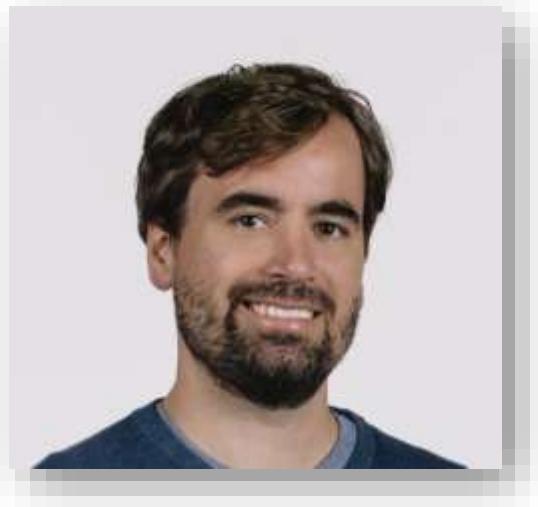


When OR meets Generative AI

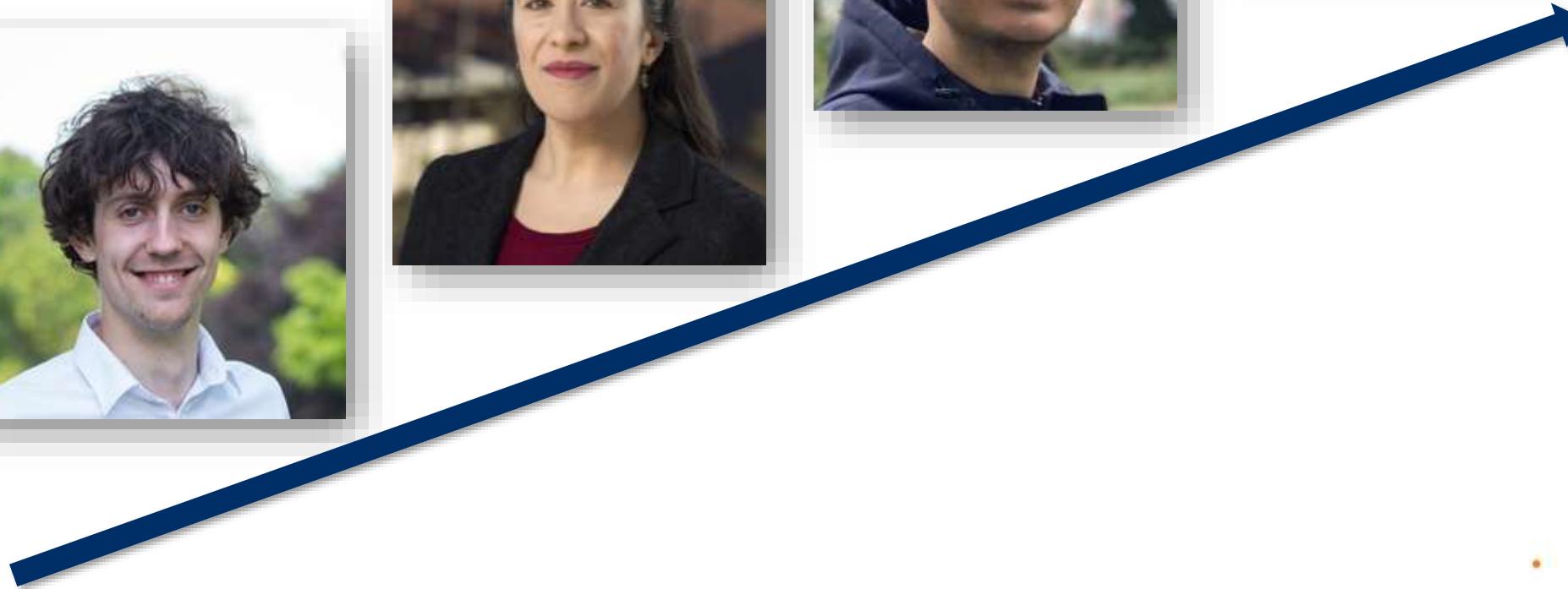
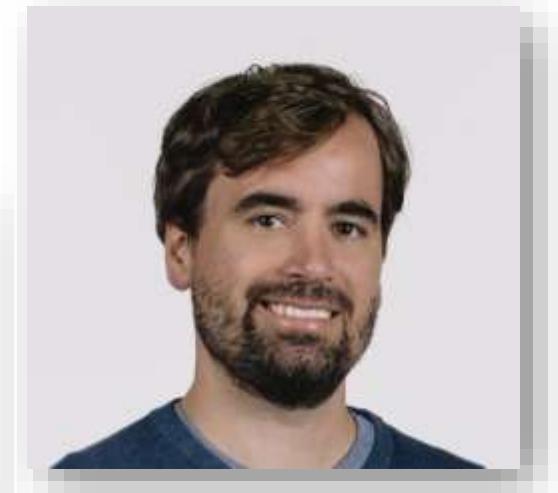
Wouter Kool

ORTC

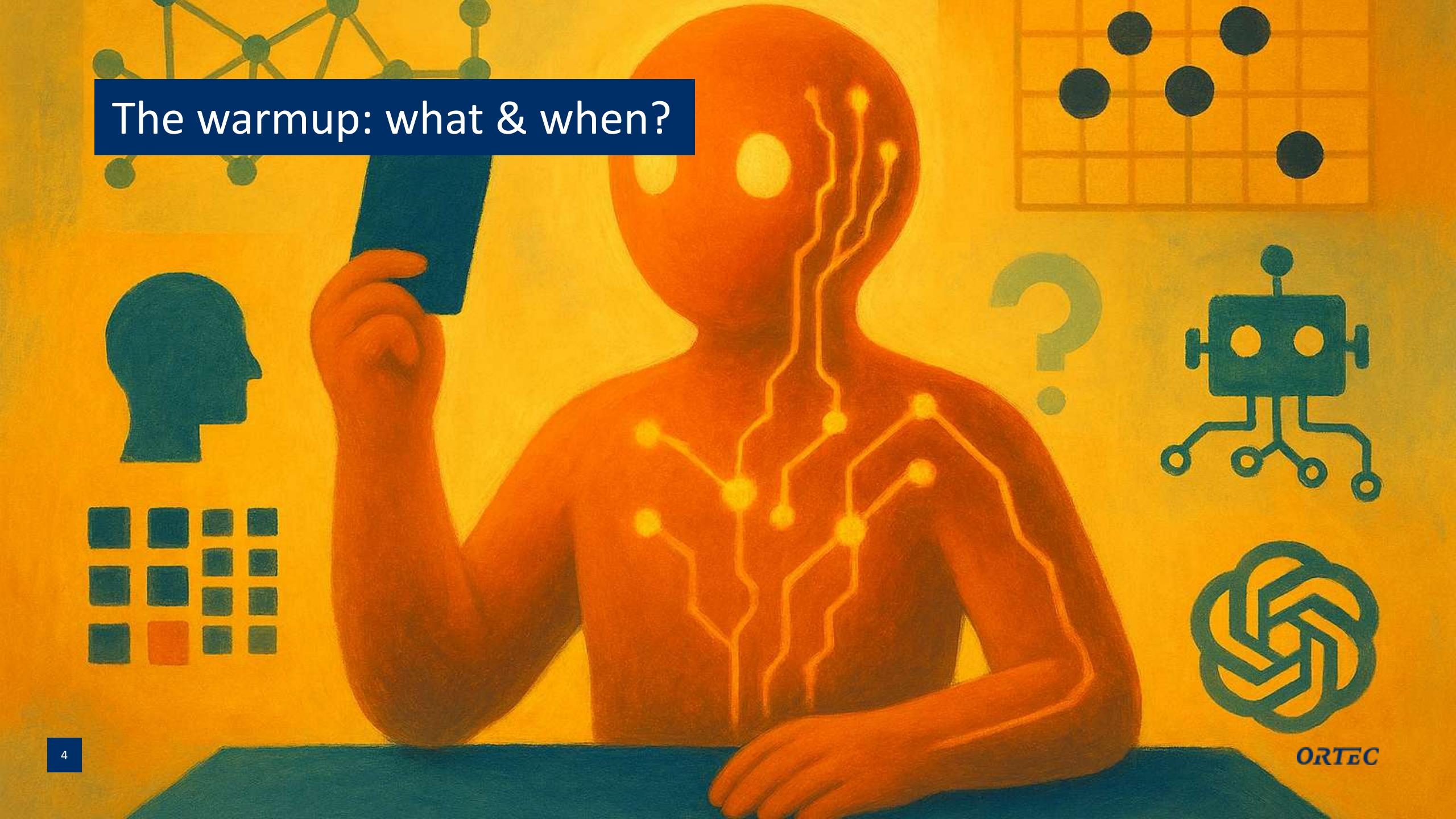
The lineup



The line up



The warmup: what & when?



