LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

CHARTING THE GLOBAL LANDSCAPE OF LANGUAGE TECHNOLOGY

June 18, 2024

A EUROPEAN COMMISSION INITIATIVE

BY NIMDZI INSIGHTS

Large Language Models

Laszlo K. Varga, Nimdzi Insights
Nadezda Jakubkova, Nimdzi Insights

LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

Housekeeping

Q&A

In-session polls

Recording

Post-event feedback survey

LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

Agenda

Time	Session
10:00-10:30	**Welcome & Keynotes** Hosts: Philippe Gelin, Laszlo K. Varga Keynotes: Renate Nikolay, Renato Beninatto
10:30-11:25	**The Landscape of Language Technology** Speakers: Laszlo K. Varga
11:30-12:00	**Multilingualism of European Websites and the Technology Solutions Supporting It** Speakers: Andrejs Vasiļjevs
12:00-13:00	**Lunch Break**
13:30-14:20	**Large Language Models and Foundational Language Technologies** Speakers: Laszlo K. Varga, Nadezda Jakubkova
14:25-15:15	**Text-based Language Technologies** Speakers: Laszlo K. Varga, Jourik Ciesielski
15:15:15:35	**Break**
15:35-16:25	**Speech-Based Language Technologies** Speakers: Laszlo K. Varga, Igor Szoke, Khalid Choudry
16:30-17:00	**Closing** Speakers: Philippe Gelin, Laszlo K. Varga

LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

Agenda

Time	Session
10:00-10:30	**Welcome & Keynotes** Hosts: Philippe Gelin, Laszlo K. Varga Keynotes: Renate Nikolay, Renato Beninatto
10:30-11:25	**The Landscape of Language Technology** Speakers: Laszlo K. Varga
11:30-12:00	**Multilingualism of European Websites and the Technology Solutions Supporting It** Speakers: Andrejs Vasiļjevs
13:30-14:20	**Large Language Models and Foundational Language Technologies** Speakers: Laszlo K. Varga, Nadezda Jakubkova
14:25-15:15	**Text-based Language Technologies** Speakers: Laszlo K. Varga, Jourik Ciesielski
15:15:15:35	**Break**
15:35-16:25	**Speech-Based Language Technologies** Speakers: Laszlo K. Varga, Igor Szoke, Khalid Choudry
16:30-17:00	**Closing** Speakers: Philippe Gelin, Laszlo K. Varga

LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

Large language models (LLM)

Laszlo K. Varga

Lead Researcher and Analyst,

Nimdzi Insights

Nadežda Jakúbková

Technology Researcher

Nimdzi Insights

LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

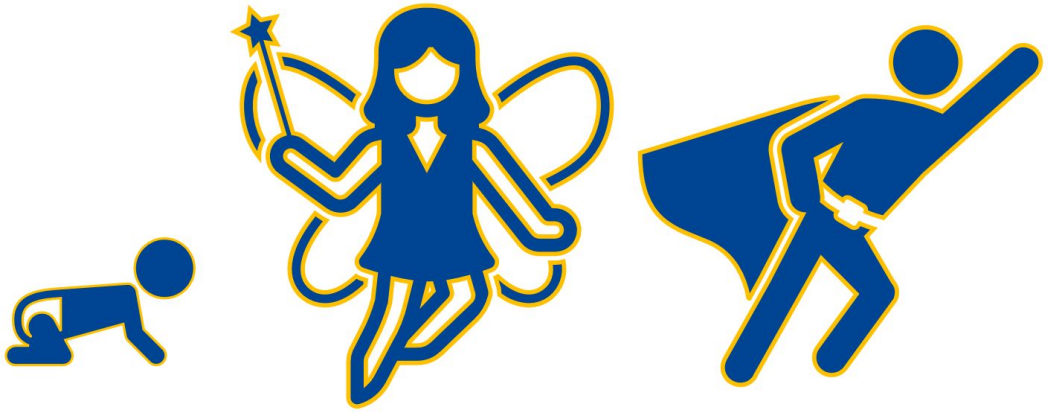
Large language models (LLM) 1

Introduction.

