



Von Daten zu Entscheidungen: KI verstehen, statt ihr zu glauben

Prof. Dr. Maximilian Moll

München, den 08. Juli 2026

Die KI trifft objektive Entscheidungen

Die KI trifft objektive Entscheidungen

Die KI trifft objektive Entscheidungen

Die KI trifft objektive Entscheidungen

Die KI trifft objektive Entscheidungen

Die KI trifft objektive Entscheidungen

- Leitung: Prof. Dr. Stefan Pickl, Prof. Dr. Maximilian Moll
- Interdisziplinäre Forschungsgruppe mit Schwerpunkt Entscheidungsunterstützung
- >15 PhD Doktoranden
- 3 Postdoktoranden als Forschungsleiter
- 2 ehemalige Generäle
- Forschungs- und Ausbildungskurse
- Seminare für Führungskräfte
- VR - Lab
- Hardware für Datenanalyse
- OSINT – Lab



Artificial intelligence refers to systems **that use technologies** such as: text mining, computer vision, speech recognition, natural language generation, machine learning, deep learning to gather and/or use **data** to predict, recommend or decide, with varying levels of **autonomy**, the **best** action to achieve **specific goals**.

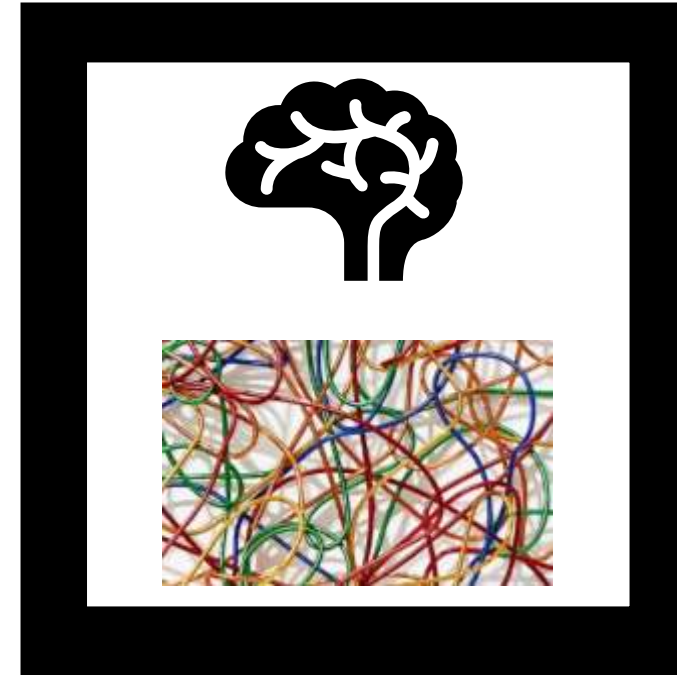
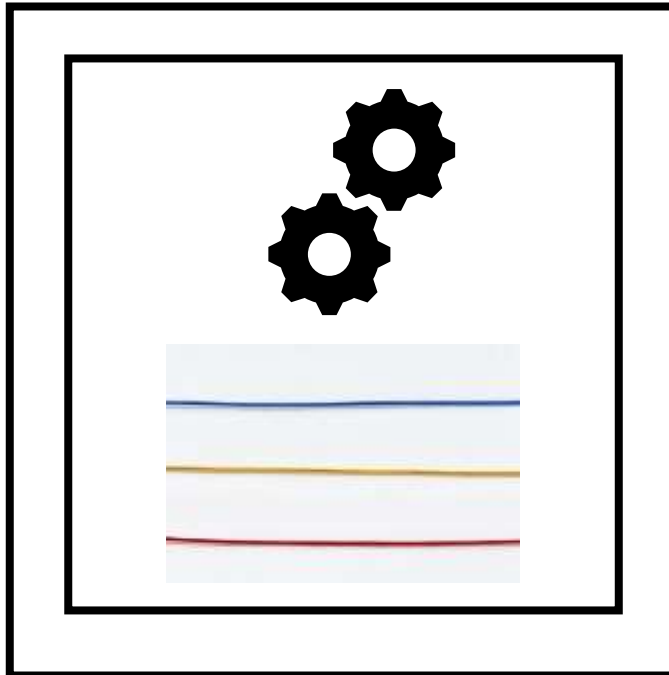
Artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given **set of human-defined objectives**, generate outputs such as content, predictions, recommendations, or decisions **influencing the environments** they interact with.

(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods.”

Artificial intelligence (AI) systems are software (and possibly also hardware) systems **designed by humans** that, given a **complex goal**, act in the physical or digital dimension **by perceiving** their environment through **data** acquisition, **interpreting** the collected structured or unstructured data, **reasoning** on the knowledge, or processing the information, derived from this data and **deciding** the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also **adapt their behaviour** by analysing how the environment is affected by their previous actions."

Artificial Intelligence refers to systems that display **intelligent behaviour** by analysing their environment and taking action — with some degree of autonomy — to achieve specific goals.

AI is a **generic term** that refers to any machine or algorithm that is capable of observing its environment, learning, and based on the knowledge and experience gained, taking **intelligent action** or proposing decisions. There are many different technologies that fall under this broad AI definition. **At the moment**, ML techniques are the most widely used







