THE UNIVERSITY OF QUEENSLAND AUSTRALIA | CREATE CHANGE

Centre for Digital Cultures & Societies

LDaCA Language Data Commons of Australia

Invited Scholar Presentation

Prof. Laurence Anthony

LLMs for HASS Research: Opportunities, Challenges, and a Human-Centred Future

This talk was partially funded by the **Centre for Digital Cultures and Societies** at the University of Queensland and the **Language Data Commons of Australia** (LDaCA). Language Data Commons of Australia (LDaCA) (DOI: 10.3565/kq2v-9g52) is a co-investment partnership with the **Australian Research Data Commons** (ARDC) through the HASS and Indigenous Research Data Commons. The ARDC is enabled by the Australian Government's **National Collaborative Research Infrastructure Strategy** (NCRIS).

Acknowledgement of Country

- The University of Queensland (UQ) acknowledges the Traditional Owners and their custodianship of the lands on which we meet.
- We pay our respects to their Ancestors and their descendants, who continue cultural and spiritual connections to Country.
- We recognise their valuable contributions to Australian and global society.

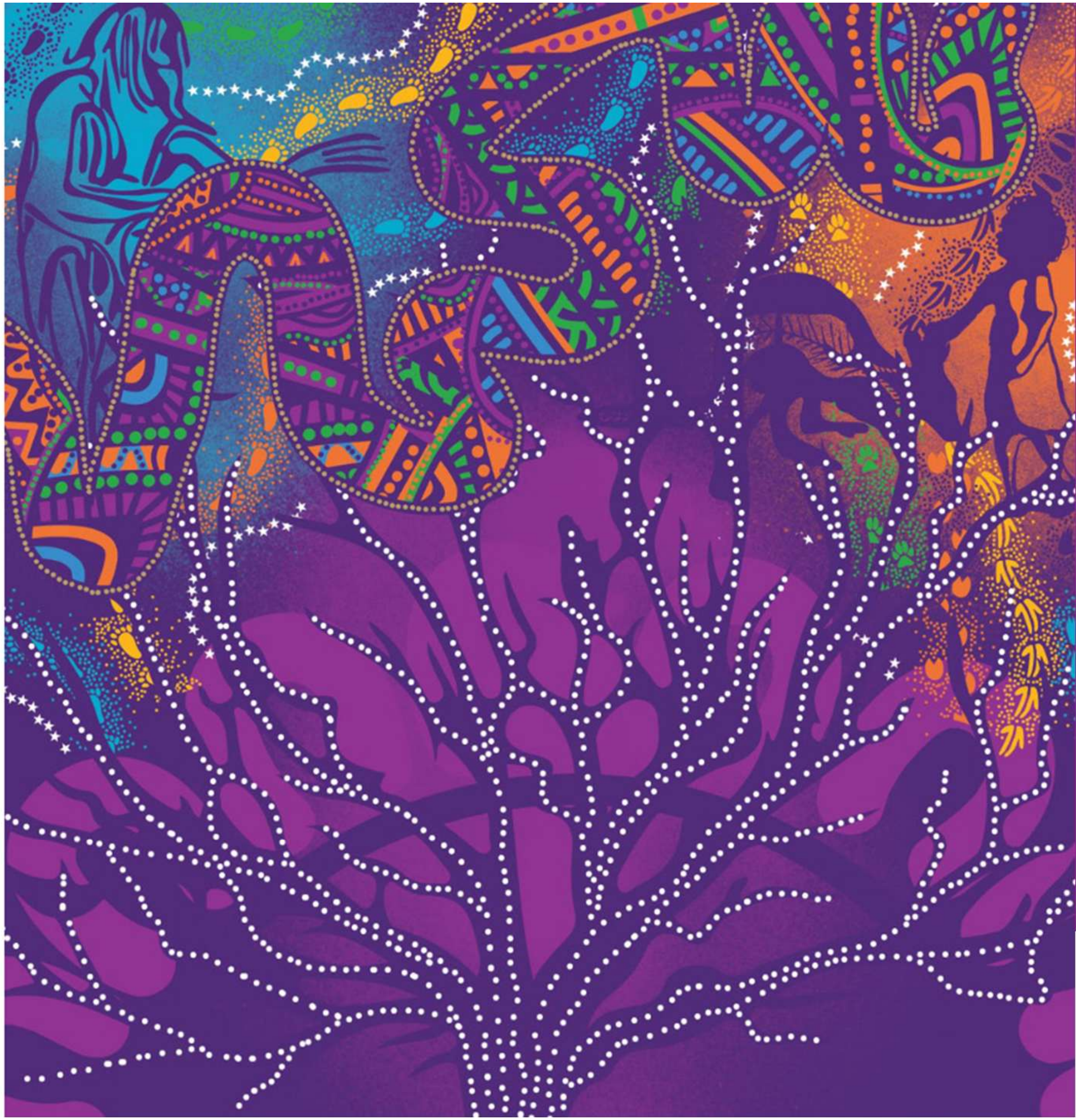
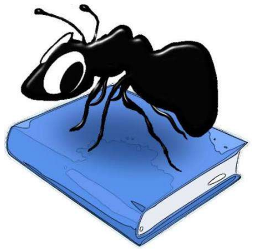


Image: Digital reproduction of *A guidance through time* by Casey Coolwell and Kyra Mancktelow

Laurence Anthony

Professor of Applied Linguistics at Waseda University in Tokyo.

- **Corpus linguist & language data scientist** working at the intersection of **corpus linguistics, AI, educational technology, and English for Specific Purposes**
- Creator of **AntConc**, one of the world's most widely used corpus analysis tools
- Academic background spanning **Mathematical Physics → TESL/TEFL → Applied Linguistics**
- Founding member of Waseda University's **Center for English Language Education in Science and Engineering** (CELESE)
- **CEO, AntLab Solutions**, developing corpus, AI, machine-learning and Big Data solutions for education and business
- Winner of several **national and international awards** e.g., the 2012 JAECS National Prize
- Has given more than 60 keynotes and plenaries, published more than 130 journal articles, 6 books, 20 book chapters with overall more than 13,000 GS citations.



Today's Talk: **LLMs for HASS Research: Opportunities, Challenges, and a Human-Centred Future**

Let’s welcome Laurence

The University of Queensland
Australia

University of Queensland, Australia
The Centre for Digital Cultures & Societies, UQ
August 18, 2026

LLMs for HASS Research

Opportunities, Challenges, and a Human-Centered Future

Laurence Anthony
Professor
Faculty of Science and Engineering
Waseda University, Tokyo, Japan
anthony@waseda.jp
http://www.laurenceanthony.net/
@antlabjp

August 18, 2026. University of Queensland, Australia

Faculty of Science and Engineering

LLMs for HASS Research

Voices on AI

"AI is corpus linguistics on steroids"

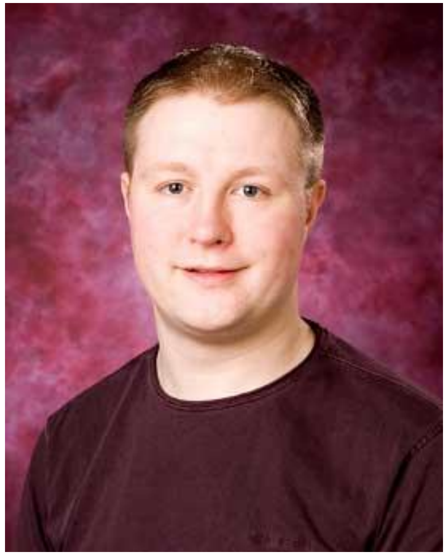
Dr. Lisa Cheung & Dr Peter Crosthwaite

CorpusChat: Where Corpus Linguistics and GenAI meet

Corpus Linguistics & Applied Linguistics Research 2025

https://www.um.es/languagecorpora/clresearch2025/

November 3, 2025



" I will address the specific prospect that the operation of LLMs can substitute in part or whole for a corpus linguist’s analysis. This proposition is **highly improbable** due to the opacity and inscrutability of black-box algorithms in comparison to well-understood corpus methods, deployed with care and skill. "

Andrew Hardie
Quandaries of corpus analysis in the age of AI hype
Corpus Linguistics 2025
https://www.cl2025.co.uk/programme
30th June - 3rd July 2025

Voices on AI

"I think the most beautiful thing about deep learning is that **it actually works**... because you got these ideas, you got the little neural network, you got the back propagation algorithm, so maybe if you make it large...and you train it on a lot of data then **it will do the same function of the brain does**, and **it turns out to be true**."

Lex Fridman podcast: May 8, 2020
https://www.youtube.com/watch?v=13CZPWmke6A

Ilya Sutskever
(Ex)-Chief Scientist of OpenAI

OpenAI

LLMs for HASS Research

Voices on AI

"LLMs and other generative **AI are enormously powerful**, because they soak up, abstract, and can mashup the work of millions of humans. But they are **only a bit more intelligent than an encyclopedia**. Central to intelligence is the ability to learn, adapt, and act in novel situations."

(Twitter: March 25, 2024)

Christopher Manning

Thomas M. Siebel Professor in Machine Learning
Professor of Linguistics and of Computer Science
Director, Stanford Artificial Intelligence Laboratory (SAIL)
Associate Director, Stanford Institute for Human-Centered Artificial Intelligence (HAI)
2024 IEEE von Neumann medal recipient

Stanford University

LLMs for HASS Research

Voices on AI

"LLMs are good at manipulating language, **but not at thinking**...
I think the shelf life of the current [AI] paradigm is fairly short, probably **three to five years**"

(India Today Tech, Jan 24, 2025)

Yann LeCun

Chief AI Scientist, Meta

Jacob T. Schwartz Professor of Computer Science, Data Science, Neural Science, and Electrical and Computer Engineering, New York University.

ACM Turing Award Laureate

Meta

Voices on AI

Emily M. Bender

Thomas L. and Margo G. Wyckoff Endowed Professor
Department of Linguistics
University of Washington

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Emily M. Bender*
ebender@uw.edu
University of Washington
Seattle, WA, USA

Timnit Gebru*
timnit@blackinai.org
Black in AI
Palo Alto, CA, USA

Angelina McMillan-Major
aymm@uw.edu
University of Washington
Seattle, WA, USA

Shmargaret Shmitchell
shmargaret.shmitchell@gmail.com
The Aether

ABSTRACT

The past 3 years of work in NLP have been characterized by the development and deployment of ever larger language models, especially for English. BERT, its variants, GPT-2/3, and others, most recently Switch-C, have pushed the boundaries of the possible both through architectural innovations and through sheer size. Using these pretrained models and the methodology of fine-tuning them for specific tasks, researchers have extended the state of the art on a wide array of tasks as measured by leaderboards on specific benchmarks for English. In this paper, we take a step back and ask: How big is too big? What are the possible risks associated with this technology and what paths are available for mitigating those risks? We provide recommendations including weighing the environmental and financial costs first, investing resources into curating and carefully documenting datasets rather than ingesting everything on the web, carrying out pre-development exercises evaluating how the planned approach fits into research and development goals and supports stakeholder values, and encouraging research directions beyond ever larger language models.

alone, we have seen the emergence of BERT and its variants [39, 70, 74, 113, 146], GPT-2 [106], T-NLG [112], GPT-3 [25], and most recently Switch-C [43], with institutions seemingly competing to produce ever larger LMs. While investigating properties of LMs and how they change with size holds scientific interest, and large LMs have shown improvements on various tasks (§2), we ask whether enough thought has been put into the potential risks associated with developing them and strategies to mitigate these risks.

We first consider environmental risks. Echoing a line of recent work outlining the environmental and financial costs of deep learning systems [129], we encourage the research community to prioritize these impacts. One way this can be done is by reporting costs and evaluating works based on the amount of resources they consume [57]. As we outline in §3, increasing the environmental and financial costs of these models doubly punishes marginalized communities that are least likely to benefit from the progress achieved by large LMs and most likely to be harmed by negative environmental consequences of its resource consumption. At the scale we are discussing (outlined in §2), the first consideration should be the

“LLMs are **not performing natural language understanding**... the tendency of human interlocutors to **impute meaning** where there is none **can mislead both NLP researchers and the general public** into taking synthetic text as meaningful." (p. 610-611)

Emily M. Bender

Thomas L. and Margo G. Wyckoff Endowed Professor

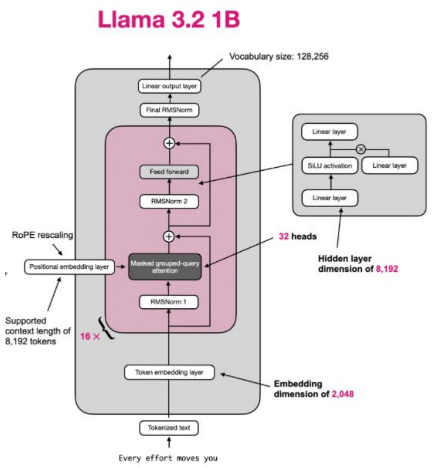
Department of Linguistics

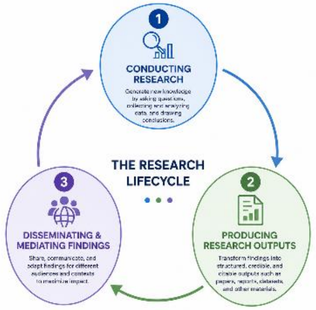
University of Washington

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, 610–623. https://doi.org/10.1145/3442188.344592

Overview

- What is a large language model?
  - The principles of LLMs
  - Modern LLM architecture and build process
  - Implications for HASS research
- Opportunities and challenges for AI in HASS research
  - conducting research
  - producing research outputs
  - disseminating and mediating findings
- A human-centered future for HASS research with AI
  - Personal Context Management (PCM) frameworks





For further information...

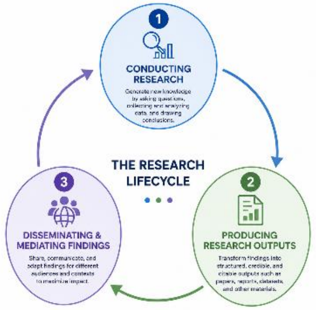
Innovations in Specialised Englishes: Emerging Trends in EAP and ESP

Eds. Lillian Wong & Ken Hyland

(expected 2027)

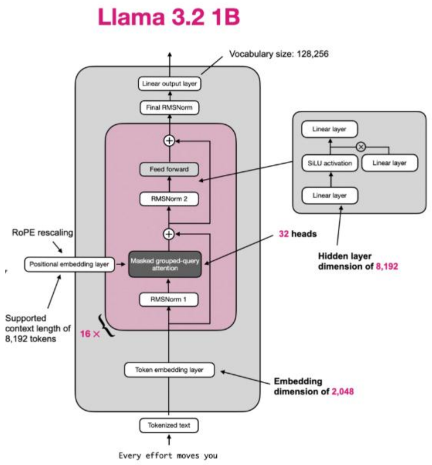
Chapter by L. Anthony

AI in EAP and ESP: Opportunities and Challenges Across the Research Lifecycle



What is a large language model?

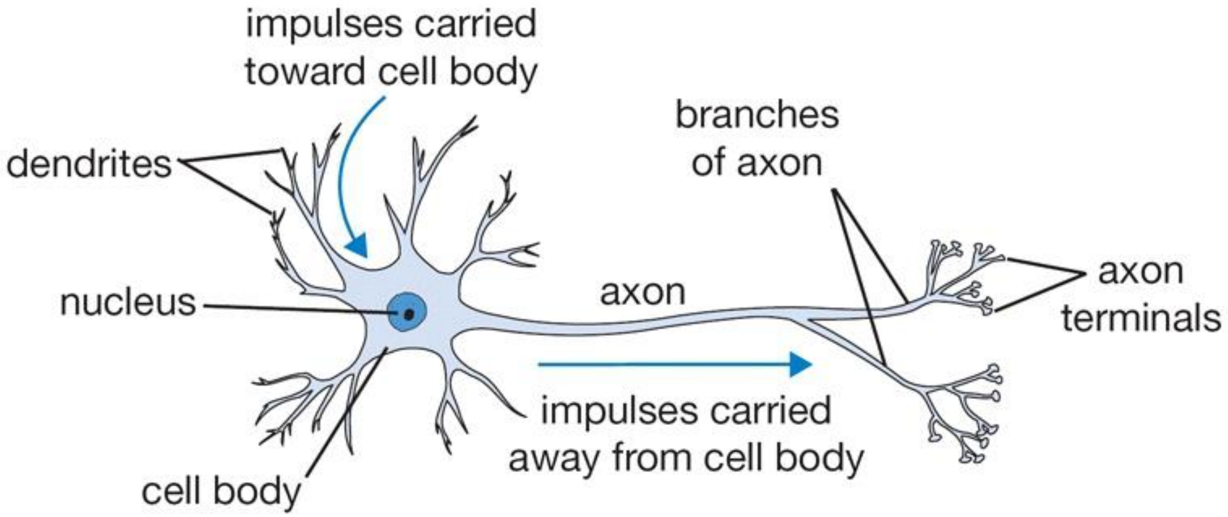
Definition; Modern LLM architecture and design
Implications for HASS research



What are Large Language Models (LLMs)?

Definition

- Large Language Models (LLMs) are (usually transformer-based) neural networks, designed to represent, process, and often generate natural language. LLMs are part of the broader family of AI models, including generative AI and discriminative AI models.

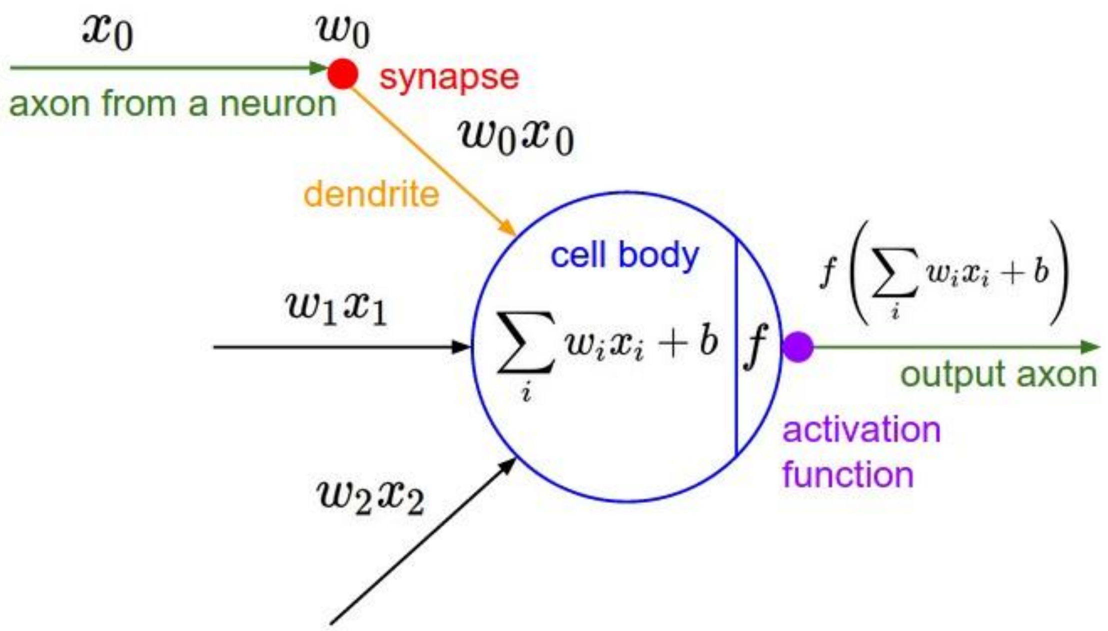


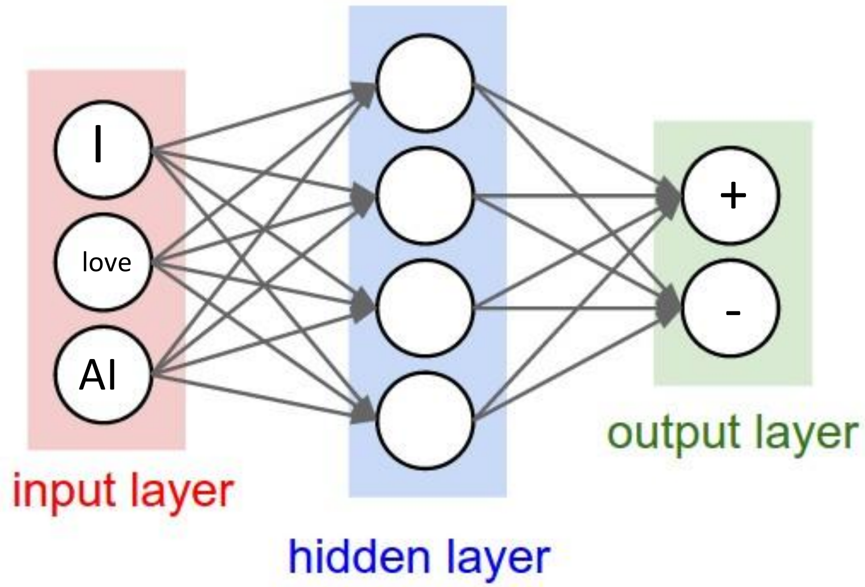
https://cs231n.github.io/neural-networks-1/

What are Large Language Models (LLMs)?

Definition

- Large Language Models (LLMs) are (usually transformer-based) neural networks, designed to represent, process, and often generate natural language. LLMs are part of the broader family of AI models, including generative AI and discriminative AI models.