AlphaGo beats Lee Sedol (2016)



The Adam optimizer (Kingma & Ba, 2015)

Require: α : Stepsize

Require: $\beta_1, \beta_2 \in [0, 1]$: Exponential decay rates for the moment estimates

Require: $f(\theta)$: Stochastic objective function with parameters θ

Require: θ_0 : Initial parameter vector

$m_0 \leftarrow 0$ (Initialize 1st moment vector)

$v_0 \leftarrow 0$ (Initialize 2nd moment vector)

$t \leftarrow 0$ (Initialize timestep)

while θ_t not converged **do**

$t \leftarrow t + 1$

$g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$ (Get gradients w.r.t. stochastic objective at timestep t)

$m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ (Update biased first moment estimate)

$v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ (Update biased second raw moment estimate)

$\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ (Compute bias-corrected first moment estimate)

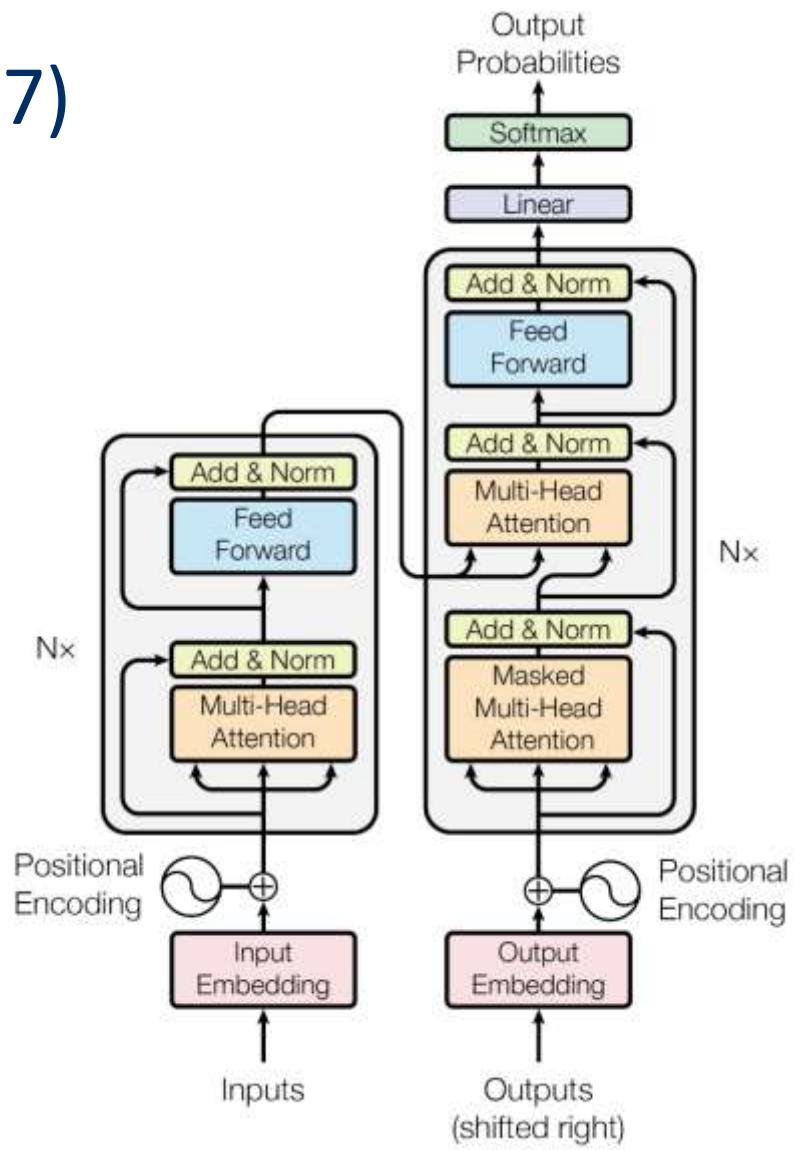
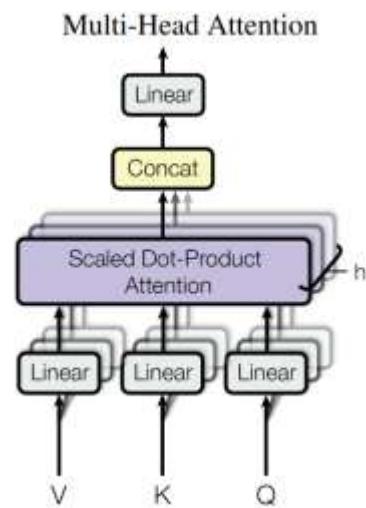
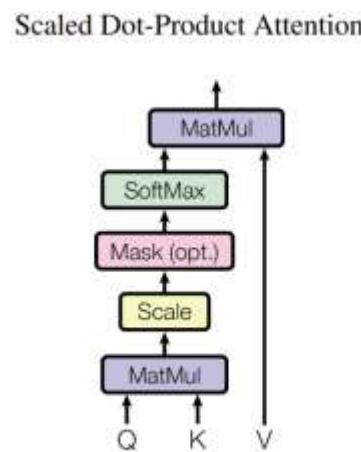
$\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$ (Compute bias-corrected second raw moment estimate)

$\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$ (Update parameters)

end while

return θ_t (Resulting parameters)

The Transformer (Vaswani et al., 2017)



ChatGPT released (2022)

ChatGPT



Examples

"Explain quantum computing in simple terms"

"Got any creative ideas for a 10 year old's birthday?"

"How do I make an HTTP request in Javascript?"



Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests



Limitations

May occasionally generate incorrect information

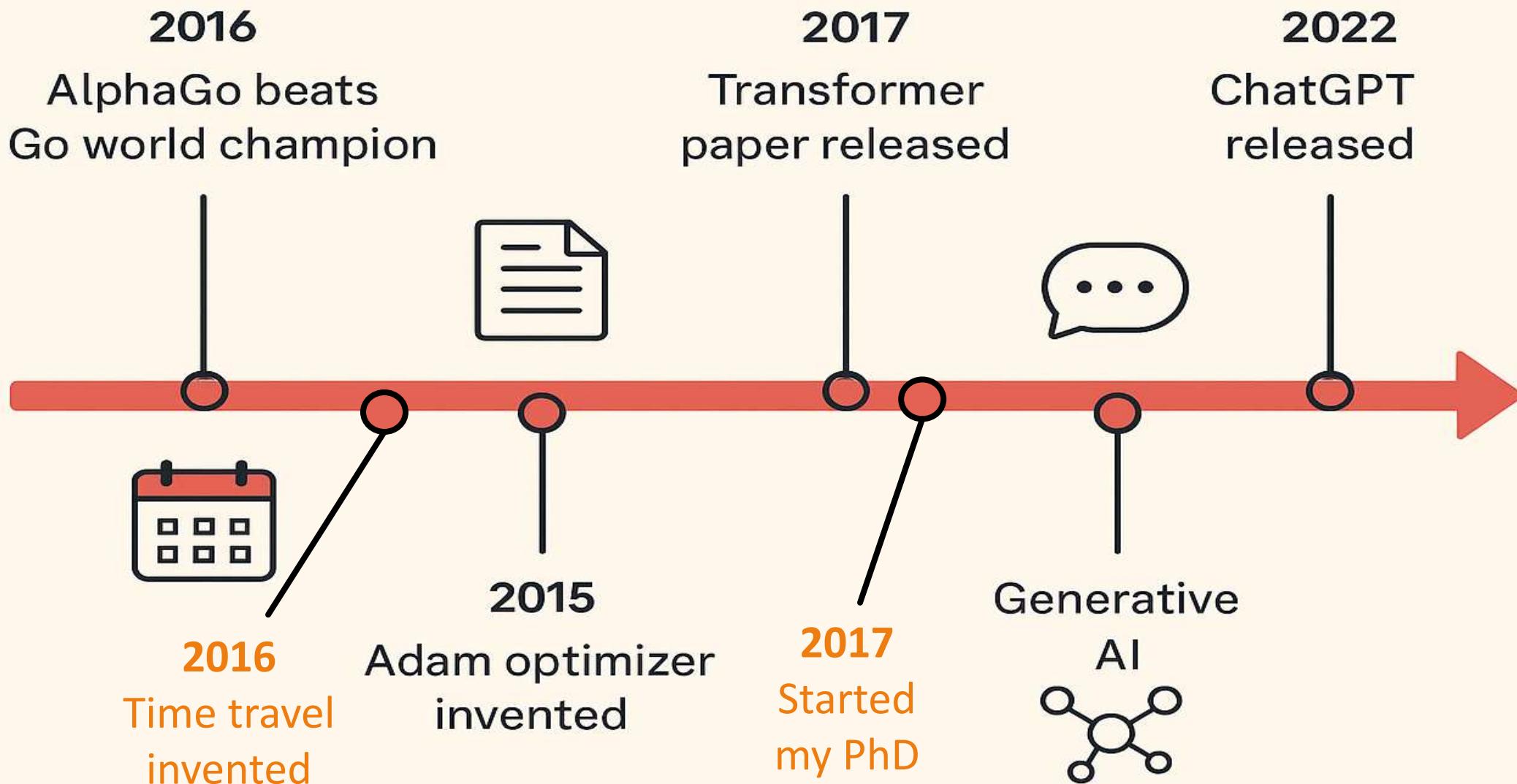
May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021



Free Research Preview: ChatGPT is optimized for dialogue. Our goal is to make AI systems more natural to interact with, and your feedback will help us improve our systems and make them s

AI TIMELINE (2016–2025)