Hund



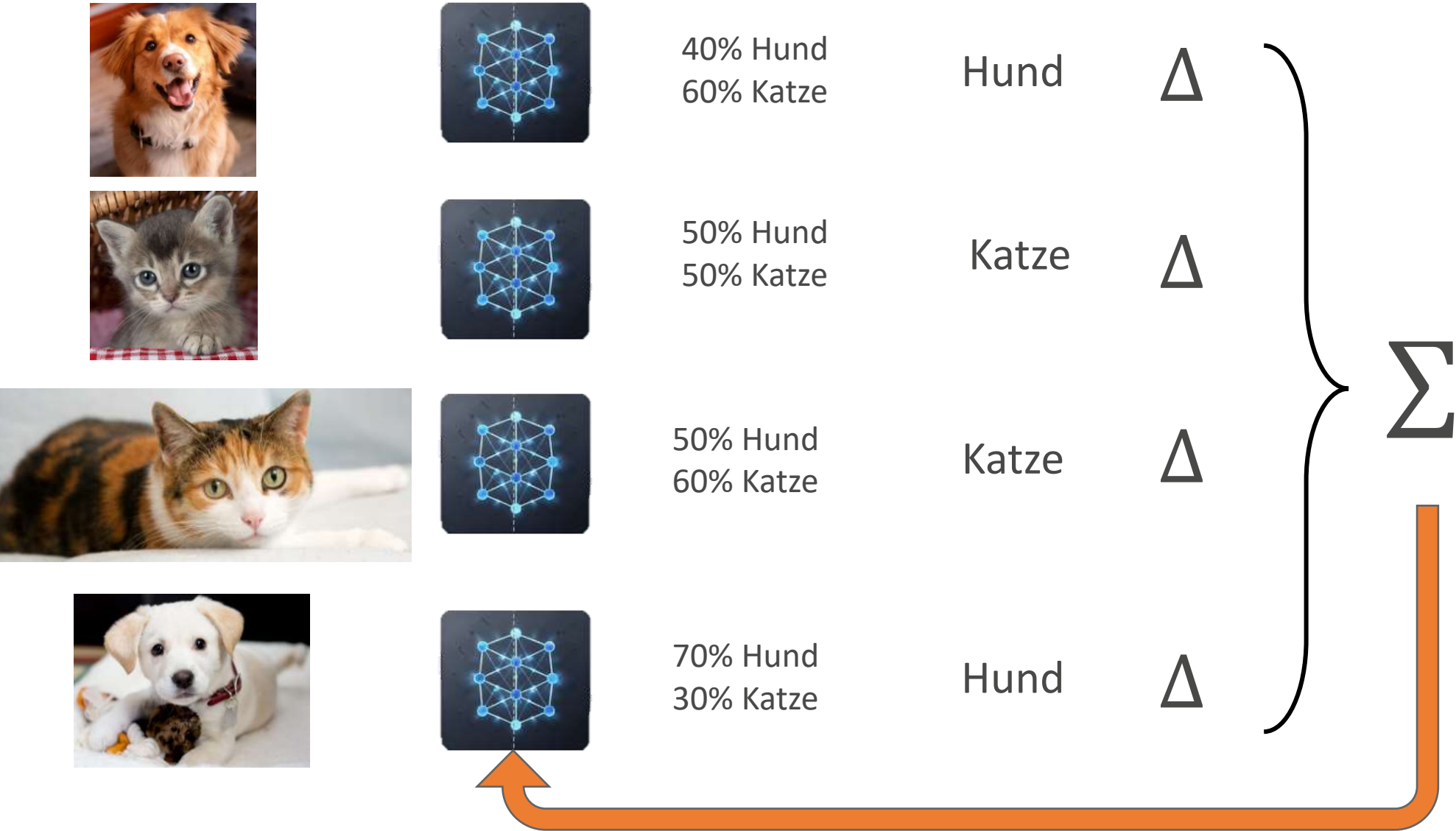
Katze

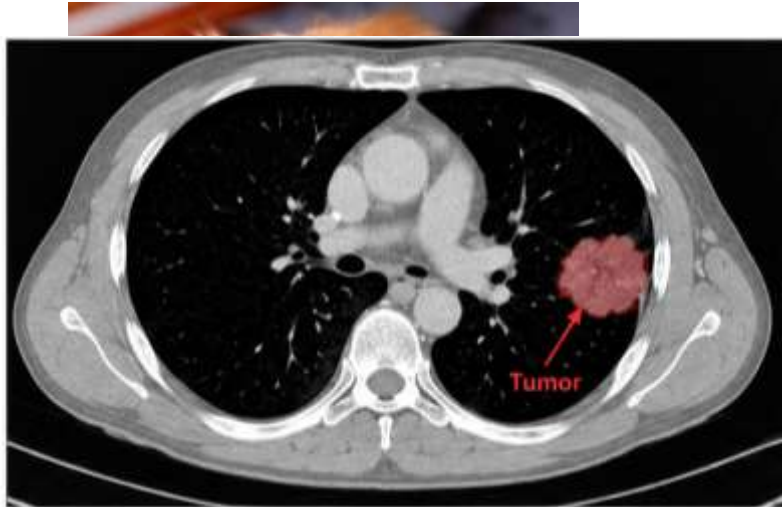


Hund



Katze

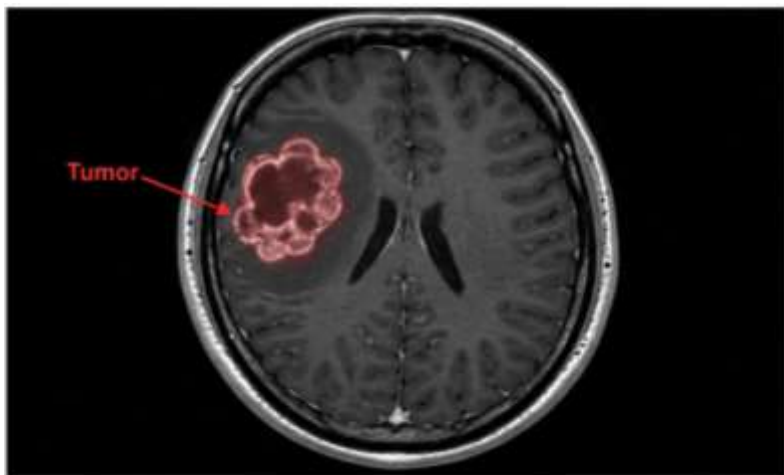




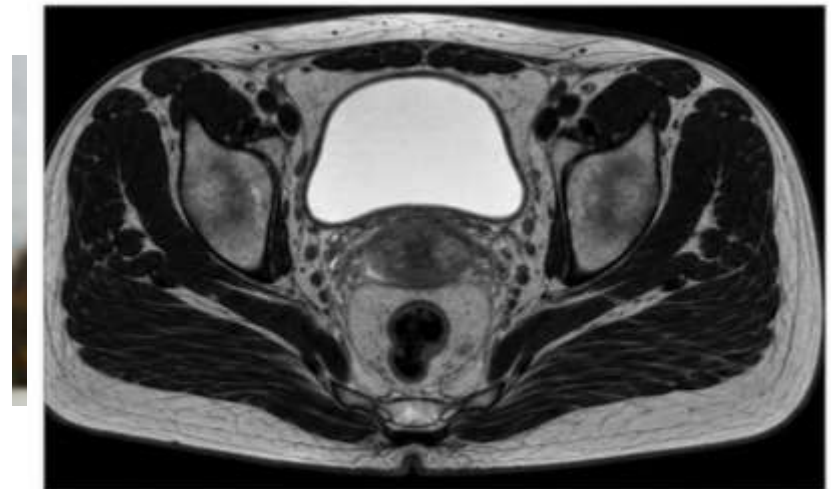
Hund

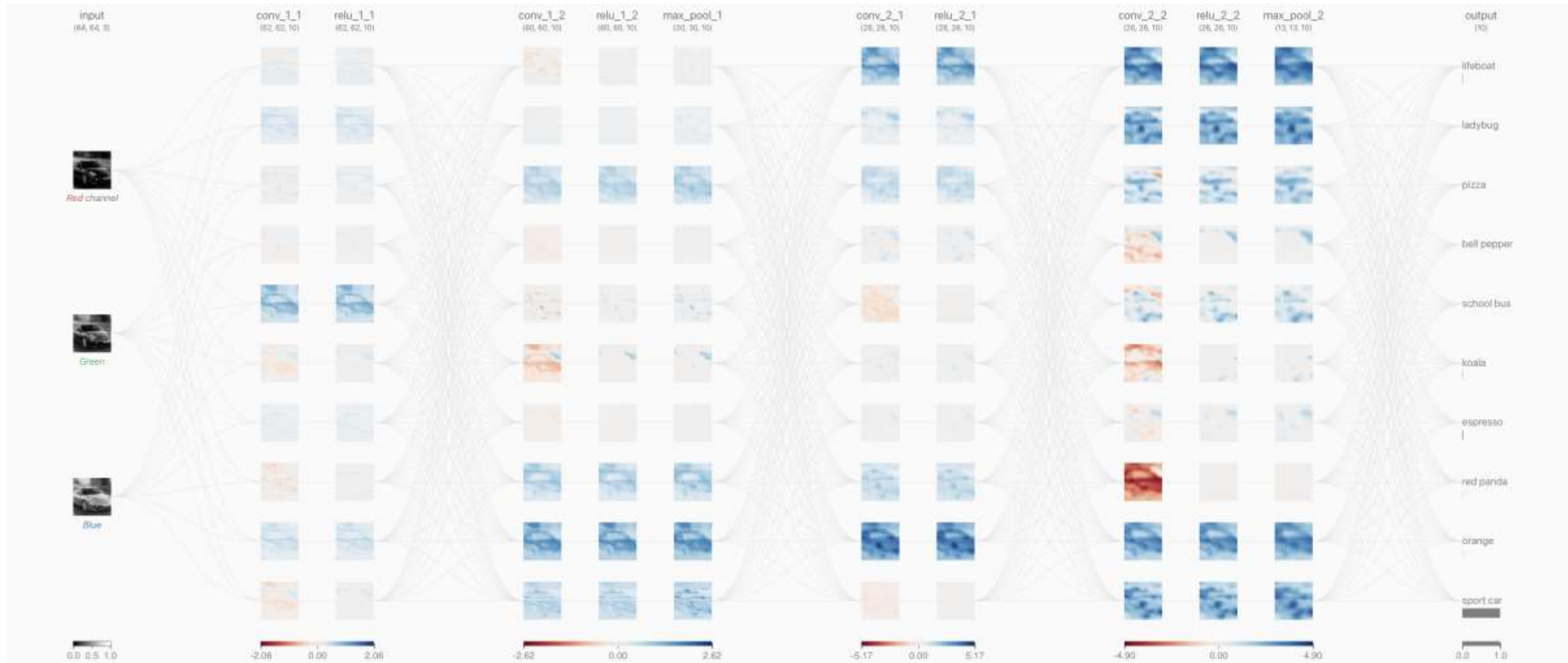


Katze

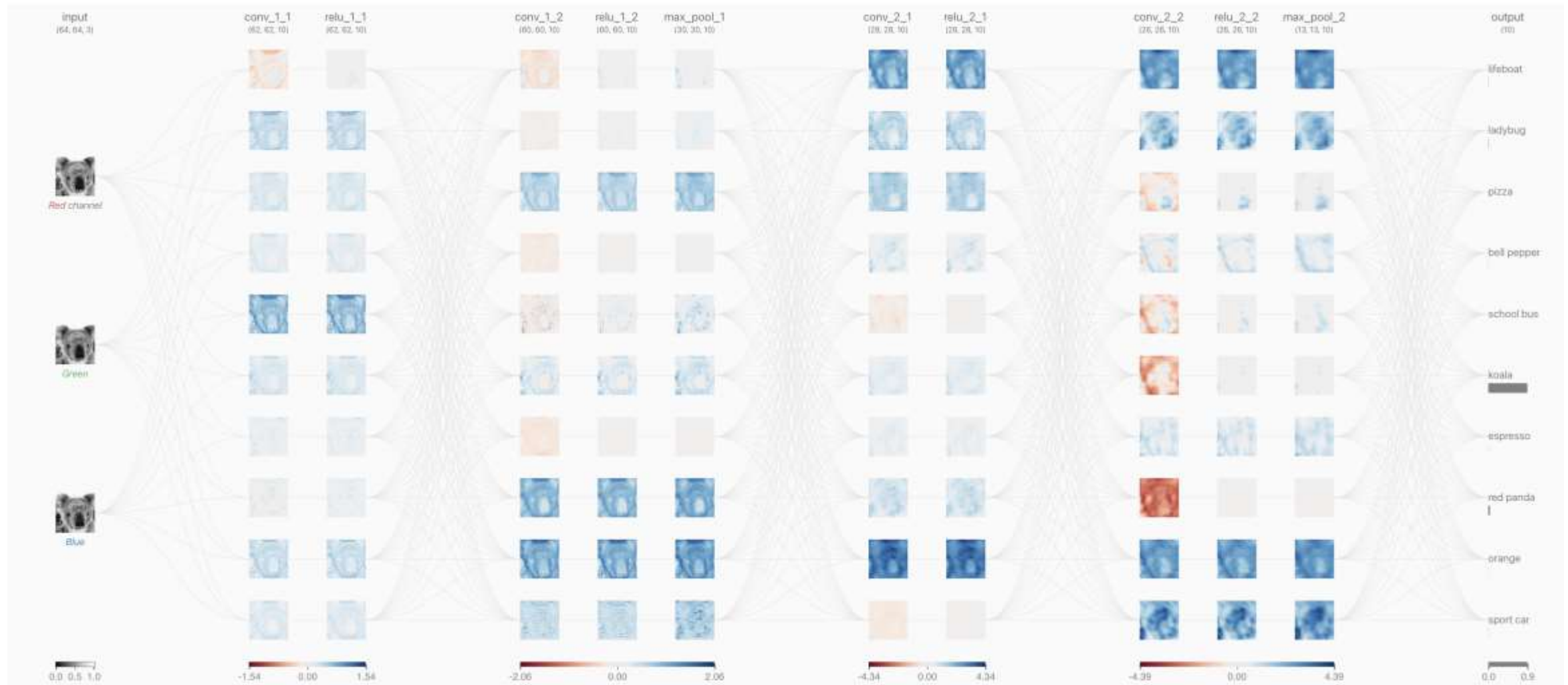