It's only 1.5 years since ChatGPT, but 7 years since the Transformer was born. The new hype is in the natural language interface and multimodality.

The real-world impact of LLMs is now evident in applications like chatbots, productivity and coding assistants, content-creation tools, and many more.

The potential economic impact of LLMs is in the EUR billions, and they are also democratising multilingual and productivity application development.



LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

Large language models (LLM) 2

Some core issues.

- Training and running LLMs is computationally **expensive**.
- Training data can lead to discriminatory, **biased** outputs.
- Languages are **not** equal.
- **Hallucinations** vs grounding in facts.
- **Privacy and security**, confidentiality and misuse concerns.
- **ROI** is hard to predict.

So why adopt now?

- Innovation signalling.
- Competitive advantage.
- Fear of missing out.
- Preparing for future wider adoption and disruption.

LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

Large language models (LLM) 3

(Some) Latest developments.

Applications:

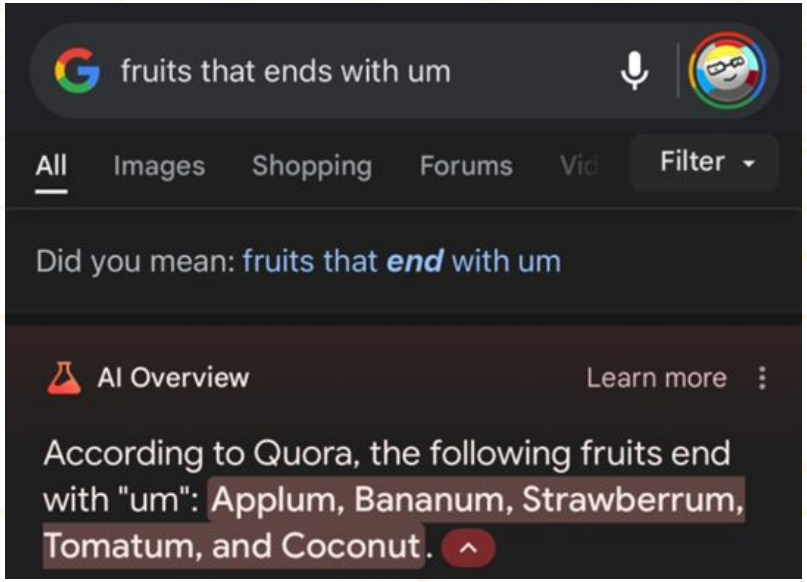
- Big tech & AI labs: further and deeper productisation.
  GPT-4o, Google AI Overviews, Microsoft Copilot+ PCs, Apple Intelligence.
- Unbabel Halo @ AI For Good
- TowerLLM for language purposes

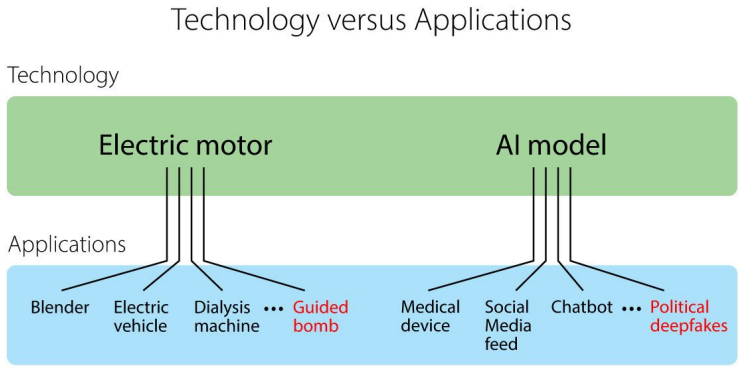
Foundations:

- Mamba in action: Jamba

AI acts & risks:

- California
  SB-1047 (Safe and Secure Innovation for Frontier Artificial Intelligence Models);
- EU
  AI Act & EU AI Office
- LLM misuse cases
  At $20 a month.