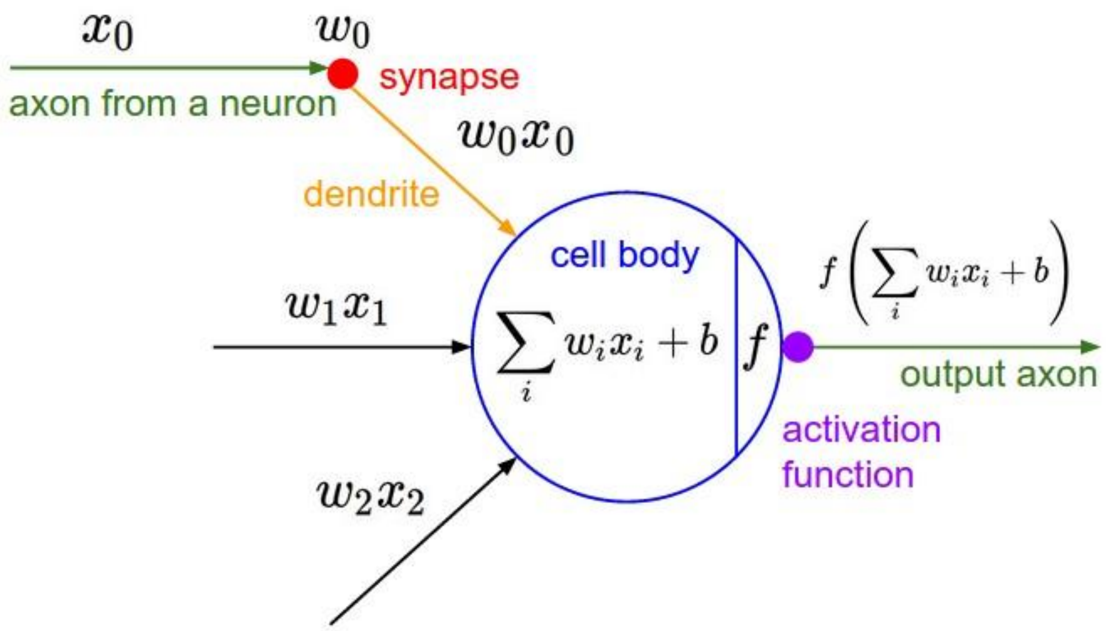


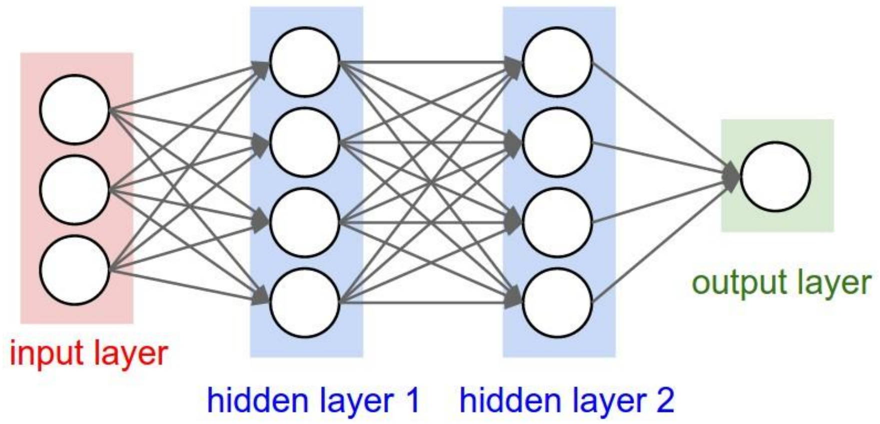
https://cs231n.github.io/neural-networks-1/

What are Large Language Models (LLMs)?

Definition

- Large Language Models (LLMs) are (usually transformer-based) neural networks, designed to represent, process, and often generate natural language. LLMs are part of the broader family of AI models, including generative AI and discriminative AI models.





https://cs231n.github.io/neural-networks-1/

What are Large Language Models (LLMs)?

LLM architecture – The Transformer Model (2017)

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

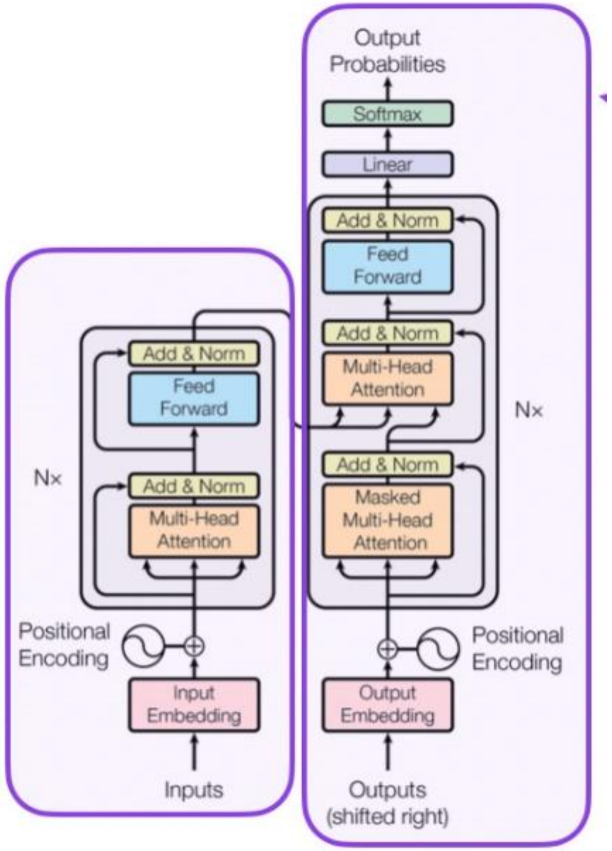


Figure 1: The Transformer - model architecture.

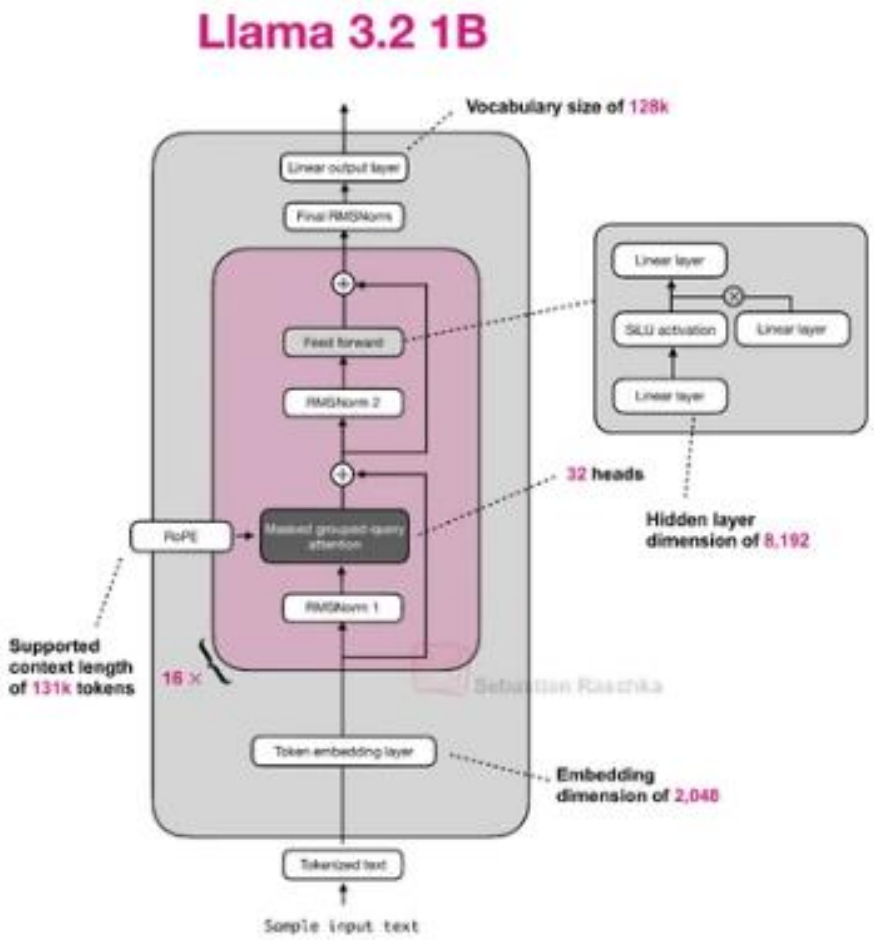
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

What are Large Language Models (LLMs)?
LLM architecture – Transformer Models (2026)

Raschka, S. (2026). LLM Architecture Gallery
https://sebastianraschka.com/llm-architecture-gallery/

What are Large Language Models (LLMs)?

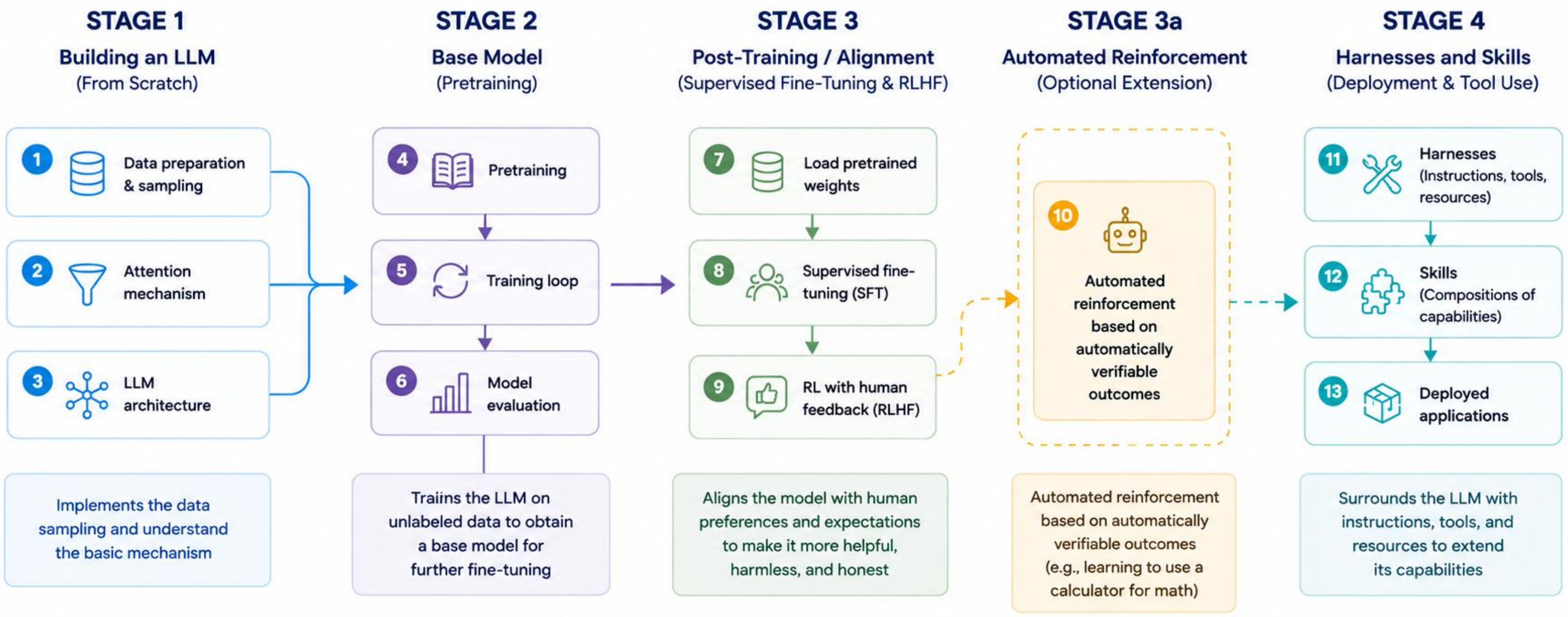
LLM architecture – Transformer Models (2026)



Raschka, S. (2026). LLM Architecture Gallery
https://sebastianraschka.com/llm-architecture-gallery/

What are Large Language Models (LLMs)?

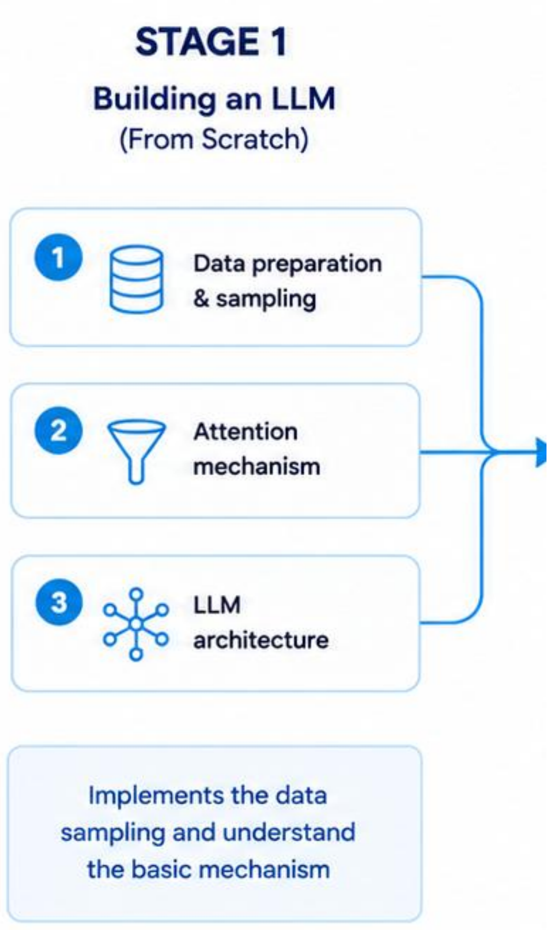
LLM build process



Adapted from Raschka, S. (2024). *Build a Large Language Model (From Scratch)*. Simon and Schuster.

What are Large Language Models (LLMs)?

LLM build process – Stage 1 – LLMs operate solely on 'tokens'

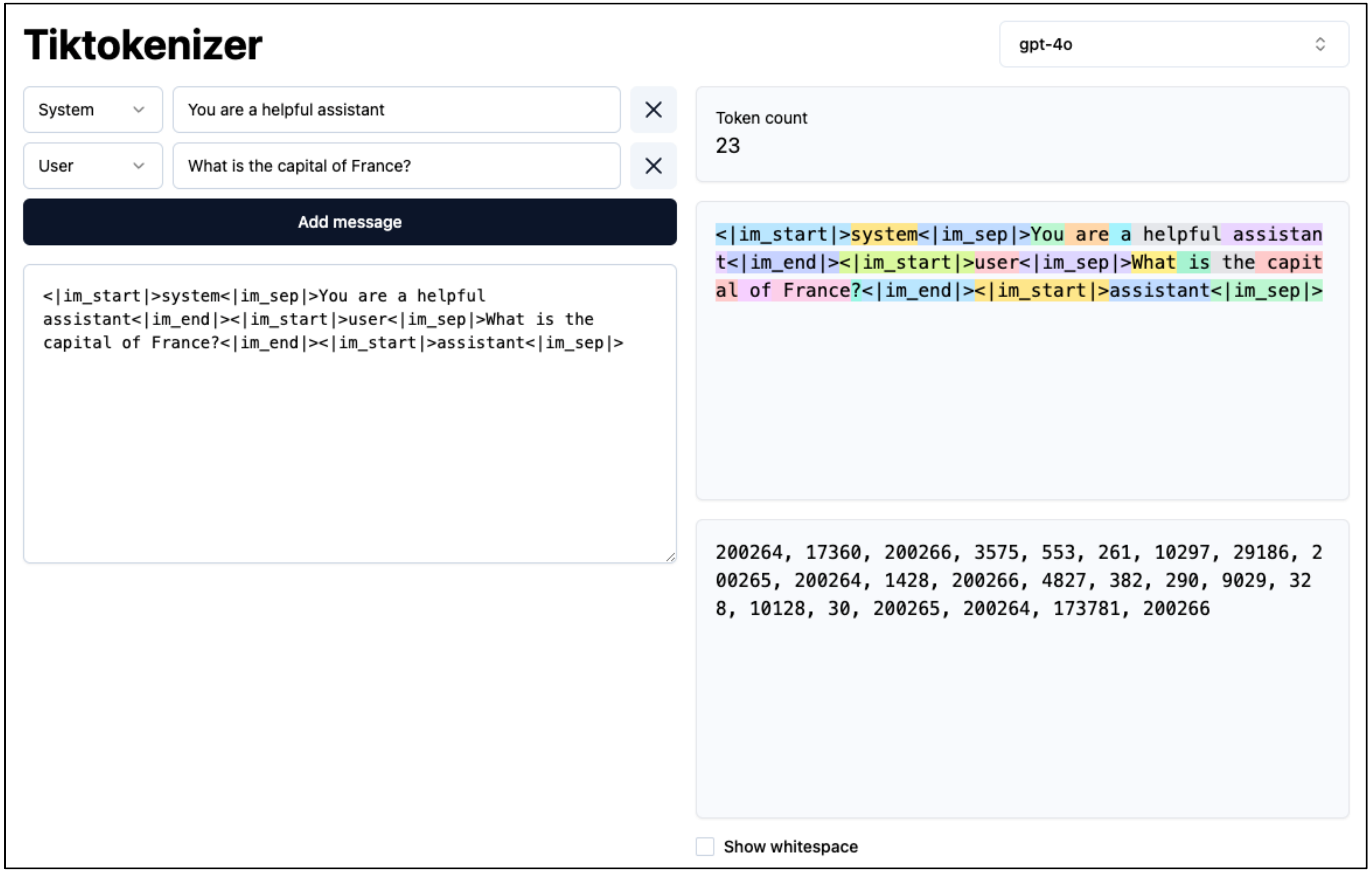


- LLMs use a non-standard lexicon (e.g., "misch" + "iev" + "ous")
  - alternative approaches use characters or bytes rather than word fragments (e.g., "d" +"o" + "g")
- LLMs are not given (explicit) word classes or grammar rules
  - e.g., subject, verbs, objects, past tense, present continuous, …
- LLMs are not given (explicit) word-level, sub-text-level, or text-level annotation
  - e.g., parts-of-speech (POS) tags, speaker descriptors, text-types descriptors, …
- LLMs are not given (explicit) register or genre distinctions
  - e.g., academic, conversational, formal, informal, science, engineering, …

[see Tiktoken: https://tiktokenizer.vercel.app/]

What are Large Language Models (LLMs)?

LLM build process – Stage 1 – LLMs operate solely on 'tokens'



https://tiktokenizer.vercel.app/?model=gpt-4o

What are Large Language Models (LLMs)?

LLM build process – Stage 1 – LLMs operate solely on 'tokens'

The cat sat on the mat.
The dangerous dog chased after the mischievous cat.

```
The cat sat on the mat.
The dangerous dog chased after the mischievous cat.
```


https://tiktokenizer.vercel.app/?model=gpt-4o

What are Large Language Models (LLMs)?

LLM build process – Stage 1 – LLMs operate solely on 'tokens'

System prompt: You are a helpful assistant.
User prompt: What is the capital of France?
Assistant response: **?**

```
<|im_start|>system<|im_sep|>You are a helpful assistan
t<|im_end|><|im_start|>user<|im_sep|>What is the capit
al of France?<|im_end|><|im_start|>assistant<|im_sep|>
```

```
200264, 17360, 200266, 3575, 553, 261, 10297, 29186, 2
00265, 200264, 1428, 200266, 4827, 382, 290, 9029, 32
8, 10128, 30, 200265, 200264, 173781, 200266
```


https://tiktokenizer.vercel.app/?model=gpt-4o

What are Large Language Models (LLMs)?

LLM build process – Stage 1 – LLMs operate solely on 'tokens'

- LLMs use a non-standard lexicon (e.g., "misch" + "iev" + "ous")
  - alternative approaches use characters or bytes rather than word fragments (e.g., "d" +"o" + "g")
- LLMs are not given (explicit) word classes or grammar rules

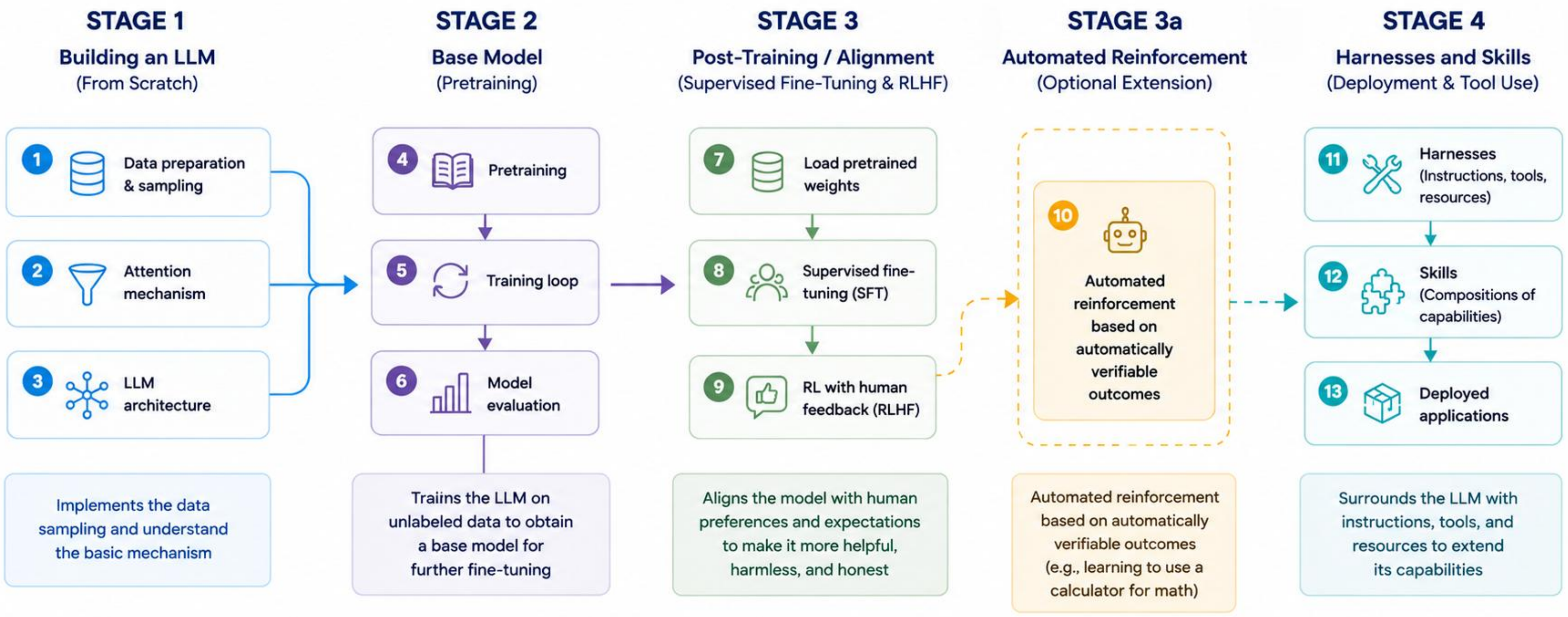
The categories that matter to HASS researchers are not explicitly encoded in LLMs.