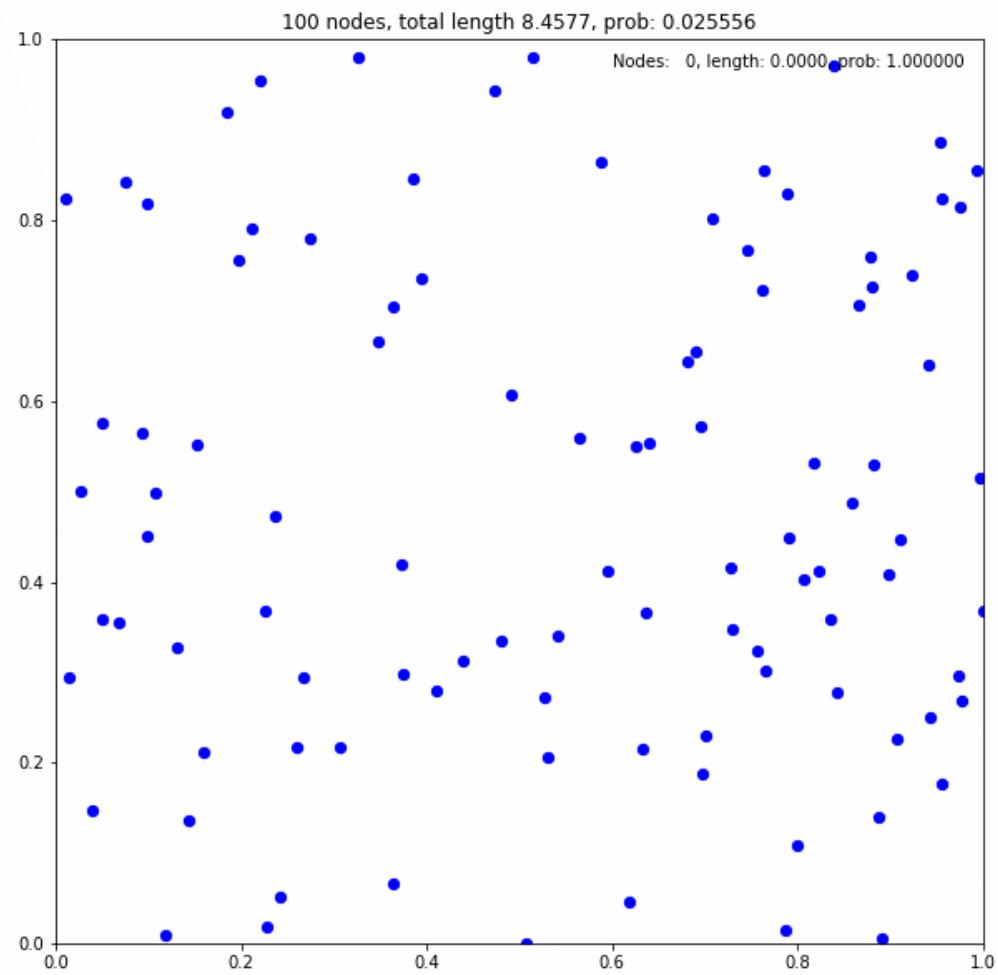
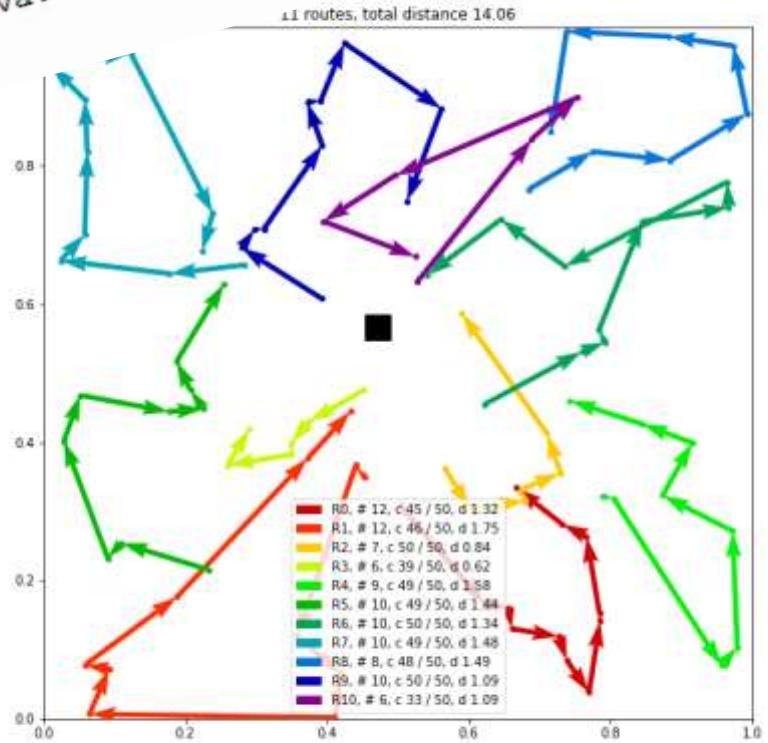


ATTENTION, LEARN TO SOLVE ROUTING PROBLEMS!

Wouter Kool
University of Amsterdam
ORTEC
w.w.m.kool@uva.nl

Herke van Hoof
University of Amsterdam
h.c.vanhoof@uva.nl

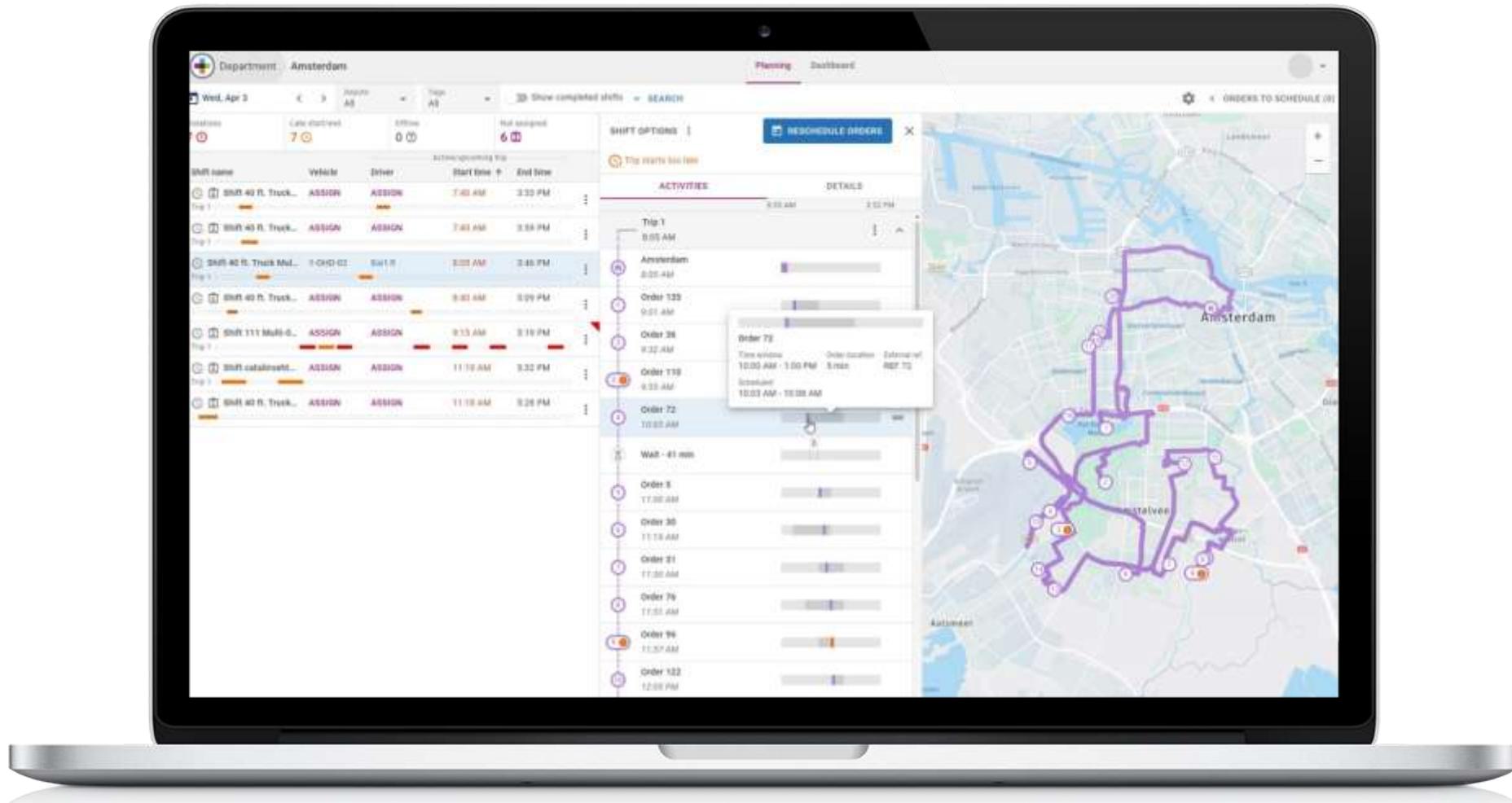
Max Welling
University of Amsterdam
CIFAR
m.welling@uva.nl



ORTEC

P

π



Real world problems...

Raeal



Real world problems are rarely as clean as our equations!

- A wise LLM that wrote my abstract -



5 years from now?