Source: https://www.deeplearning.ai/the-batch/issue-252/

LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

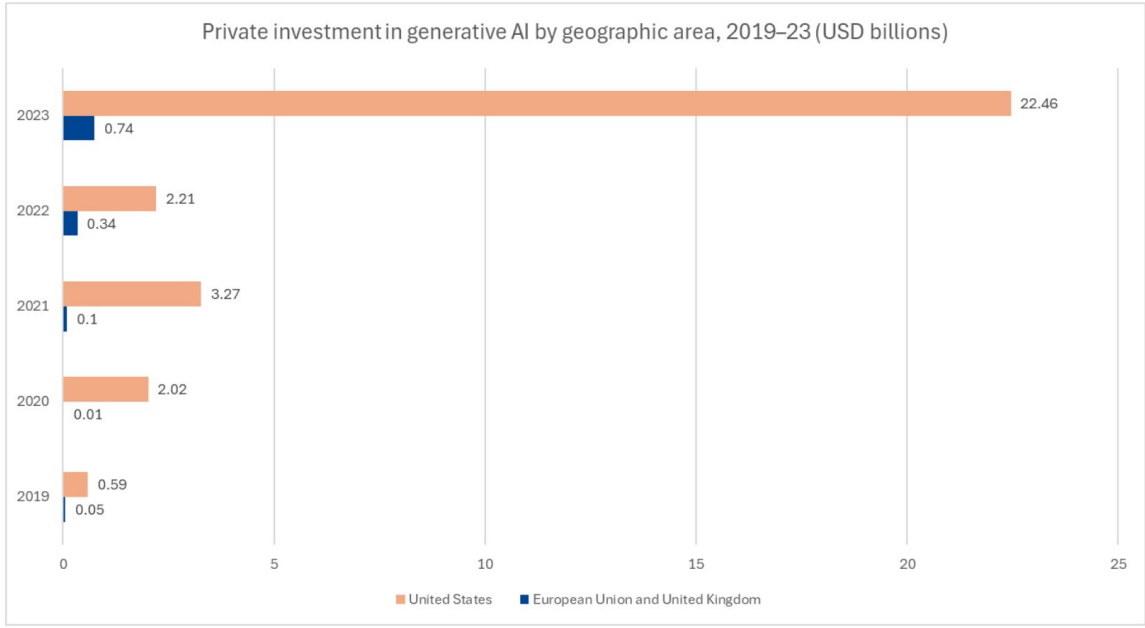
Large language models (LLM) 4

Investment.

While **Microsoft** has invested more than USD 11 billion (~EUR 10 billion) into **OpenAI**, and is allegedly planning to invest USD 100 billion (! ~EUR 90 billion) into Stargate for GenAI by 2028, **Google** Deepmind's CEO (who allegedly has the largest computing power) also said they are ready to invest USD 100 billion into AI (~EUR 90 billion) - not exclusively to LLMs.

Saudi Arabia plans to invest USD 40 billion (~EUR 35 billion) into AI technologies, and **Amazon** is allegedly developing their new largest foundational model (Olympus), and has invested more than USD 4 billion (~EUR 3.5 billion) in **Anthropic**.

Reportedly, the cost of development of OpenAI's GPT-4 was in the USD 80 million (~EUR 75 million) range, while Google Gemini Ultra likely entailed more than double of that.