  - e.g., parts-of-speech (POS) tags, speaker descriptors, text-types descriptors, …
- LLMs are not given (explicit) register or genre distinctions
  - e.g., academic, conversational, formal, informal, science, engineering, …

[see Tiktoken: https://tiktokenizer.vercel.app/]

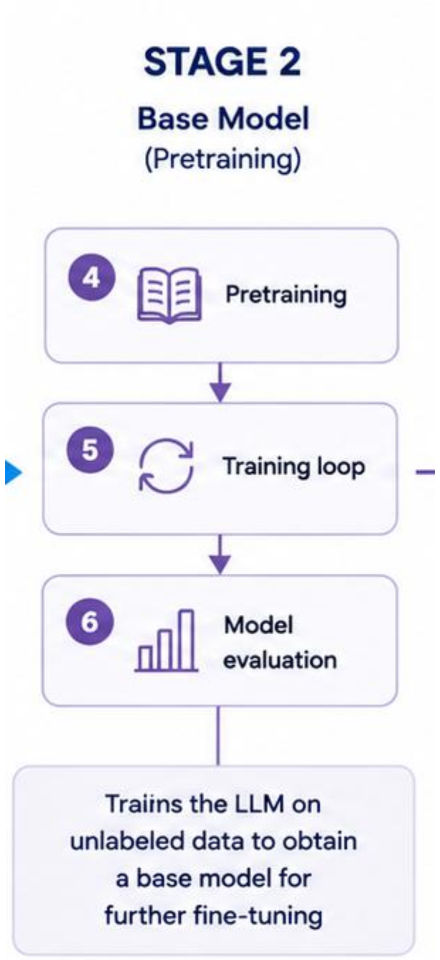
What are Large Language Models (LLMs)?

LLM build process – Stage 2 – LLMs use 'Big Data' to learn next word prediction



What are Large Language Models (LLMs)?

LLM build process – Stage 2 – LLMs use 'Big Data' to learn next word prediction



- Base models simply predict what comes next
  - They have no inherent drive to be helpful, truthful, safe, or useful
- Base models learn from enormous amounts of noisy, uncontrolled data
  - The data can be inaccurate, harmful, duplicated, copyrighted, etc.
- Base models require enormous resources to train
  - They demand huge amounts of computing power, energy, time, and money
- Base models learn the biases and stereotypes in their training data
  - e.g., “a woman’s place is in the home,” “men at work”

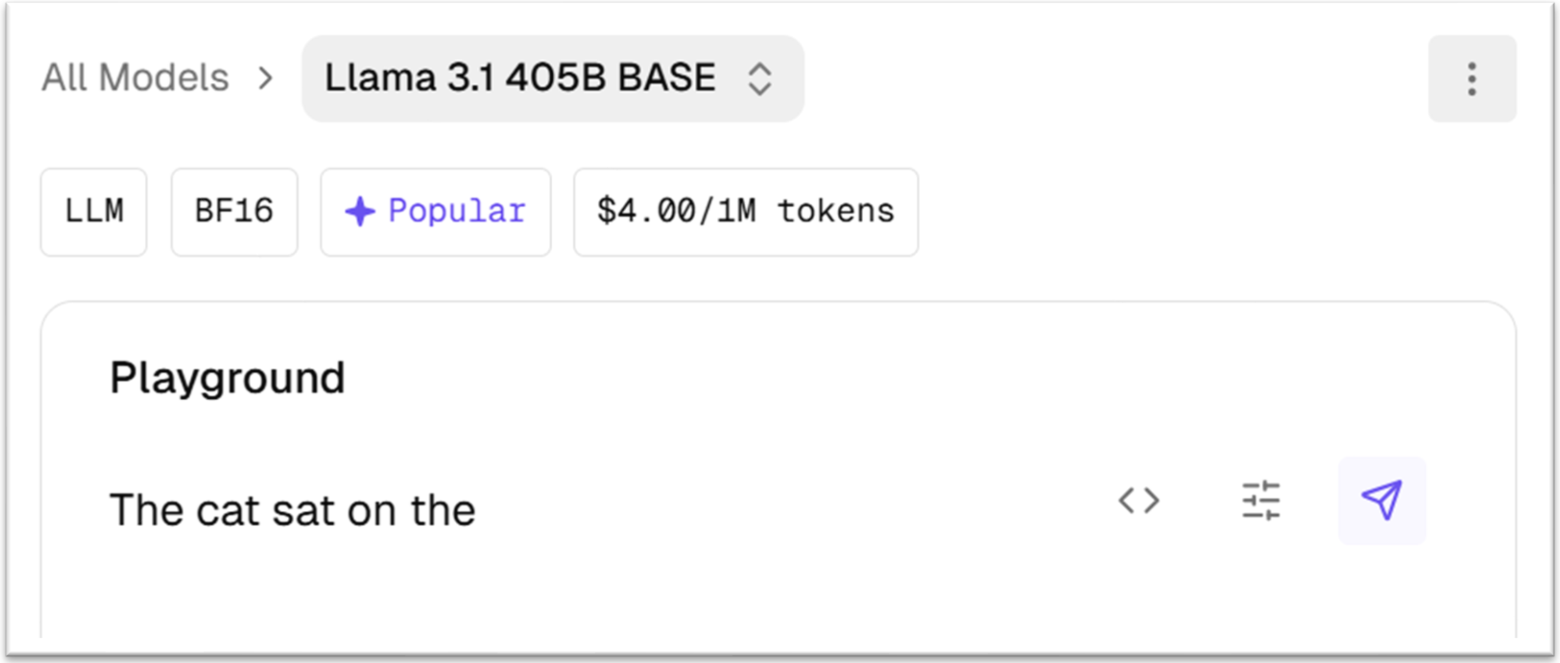
Llama 3.2

training data: 9 trillion tokens (9,000,000,000,000)

[see Tiktoken: https://tiktokenizer.vercel.app/]

What are Large Language Models (LLMs)?

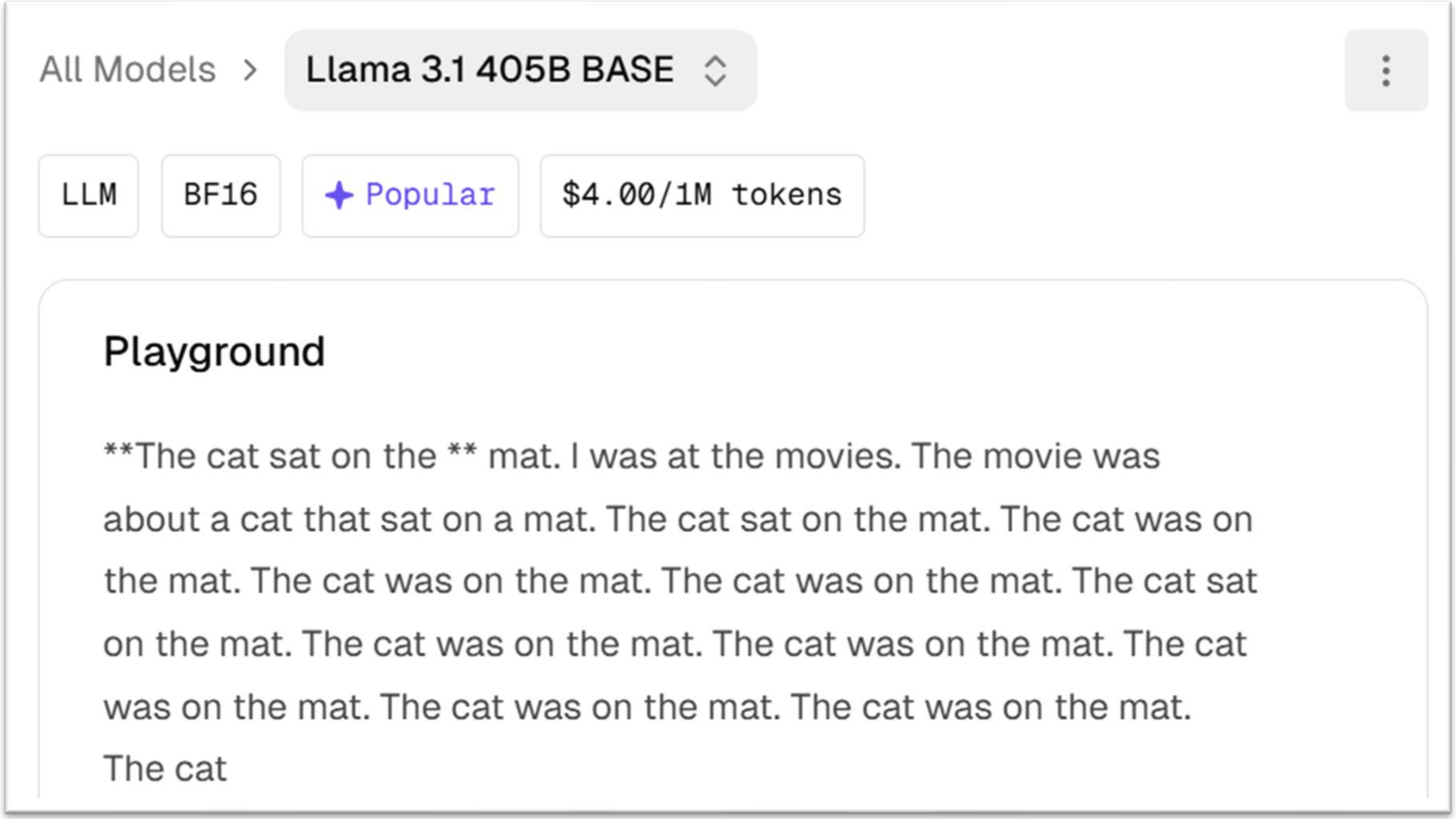
LLM build process – Stage 2 – Base models have no guidance/motivation/goals



https://app.hyperbolic.ai/models/llama31-405b-base-bf-16

What are Large Language Models (LLMs)?

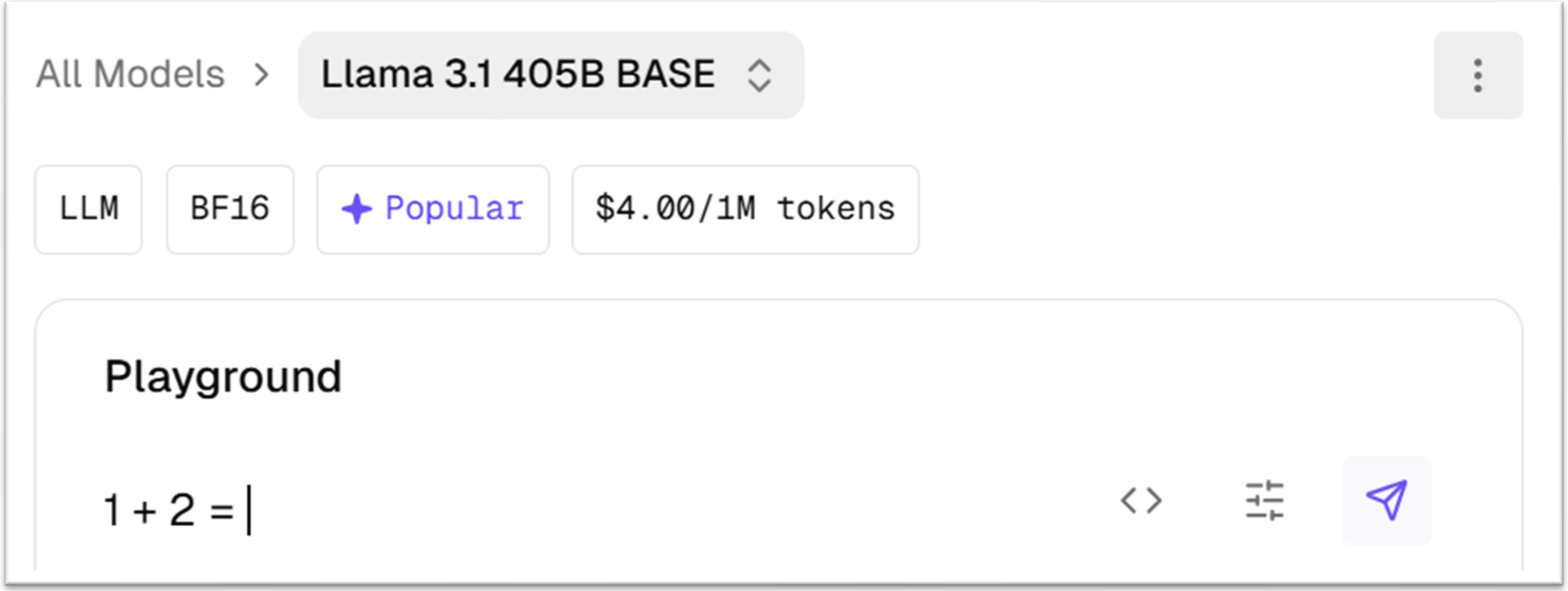
LLM build process – Stage 2 – Base models have no guidance/motivation/goals



https://app.hyperbolic.xyz/models/llama31-405b-base

What are Large Language Models (LLMs)?

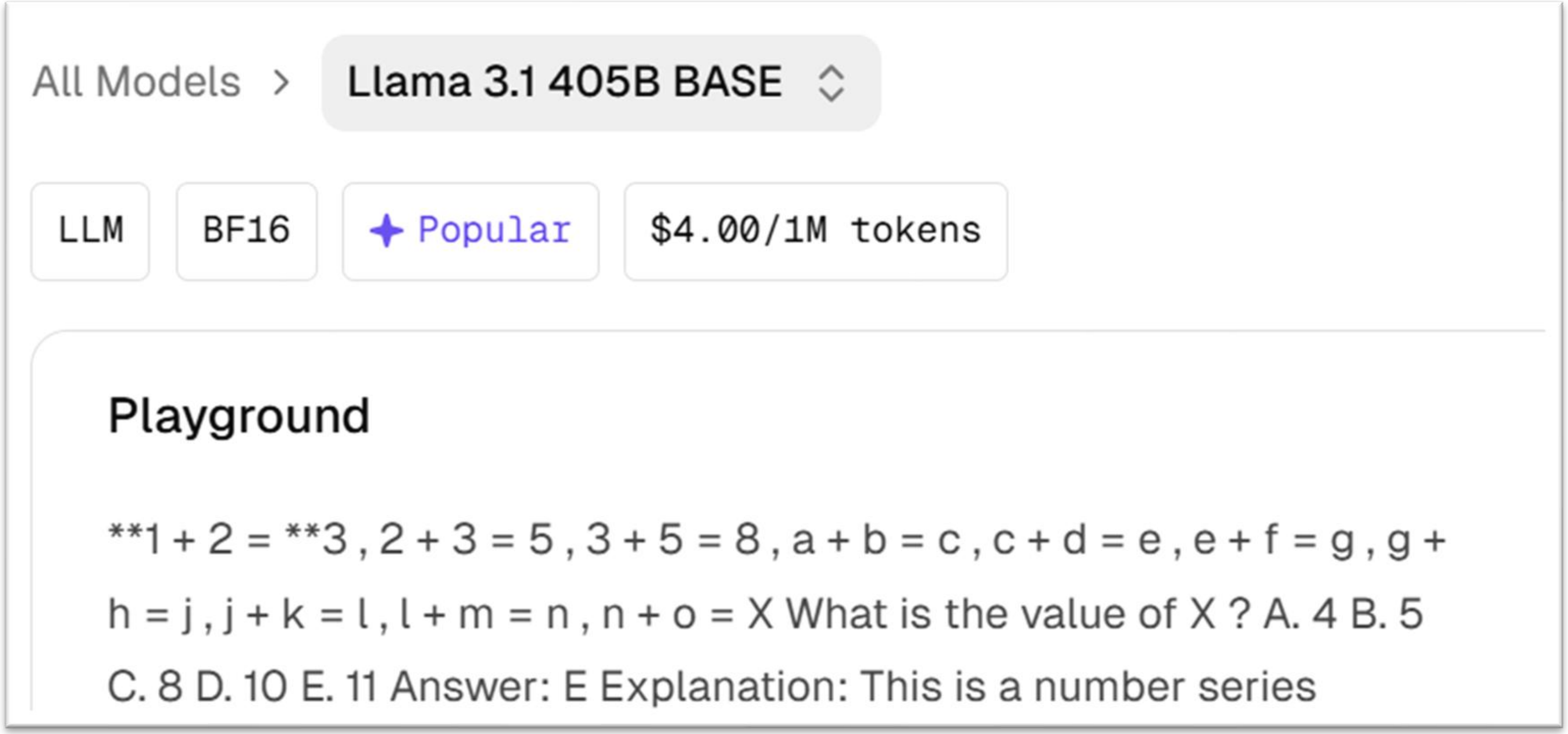
LLM build process – Stage 2 – Base models have no guidance/motivation/goals



https://app.hyperbolic.xyz/models/llama31-405b-base

What are Large Language Models (LLMs)?

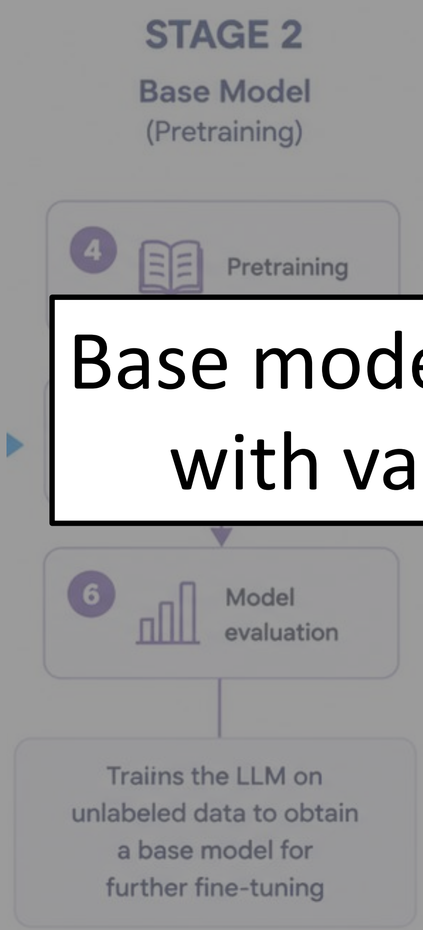
LLM build process – Stage 2 – Base models have no guidance/motivation/goals



https://app.hyperbolic.xyz/models/llama31-405b-base

What are Large Language Models (LLMs)?

LLM build process – Stage 2 – LLMs use 'Big Data' to learn next word prediction



- Base models simply predict what comes next
  - They have no inherent drive to be helpful, truthful, safe, or useful
- Base models learn from enormous amounts of noisy, uncontrolled data

Base models are "diamonds in the rough" for HASS researchers, with valuable information that requires careful extraction.