Hund





<https://poloclub.github.io/cnn-explainer/>



<https://poloclub.github.io/cnn-explainer/>



Hund



Katze



Hund



Katze









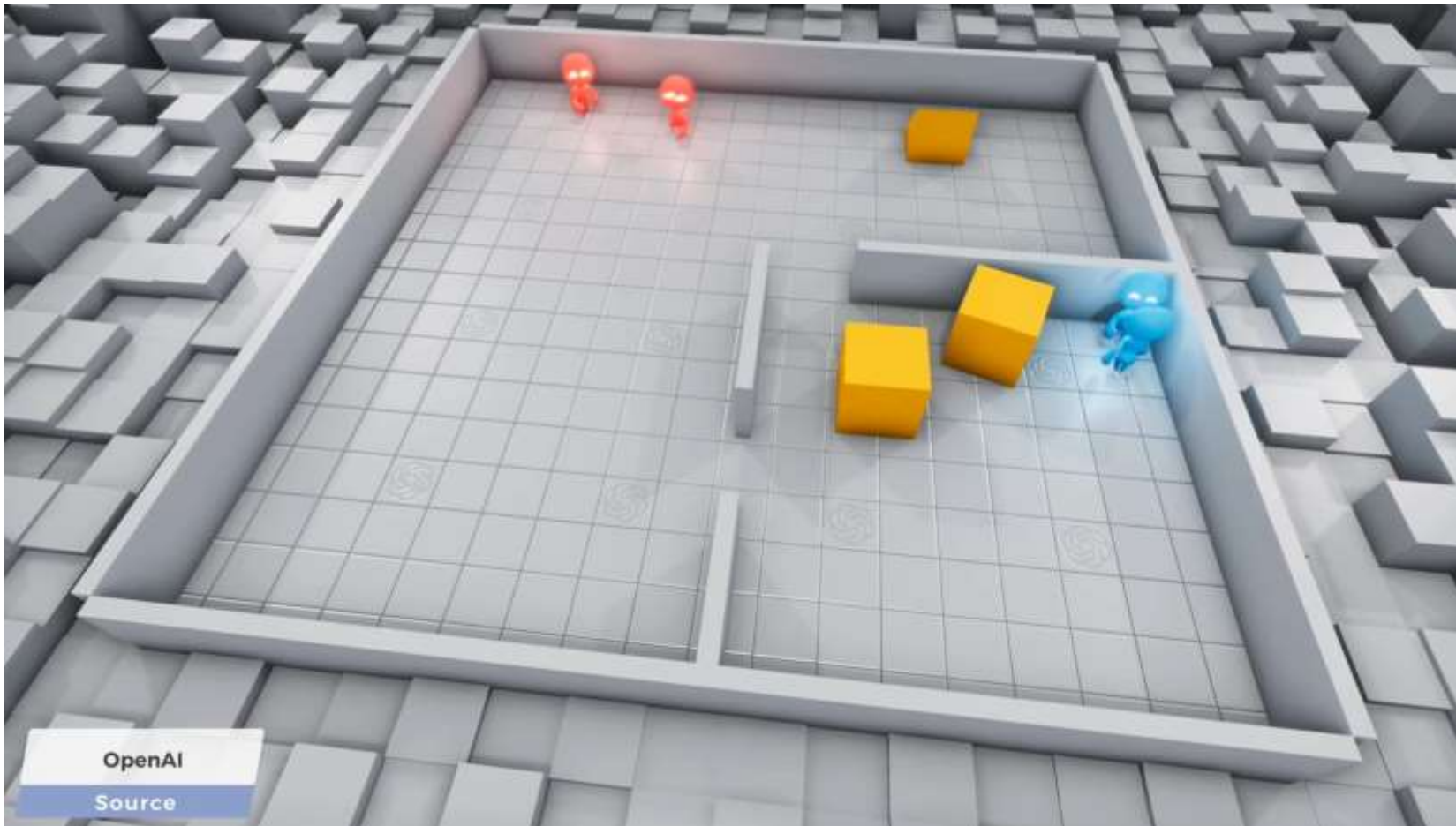
Hund

Katze

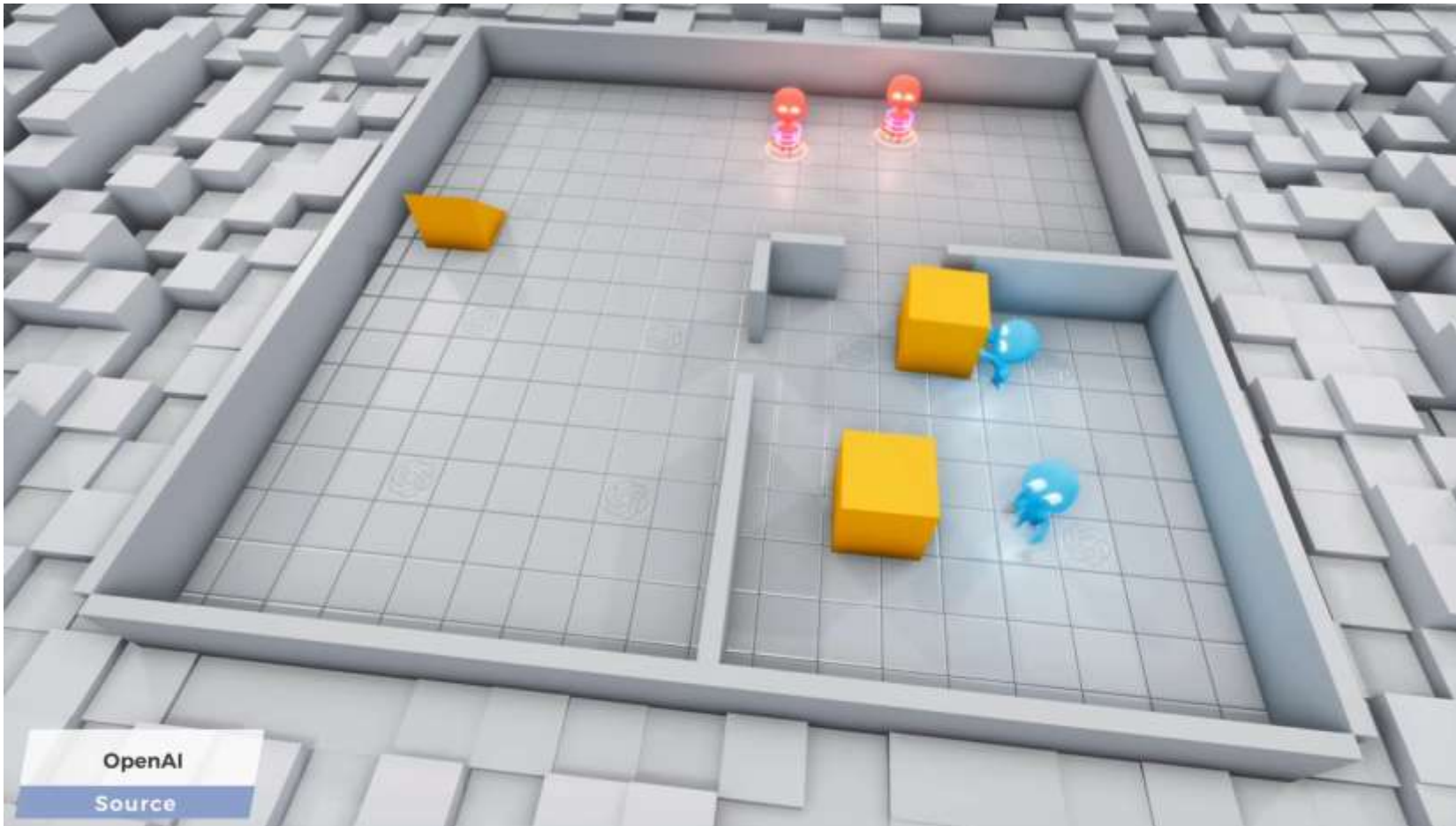


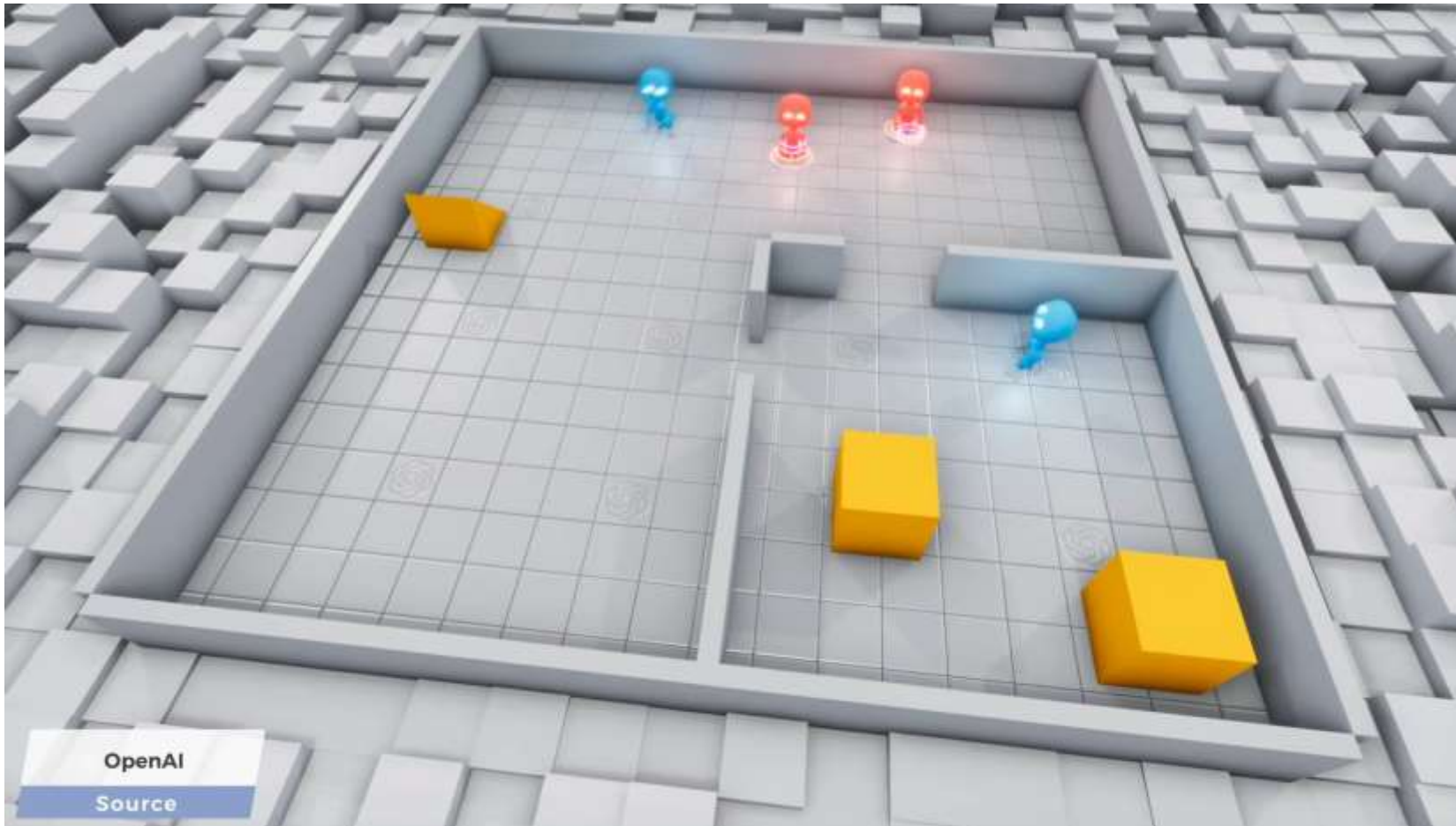