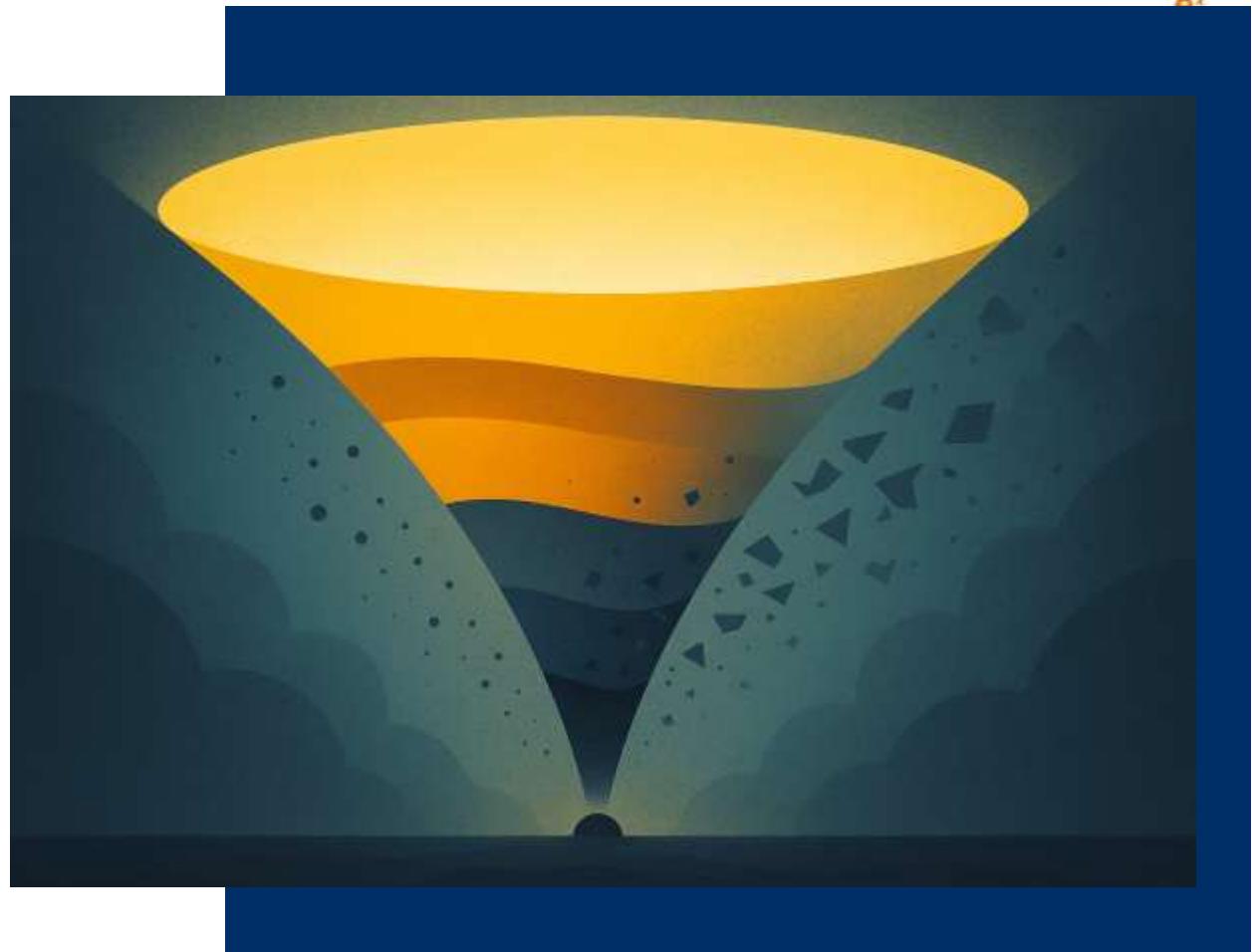


In a typical OR application

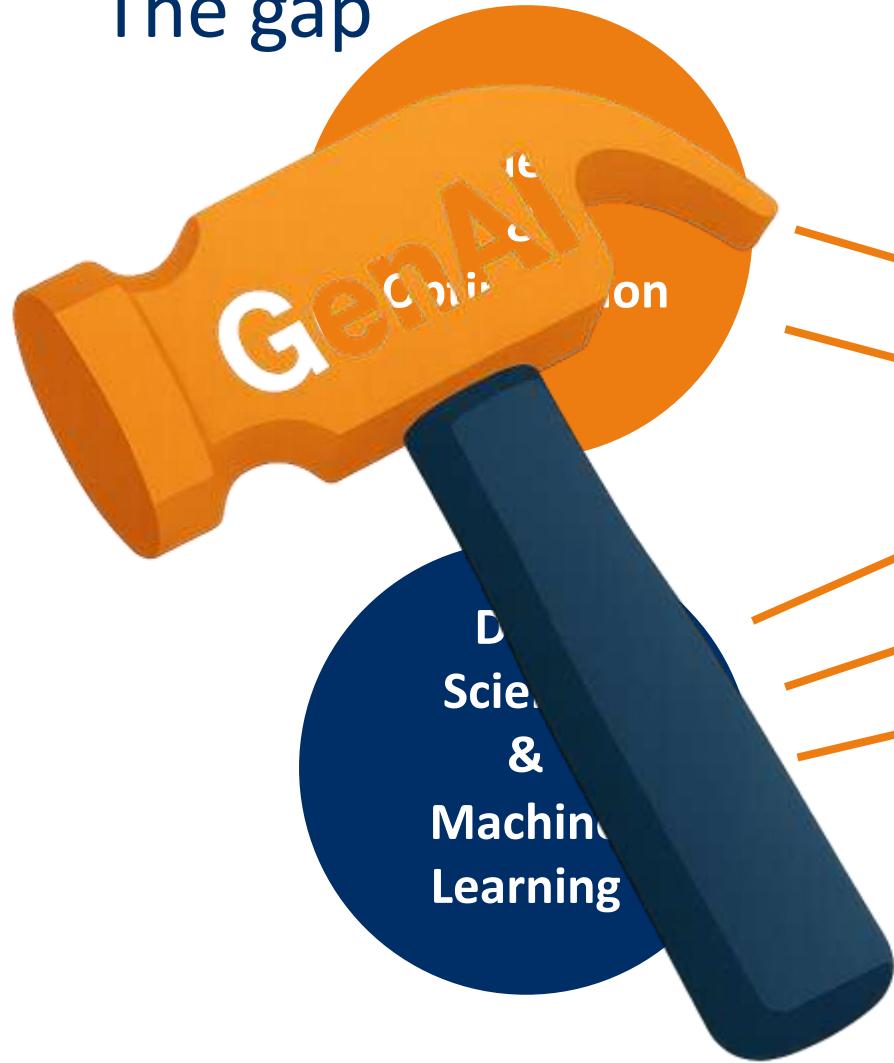
We

- find an **approximate** solution
- to a **simplified** problem
- defined in terms of **inaccurate** data
- and/or **estimated** parameters
- while **ignoring** uncertainty*

* Not all may be (equally) applicable



The gap



The solution
to our model...

- find an **approximate** solution
- to a **simplified** problem
- defined in terms of **inaccurate** data
- and/or **estimated** parameters
- while **ignoring** uncertainty



and its realization
in the real world!

Generative AI



The solution
to our model...

- find a **better** solution
- to a **more realistic** problem
- defined in terms of **more accurate** data
- and/or **better estimated** parameters
- while **considering** uncertainty

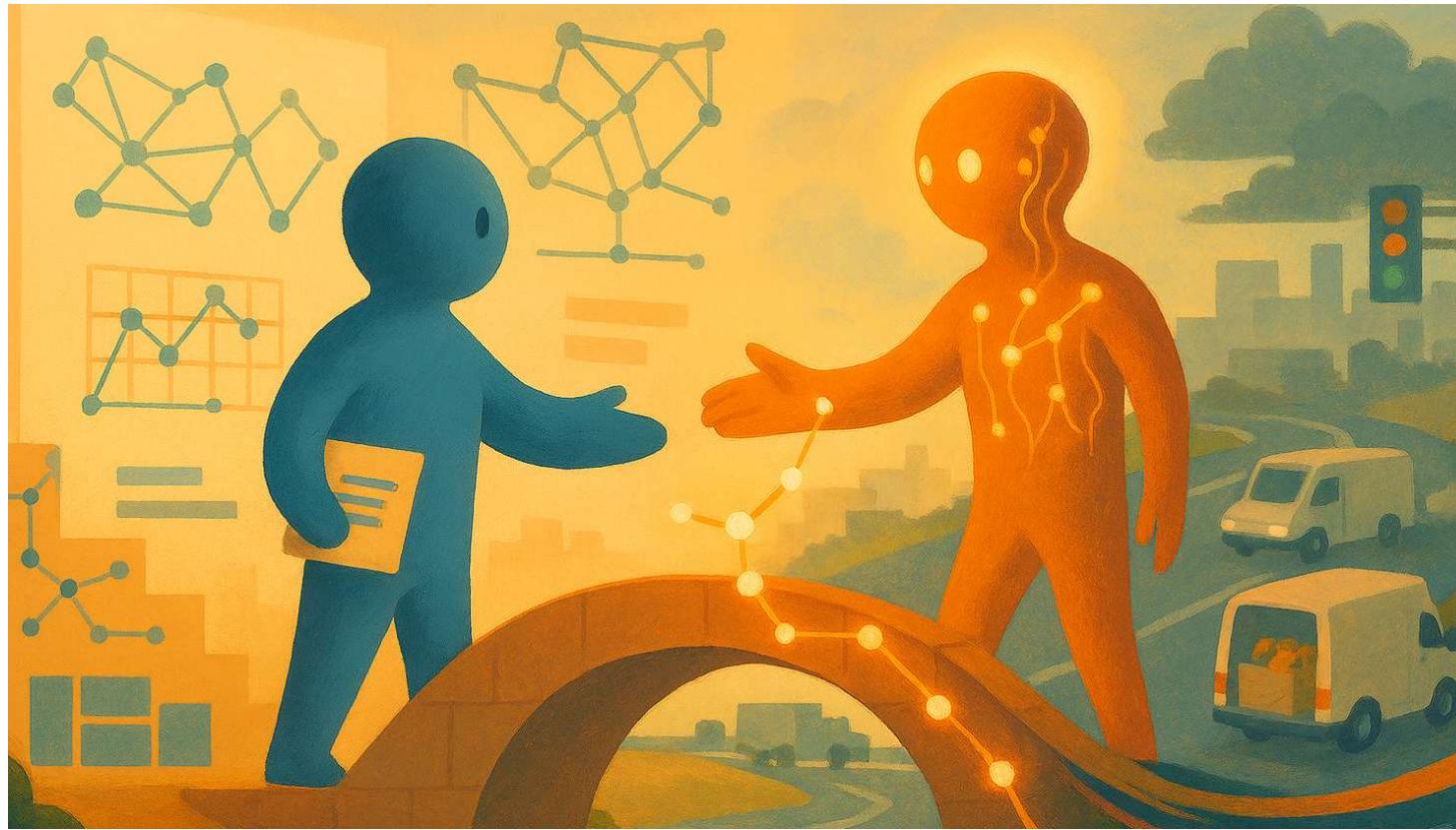


and its realization
in the real world!

Generative AI can help us bridge...



The model



The real world

Route optimization



Route when taking it slow, legally requiring an extra break.

Driver

Perfectly optimized (by ORTEC*) route with no break.

Workforce scheduling



“I'd prefer to work Saturday or Sunday but not both.”



But... how?