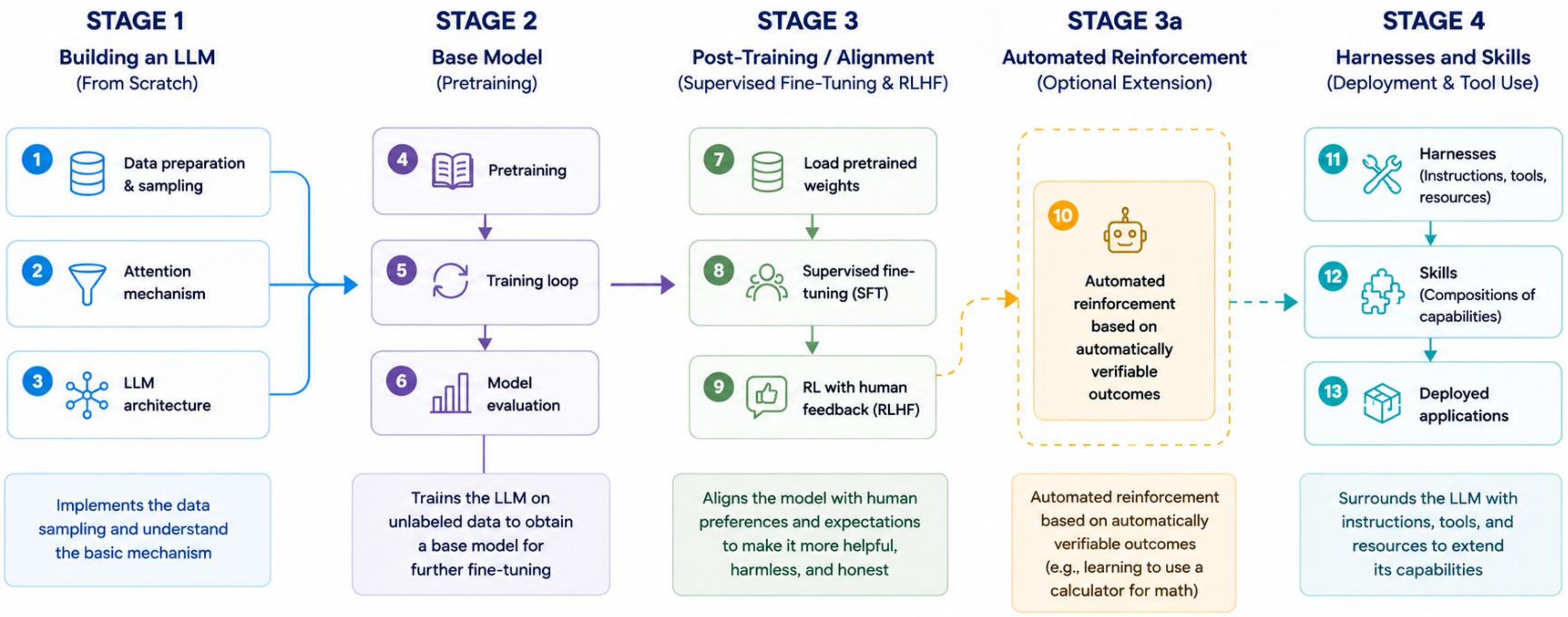
oney

- Base models learn the biases and stereotypes in their training data
  - e.g., "a woman's place is in the home," "men at work"

[see Tiktoken: https://tiktokenizer.vercel.app/]

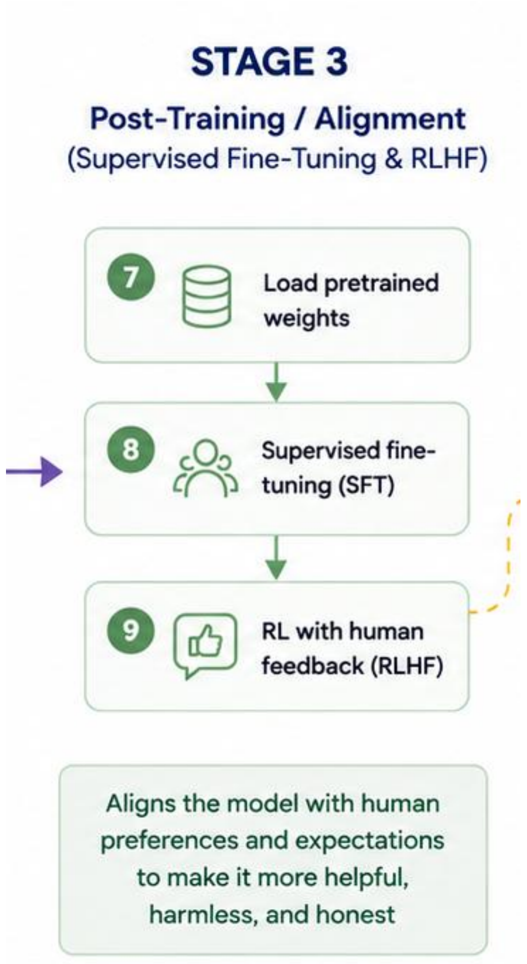
What are Large Language Models (LLMs)?

LLM build process – Stage 3 – LLMs receive guidance, intentions, and goals



What are Large Language Models (LLMs)?

LLM build process – Stage 3 – LLMs receive guidance, intentions, and goals



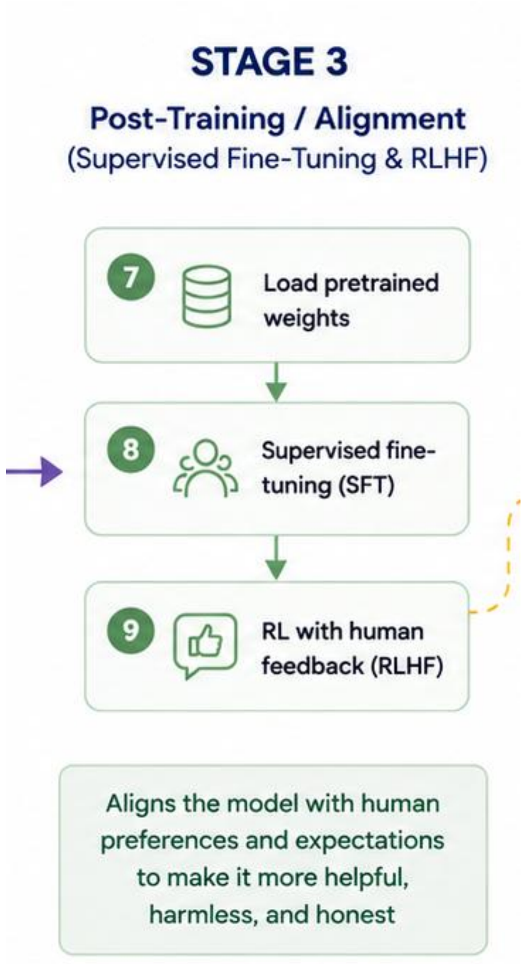
- Llama 3.2 fine tuning
  - "In post-training, we ... engage in synthetic data generation that goes through careful data processing and filtering to ensure high quality. We carefully blend the data to optimize for high quality across multiple capabilities like summarization, rewriting, instruction following, language reasoning, and tool use."

Meta

https://ai.meta.com/blog/llama-3-2-connect-2024-vision-edge-mobile-devices/

What are Large Language Models (LLMs)?

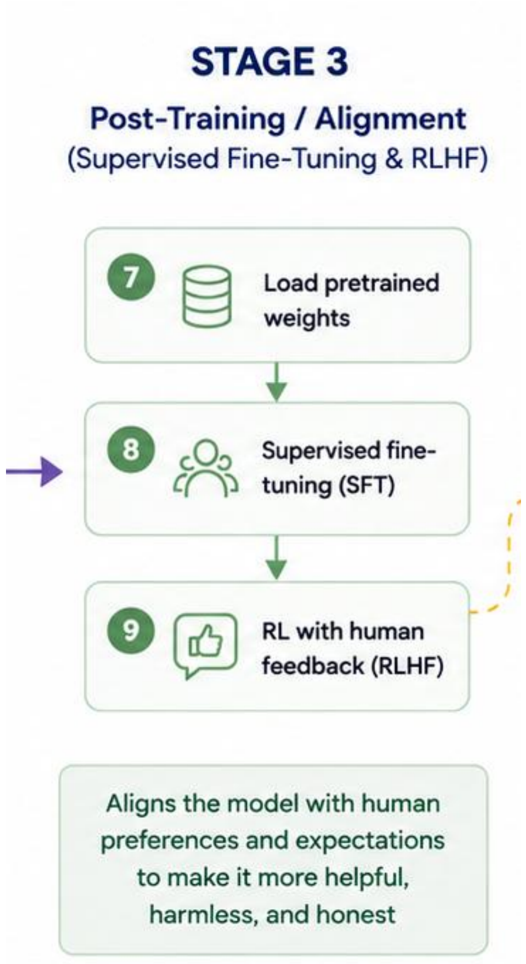
LLM build process – Stage 3 – LLMs receive guidance, intentions, and goals

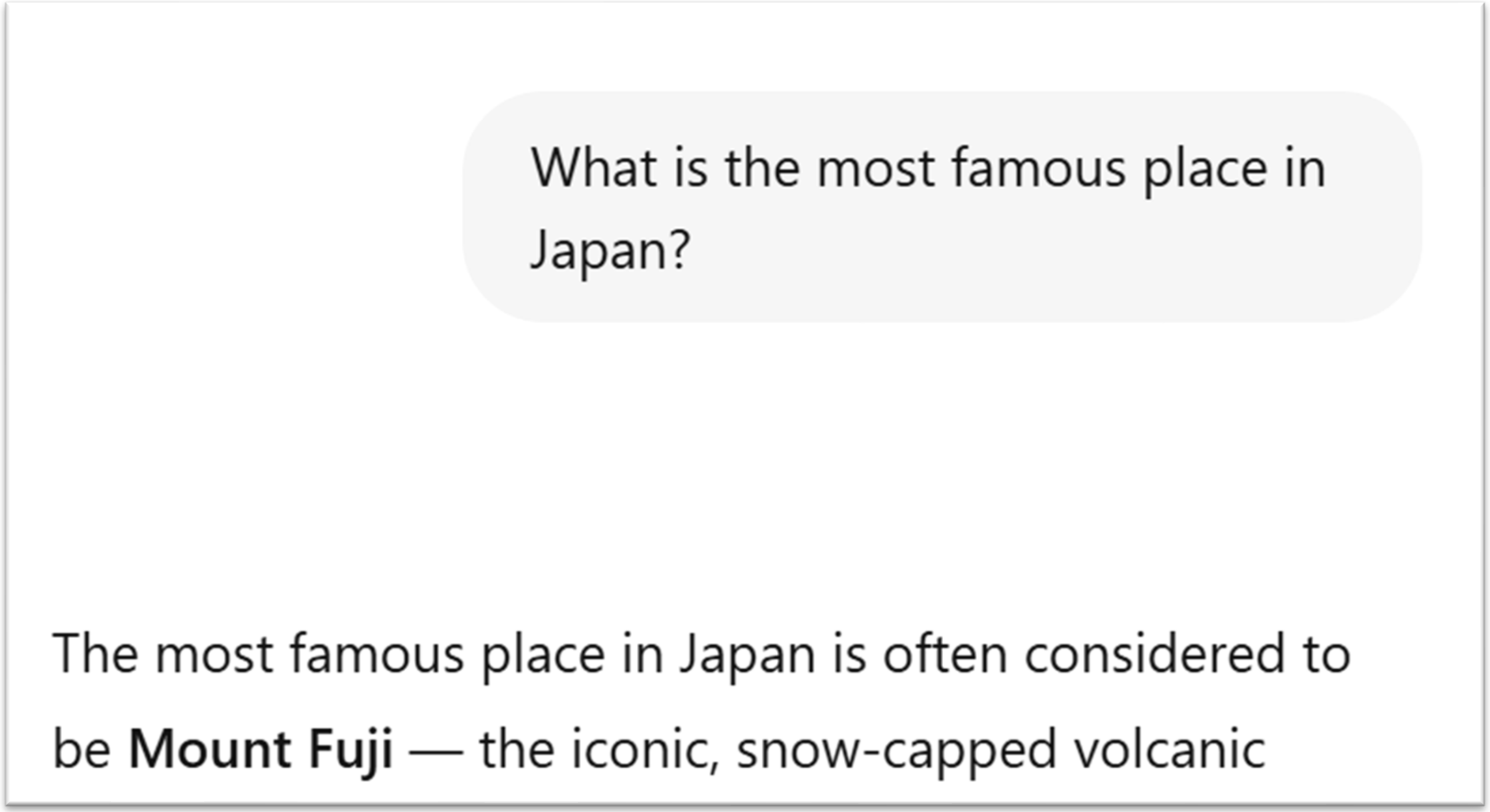


- Pseudo example of fine tuning
  - **System:** You are a helpful assistant.
  - **Human**: What is the capital of France?
  - **Assistant**: Paris
- Real example of fine tuning (Anthropic/hh-rlhf)
  - **Human**: What kind of noises did dinosaurs make?
  - **Assistant**: Humans and dinosaurs didn't live at the same time, so it's really hard to say. The best place to find out what noises dinosaurs made would be...

What are Large Language Models (LLMs)?

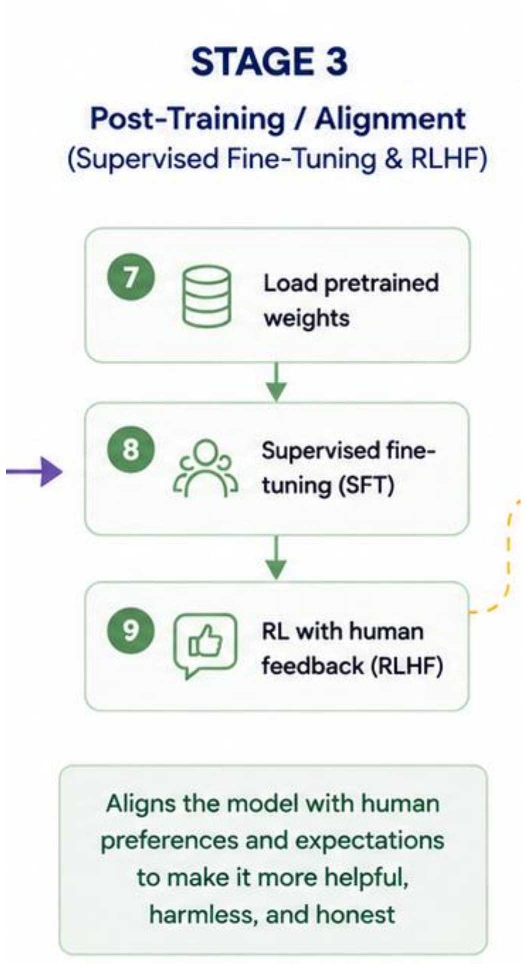
LLM build process – Stage 3 – LLMs receive guidance, intentions, and goals





What are Large Language Models (LLMs)?

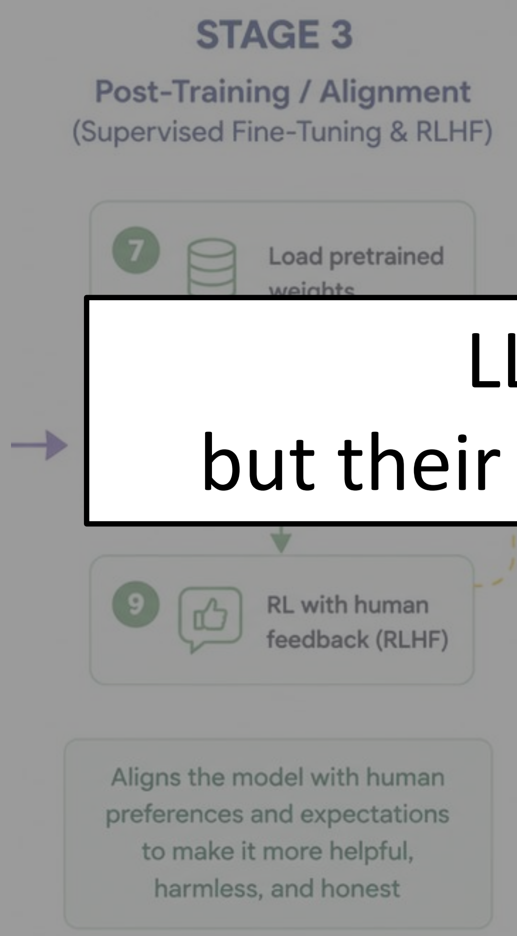
LLM build process – Stage 3 – LLMs receive guidance, intentions, and goals



- Fine-tuning fundamentally changes model behavior
  - Fine-tuning turns base models into useful tools, e.g., classifiers, chatbots, assistants
- Fine-tuning embeds human choices and values in the model
  - RLHF rewards behaviors that humans prefer
    - helpful, truthful, appropriate, safe responses
  - RLHF discourages behaviors that humans do not want
    - harmful, dangerous, biased, socially inappropriate responses

What are Large Language Models (LLMs)?

LLM build process – Stage 3 – LLMs receive guidance, intentions, and goals

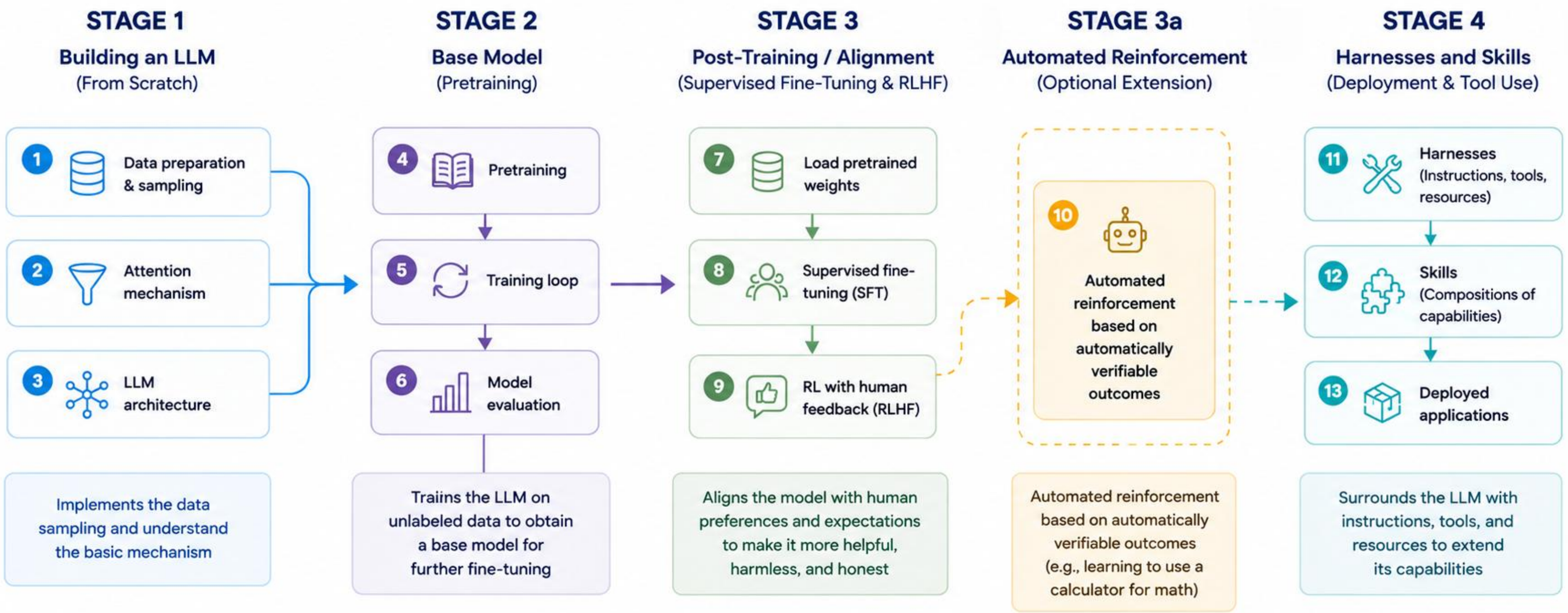


- Fine-tuning fundamentally changes model behavior
  - Fine-tuning turns base models into useful tools, e.g., classifiers, chatbots, assistants
- Fine-tuning embeds human choices and values in the model
    - harmful, dangerous, biased, socially inappropriate responses

LLMs are useful tools for HASS researchers, but their behavior is shaped by human choices and values.

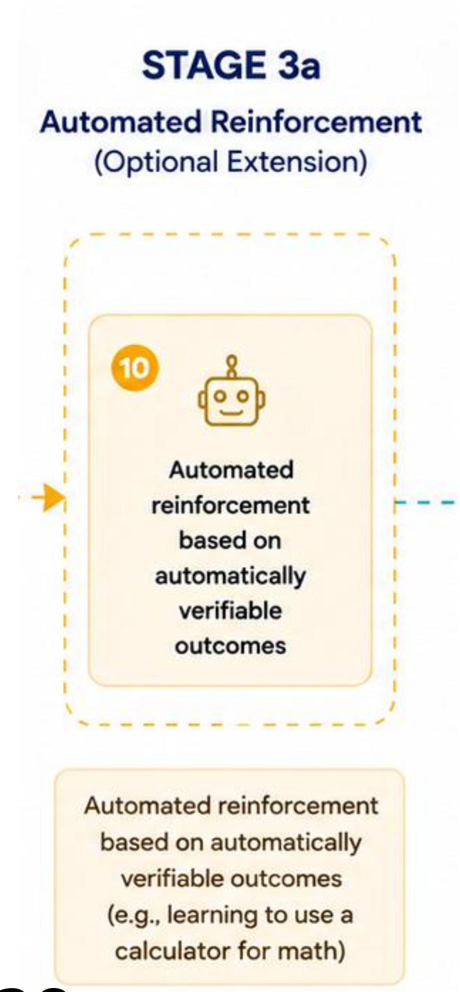
What are Large Language Models (LLMs)?

LLM build process – Stage 3a – Building 'super intelligent' systems?