Data source: Stanford AI Index Report 2024[71]

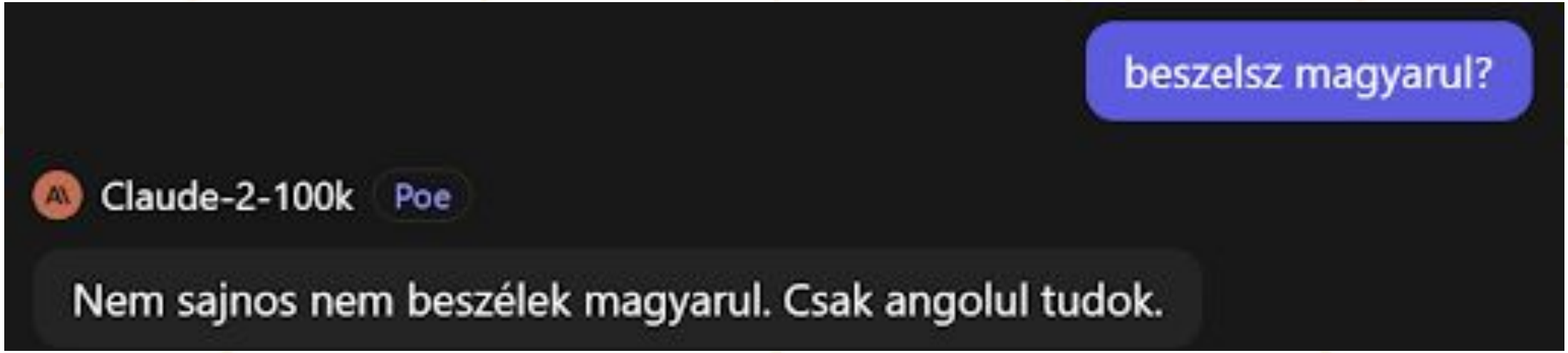
The LLM investment hype is less about solving intra-language issues and more about productivity and efficiency gains.

Time-saver tools and applications are prime movers of investment.

LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

Large language models (LLM) 5

Quality and language coverage.



LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

Large language models (LLM) 5

Quality and language coverage.

Benchmarks are more eye candies than real predictors of performance, especially because of cherry-picking. Leaderboards exist, more transparency is needed. Purpose-driven, **custom evaluation is needed for actual implementation**. Proof of concept is easy to create, real-life implementations are more complex.

Language support by LLMs is empirical, primarily because of lack of data availability and training transparency. Even Meta doesn't fully disclose their training data, other main providers obscure their methods.

English-bias is inherent in most models: "*will perform best for English-language use cases*" is an often-read disclaimer. Google allegedly puts effort into bilingual data (via LSPs). Silo.ai's models are a welcome addition to support EU languages.

Language	Percent	Language	Percent
en	89.70%	uk	0.07%
unknown	8.38%	ko	0.06%
de	0.17%	ca	0.04%
fr	0.16%	sr	0.04%
sv	0.15%	id	0.03%
zh	0.13%	cs	0.03%
es	0.13%	fi	0.03%
ru	0.13%	hu	0.03%
nl	0.12%	no	0.03%
it	0.11%	ro	0.03%
ja	0.10%	bg	0.02%
pl	0.09%	da	0.02%
pt	0.09%	sl	0.01%
vi	0.08%	hr	0.01%

Llama 2 training data language distribution

LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

Large language models (LLM) 6

Market sizing.
Demand.

LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

Large language models (LLM) 6

Market sizing.
Demand.

As an emerging tech, monetisation of LLMs is not solved, but the potential is huge for all main LLM actors. **Though investment into LLMs is massive, direct revenues are lagging.**

Demand is tangible and loud, hype-driven, but not fully visible. Many large corporations make LLM commitments in their AI strategies. Growth potential is VERY high, for productivity incl. coding, and newly multimodal for multilingual communications.

Market character.

Everyone wants AI and LLMs, for everything. Use cases are easy to create and test, at-scale enterprise deployment is non-trivial at this point. Company-wide “AI initiatives” are demanded from top-level, often without consideration of other risks - compliance, AI Ethics, misuse, etc. ROI is hard to determine: “*IT must have figured it out*”.

Paradox of choice is present, and platform lock-in is a risk.