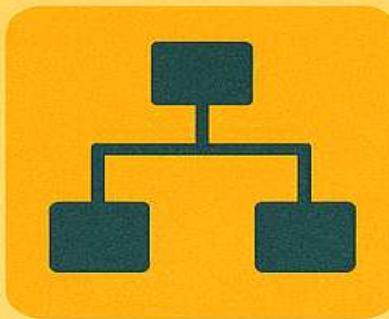


AI TAXONOMY



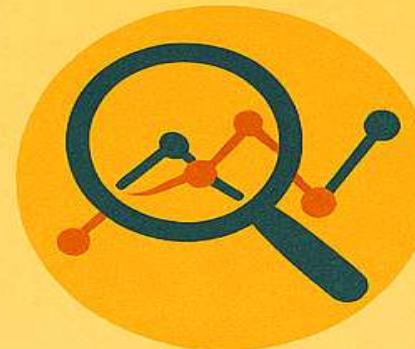
AI

Intelligent systems



OR

Decision optimization



ML

Learn from data



DL

Neural networks



RL

Learn by feedback



LLMs

Language models



GenAI

Content creation

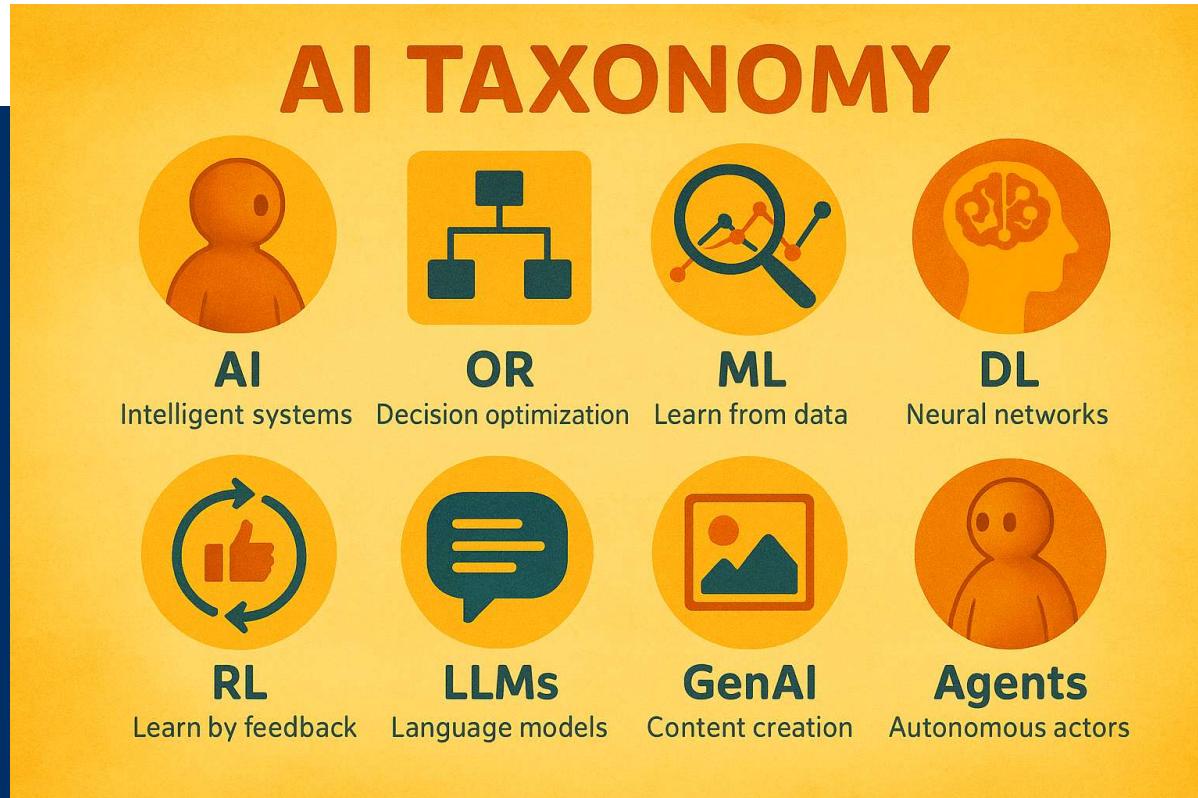


Agents

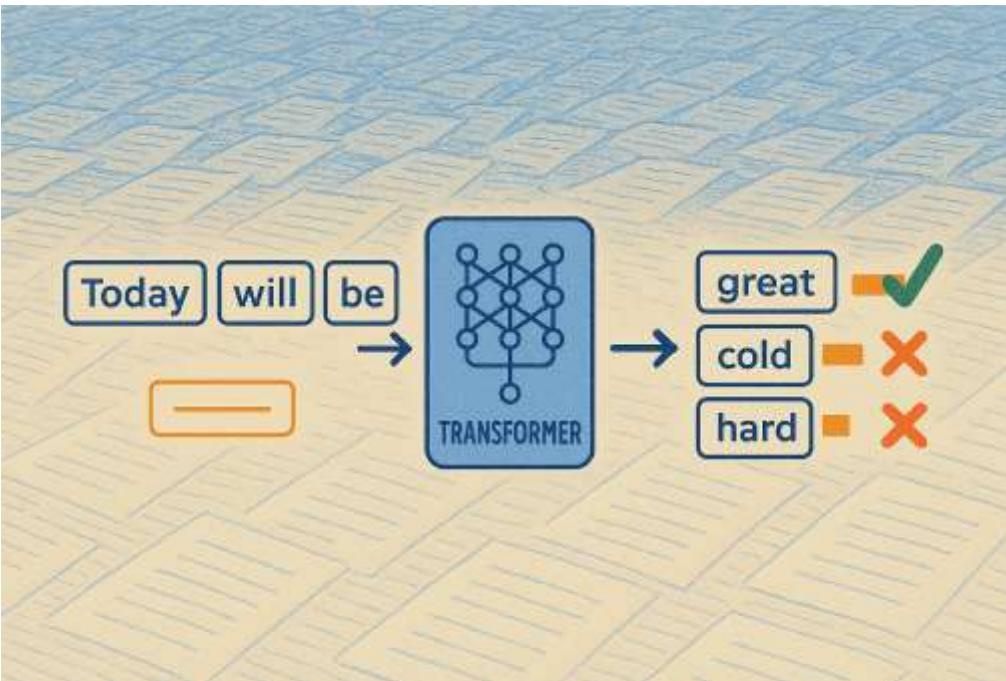
Autonomous actors

A taxonomy

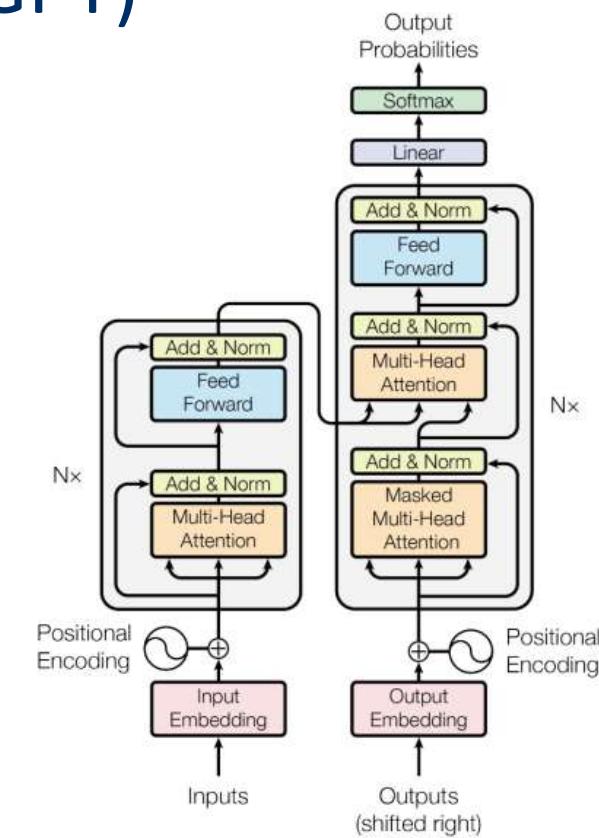
- **AI:** Intelligent systems
- **OR:** Decision optimization
- **ML:** Learn from data
- **DL:** Neural networks
- **RL:** Learn by feedback
- **LLMs:** Language models
- **GenAI:** Content creation
- **Agents:** Autonomous actors



Generative Pre-trained Transformer (GPT)

