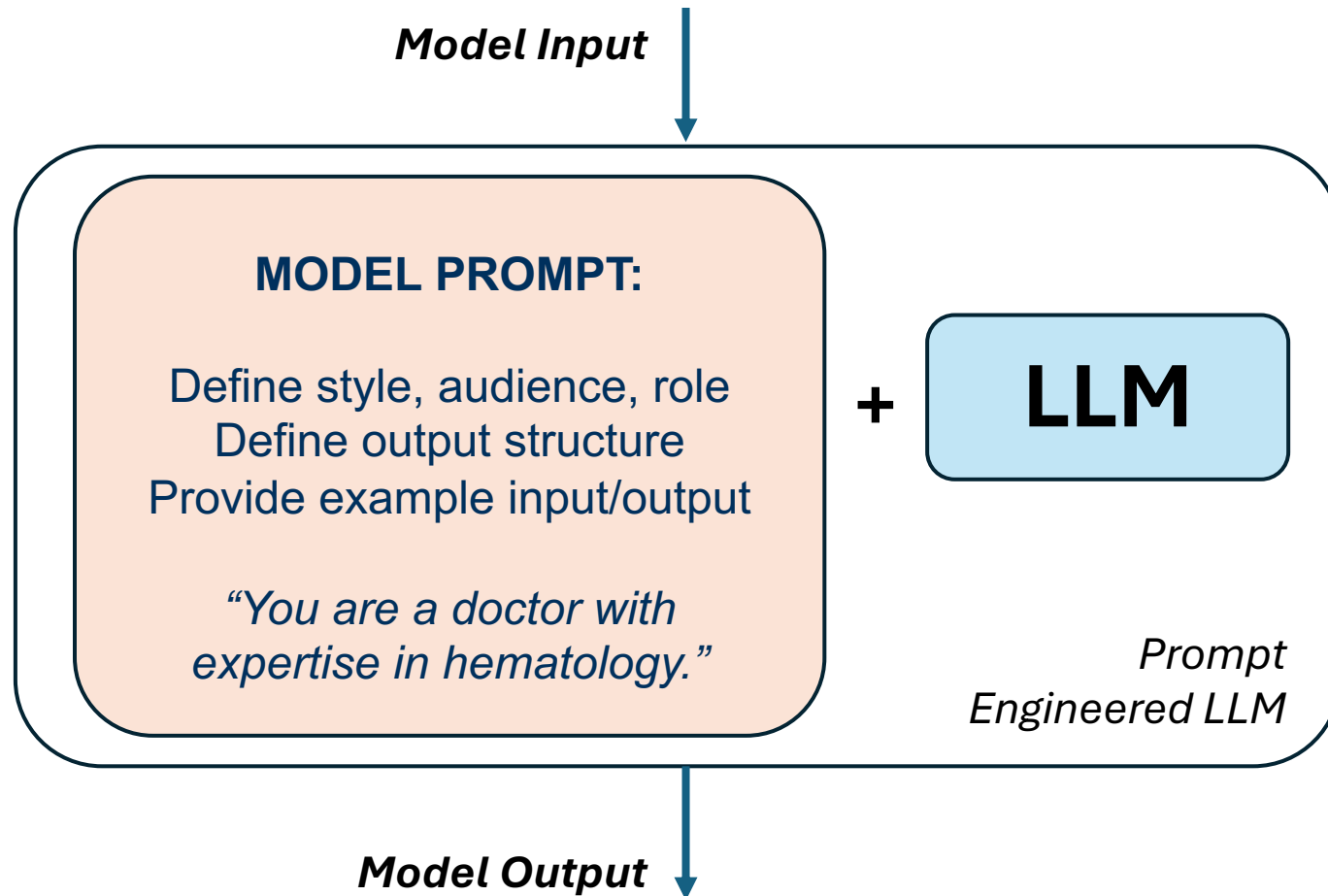


*Ethical
Concerns for
Patient
Fairness*

The field of LLMs is continuously evolving



Prompt Engineering



Zero-Shot LLMs:

Model hallucinations

Inconsistent outputs



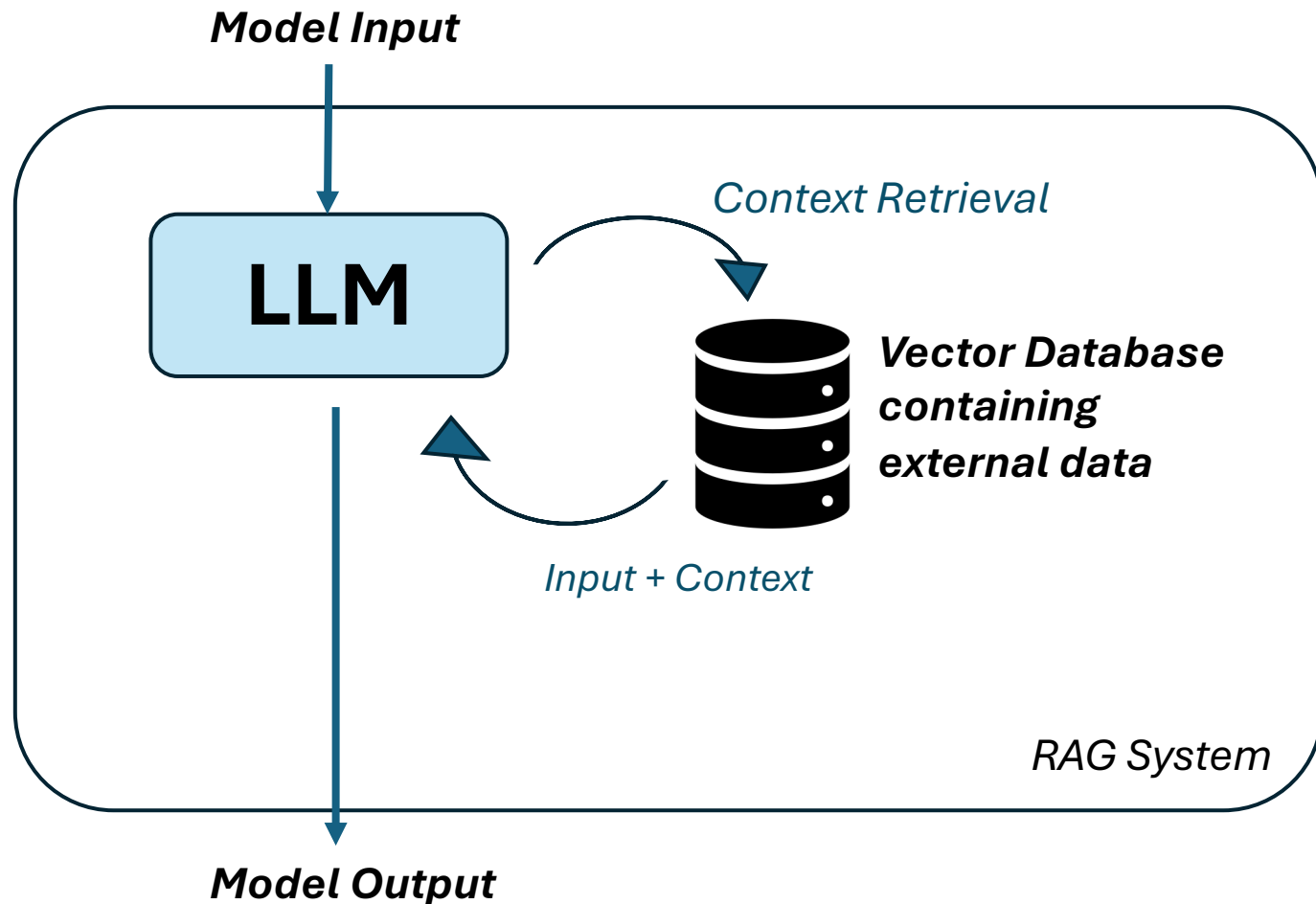
Prompt Engineering

*Improved
instruction following*

*Improved output
consistency*

Domain adaptation

Retrieval Augmented Generation (RAG)



Zero-Shot LLMs:

Factual Inaccuracies

Knowledge Cutoffs



RAG

Leverages LLM for reasoning

External Grounding

Multimodal data integration

Curation of trusted data sources

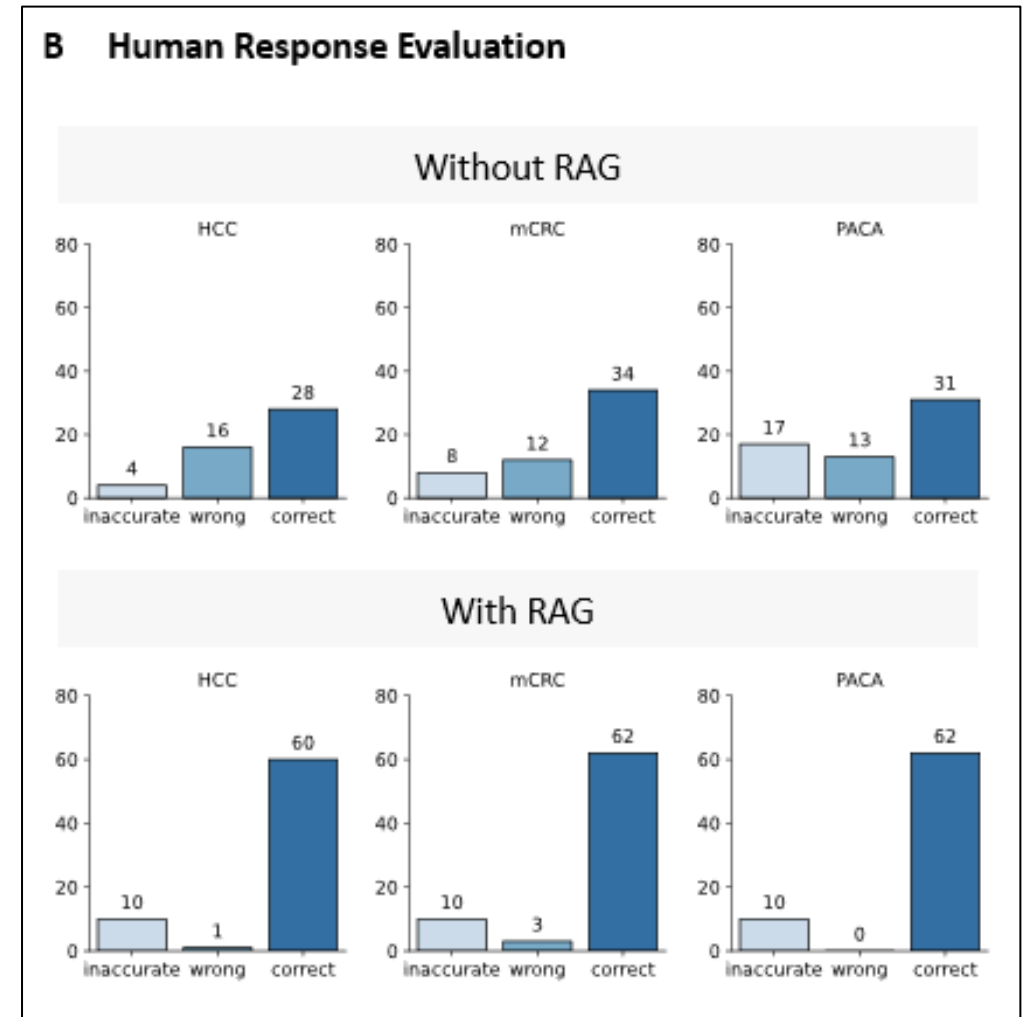
Retrieval-Augmented Generation (RAG) Clinical Example