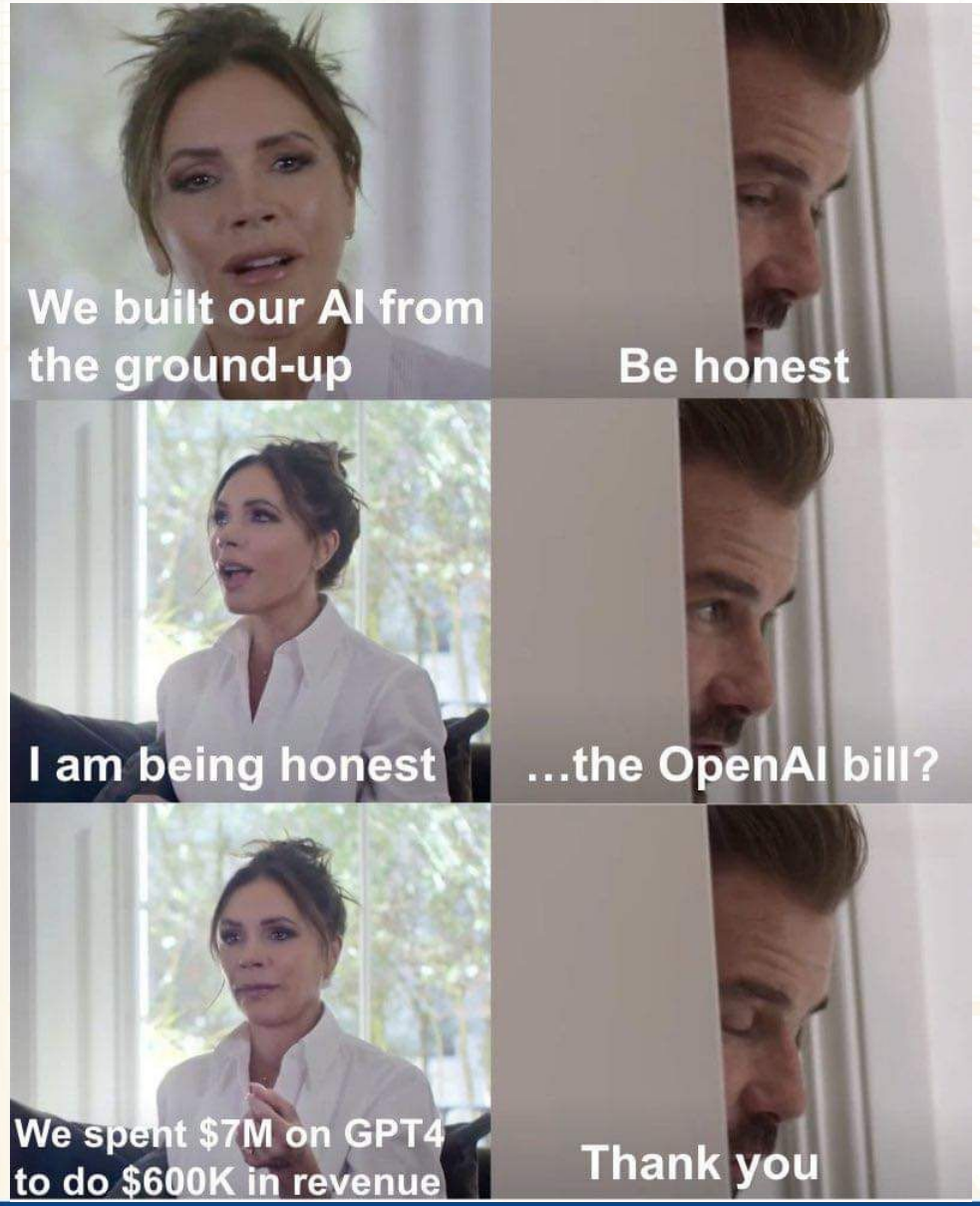
Additional data, customisation, and deployment services will drive additional growth of LLM revenues.

