Mentimeter

BOMOS & AI

5 maart ‘25

Wouter van den Berg | Beeld & Geluid

TNO innovation for life

23

Contact

Wouter van den Berg

Consultant
Semantic Interoperability
@ TNO

Wouter van den berg

wouter.vandenberg@tno.nl

TNO Data Ecosystems

https://semantic-treehouse.nl

Recente activiteiten

- BOMOS klankbordgroep
- Europese data space projecten
  - Energy (BD4NRG, Enershare, Hedge-IoT)
  - Agriculture (ZeroW)
  - Best practices (DSSC, Data Act standardisation request)
- Centre of Excellence (CoE) for Data Sharing & Cloud
- EDI standaardisatie & Semantic Treehouse
  - E-factuur
  - SETU
  - Ketenstandaard
  - NEN

Mentimeter

Strategy

- Governance
- Vision
- Finance

Implementation support

- Training
- Pilot
- Helpdesk
- Validation & Certification
- Module development

Tactical

- Community
- Architecture
- Quality policy benchmarking
- Rights policy
- Adoption & recognition

Operational

- Initiation
- Wishes and requirements
- Documentation
- Development
- Implementation

Communication

- Promotion
- Publication
- Complaint handling

1

Mentimeter

AI → Generatieve AI → Large Language Models

Mentimeter

Agenda

1. Introductie
2. Peiling (mentimeter)
3. AI landschap
4. Mapping experiment
5. Potentieel van AI voor BOMOS (mentimeter)
6. Q&A en afsluiting

Mentimeter

Wie bent u?

2

36

Mentimeter

Wie bent u?

2

36

Mentimeter

Wie bent u?

2

36

Mentimeter

Wie bent u?

2

36

Mentimeter

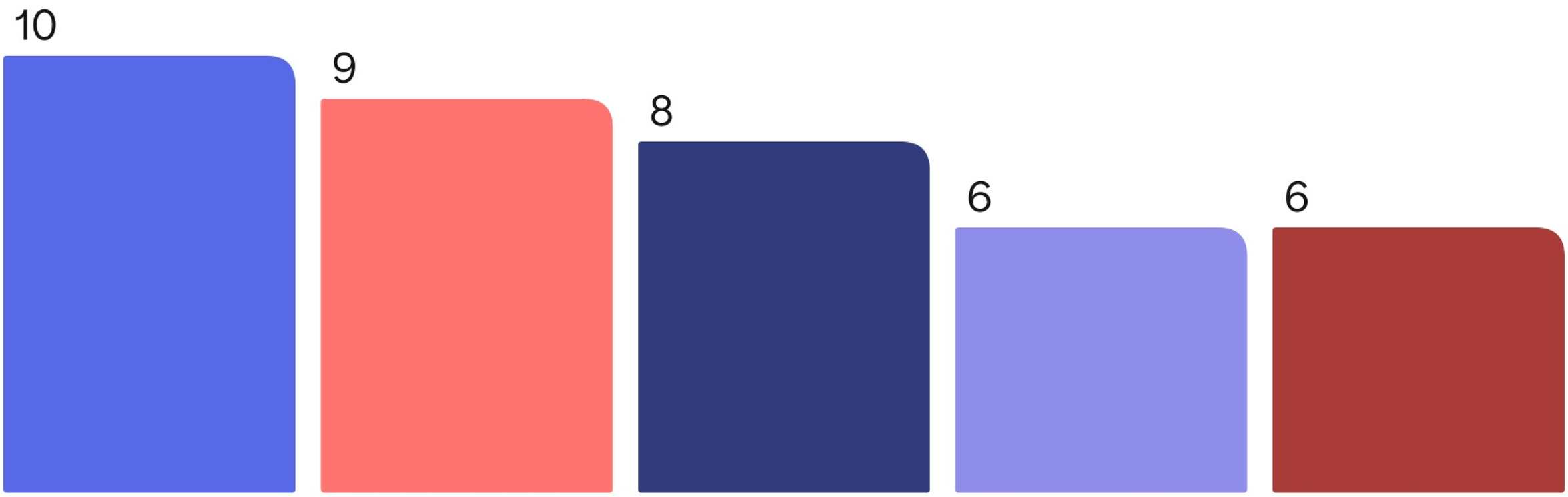
Wie bent u?