NEJM AI
CASE STUDY

GPT-4 for Information Retrieval and Comparison of Medical Oncology Guidelines

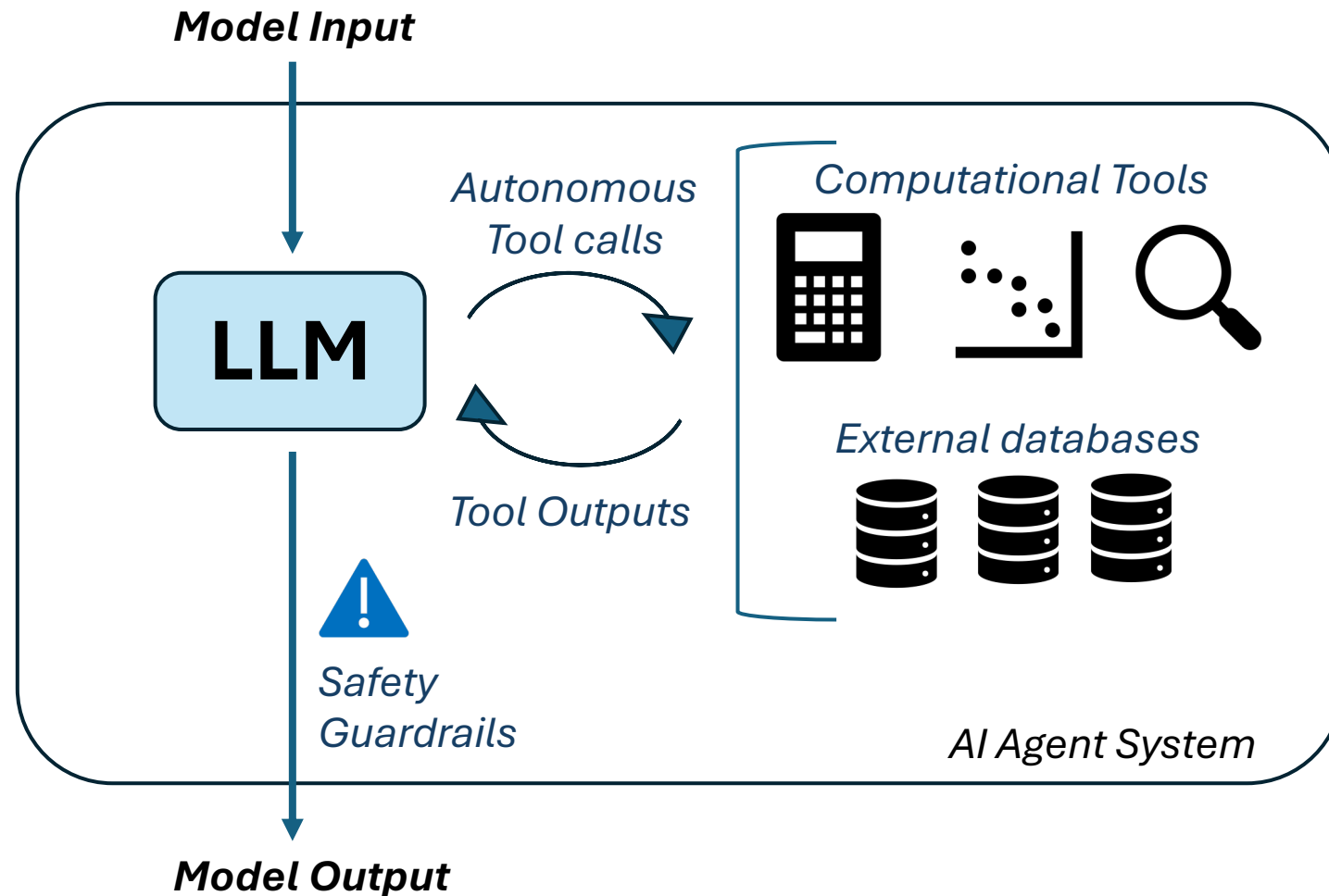
Authors: Dyke Ferber ^{ID}, Isabella C. Wiest, M.D., M.Sc. ^{ID}, Georg Wölflein, M.Sc. ^{ID}, Matthias P. Ebert, M.D. ^{ID}, Gernot Beutel, M.D. ^{ID}, Jan-Niklas Eckardt, M.D., M.Sc. ^{ID}, Daniel Truhn, M.D., M.Sc. ^{ID}, Christoph Springfeld, M.D., Ph.D. ^{ID}, Dirk Jäger, M.D. ^{ID}, and Jakob Nikolas Kather, M.D., M.Sc. ^{ID} [Author Info & Affiliations](#)

Published May 17, 2024 | NEJM AI 2024;1(6) | DOI: 10.1056/AIcs2300235 | VOL. 1 NO. 6 | Copyright © 2024



Ferber, D., Wiest, I. C., Wölflein, G., Ebert, M. P., Beutel, G., Eckardt, J. N., ... & Kather, J. N. (2024). GPT-4 for Information Retrieval and Comparison of Medical Oncology Guidelines. NEJM AI, Aics2300235.