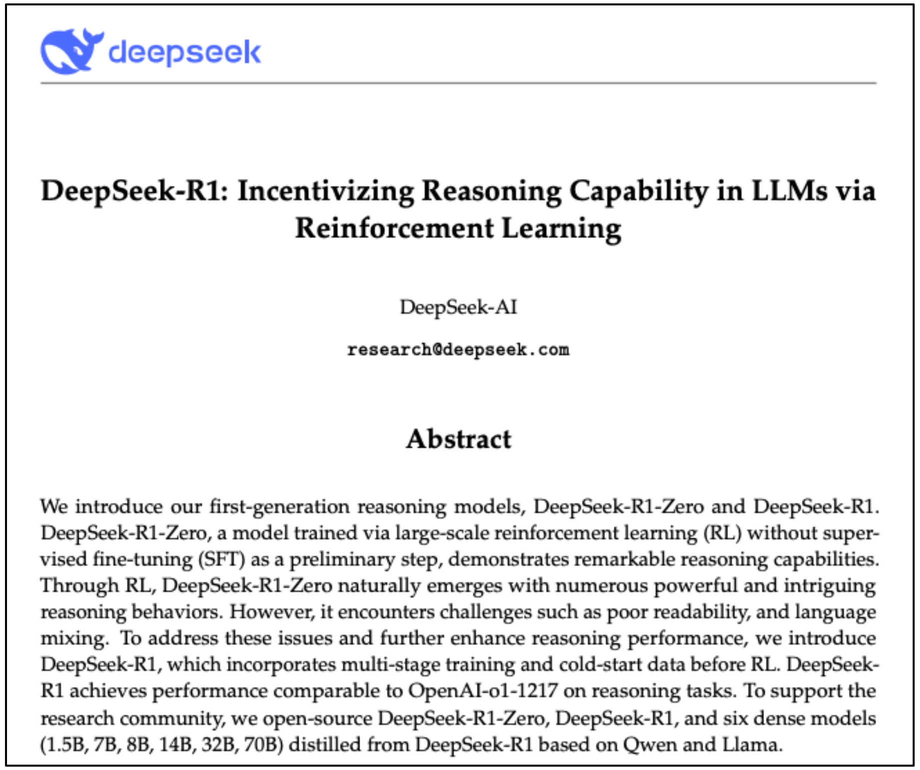


What are Large Language Models (LLMs)?

LLM build process – Stage 3a – Building 'super intelligent' systems?



- Automated reinforcement based on automatically verifiable outcomes
  - e.g., learning to use a calculator for math, ...



deepseek

DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning

DeepSeek-AI

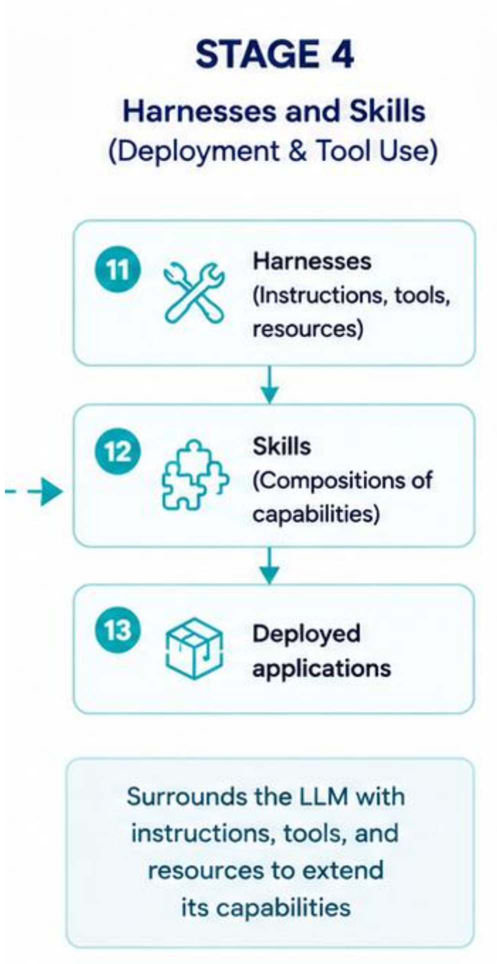
research@deepseek.com

Abstract

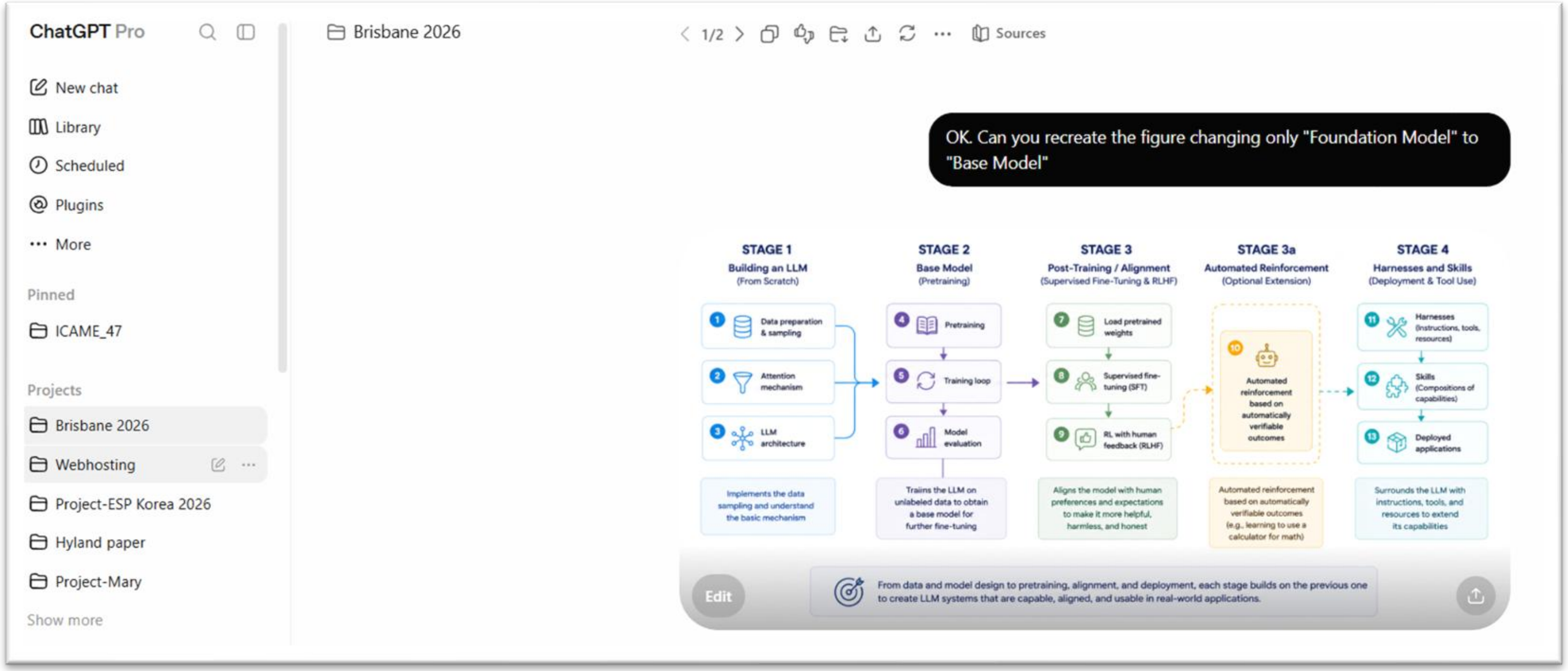
We introduce our first-generation reasoning models, DeepSeek-R1-Zero and DeepSeek-R1. DeepSeek-R1-Zero, a model trained via large-scale reinforcement learning (RL) without supervised fine-tuning (SFT) as a preliminary step, demonstrates remarkable reasoning capabilities. Through RL, DeepSeek-R1-Zero naturally emerges with numerous powerful and intriguing reasoning behaviors. However, it encounters challenges such as poor readability, and language mixing. To address these issues and further enhance reasoning performance, we introduce DeepSeek-R1, which incorporates multi-stage training and cold-start data before RL. DeepSeek-R1 achieves performance comparable to OpenAI-o1-1217 on reasoning tasks. To support the research community, we open-source DeepSeek-R1-Zero, DeepSeek-R1, and six dense models (1.5B, 7B, 8B, 14B, 32B, 70B) distilled from DeepSeek-R1 based on Qwen and Llama.

What are Large Language Models (LLMs)?

LLM build process – Stage 3a – Building 'super intelligent' systems?

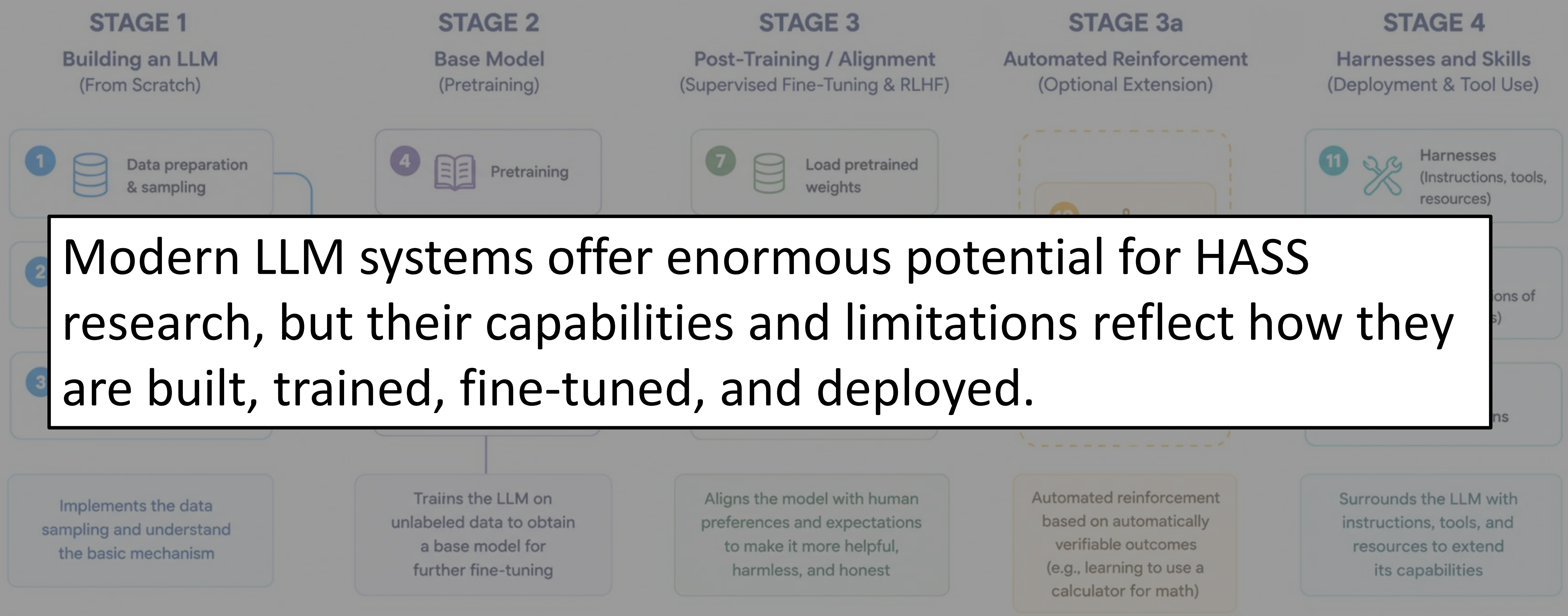


- Surround the LLM with instructions, tools, and resources
  - *e.g., web search, code execution, file access, ...*



What are Large Language Models (LLMs)?

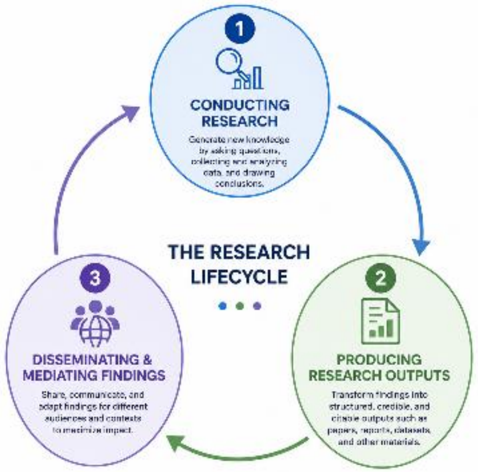
Interim summary



Modern LLM systems offer enormous potential for HASS research, but their capabilities and limitations reflect how they are built, trained, fine-tuned, and deployed.

Opportunities and challenges for AI in HASS research

conducting research; producing research outputs; disseminating and mediating findings



Faculty of Science and Engineering

Opportunities and challenges for AI in HASS research

Cloud-based AI

Proprietary; Public; Commercial;
Better performance (fast)
More reliable (less hallucinations)

ChatGPT

Claude

Gemini

Grok

Microsoft Copilot

Local AI

Open Source/Open Weight; Private; Free;
Worse performance (slow)
Less reliable (more hallucinations)

Hugging Face
https://huggingface.co/

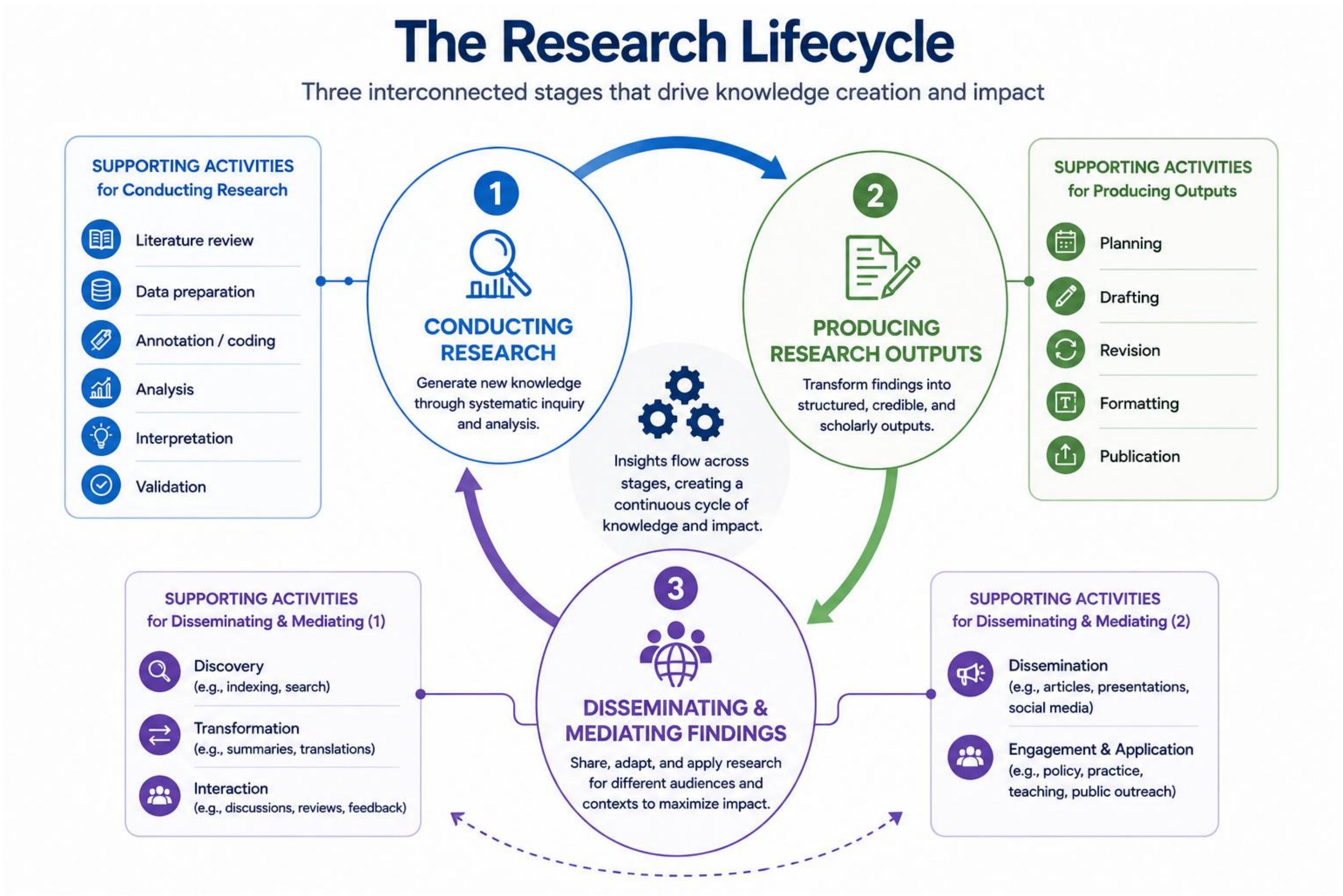
LM Studio
https://lmstudio.ai/

https://ollama.com/

https://github.com/mozilla-ai/llamafile

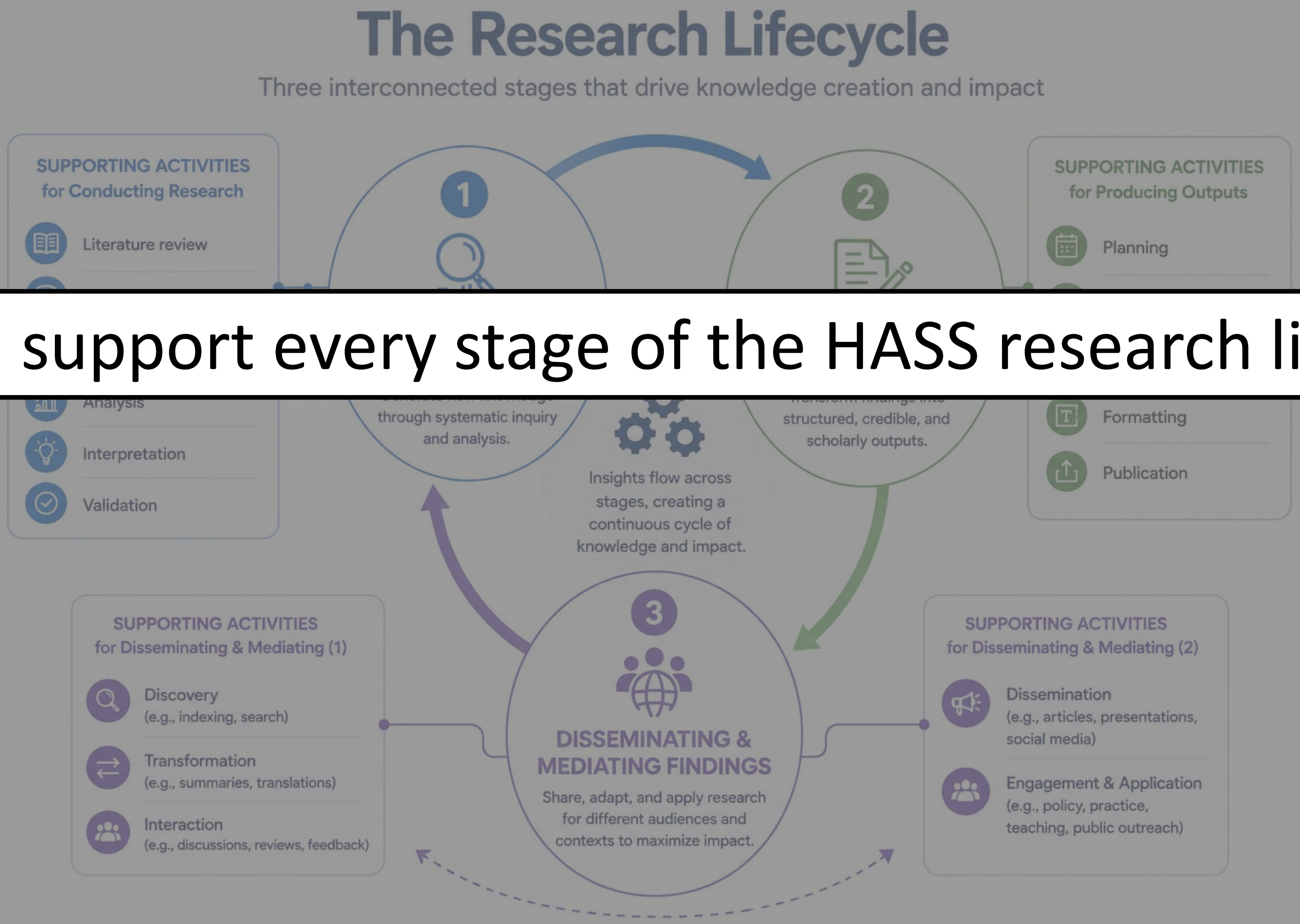
Opportunities and challenges for AI in HASS research

Three stages of the research lifecycle



Opportunities and challenges for AI in HASS research

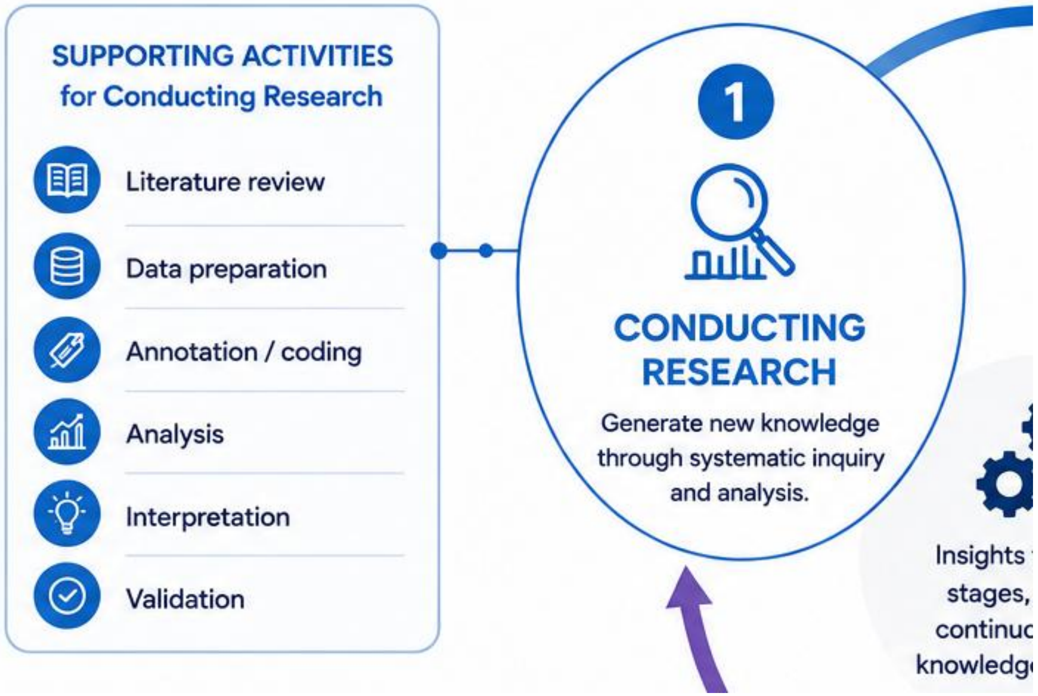
Three stages of the research lifecycle



LLMs can support every stage of the HASS research lifecycle.