2

36

Mentimeter

Wat is uw bekendheid met BOMOS?



Zeer bekend; het is onderdeel van mijn werk of was dat lange tijd

Bekend; ik heb een basis begrip van wat BOMOS is of wel eens toegepast

Neutraal; ik heb de documentatie wel eens bekeken of weet die te vinden

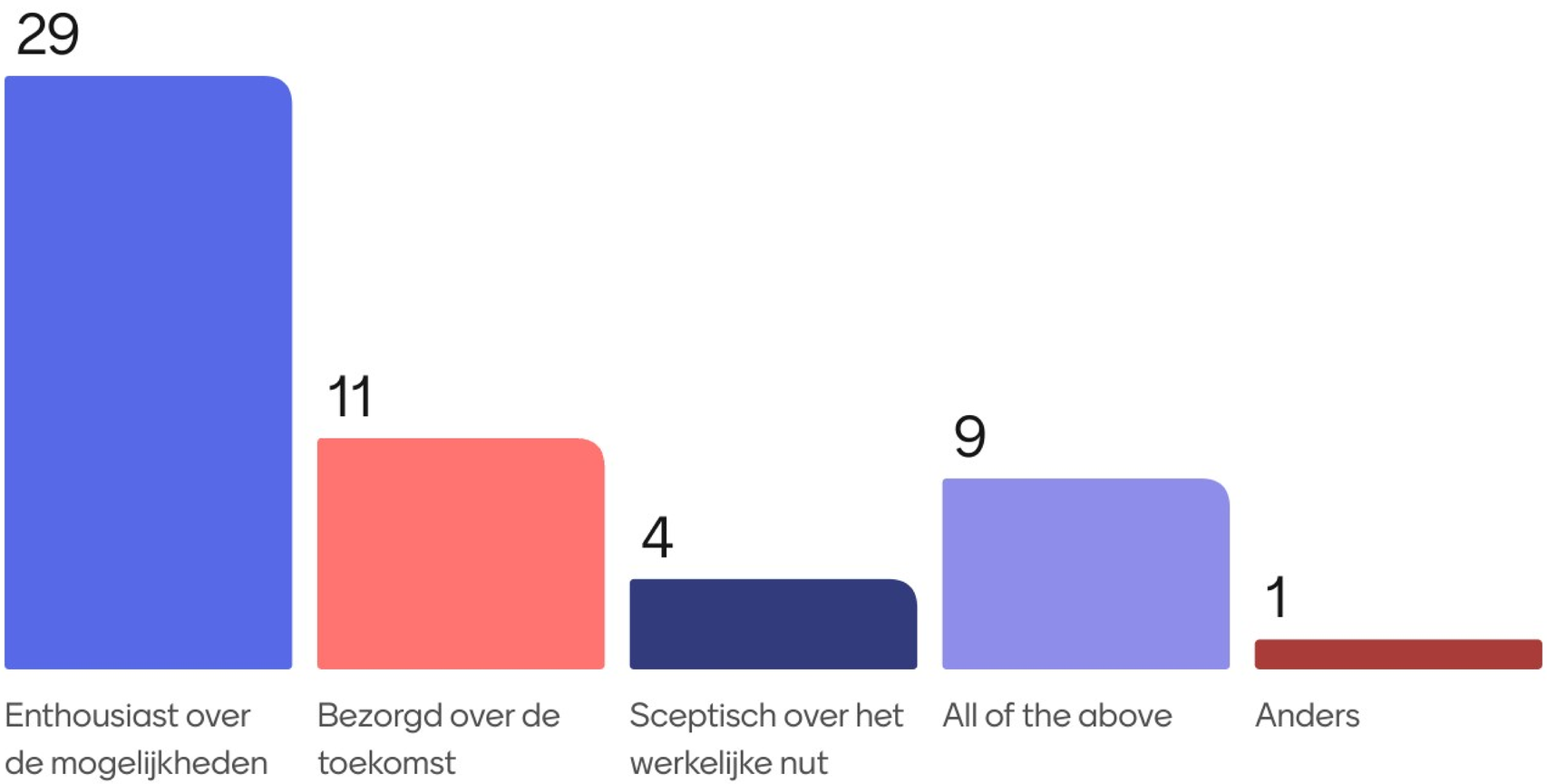
Een beetje onbekend; ik heb van BOMOS gehoord maar niet veel meer dan datnd