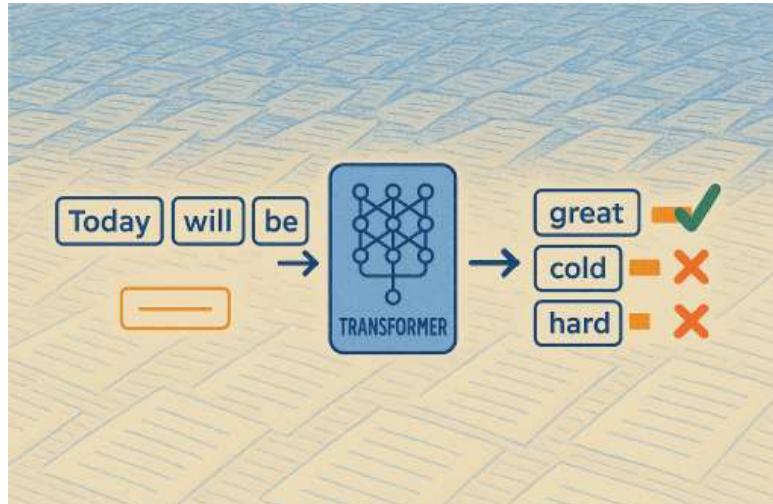


Generative pre-training

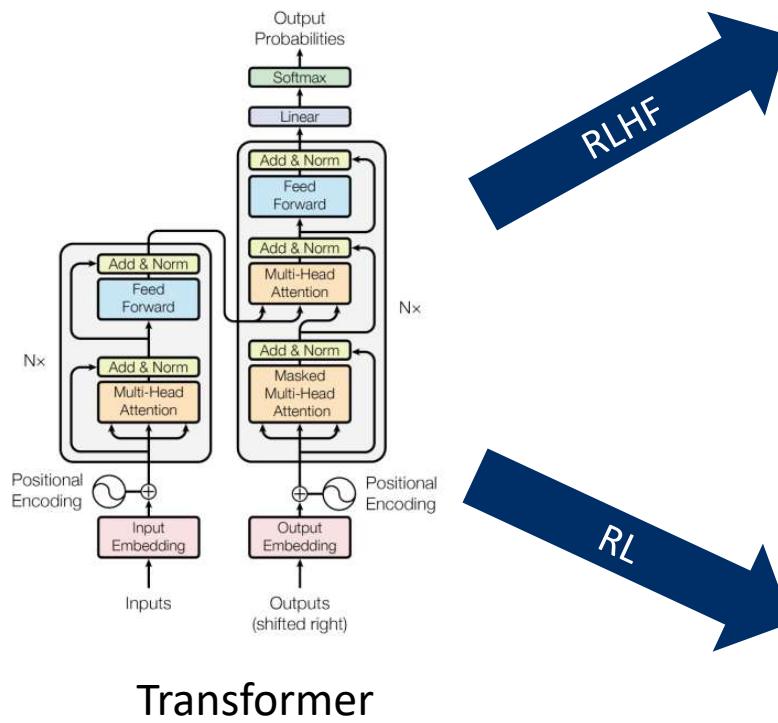


Transformer

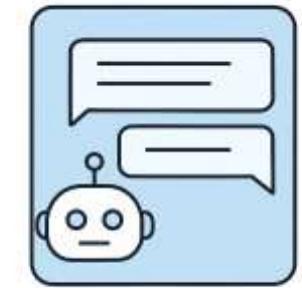
Fine-tuning



Generative pre-training



Transformer

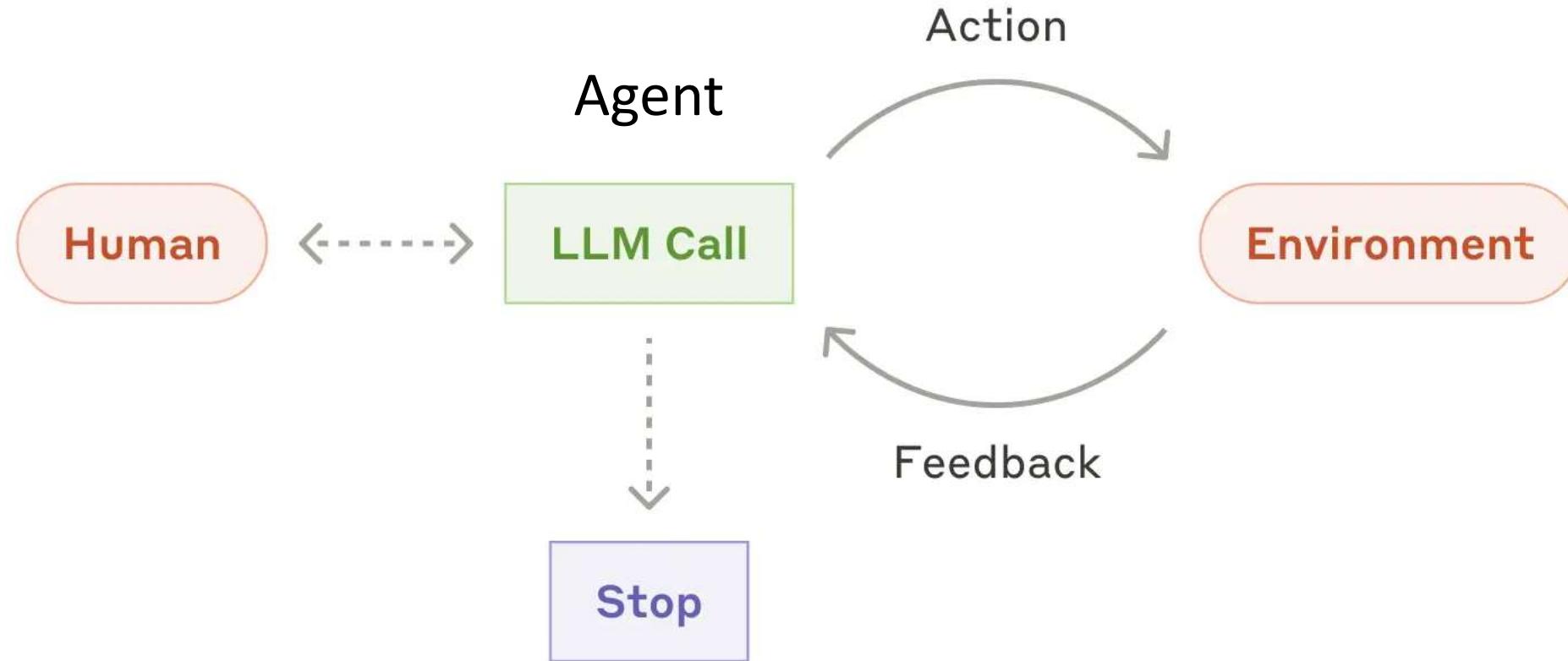


CHATBOT



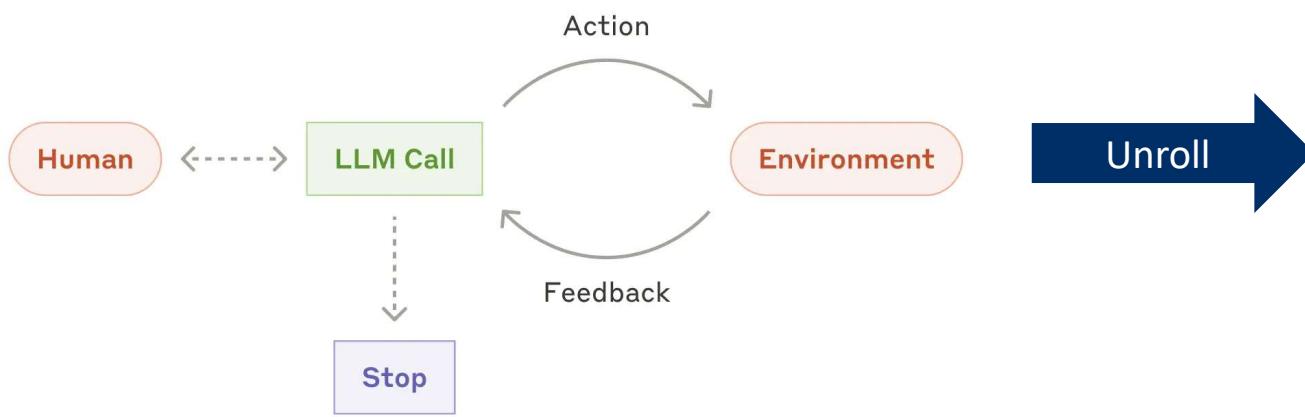
AGENT

AI Agent: LLM in a loop

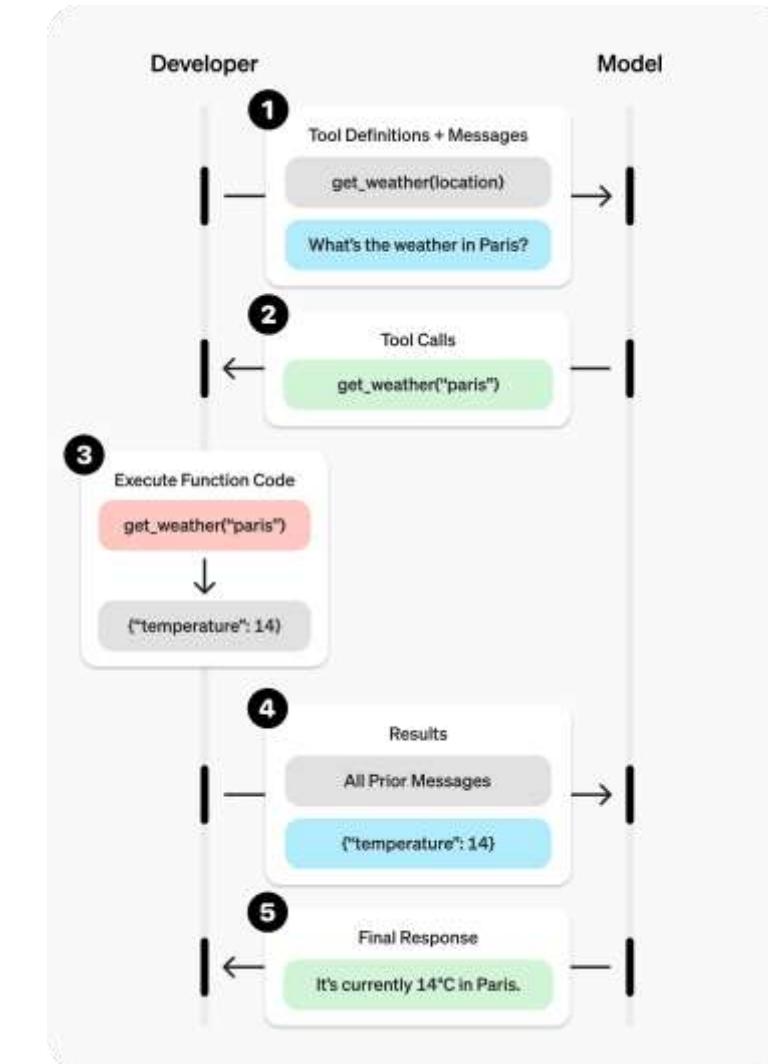


<https://www.anthropic.com/engineering/building-effective-agents>

Function calling



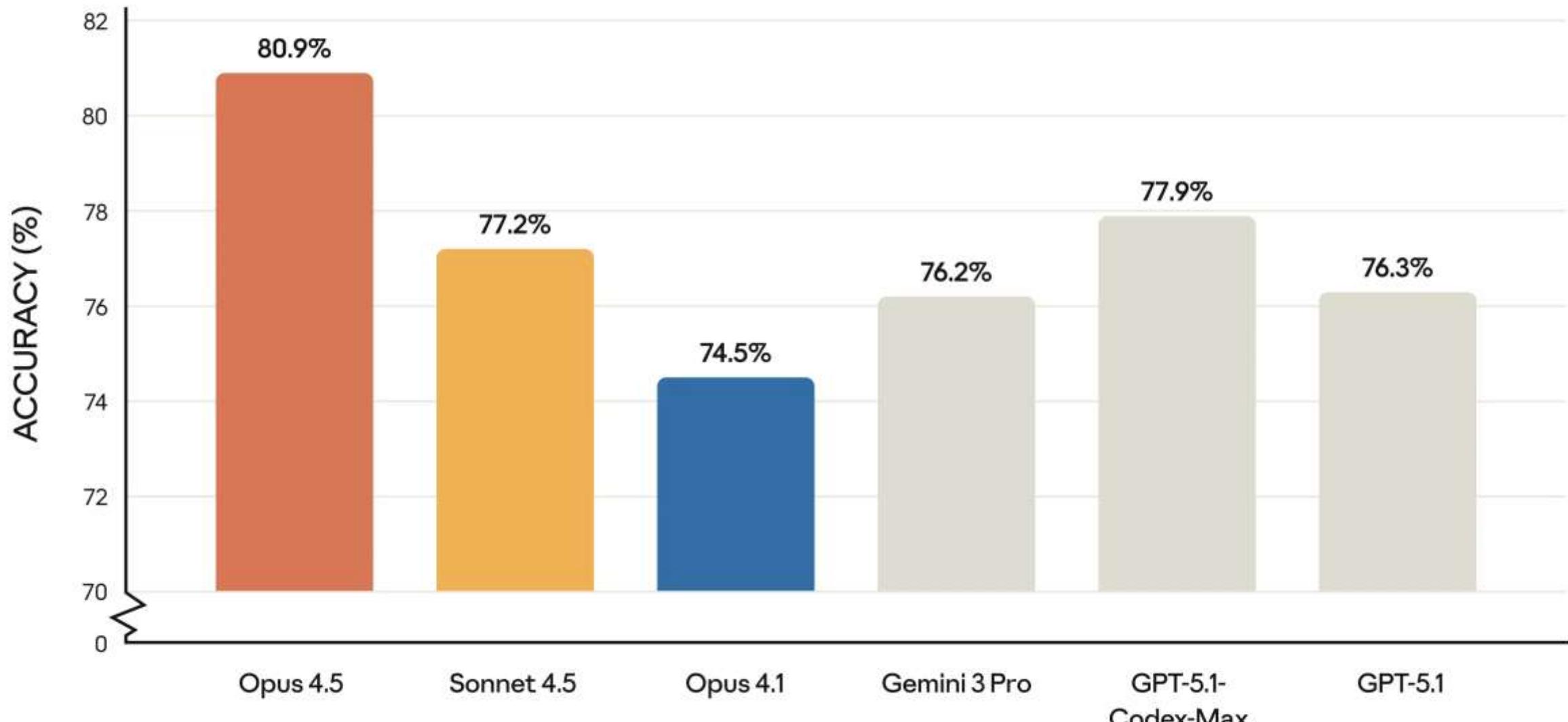
Unroll



<https://platform.openai.com/docs/guides/function-calling>

Software engineering

SWE-bench Verified (n=500)