Hund

Katze

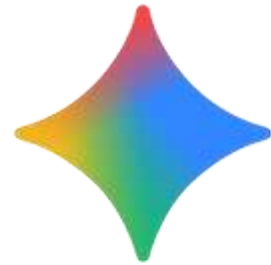




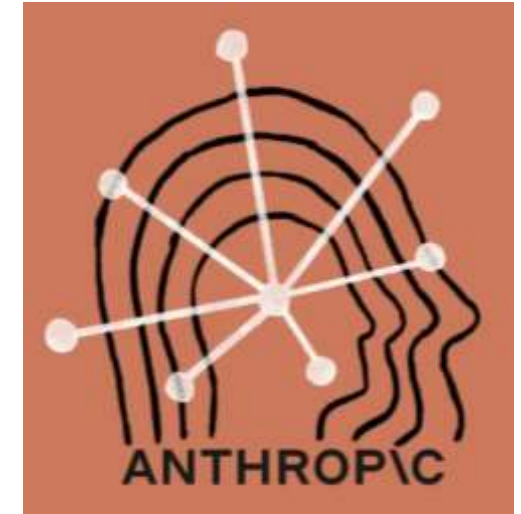






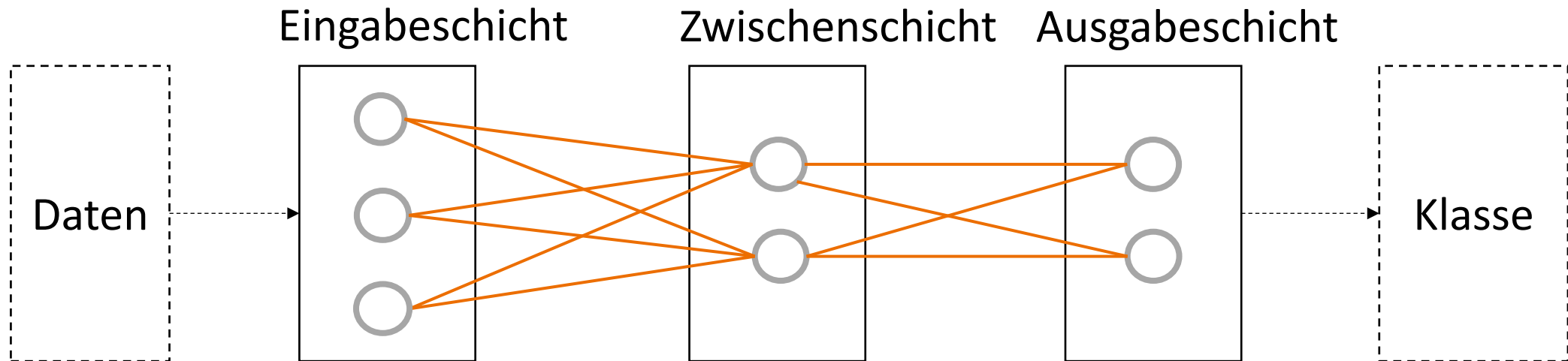


Gemini

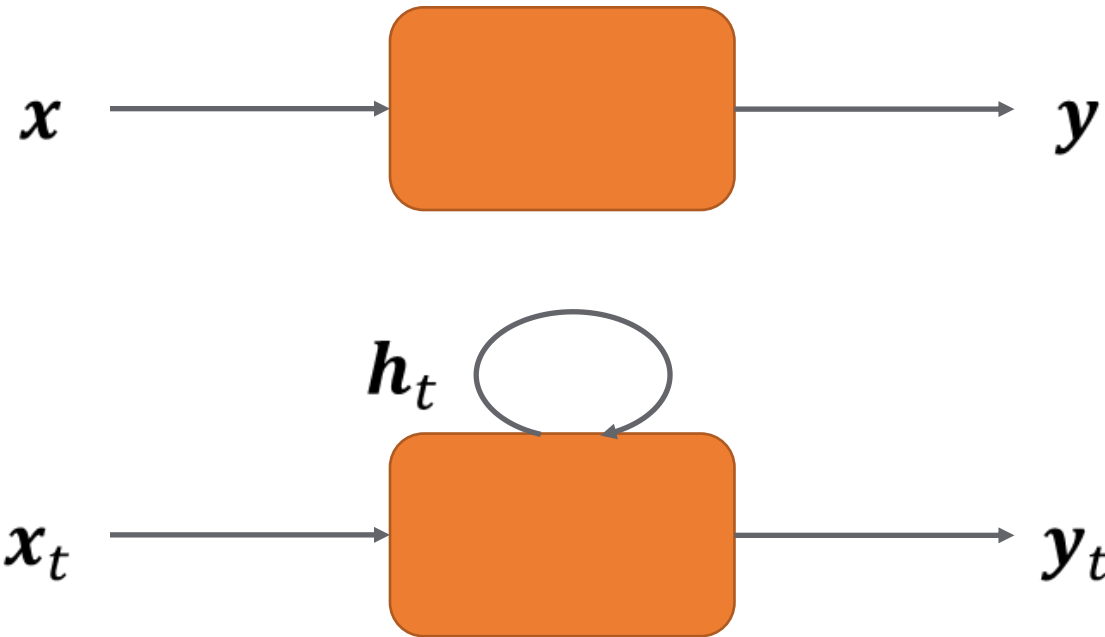


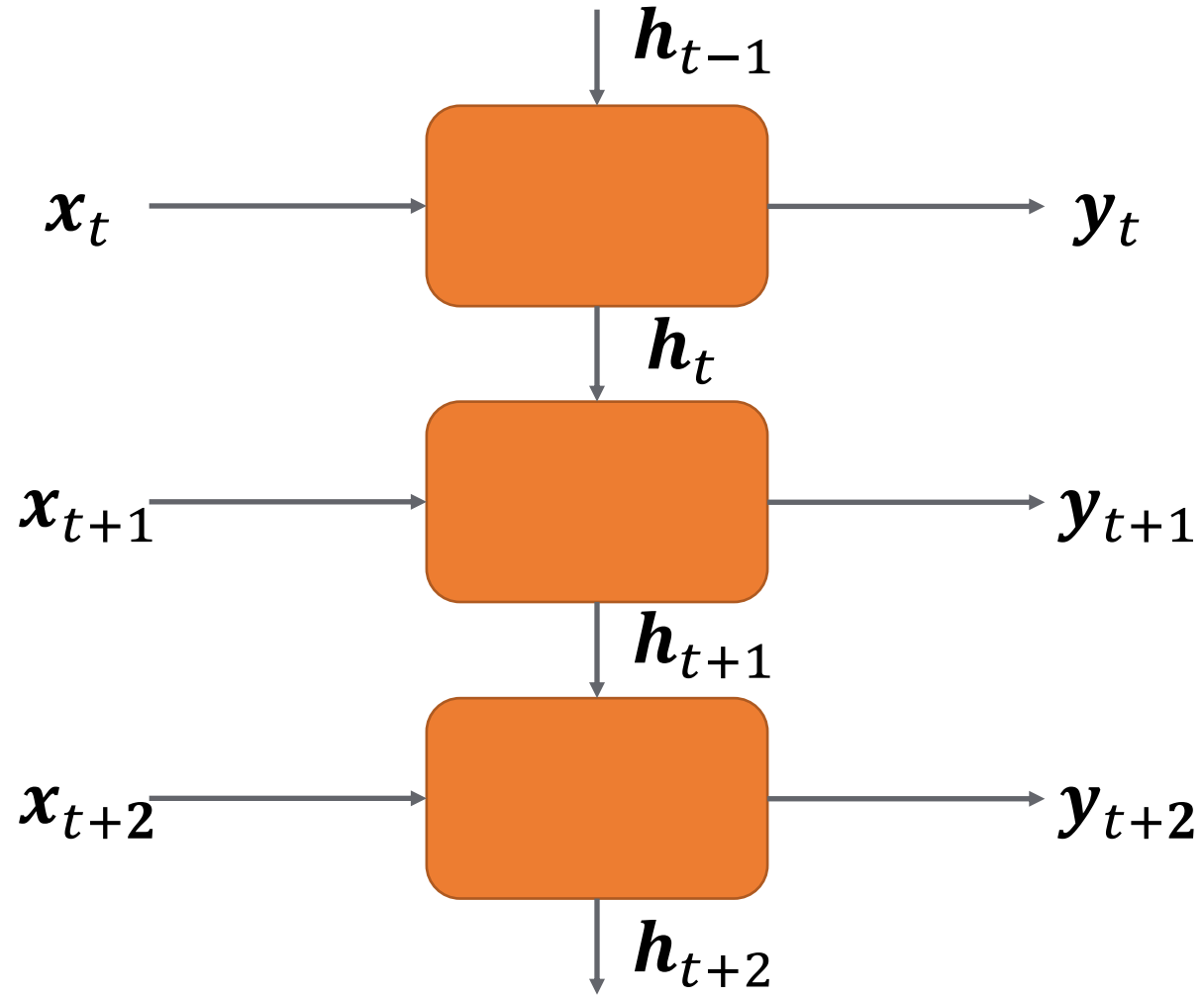
Qwen3

Idee des neuronalen Netzes:



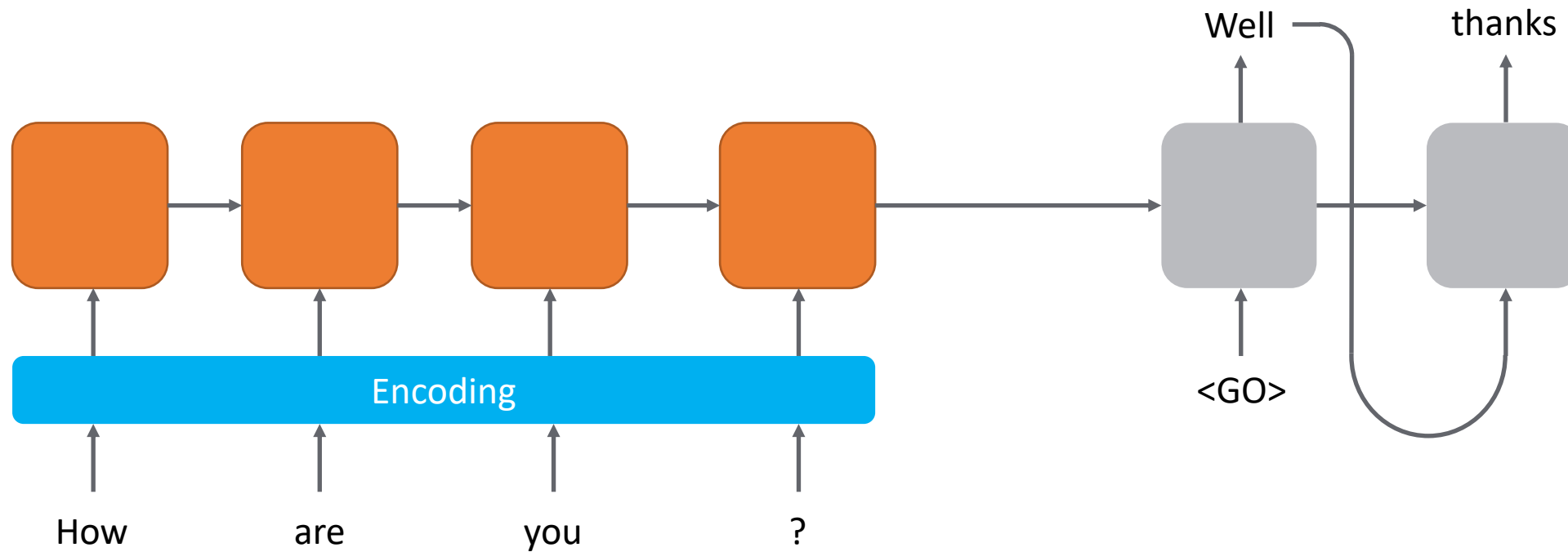
In einem Graphen sind **Neuronen** (Knoten) über **Synapsen** (Kanten) miteinander verbunden. Jedes Neuron verarbeitet Eingangssignale zu Ausgangssignalen.

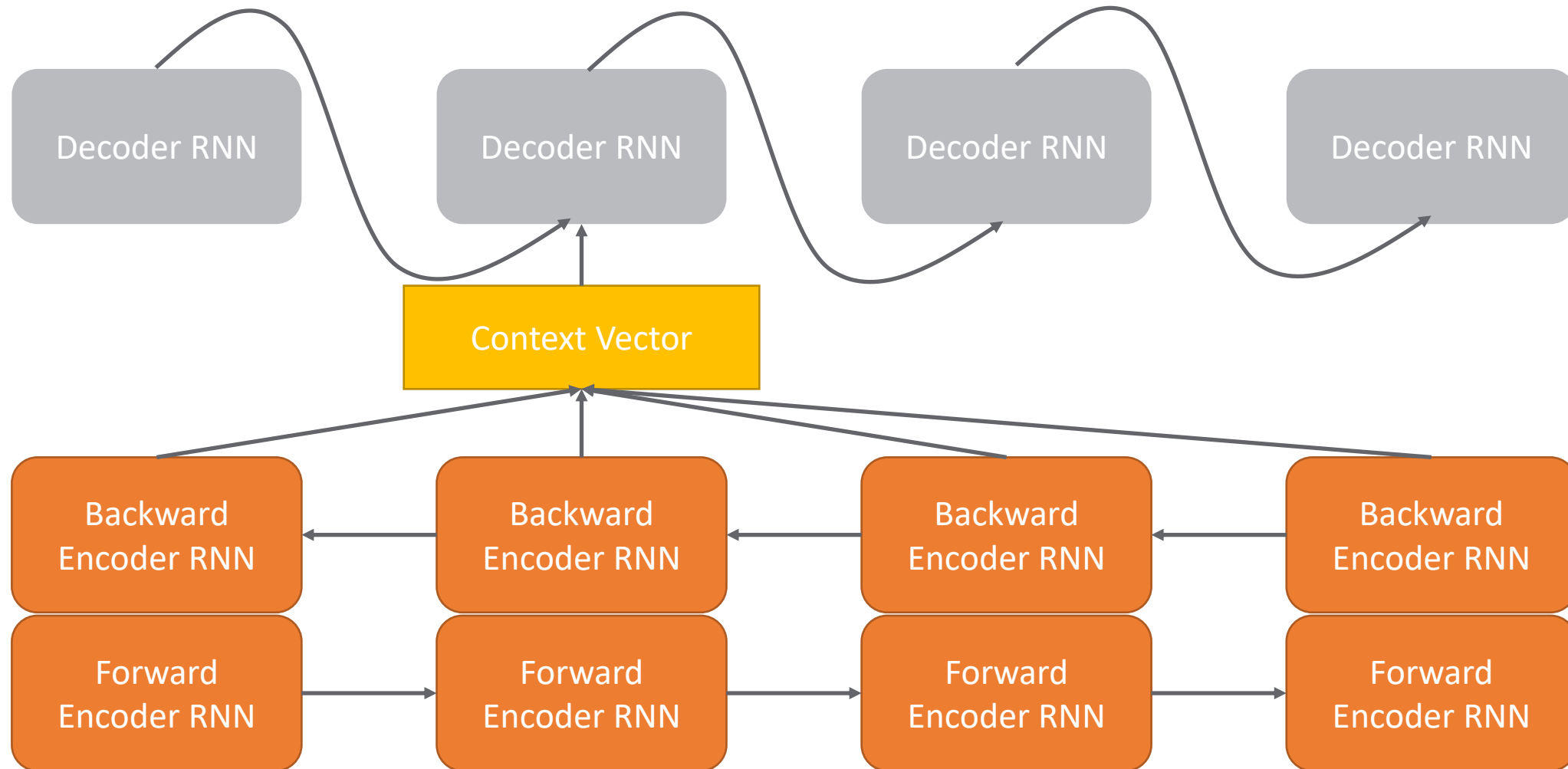




Encoder

Decoder





GPT-3.5 & GPT-4 GPT-3 (Legacy)

Many words map to one token, but some don't: indivisible.

Unicode characters like emojis may be split into many tokens containing the underlying bytes: 🍌

Sequences of characters commonly found next to each other may be grouped together: 1234567890



Clear

Show example

Tokens
57

Characters
252

Many words map to one token, but some don't: indivisible.

Unicode characters like emojis may be split into many tokens containing the underlying bytes: 🍌

Sequences of characters commonly found next to each other may be grouped together: 1234567890

TEXT TOKEN IDS

A helpful rule of thumb is that one token generally corresponds to ~4 characters of text for common English text. This translates to roughly ¾ of a word (so 100 tokens ~= 75 words).

GPT-3.5 & GPT-4 GPT-3 (Legacy)

Viele Wörter lassen sich einem Token zuordnen, einige jedoch nicht: unteilbar.