Onbekend; ik heb recent voor het eerst van BOMOS gehoord

39

Mentimeter

Wat is uw gevoel t.a.v. AI ontwikkelingen?



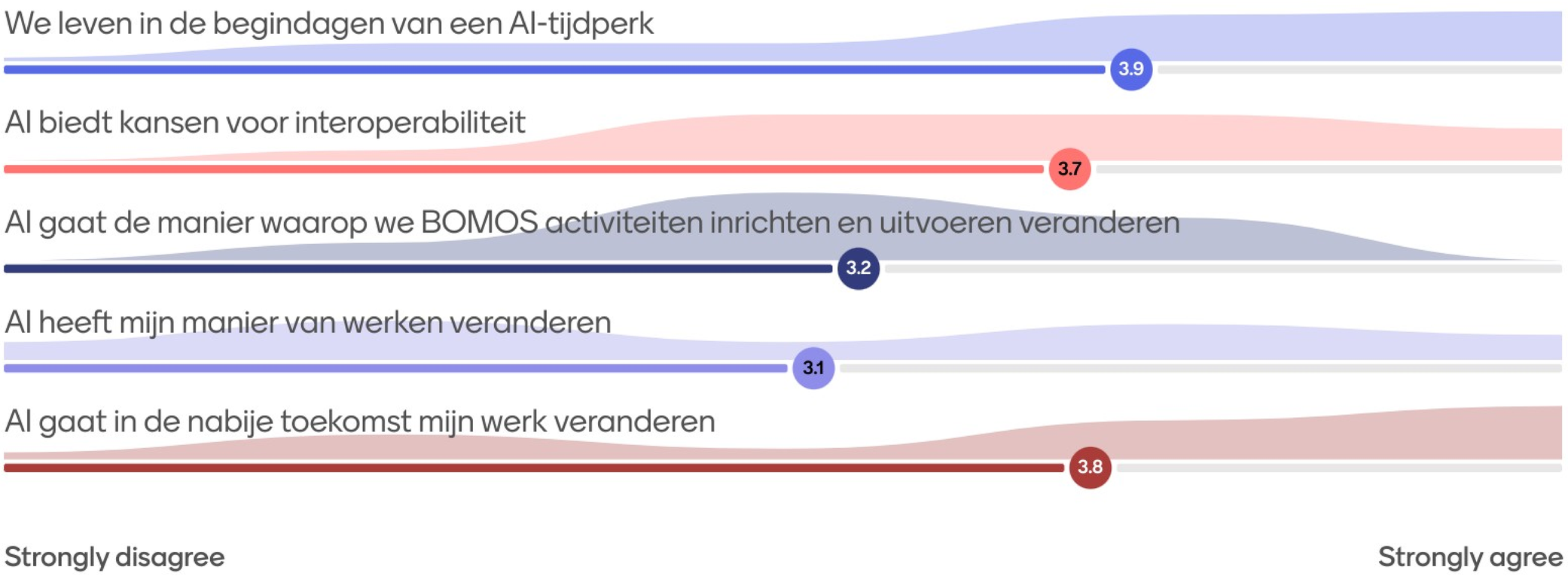
eleid

1

39

Mentimeter

Wat is uw houding t.a.v. AI ontwikkelingen?