Main LT category	Generative AI - language models (LLM)
Market size estimate (2023)	EUR 1-2 billion
Growth potential	Very high; EUR 5-10 billion in 2024
Investment interest	Very high
Market character	Core tech is new, emerging market Big-tech dominated with strong open-source ecosystem for local/enterprise customizations As foundational tech, fuels growth and changes in other LT and connected markets (NLP, MT, Chatbots, MLCG)
AI / ML adoption / disruption level	Fundamental
Technology maturity level	Emerging foundation technology Fast growing with high impact on other LTs

LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

Large language models (LLM) 7

Main actors.



LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

Large language models (LLM) 7

Main actors.

Big tech and major AI labs dominate with proprietary solutions on their cloud platforms (MS, Google, Amazon, Meta, IBM, Nvidia).

Many US-HQ companies run **European labs for innovation**, and Europe is on the LLM foundation map.

There are contenders & **open-model initiatives are also strong**. Meta is big proponent of this, and IBM just opened their models. Local and on-premise solutions typically rely on main open models (Llama, Mistral, etc) via fine-tuning.

The money-making problem is severe (temporal?). Some experts expect that even big players can go out of business in 1-2 years.

Main actors are interconnected with investments. VC funding is strong, but big-tech also invests heavily.

Company	Country of origin	Founded	Foundational LLMs	Estimated LLM revenue (2023)	Investment / funding (till March, 2024)
Anthropic	USA	2021	Claude	EUR 100 million	EUR 6.5 billion in 2013
Cohere	Canada	2019	Command	EUR 22 million	EUR 300 million
Hugging Face	USA	2016	various open source	(not core)	EUR 400 million
Google (Alphabet)	USA	1998	Gemini, Gemma	(not core)	(not core)
Meta	USA	2004	Llama	(not core)	(not core)
OpenAI	USA	2015	GPT4, GPT4-V	EUR 0.5-1 billion	EUR 10 billion
Mistral	France	2023	Mistral, Mixtral	undisclosed	EUR 400 million
Silo	Finland	2017	Poro, Viking	EUR 40 million	EUR 10 million (and HPLT/LUMI collaboration)
Aleph Alpha	Germany	2019	Luminous	EUR 8 million	EUR 500 million
Upstage	Korea	2020	Solar	EUR 20 million	EUR 50 million
01.ai	China	2016	Yi	undisclosed	EUR 300 million
DeepSeek	China	2023	DeepSeek	undisclosed	undisclosed
IBM	USA	1911	Granite	(not core)	(not core)
Amazon	USA	1994	Titan (Olympus)	(not core)	(not core)
Databricks	USA	2013	DBRX	EUR 1.5 billion (LLM part undetermined)	EUR 2 billion
Baidu	China	2000	ERNIE	(not core)	(not core)
Alibaba	China	1999	Qwen	(not core)	(not core)
AI21	Israel	2017	Jurassic	EUR 40 million	EUR 300 million
TII UAE	UAE	2020	Falcon family	undisclosed	undisclosed
Reka.ai	USA	2022	Reka	undisclosed	EUR 60 million

Large language models (LLM) 8

Technology outlook.

Small Language Models (SLM)

- SLM are more efficient at certain tasks and can run on smaller devices. Suitable for efficient purpose-built applications.
- Foundational model creators purposefully launch multiple size variations of their models that include smaller ones as well.

Mixture of Experts (MoE)

- Runtime efficiencies with an orchestrated group of SLMs.

Grounding

- Hallucination, training cut-off, and (some) data security. Options: RAG, function calling, prompting, fine-tuning, quality-focused data.

Multimodality

- Working with multiple input and output modalities - on top of text, image, audio, and video. Common embeddings and shared transformer layers - “native” multimodality, instead of orchestration via routing (e.g. GPT-4V).

Custom, task- and domain-specific training and fine-tuning

- Whether SLMs will become the mainstream for task-specific LLM implementations versus the largest and most generally capable models via cloud access is an active research and market development question.

LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

Q&A

Thank you.

Have more questions later?
Find me on LinkedIn or reach out to the project team at ltsurvey@nimdzi.com.

Feedback survey