Unicode-Zeichen wie Emojis können in viele Token aufgeteilt werden, die die zugrunde liegenden Bytes enthalten: 🍌

Sequenzen von Zeichen, die häufig nebeneinander vorkommen, können in Gruppen zusammengefasst werden: 1234567890

Clear

Show example

Tokens
90

Characters
309

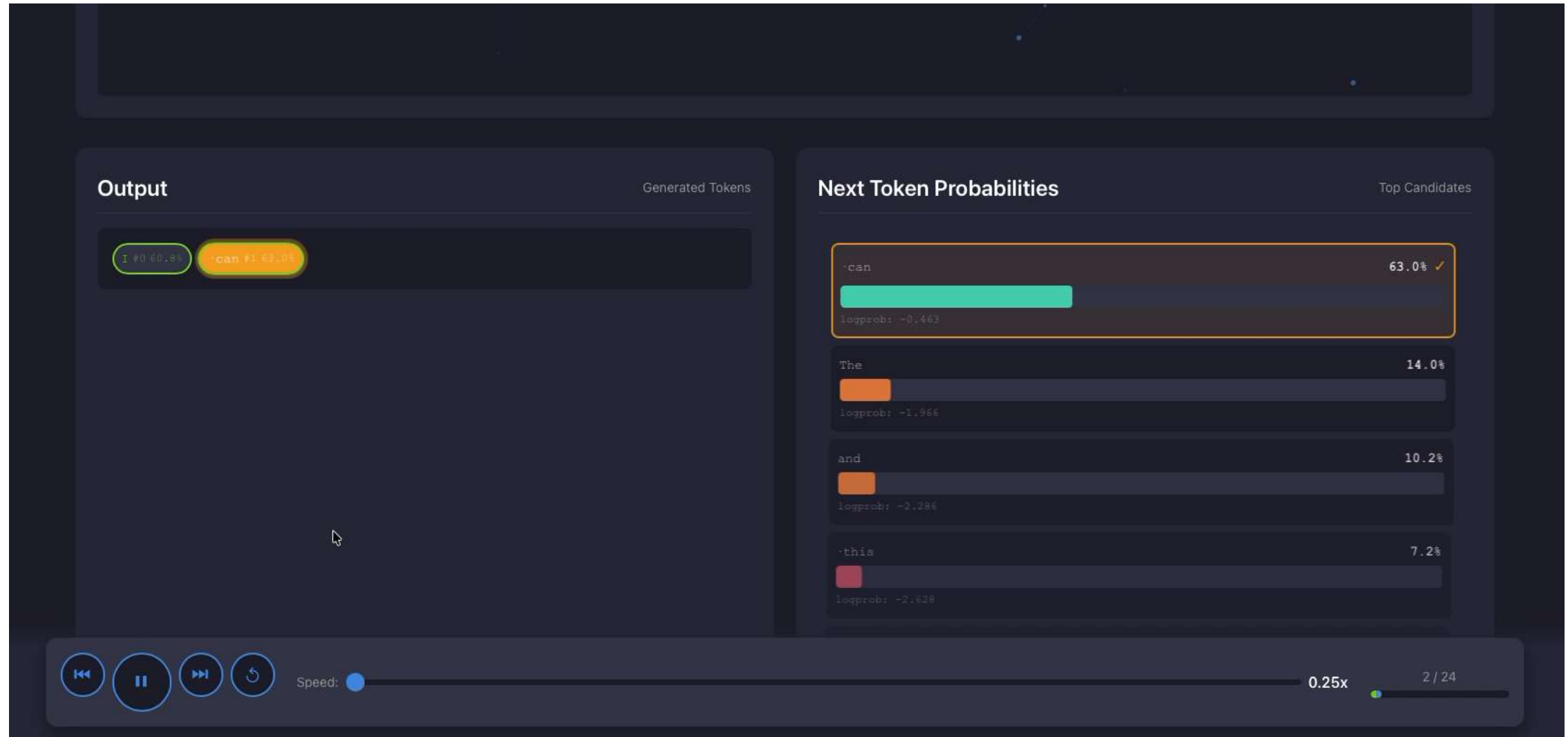
Viele Wörter lassen sich einem Token zuordnen, einige jedoch nicht: unteilbar.

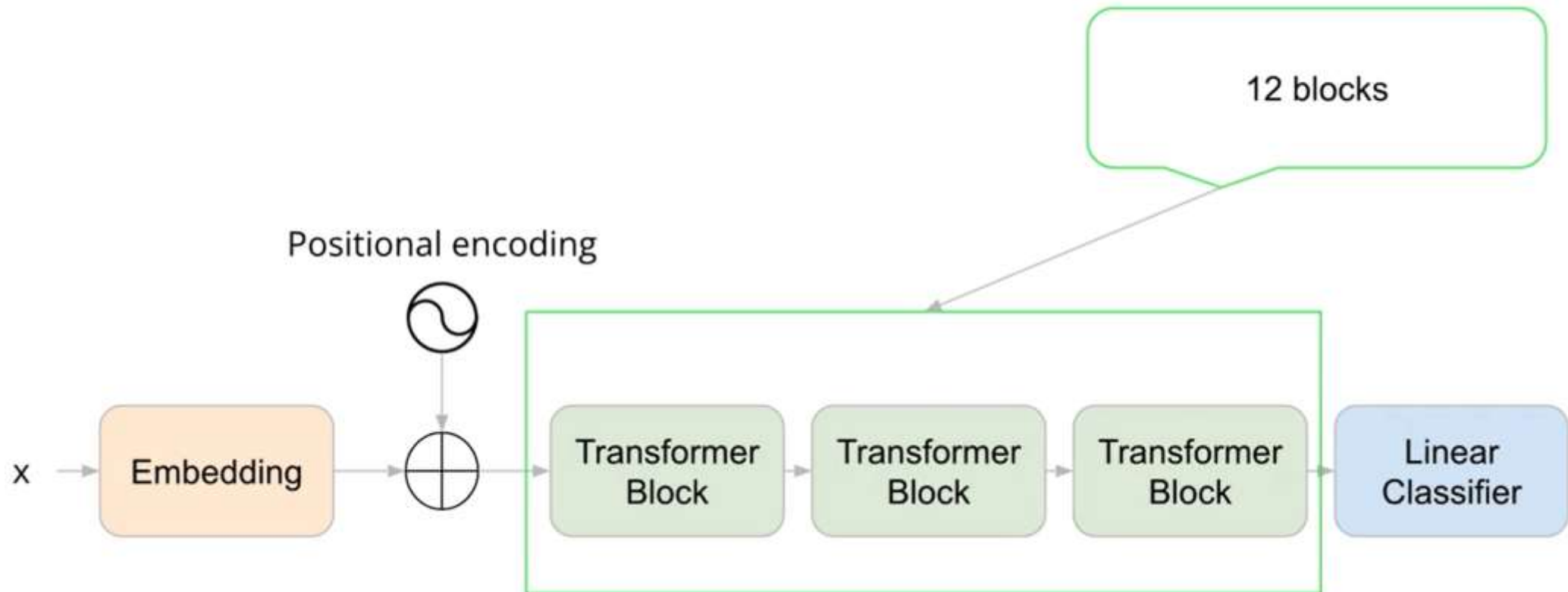
Unicode-Zeichen wie Emojis können in viele Token aufgeteilt werden, die die zugrunde liegenden Bytes enthalten: 🍌

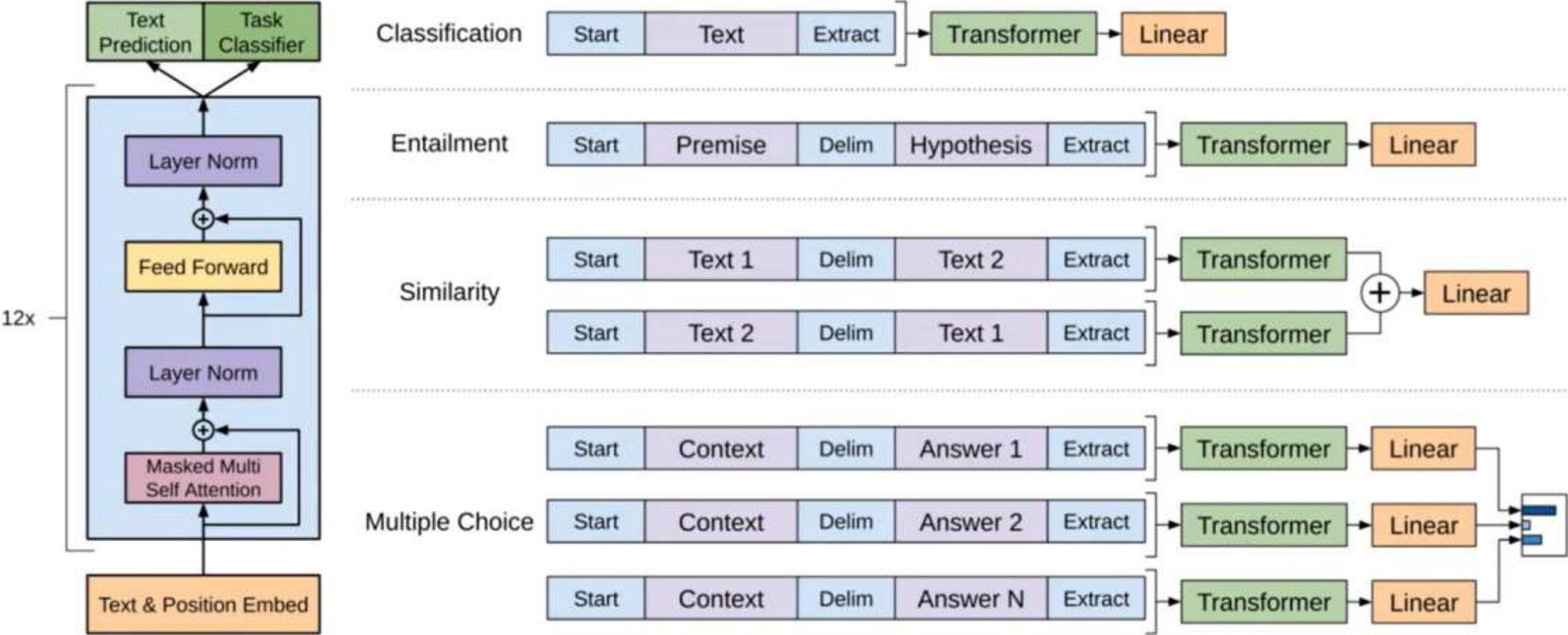
Sequenzen von Zeichen, die häufig nebeneinander vorkommen, können in Gruppen zusammengefasst werden: 1234567890