Agentic AI



✓ Agentic AI

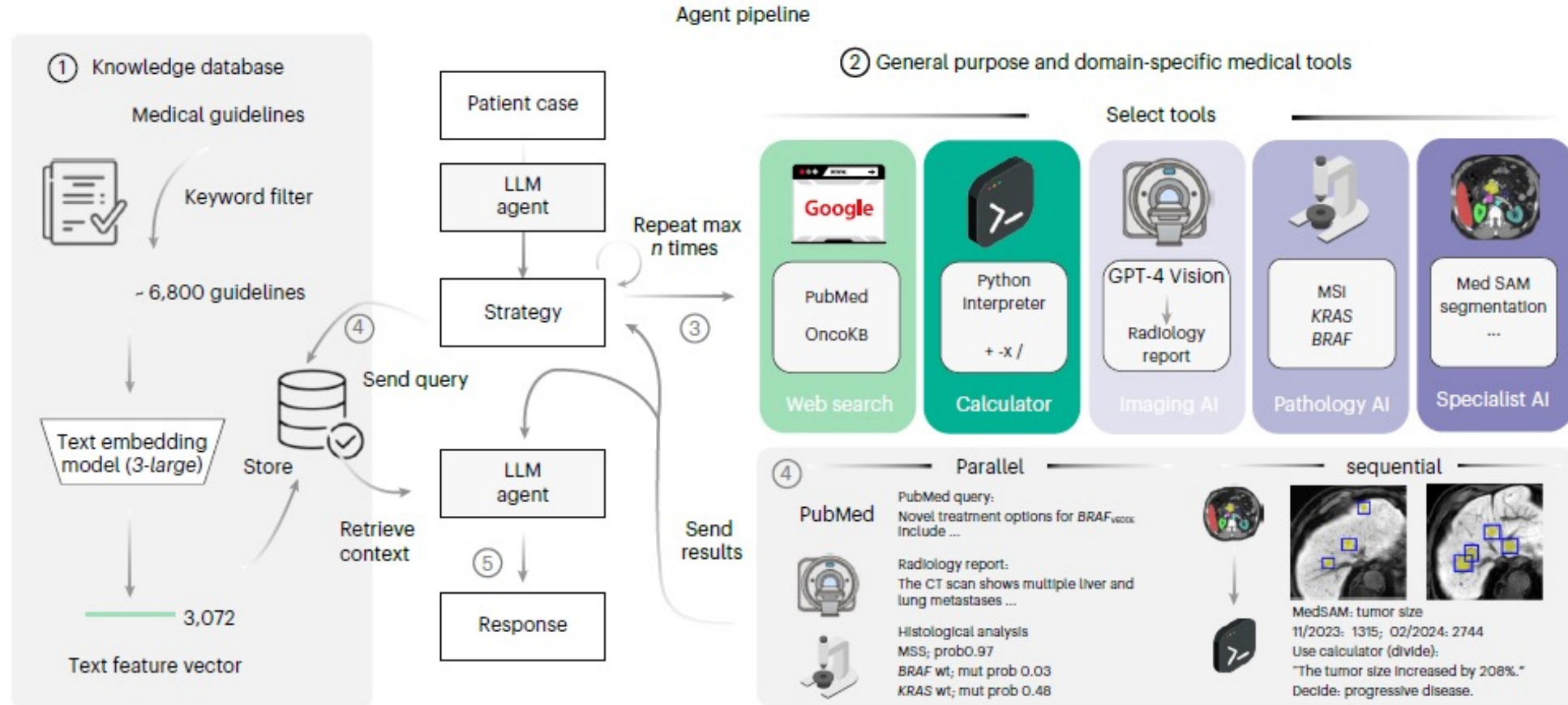
Leverages LLM reasoning capability for autonomous decision-making

Integrates multiple data modalities

Utilizes validated advanced computational methodologies

Implements guardrails for safer responses

Example of LLM Agents for Clinical Decision Support



Ferber, D., El Nahhas, O.S.M., Wölflein, G. et al. Development and validation of an autonomous artificial intelligence agent for clinical decision-making in oncology. Nat Cancer 6, 1337–1349 (2025). <https://doi-org.ezproxy.medma.uni-heidelberg.de/10.1038/s43018-025-00991-6>

LLMs are not perfect but they provide building blocks for robust medical applications



Guardrails for
response safety



Regulatory-
Compliant Systems



Trustworthy and Fair
Systems



Clinically-relevant
applications



Evaluation and
Validation of Systems

Hands on exercise

- Please save a copy of the Google Colab
- Setting up the LLM API Client:

```
OpenAI LLM Client Initialization

[3] # Please enter the OPENAI_API_KEY that we shared with you.

os.environ["OPENAI_API_KEY"] = getpass("Enter your API key: ")
client = OpenAI(base_url="https://api.kather.ai/v1", api_key=os.environ["OPENAI_API_KEY"])

... Enter your API key: 
```

API BASE: https://api.kather.ai/v1
Leipzig00: sk-...

- We recommend working in pairs 😊