DECIPHERING THE AI LANDSCAPE: BUSINESS, MARKET TRENDS, AND EMERGING RISKS

Goran S. Milovanović
Lead Data Scientist/Owner
smartocto
DATAKOLEKTIV

On the nature of revolution

Natural Intelligence

System1|System2

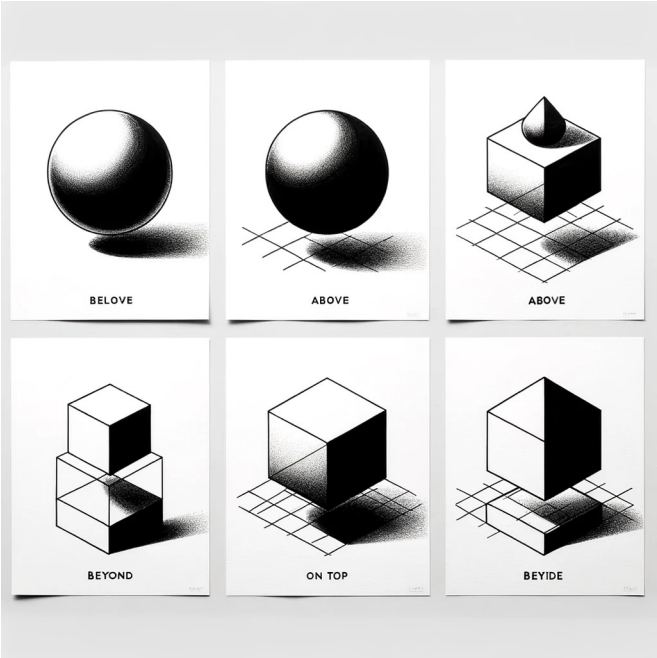
HOSPITAL DOCTOR
NURSE PHILOSOPHY
DOCTOR UNIVERSITY

System1 == Associative Machinery

System2 == Formal Semantics

The Language of Thought

Understanding intuitive physics, the visual grammar of the world (above, below, behind, etc.)



Understanding agency and causality (folk-psychology) and the *necessity* of some regularities in the world

System2 Assumption: Compositionality

$$\boldsymbol{M(\alpha \circ \beta) = f(M(\alpha), M(\beta))}$$

DOG PSYCHIATRIST

PROMPT ENGINEER

A horse riding another horse.

A unicorn riding an astronaut.