TEXT TOKEN IDS

<https://platform.openai.com/tokenizer>



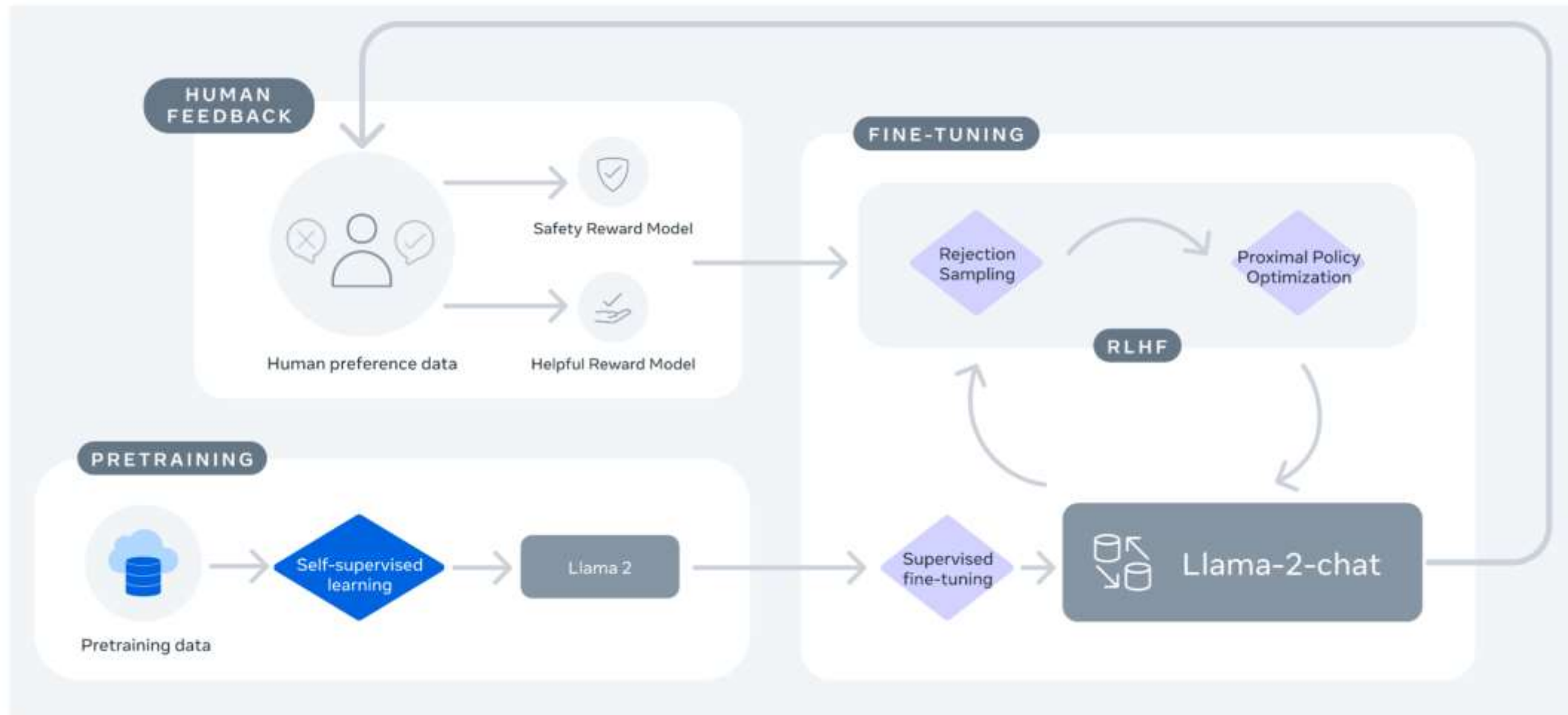


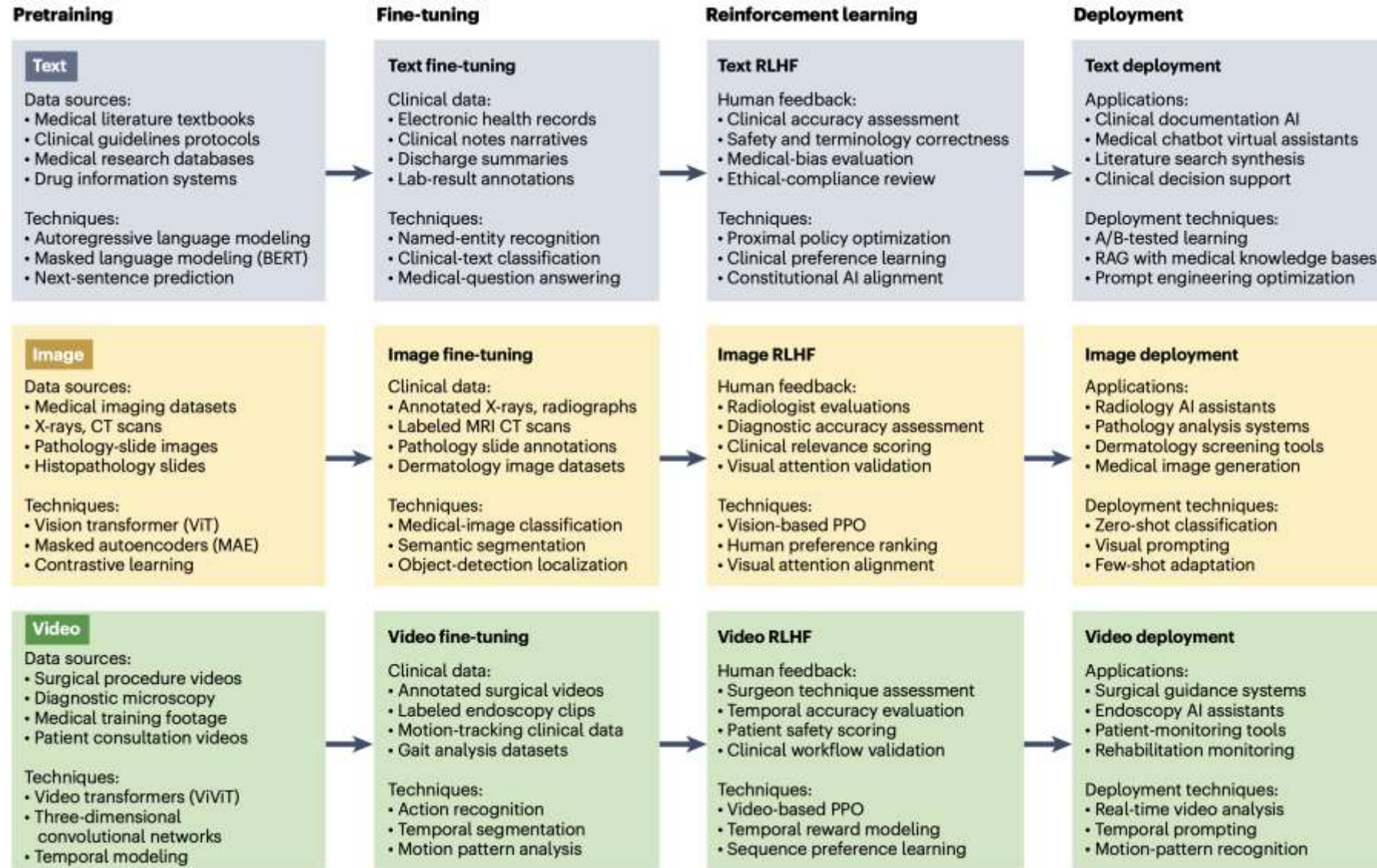


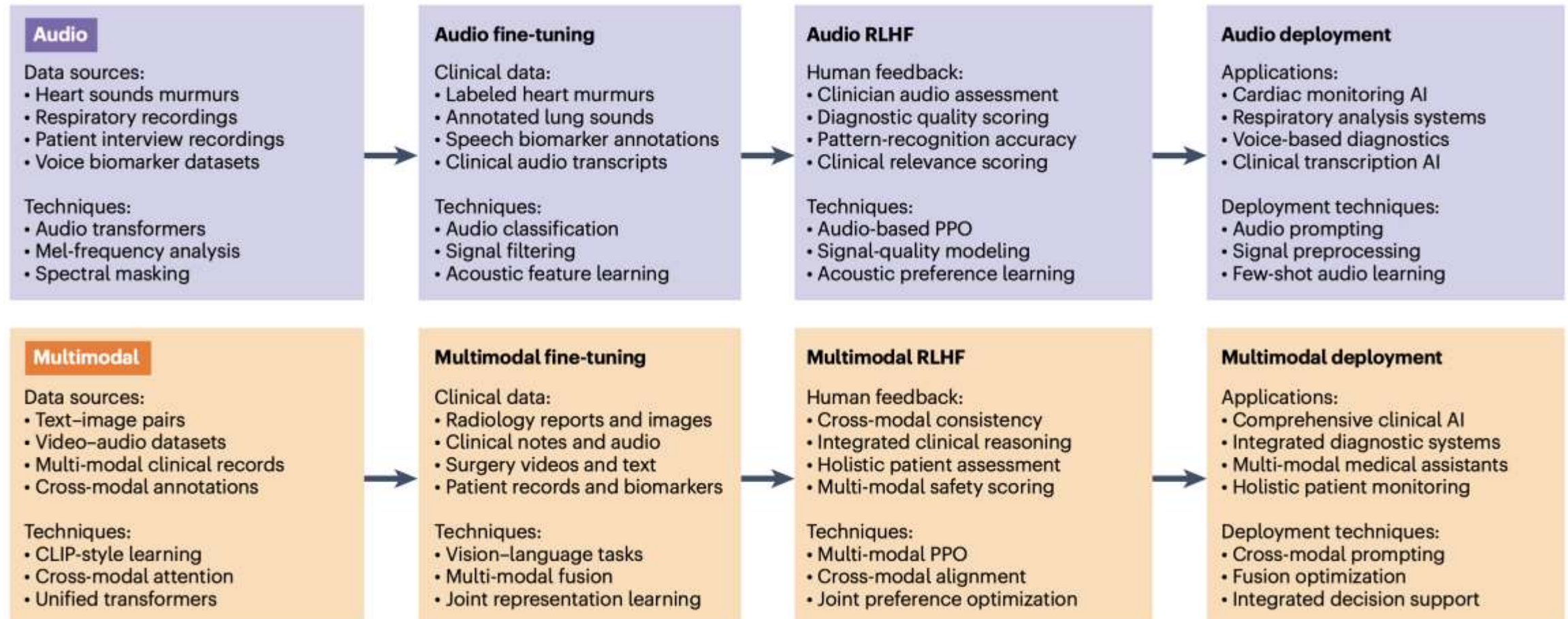
- Pretraining mit Unsupervised Task: nächsten Token basierend auf vorherigen vorhersagen
- Datensatz: BooksCorpus
- #Parameters: 110 Millionen