Opportunities and challenges for AI in HASS research

Stage 1 – Conducting research – literature review

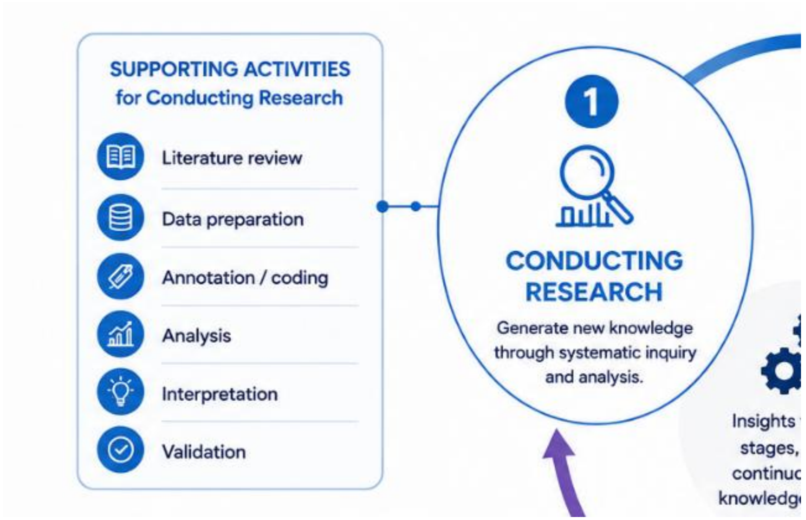


- Demo 1 – AI-assisted literature review
  - What does recent research tell us about the effects of generative AI on students' academic writing?
- Prompt
  - Find recent peer-reviewed research investigating the effects of generative AI on university students' academic writing. Identify the major findings, areas of agreement and disagreement, methodological approaches, and important limitations in the existing research. Provide links to the original sources and distinguish empirical studies from reviews or commentary.

Opportunities and challenges for AI in HASS research

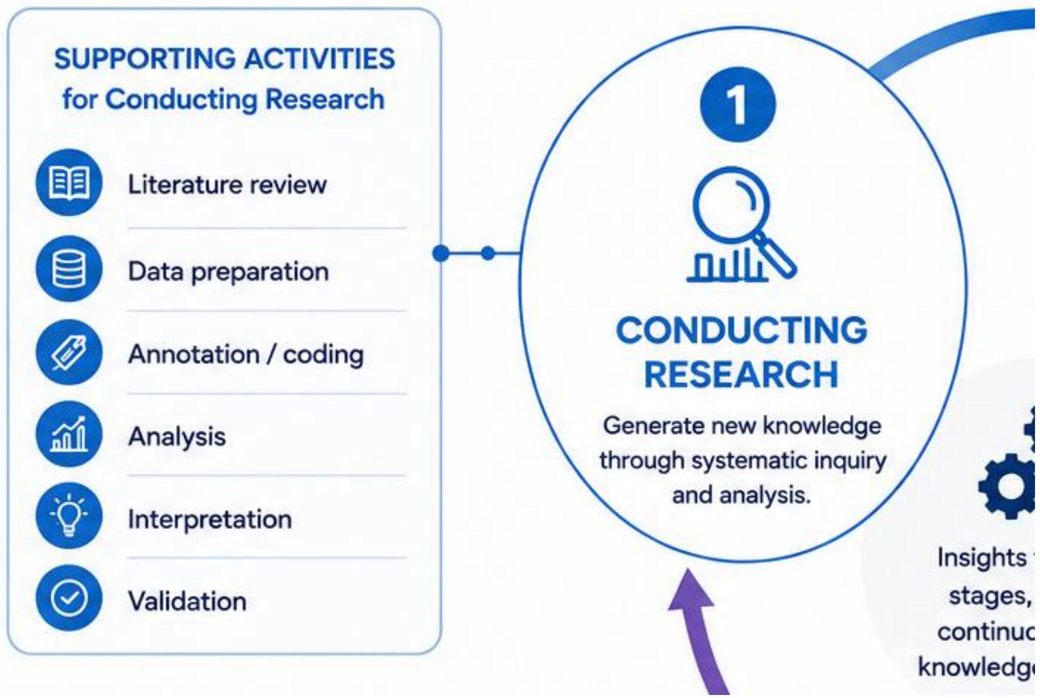
Stage 1 – Conducting research – literature review

- Would you cite the AI response?
  - Probably not.
- Would you use it to orient yourself rapidly in an unfamiliar literature?
  - Probably yes.
- Would you check the important papers yourself?
  - Certainly yes!



Opportunities and challenges for AI in HASS research

Stage 1 – Conducting research – data preparation



- Demo 2 – Preparing speech data for analysis
  - **Original data:** speech transcripts represented using eye-dialect
  - **Research need:** a standardized English version suitable for subsequent analysis Prompt
- Prompt
  - Convert the eye-dialect representation to standard written English while preserving the original lexical content, meaning, grammatical structure, hesitations, repetitions, and other features of the speech. Do not summarize, improve, or otherwise rewrite the speaker's language.

Opportunities and challenges for AI in HASS research

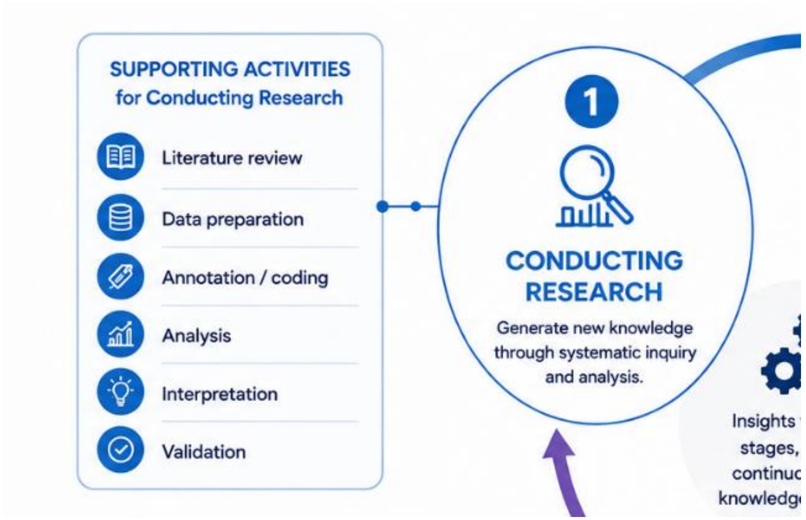
Stage 1 – Conducting research – data preparation

EYE-DIALECT TRANSCRIPT		STANDARDIZED TRANSCRIPT	
MsA:	'Ello	MsA:	Hello
Guy:	'Ello is Curly there	Guy:	Hello is Curly there
MsA:	Oo jis Who	MsA:	Oh just Who
Guy:	Johnnyh An sin	Guy:	Johnny An sin
MsA:	Oo j ista minnih	MsA:	Oh just a minute
MsA:	It's fer you dear	MsA:	It's for you dear
MsA:	I cutcher hair	MsA:	I cut your hair
MsA:	Kleenex This c'n go back in the trash can	MsA:	Kleenex This can go back in the trash can
Kid:	Okay	Kid:	Okay
MsA:	Watch ou- Getcher han'down	MsA:	Watch ou- Get your hand down
Joh:	Hello	Joh:	Hello
Guy:	Johnny	Guy:	Johnny
Joh:	Yeh	Joh:	Yeah
Guy:	Guy Detweiler	Guy:	Guy Detweiler
Joh:	Hi Guy how you doin	Joh:	Hi Guy how you doing

Opportunities and challenges for AI in HASS research

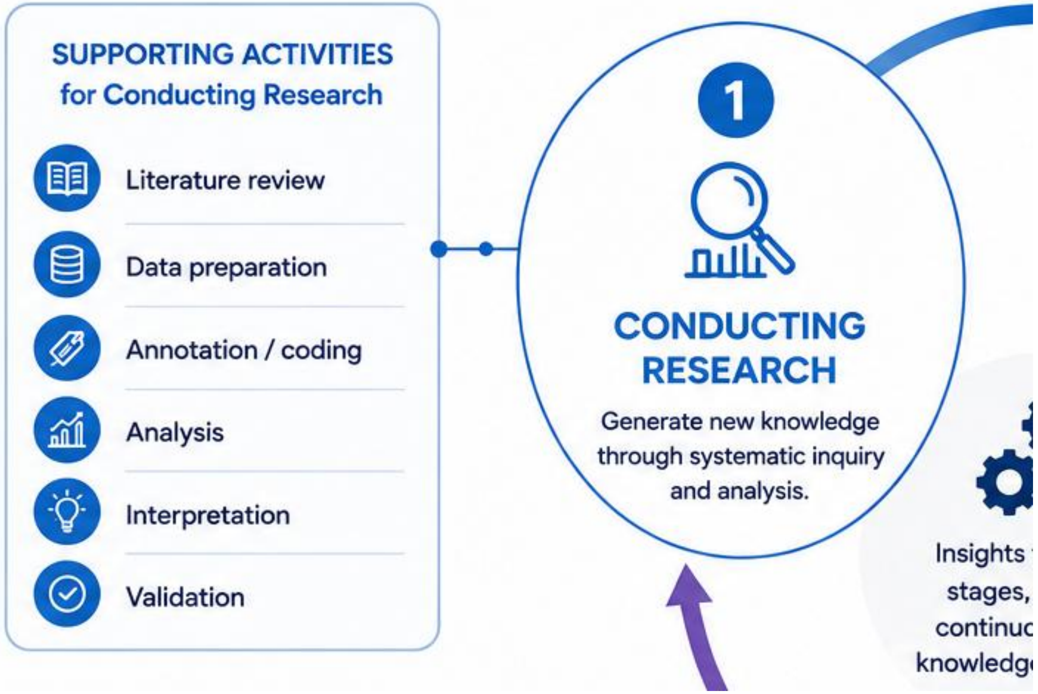
Stage 1 – Conducting research – data preparation

- Would you let the AI transform your research data?
  - Probably yes.
- Would you assume that nothing important was changed or lost?
  - Probably not.
- Would you validate the transformed data against the original?
  - Certainly yes!



Opportunities and challenges for AI in HASS research

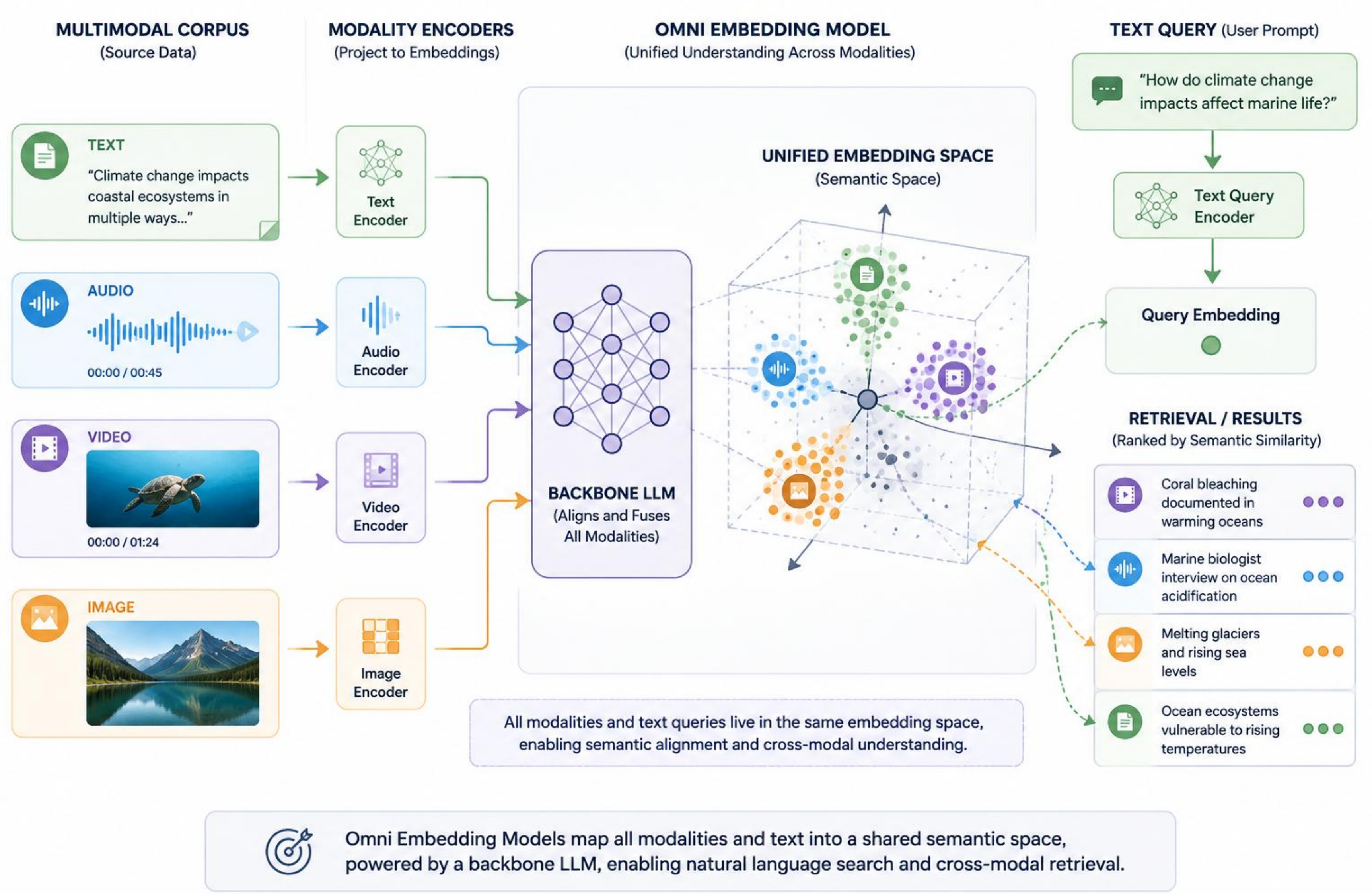
Stage 1 – Conducting research – data analysis



- Demo 3 – Semantic exploration of multimodal research data
  - **Omni embedding AI models** create
    A single semantic space for text, images, audio, and video exploration

Opportunities and challenges for AI in HASS research

Stage 1 – Conducting research – data analysis (Omni Embedding Models)

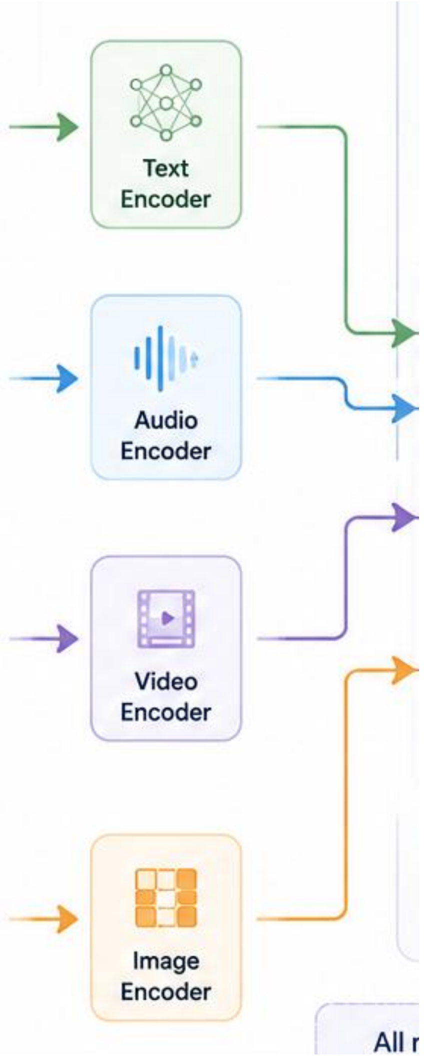


Opportunities and challenges for AI in HASS research

Stage 1 – Conducting research – data analysis (Omni Embedding Models)

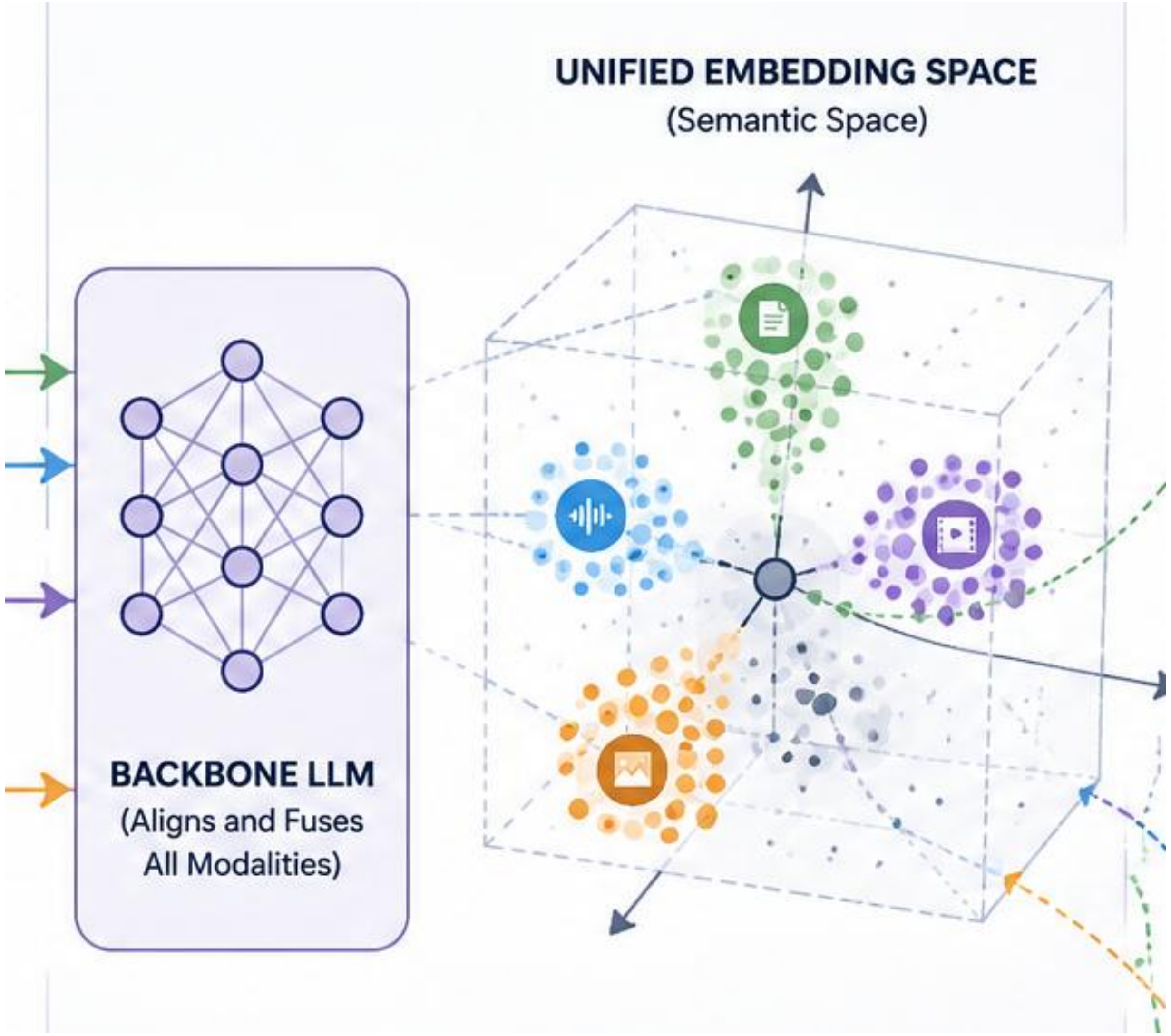
Opportunities and challenges for AI in HASS research

Stage 1 – Conducting research – data analysis (Omni Embedding Models)



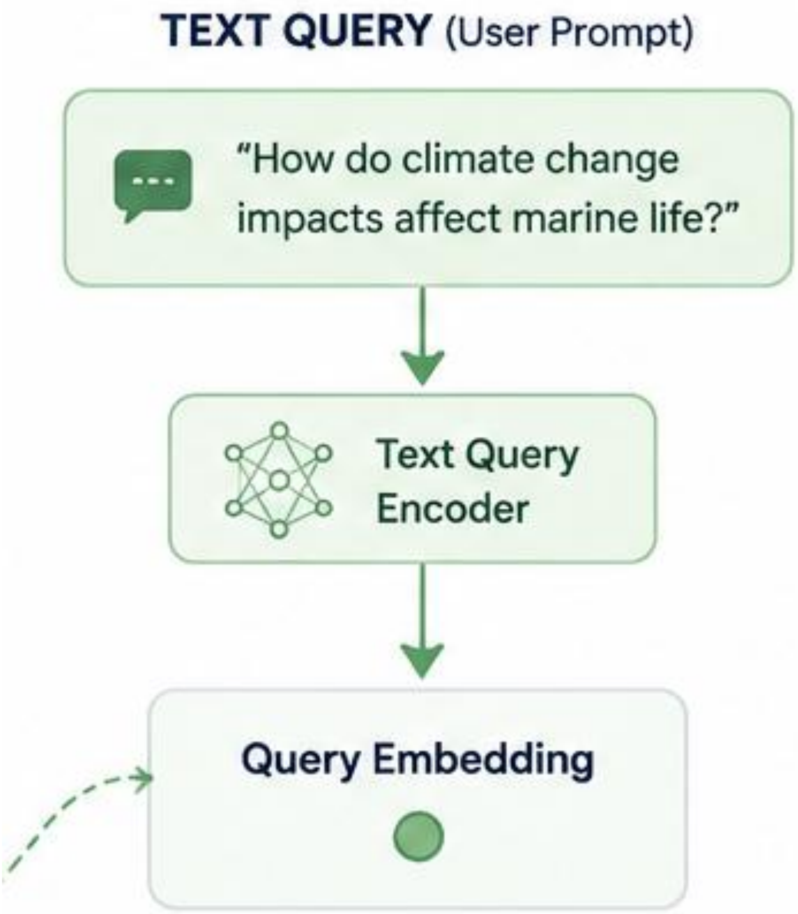
Opportunities and challenges for AI in HASS research