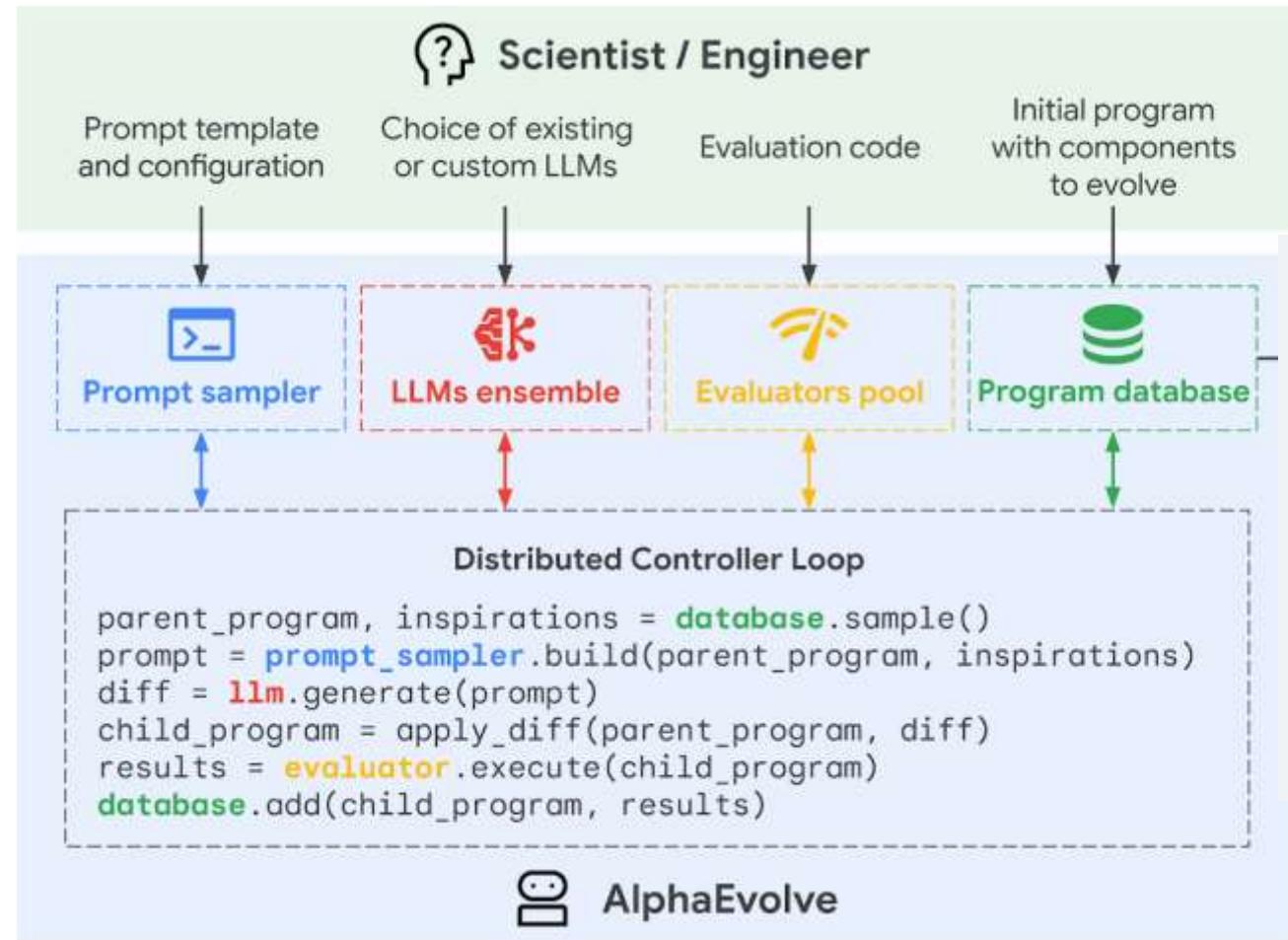


<https://www.anthropic.com/news/claude-opus-4-5>

Back to OR...



DeepMind: AlphaEvolve



AlphaEvolve: A Gemini-powered coding agent for designing advanced algorithms

AlphaEvolve: A Gemini-powered coding agent for designing advanced algorithms - Google DeepMind

Microsoft: AI for supply chain optimization

Here is the plan. You can hover your mouse on each node or edge to see more details.

```

graph LR
    S1 -- 80 --> R1
    S2 -- 50 --> R1
    S3 -- 100 --> R1
    S1 -- 100 --> R2
    S2 -- 50 --> R2
    S3 -- 80 --> R2
    R1 -- "20 L + 20 D" --> C1
    R1 -- "30 L + 20 D" --> C2
    R2 -- "10 L" --> C2
    R2 -- "30 L + 100 D" --> C3
  
```

Is it possible for Roastery 1 to be exclusively used by Cafe 2?

If we exclusively transport beans from Roastery 1 to Cafe 2, the cost would amount to \$2570, representing an approximate increase of $\frac{2570 - 2470}{2470} \approx 4\%$ compared to the current plan, which costs \$2470. Would you like to implement this change for future planning purposes?

Figure 1: *OptiGuide* interface.

Is it possible for Roastery 1 to be exclusively used by Cafe 2?

We need to write Python code to test the what-if scenario.

```
for c in cafes:
    if c != "cafe2":
        m.addConstr(
            y_light["roastery1", c] == 0, "")
        m.addConstr(
            y_dark["roastery1", c] == 0, "")
```

The result has $m.objVal = \$2570$, which is higher than $\$2470$ from the previous solution.

If we exclusively transport beans from Roastery 1 to Cafe 2, the cost would amount to $\$2570$, representing an approximate increase of $\frac{2570 - 2470}{2470} \approx 4\%$ compared to the current plan, which costs $\$2470$. Would you like to implement this change for future planning purposes?

No. Who is the contact person for supplier 1?

We can use SQL command to find out the information.

```
SELECT contact FROM suppliers WHERE id="supplier1";
```

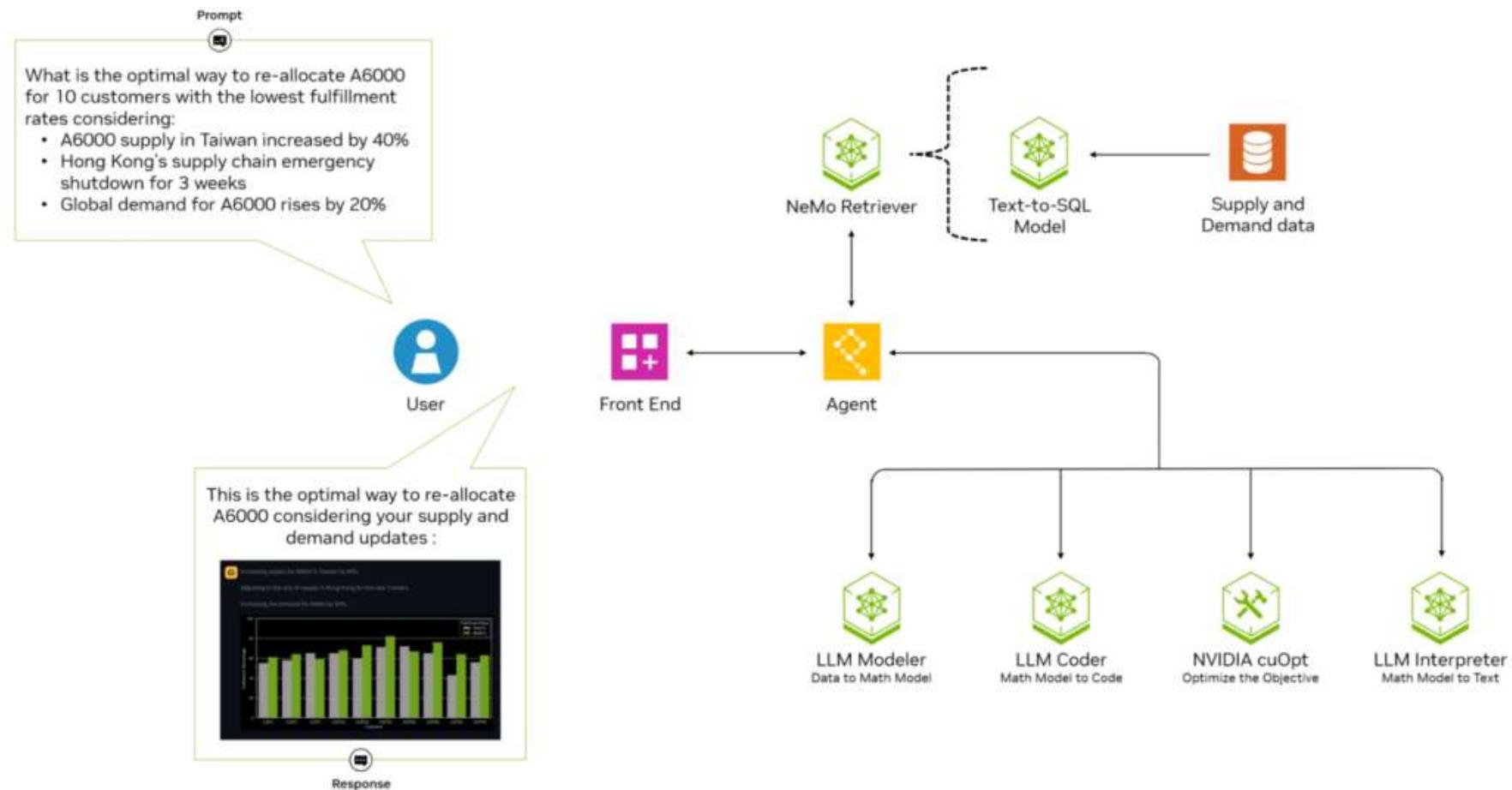
Safeguard Alert: sensitive information. Approval required!