- Primär: Mehr Daten, mehr Parameter
 - #Params: 1.5 Milliarden
 - Kontext 512 -> 1024
 - Qualitätsprobleme mit CommonCrawl
 - Generierung von WebText basierend auf Links von Reddit mit 3+ Karma -> 8 Millionen Dokumente, 40GB
- Zum ersten Mal: Produktion von langen kohärenten Texten
- Kein Fine-Tuning für spezifische Tasks (Übersetzung, Zusammenfassung, etc.) nur Vorhersage von nächstem Token
- Input durch Prompts – z.B. Zusammenfassung: <Text> TL;DR:

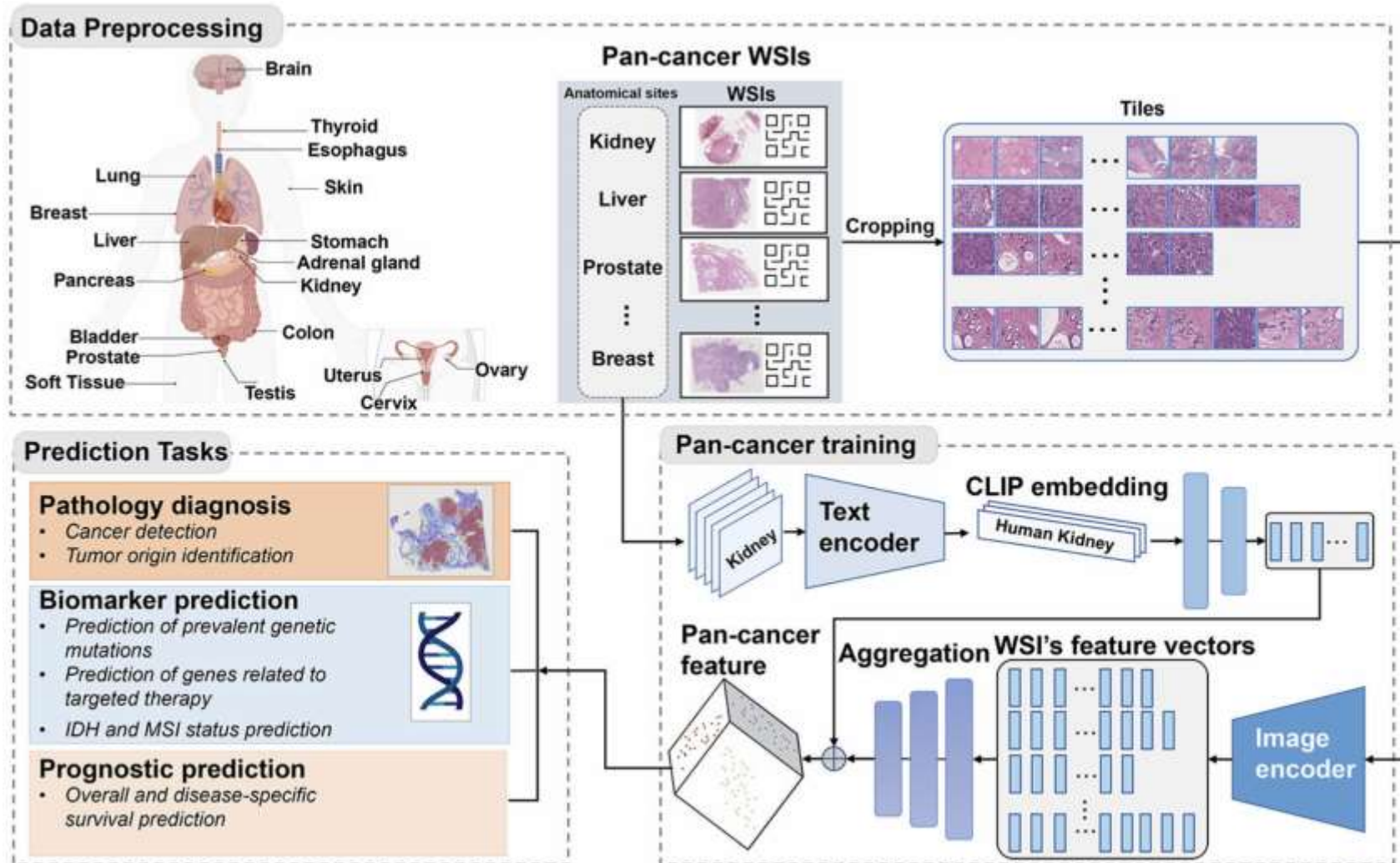
- Und so weiter:
 - GPT-2 = 10x GPT, GPT-3 = 100x GPT-2: 175 Milliarden Parameter
 - Kontext: 1024 -> 2048
- Daten:
 - CommonCrawl (410 Milliarden Tokens), 60%
 - WebText2 (19 Milliarden Tokens), 22%
 - Books1 (12 Milliarden Tokens), 8%
 - Books2 (55 Milliarden Tokens), 8%
 - Wikipedia (3 Milliarden Tokens), 3%



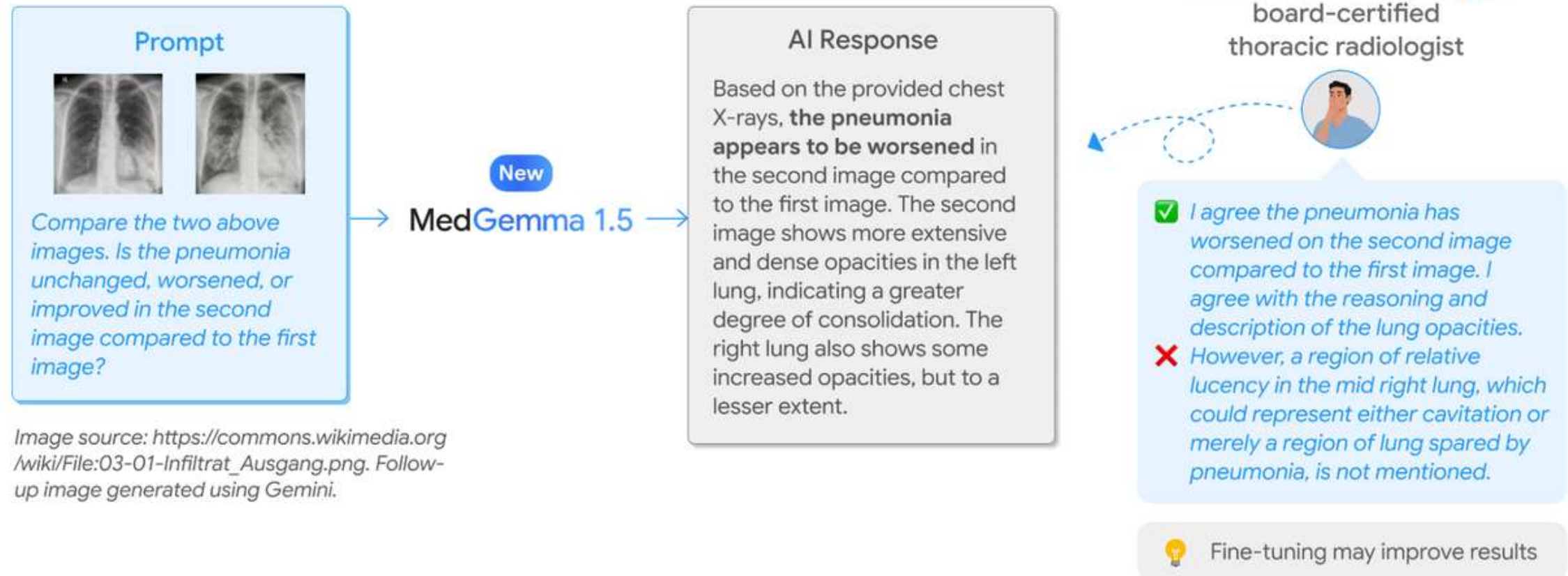




Teo, Z. L., Thirunavukarasu, A. J., Elangovan, K., Cheng, H., Moova, P., Soetikno, B., ... & Ting, D. S. W. (2025). Generative artificial intelligence in medicine. *Nature medicine*, 31(10), 3270-3282.



Wang, X., Zhao, J., Marostica, E., Yuan, W., Jin, J., Zhang, J., ... & Yu, K. H. (2024). A pathology foundation model for cancer diagnosis and prognosis prediction. *Nature*, 634(8035), 970-978.



<https://research.google/blog/next-generation-medical-image-interpretation-with-medgemma-15-and-medical-speech-to-text-with-medasr>





Human oversight. Human oversight helps ensuring that an AI system does not undermine human autonomy or causes other adverse effects. Oversight may be achieved through governance mechanisms such as a **human-in-the-loop (HITL)**, **human-on-the-loop (HOTL)**, or **human-in-command (HIC)** approach. HITL refers to the capability for human intervention in every decision cycle of the system.

<https://ec.europa.eu/futurium/en/ethics-guidelines-trustworthy-ai/stakeholder-consultation-guidelines-first-draft.html>