Stage 1 – Conducting research – data analysis (Omni Embedding Models)



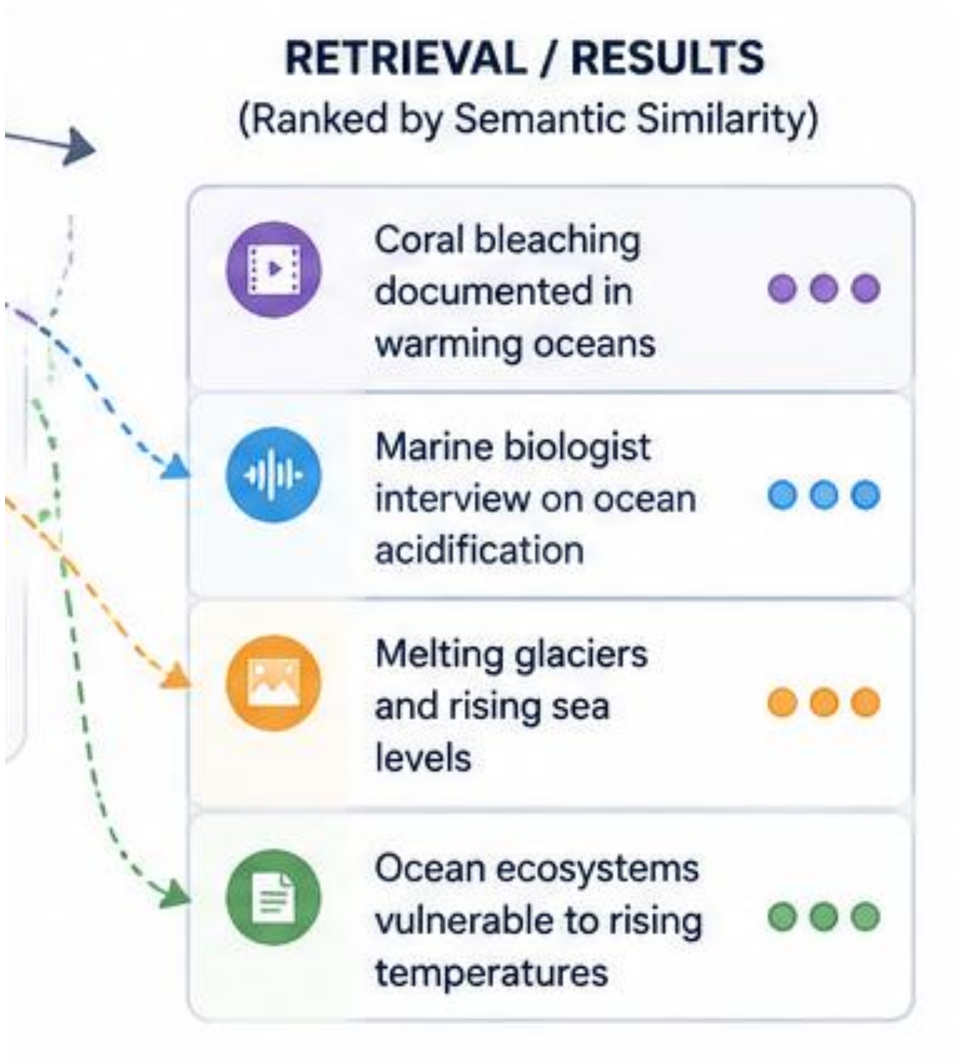
Opportunities and challenges for AI in HASS research

Stage 1 – Conducting research – data analysis (Omni Embedding Models)



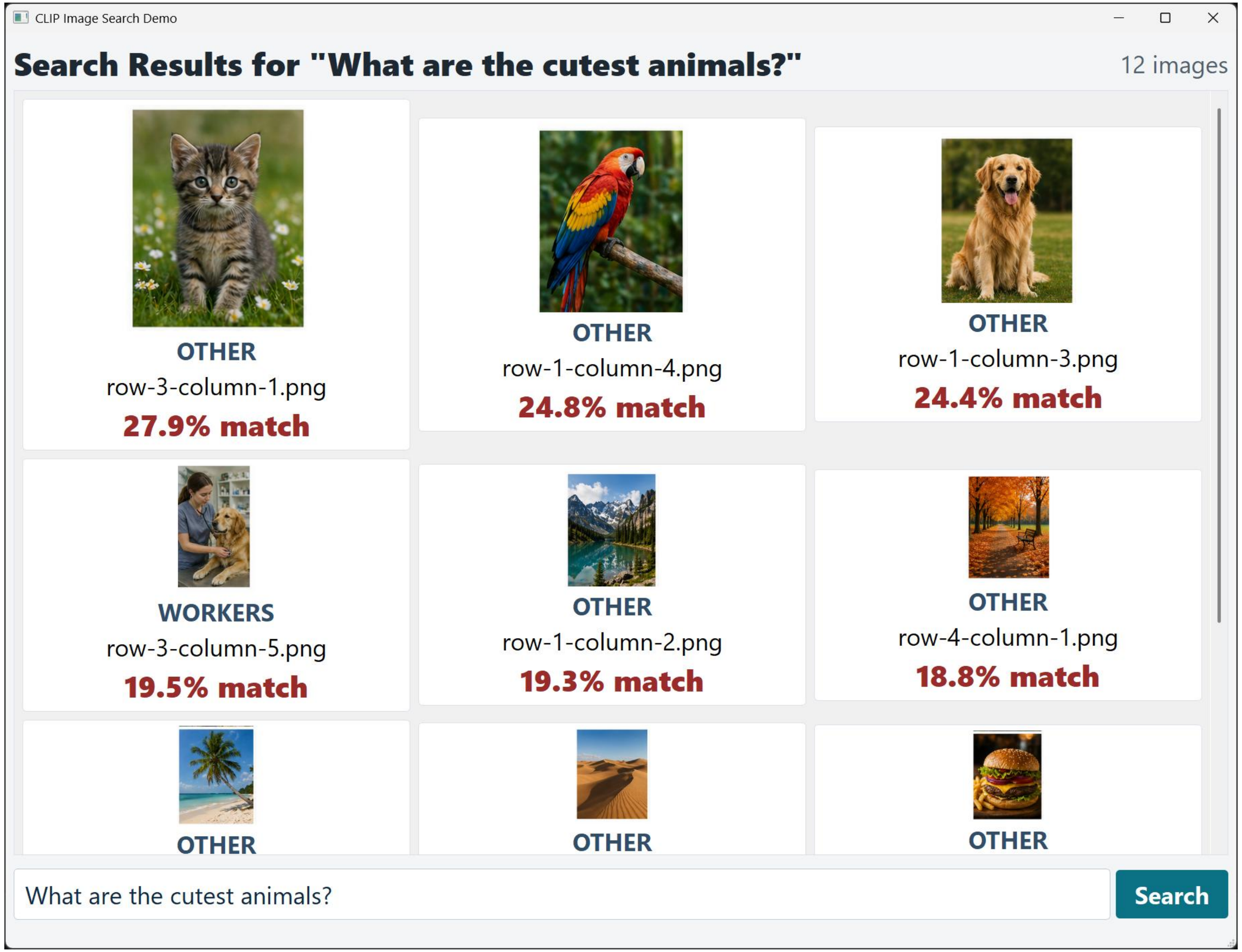
Opportunities and challenges for AI in HASS research

Stage 1 – Conducting research – data analysis (Omni Embedding Models)



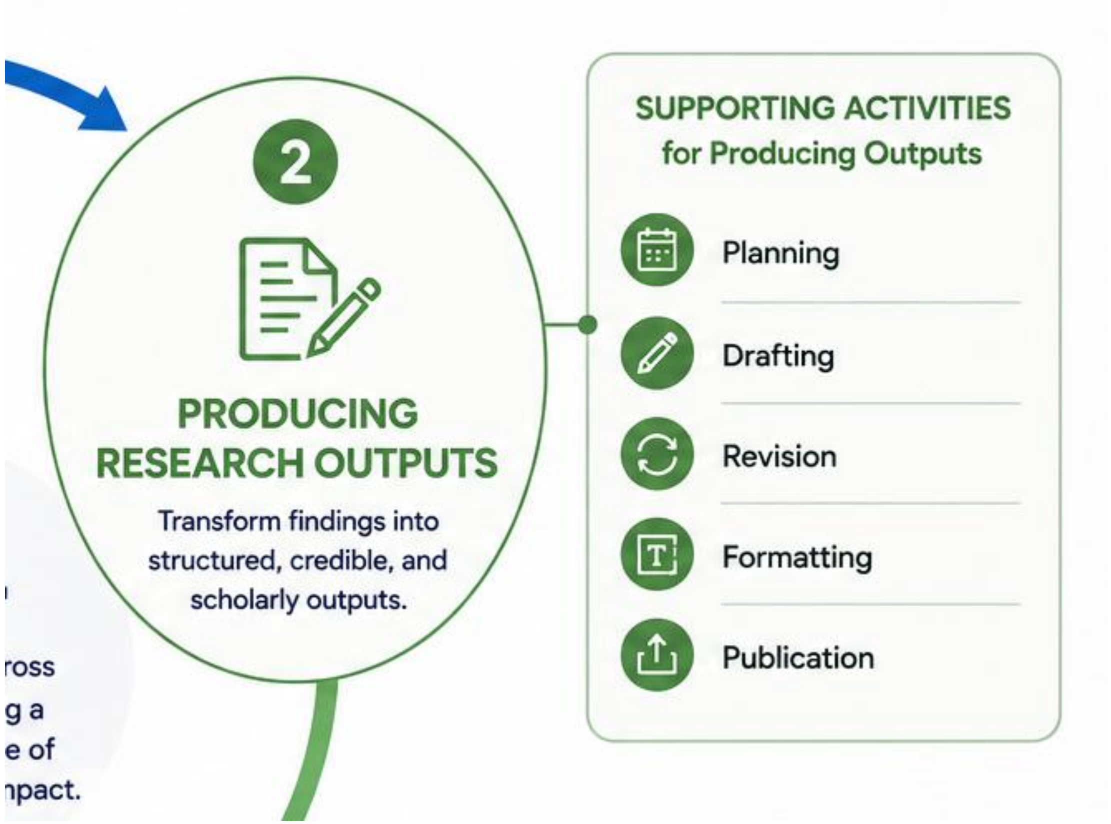
Opportunities and challenges for AI in HASS research

Stage 1 – Conducting research – data analysis (Omni Embedding Models)



Opportunities and challenges for AI in HASS research

Stage 2 – Producing research outputs – revising and proofing



- Demo 4 – Revising your writing style
  - LLMs can revise text for different styles, registers, and audiences
- Prompt
  - Revise the following paragraph to make it clearer, more concise, and more appropriate for an academic audience while preserving my meaning and writing style.

Opportunities and challenges for AI in HASS research

Stage 2 – Producing research outputs – revising and proofing

Todays abstract (before):
Large language models (LLMs) are rapidly changing what is possible in humanities and social sciences (HASS) research, but their appropriate role **is still not really clear**. This talk **look at** their influence across three interconnected stages of the research lifecycle: conducting research, producing research outputs, and disseminating and mediating findings for different audiences. First, I will explain the underlying **architechture** of modern LLMs and the implications of this design **in** HASS research. Next, I will consider how LLMs can support **many different kinds of activities**, ranging from literature review, data preparation, and qualitative analysis to writing, revision, **summarisation**, and translation. I will also **look at some of the important challenges with using these systems**, focusing on reliability, transparency, privacy, and the potential loss of scholarly agency. Finally, I will propose a human-centered future **where** researchers use LLMs to extend their capabilities while **still keeping control and being responsible for** every stage of the research lifecycle.

Todays abstract (after):
Large language models (LLMs) are rapidly changing what is possible in humanities and social sciences (HASS) research, but their appropriate role **remains unclear**. This talk **examines** their influence across three interconnected stages of the research lifecycle: conducting research, producing research outputs, and disseminating and mediating findings for different audiences. First, I will explain the underlying **architecture** of modern LLMs and the implications of this design **for** HASS research. Next, I will consider how LLMs can support **activities** ranging from literature review, data preparation, and qualitative analysis to writing, revision, **summarization**, and translation. I will also **explore key challenges**, focusing on reliability, transparency, privacy, and the potential loss of scholarly agency. Finally, I will propose a human-centered future **in which** researchers use LLMs to extend their capabilities while **retaining control over and responsibility for** every stage of the research lifecycle.

Opportunities and challenges for AI in HASS research

Stage 3 – Disseminating and mediating findings – adapting research outputs

- Demo 5 – Adapting research outputs
  - LLMs can transform research outputs for different audiences, purposes, and languages
- Prompt
  - Rewrite this abstract as a 100-word summary for a general audience with no specialist knowledge of AI or HASS. Preserve the main claims and avoid adding information that is not in the original.

Opportunities and challenges for AI in HASS research

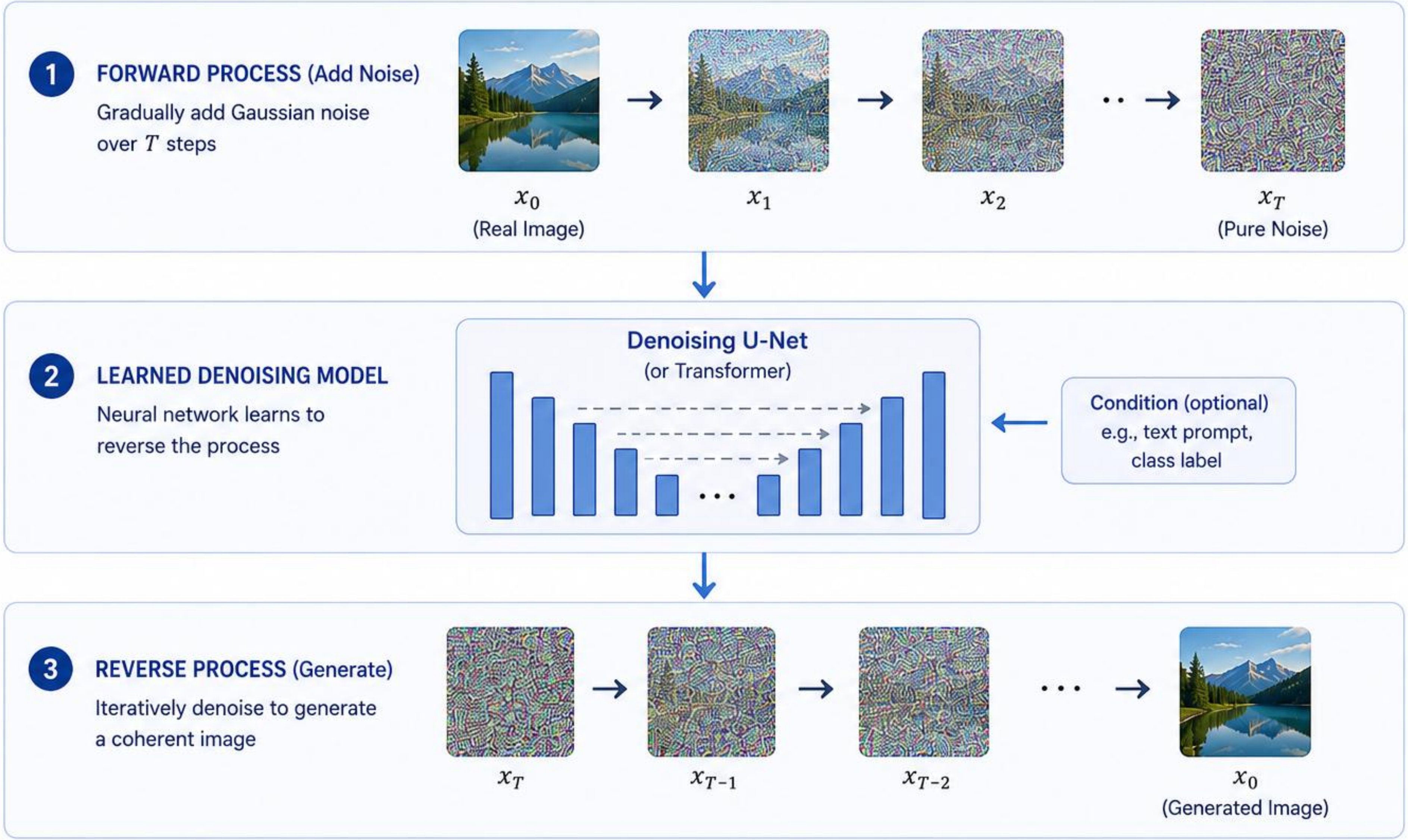
Stage 3 – Disseminating and mediating findings – adapting research outputs

Todays abstract (before):
Large language models (LLMs) are rapidly changing what is possible in humanities and social sciences (HASS) research, but their appropriate role remains unclear. This talk examines their influence across three interconnected stages of the research lifecycle: conducting research, producing research outputs, and disseminating and mediating findings for different audiences. First, I will explain the underlying architecture of modern LLMs and the implications of this design for HASS research. Next, I will consider how LLMs can support activities ranging from literature review, data preparation, and qualitative analysis to writing, revision, summarization, and translation. I will also explore key challenges, focusing on reliability, transparency, privacy, and the potential loss of scholarly agency. Finally, I will propose a human-centered future in which researchers use LLMs to extend their capabilities while retaining control over and responsibility for every stage of the research lifecycle.

Todays abstract (after):
Large language models (LLMs), **such as those behind many current AI tools**, are changing **how research in the humanities and social sciences is carried out and communicated**. This talk explores how LLMs can **assist researchers throughout the research process**, from **finding and analyzing information** to writing, revising, summarizing, translating, and **sharing their findings**. It also examines important concerns, including **whether AI-generated information can be trusted, how transparently these systems operate, how research data can be kept private, and whether researchers risk giving up too much control**. I argue that researchers should use LLMs to **enhance their capabilities while remaining responsible for their work**.