Ticket submitted.

To answer this question, we need approval from administrator to access privileged information. We just created a ticket and will get back to you in 12 hours...

Waiting for response...

NVIDIA: supply chain agent



<https://developer.nvidia.com/blog/building-an-ai-agent-for-supply-chain-optimization-with-nvidia-nim-and-cuopt/>

Gurobi: modeling assistant



Search

INTRODUCTION

[About this repository](#)

[Getting started](#)

[Use cases](#)

LLM PROMPTING

[Writing your first prompt](#)

[Example prompts](#)

[A deeper look](#)

[Custom GPTs](#)

[Tips and pitfalls](#)

LEARNING OPTIMIZATION

[Comparing to Machine Learning](#)

[Modeling](#)

Welcome to Gurobi AI Modeling!



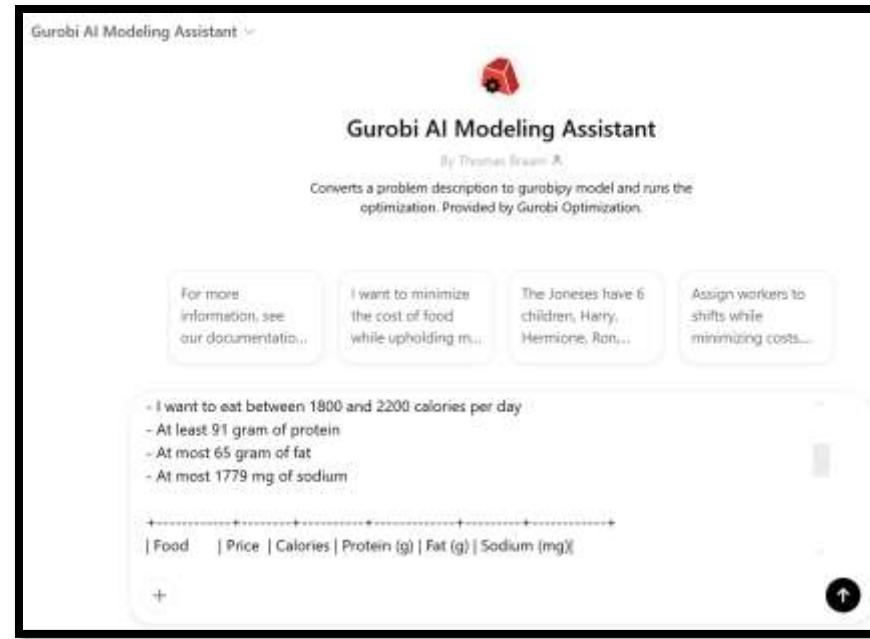
`gurobi-ai-modeling` aims to familiarize users with using generative AI for mathematical modeling. This repository contains documentation that features the following:

- Help users understand the capabilities and limitations of the state-of-the-art generative AI tools in the context of optimization.
- Propose best practices for using generative AI to build and solve optimization models and describe common pitfalls.
- Share examples of prompts and expected outcomes.
- Describe tools that can help create and solve models.

Contents

INTRODUCTION

- [About this repository](#)
 - [How to utilize this repository](#)
 - [Structure](#)
 - [What to take away from this](#)
- [Getting started](#)
 - [Using ChatGPT as a translator for mathematics](#)
- [Use cases](#)
 - [Specialties](#)



ORTEC: assistant

Home Zoetermeer test department AI Test

Filter: All territories

Territory assignment

Territories (4)

Name	Work time	Pal
Territory A	55 h 39 min	732
Territory Y	48 h 15 min	592
Territory C	58 h 43 min	712
Territory D	49 h 23 min	336

Customers (100)

Tags: All Unassigned Assigned

Search customers:

<input type="checkbox"/>	Name	Territory	Tot. duration	Tot. pal	Location	Tags
	Customer 1	-	1 h 0 min	20	Location 1	<input type="checkbox"/>
	Customer 2	Territory A	34 min	24	Location 2	<input type="checkbox"/>
	Customer 3	Territory Y	1 h 16 min	28	Location 3	<input type="checkbox"/>
	Customer 4	Territory A	40 min	20	Location 4	<input type="checkbox"/>
	Customer 5	Territory D	5 h 12 min	20	Location 5	<input type="checkbox"/>

Total work time: 212 h 0 min

ORTEC Assistant

My solution is not clustered enough, what can we do about that?

+

Thanks for letting me know you're concerned about clustering! When a solution isn't "clustered enough," it usually means that customers in the same territory aren't located close to each other, or the territories aren't as geographically compact as you'd like.

Here are some ways we can address this:

- We can run territory optimization prioritizing clustering, so territories become more geographically focused—helping ensure customers in the same territory are closer together.
- This will aim to improve clustering, but there may be minor trade-offs, like a slightly less balanced workload or a small increase in travel time.

Would you like me to start an optimization that puts more emphasis on clustering to create more compact territories for you?

[Message ORTEC Assistant](#)

AI-generated content may be inserted.



The use of Generative Artificial Intelligence (GenAI) in operations research: review and future research agenda

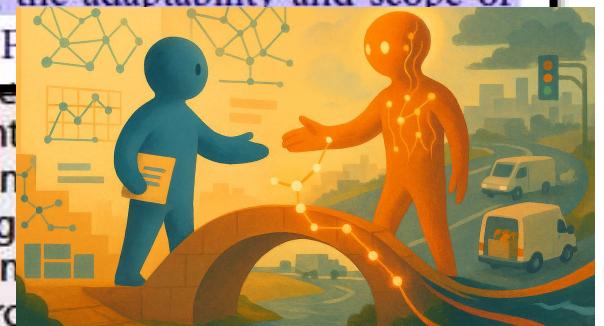
Qin Zhou^a and Jiuh-Biing Sheu^b

^aUniversity of Southampton Business School, Southampton, UK; ^bDepartment of Industrial Engineering and Management, National Taiwan University, Taiwan, R.O.C

ABSTRACT

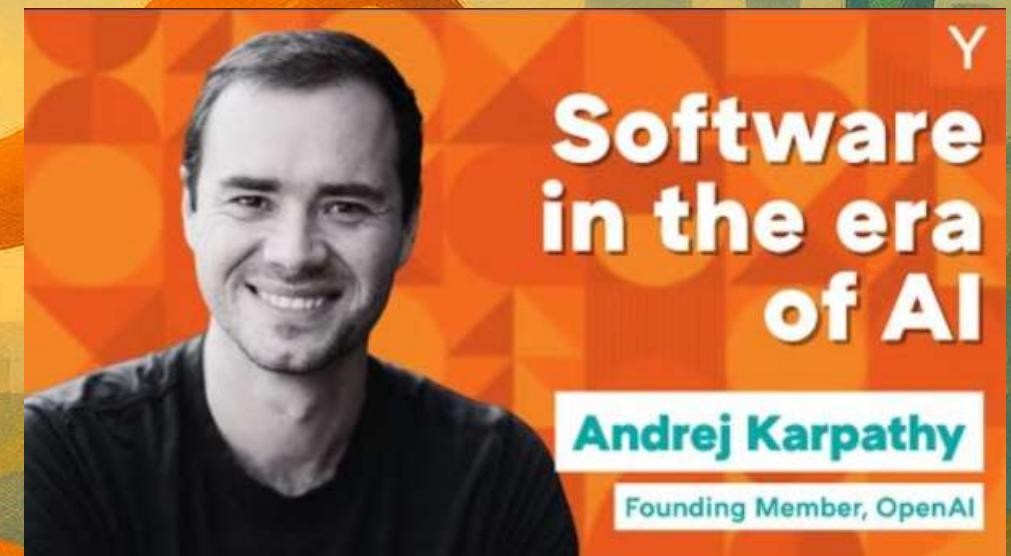
The emergence of Generative Artificial Intelligence (GenAI) represents a significant advancement in computational capabilities, offering transformative opportunities for Operations Research (OR). This study explores the role of GenAI in OR through a systematic literature review. Following a careful analysis of the collected papers, they are classified into two main categories based on the nature of the study: (1) application papers and (2) review and position papers. The latter provide a conceptual overview of GenAI's broader implications for OR, while the application papers are organized into a taxonomy encompassing three core dimensions: (1) GenAI for mathematical programming and optimization, (2) GenAI for stochastic systems, and (3) GenAI for simulation, strategic analysis, game theory, and risk management. Drawing insights from both conceptual and empirical studies, this review identifies cross-cutting themes and outlines a future research agenda to guide continued exploration at the intersection of GenAI and OR.

the OR workflow.² GenAI is now being applied to tasks such as translating natural language³ into formal problem formulations, generating novel solution algorithms, synthesizing high-fidelity data for robust simulation, and modelling complex strategic interactions with unprecedented realism.⁴ Its unique strength lies in bridging the gap between the unstructured, human-centric world of language and the structured, formal domain of mathematical optimization, thereby enhancing the adaptability and scope of operational processes.⁵ Future research





Software



Software Operations Research in the era of AI

- LLMs = operating systems
 - We're in the 70s: terminal with 100k memory
- LLM psychology
 - Encyclopedic knowledge with severe amnesia
- Opportunities
 - Partial autonomy AI applications:
 - Context management
 - LLM orchestrations
 - Integrated GUI
 - Autonomy slider
 - Human-AI verification, vibe-coding, demo (works.any()) vs product (works.all()), build for agents



[Andrej Karpathy: Software Is Changing \(Again\)](#)

The Takeaway

“Real world problems are rarely as clean as our equations!”



Y

**Software
in the era
of AI**

Andrej Karpathy

Founding Member, OpenAI