No compositionality→ hallucinations stay

No System 2 Reasoning → No Artificial General Intelligence

Face it:
Artificial General Intelligence is nowhere near

On the nature of revolution

Achieved System1 Superintelligence

On the nature of revolution

Artificial General Intelligence → NULL

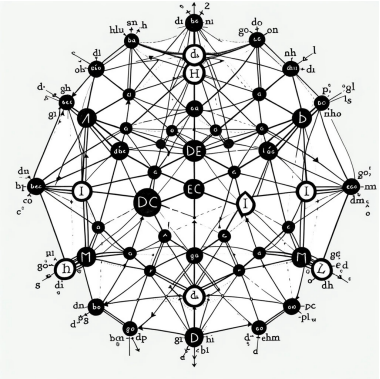
Achieving System1 Superintelligence

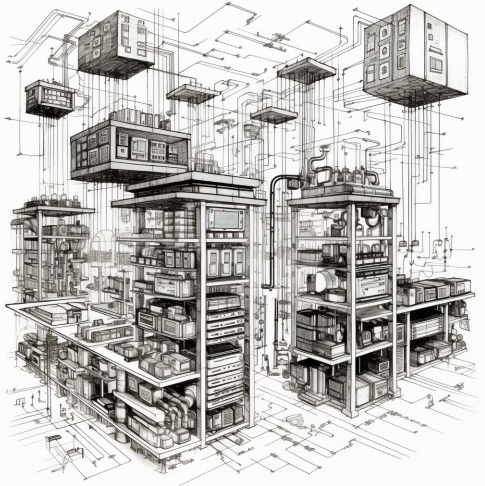
is super useful

Consequences

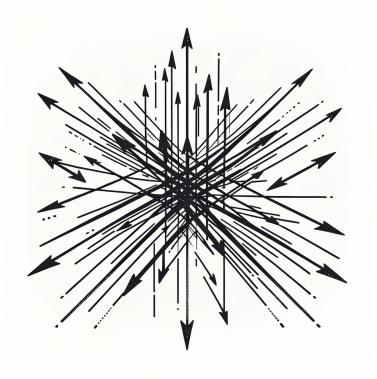
Large Language Models

Knowledge Graphs





Knowledge Vectors



Internet Search ⟷ Promp Engineering ⟷ External Sources

Client

AIaaS|DevsAI

AIaaS value propostion: scale

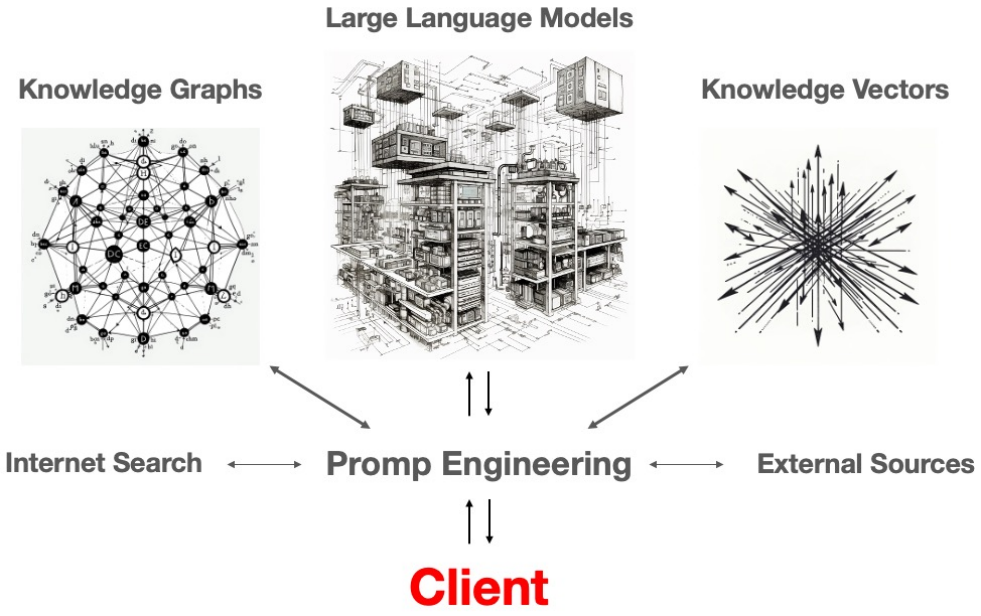
Infra cost
Training cost
Data cost
Inference cost

DevsAI value propostion: specialization

Infra cost

Compute cost

Knowledge Engineering cost



Strategies

AIaaS Strategy: Train specialized LLMs

Infra cost
Training cost
Data costs
Inference cost

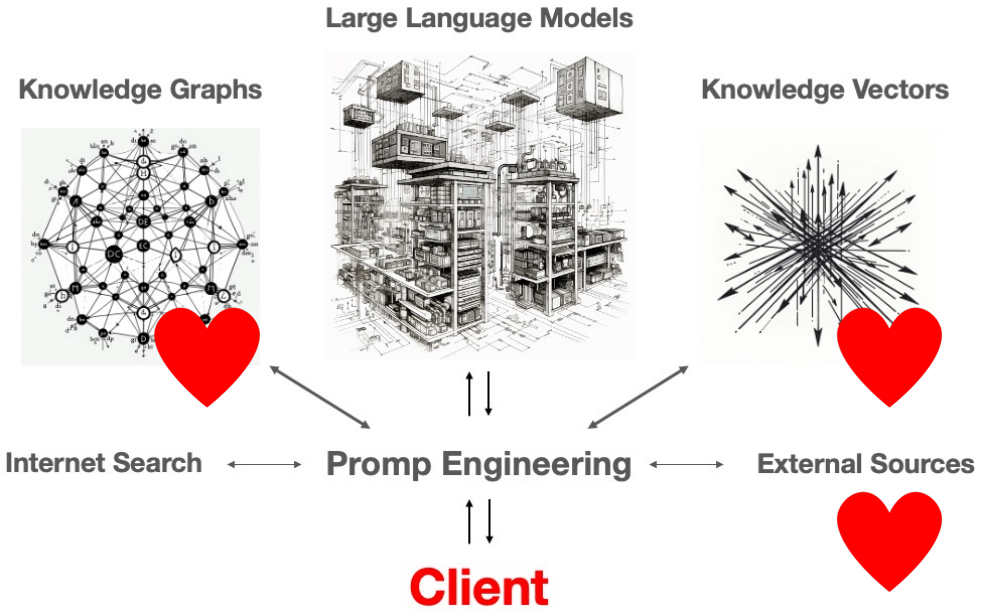
How many domain-specific models can be trained at what respective risk?

BloombergGPT

- Trained on AWS SageMaker
- ~700 billion tokens dataset, only **50 billion parameters**
- 64 x p4d.24xlarge instances → 64 x 8 Nvidia 40GB A100 GPUs → 512 GPUs total
- ~$33 per instance per hour → $2112 per hour for complete setup
- **Trained for 53 days → ~$2.7 million (~$1 million with spot-pricing)**

DevsAI value propostion: specialization

Infra cost
Compute cost
Knowledge Engineering costs



Training a general purpose LLM is OK.

If you are OpenAI, Microsoft, Google, xAI, Anthropic, Amazon, or somewhere near.

Training a specialized LLM is risky:

DevsAI might beat you.

Trends

Not too many domain-specific LLMs: BloombergGPT, Med-PaLM2, ClimateBERT, BioBERT, KAI-GPT, FinGPT…

On the other hand:

- OpenAI introduced GPT-4 Turbo with 128K context,
- JSON response format,
- Reproducible outputs,
- Stateful API (i.e. conversation memory)…

- … and Anthropic Claude 2.1 now has a 200K context window, search and retrieval capabilities over a variety of knowledge bases (Elasticsearch, vector databases, web search, Wikipedia), tool use, etc.

→ AIaaS is targeting the DevsAI business

Risks

The emerging market risk is in the *P*(match):

AIaaS	DevsAI
(high cost) **Infra**	**Infra** (low cost)
(high cost) **Training**	
(high cost) **Data**	**Knowledge** (moderate cost)
(high cost) **Inference**	**Compute** (high cost) **?**

DECIPHERING THE AI LANDSCAPE: BUSINESS, MARKET TRENDS, AND EMERGING RISKS