Opportunities and challenges for AI in HASS research

Stage 3 – Disseminating and mediating findings – diffusion-based image generation

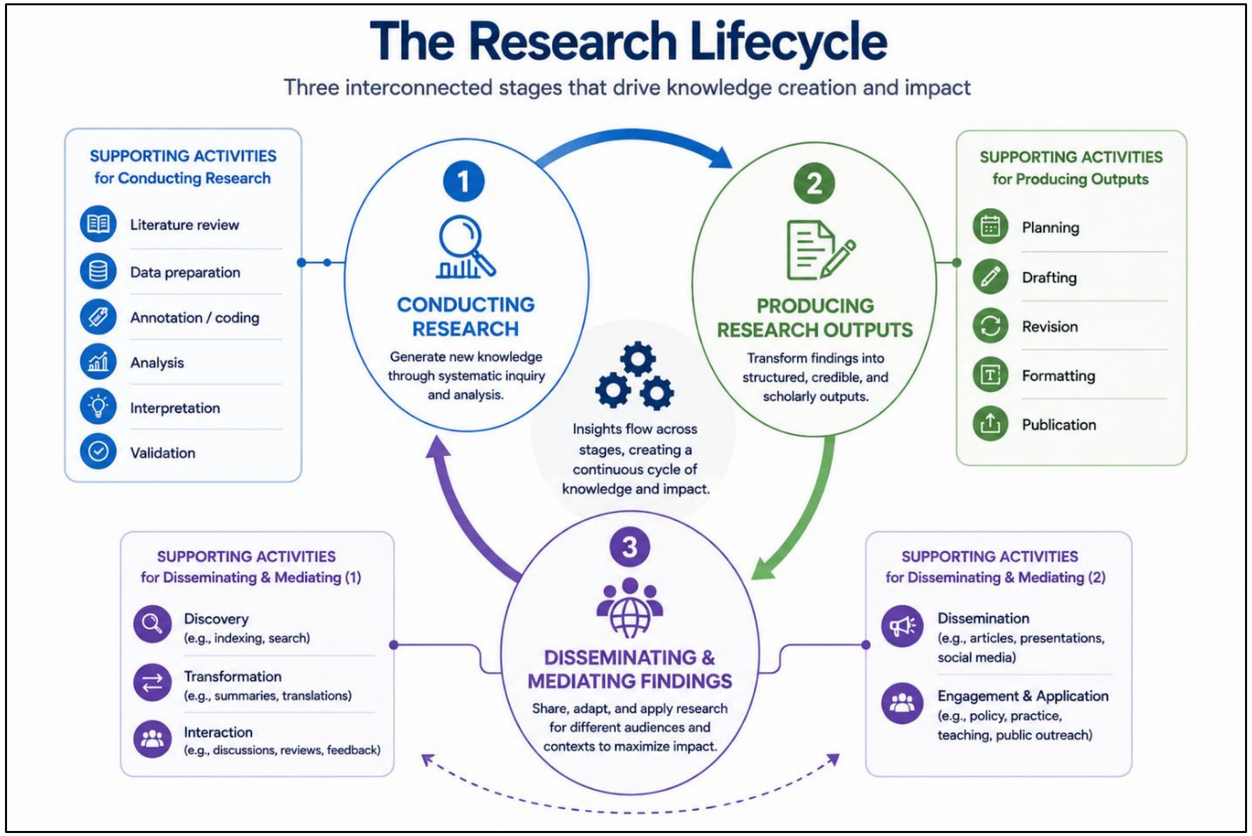


Opportunities and challenges for AI in HASS research

Stage 3 – Disseminating and mediating findings – diffusion-based image generation

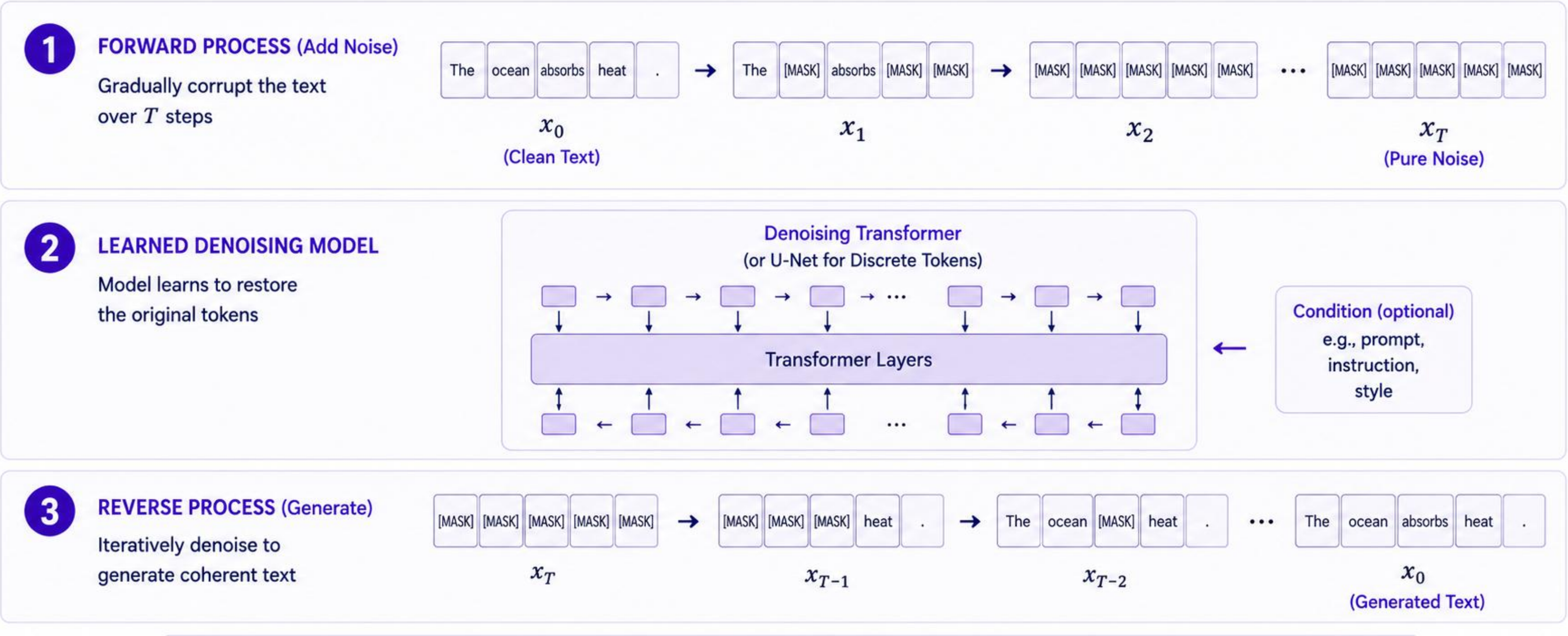
Three-part research lifecycle

Research	**Production**	**Mediation**
Literature review	Planning	Representation
Data preparation	Drafting	Discovery
Annotation/coding	Revision	Transformation
Analysis	Publication	Interaction
Interpretation		Dissemination
Validation		



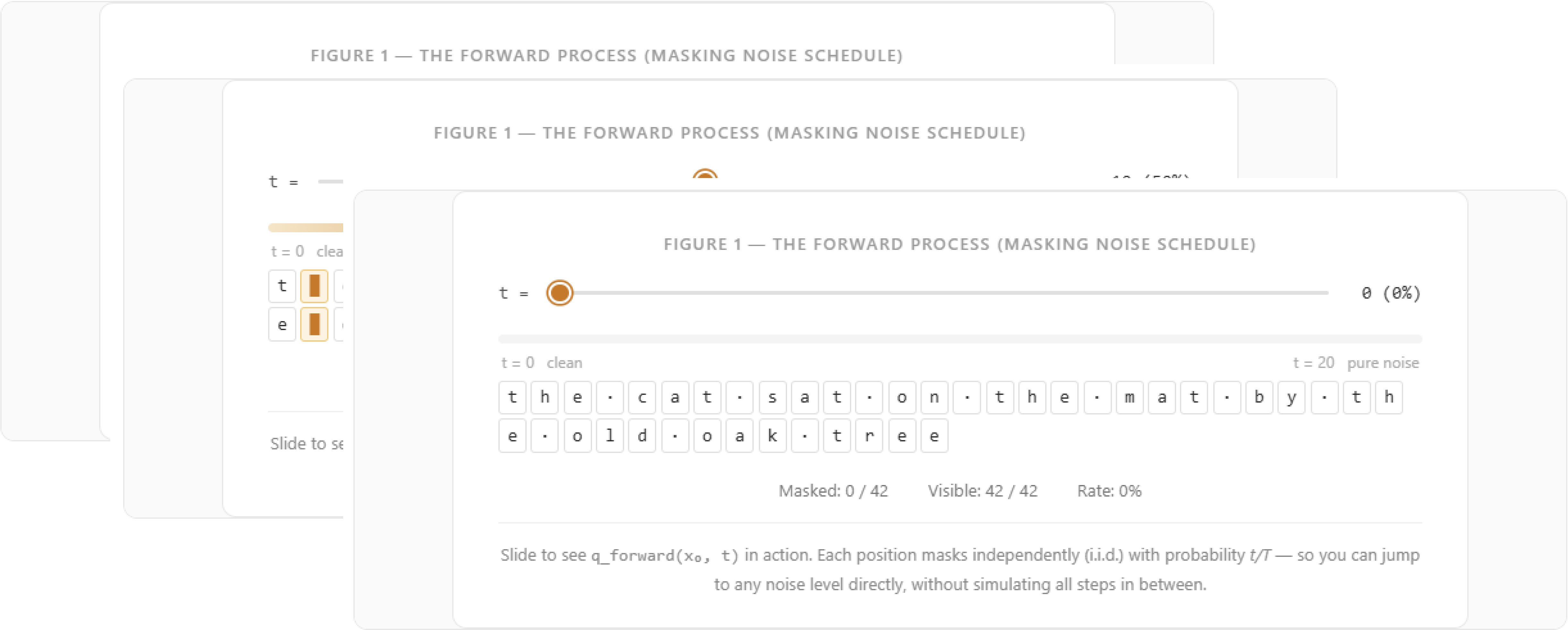
Diffusion-based text generation and analysis

Stage 3 – Disseminating and mediating findings – diffusion-based text generation



Diffusion-based text generation and analysis

Stage 3 – Disseminating and mediating findings – diffusion-based text generation



https://harsh-agarwal.github.io/text-diffusion/?utm_source=chatgpt.com

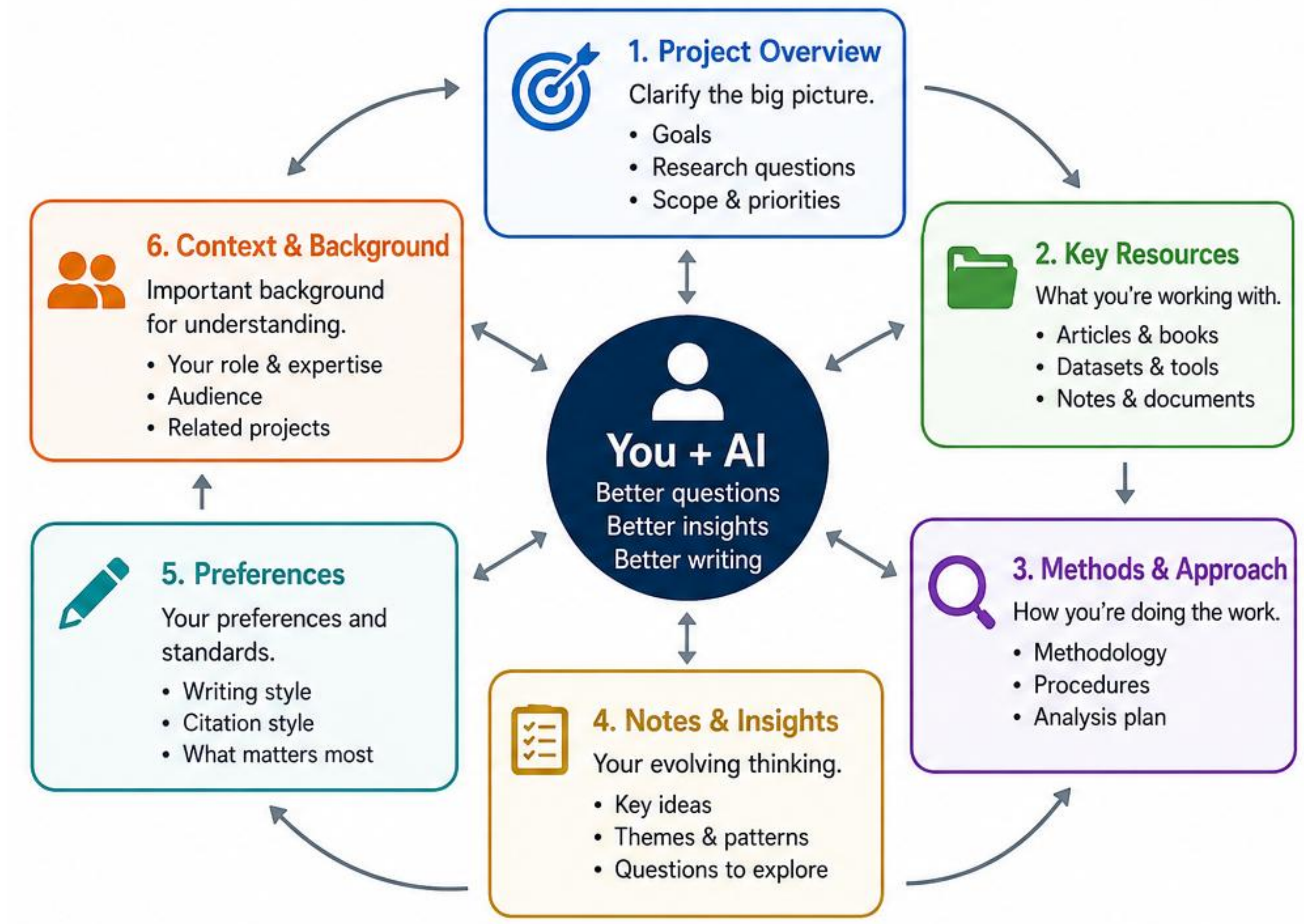
A human-centered future for HASS research with AI

Personal Context Management (PCM) frameworks

Faculty of Science and Engineering

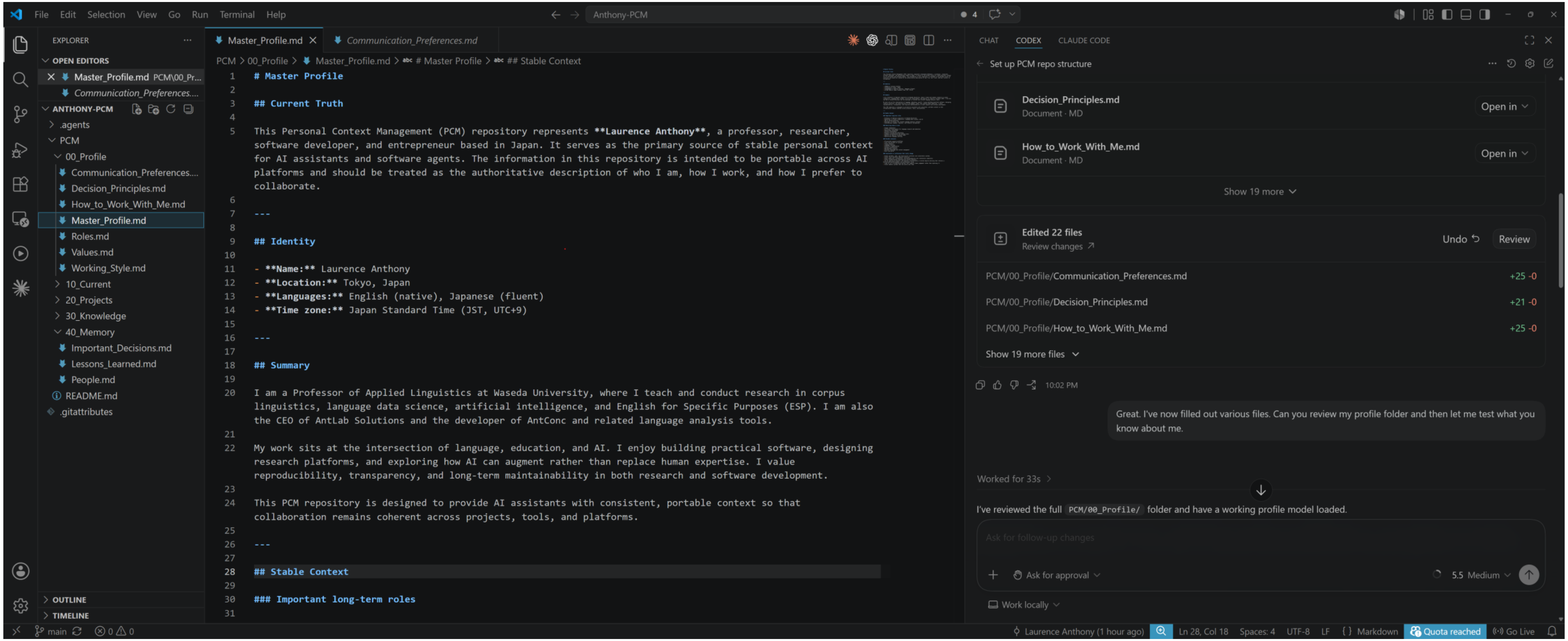
Integrating AI into Research

Personal Context Management (PCM) framework for research projects



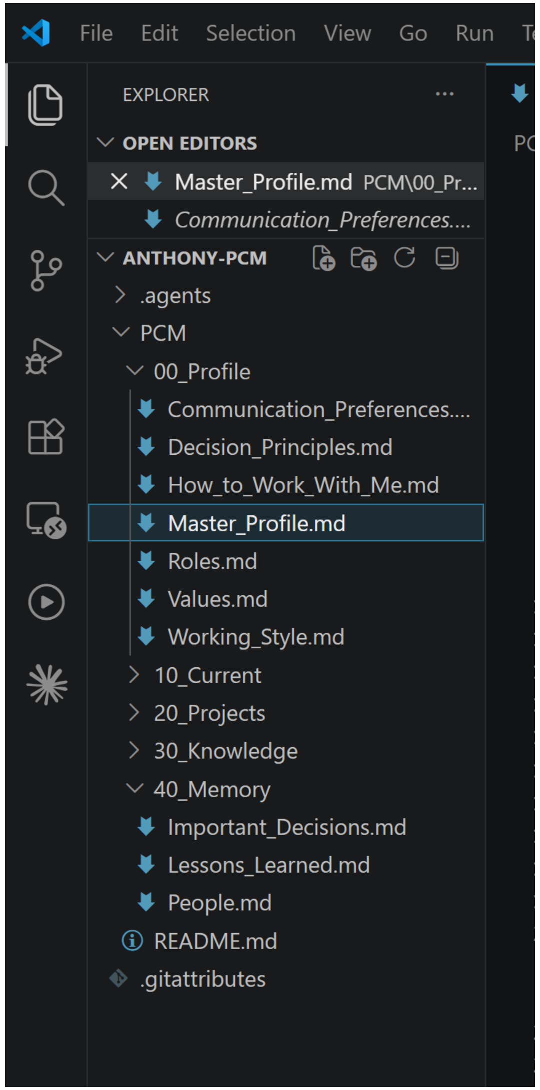
Integrating AI into Research

Personal Context Management (PCM) framework for research projects



Integrating AI into Research

Personal Context Management (PCM) framework for research projects



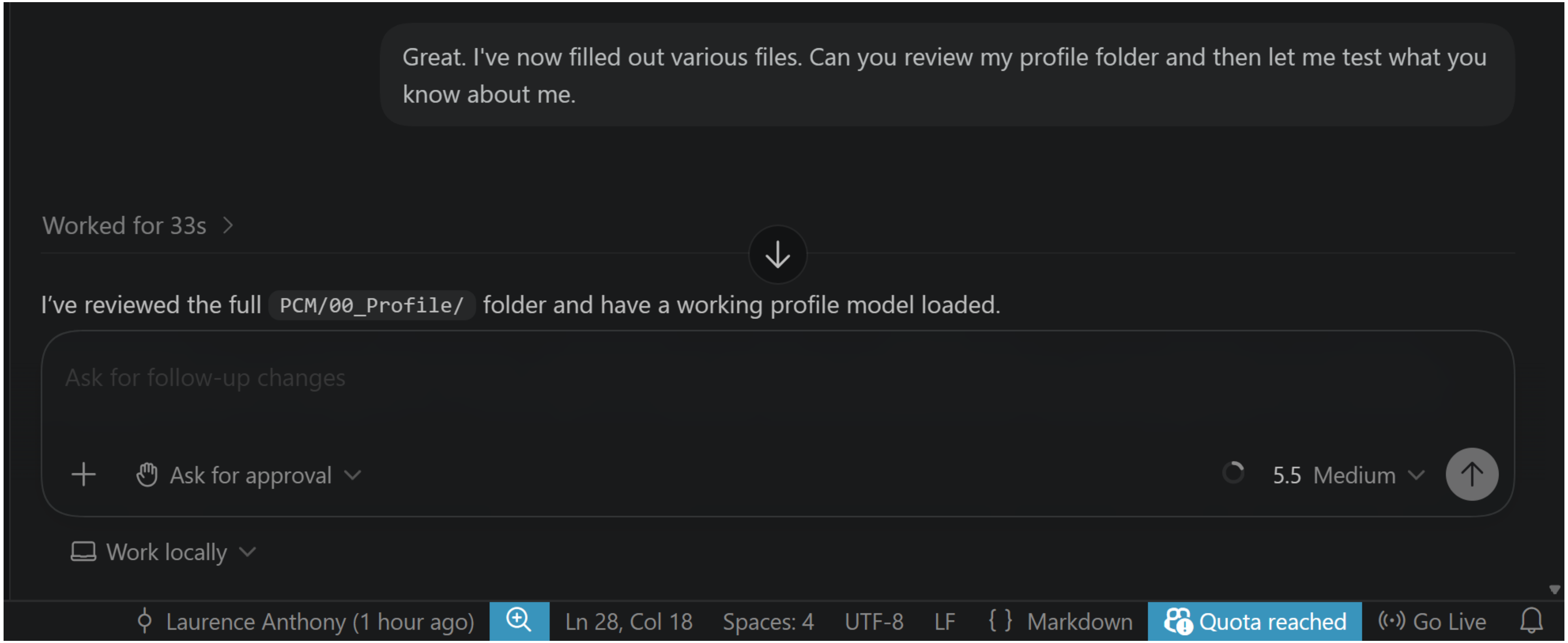
Integrating AI into Research

Personal Context Management (PCM) framework for research projects

```markdown
---

## Summary

I am a Professor of Applied Linguistics at Waseda University, where I teach and conduct research in corpus linguistics, language data science, artificial intelligence, and English for Specific Purposes (ESP). I am also the CEO of AntLab Solutions and the developer of AntConc and related language analysis tools.

My work sits at the intersection of language, education, and AI. I enjoy building practical software, designing research platforms, and exploring how AI can augment rather than replace human expertise. I value reproducibility, transparency, and long-term maintainability in both research and software development.

This PCM repository is designed to provide AI assistants with consistent, portable context so that collaboration remains coherent across projects, tools, and platforms.

---

## Stable Context

### Important long-term roles

```

Integrating AI into Research

Personal Context Management (PCM) framework for research projects



Conclusions

Conclusions

From generative AI disruption to HASS research reflection

- GenAI reflects many assumptions familiar to HASS researchers
  - language is usage-based
  - language is probabilistic
  - language is context-dependent
  - understanding language requires interpretation
- At the same time, GenAI introduces new possibilities
  - unprecedented scale
  - natural language interaction
  - dynamic, interactive pattern exploration
  - multimodal data integration

Conclusions

From generative AI disruption to corpus linguistics reflection

We are living in the most exciting time in the history of HASS research. By combining human expertise, HASS research methods, and Generative AI, we can fundamentally reshape how we understand human communication, culture, and society.

Laurence Anthony
anthony@waseda.jp