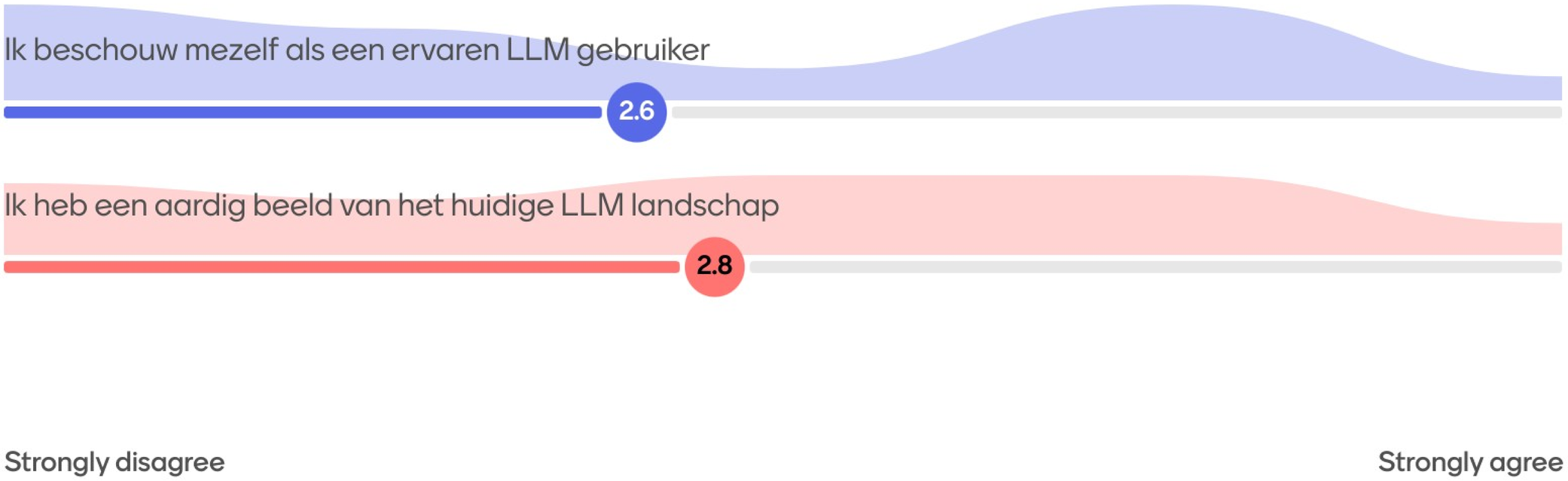


1

38

Mentimeter

Wat is uw kennis en ervaring met LLMs?



1

40

Mentimeter

Agenda

1. Introductie
2. Peiling (mentimeter)
3. AI landschap
4. Mapping experiment
5. Potentieel van AI voor BOMOS (mentimeter)
6. Q&A en afsluiting

Mentimeter

AI landschap (verouderd)

Service	Best Model	Live Mode	"Reasoning"	Web Access	Generates Images	Deep Research	Executes Code	Data Analysis	Sees images	Sees video	Reads Docs
OpenAI ChatGPT	GPT-4o	✓ Full multimodal	✗	✓	✓ DALL-E3	✓	✓	✓	✓	In Live Mode	✓
	o1/o3 family	✗	✓	✗	✗	✓	✓	✗	✓	✗	✗
Microsoft Copilot	"Copilot"	Voice only	✓	✓	✓ DALL-E3	✗	Limited	✗	✓	✗	✓
Anthropic Claude	Claude 3.5	✗	✗	✗	✗	✗	✓	Limited	✓	✗	✓
Google Gemini	Gemini family	Voice only	✓	✓	✓ Imogen-3	✓	Limited	Limited	✓	✓	✓
X.ai Grok	Grok-2	✗	✗	✓ Mostly X	✓ Aurora	✗	✗	✗	✓	✗	✓
DeepSeek	DeepSeek v3	✗	✓	✓	✗	✗	✗	✗	✓	✗	Limited

Bron: oneusefulthing.org

AI landschap (namen)

When picking among the 9 AI models that are now available from OpenAI, the rules are easy:

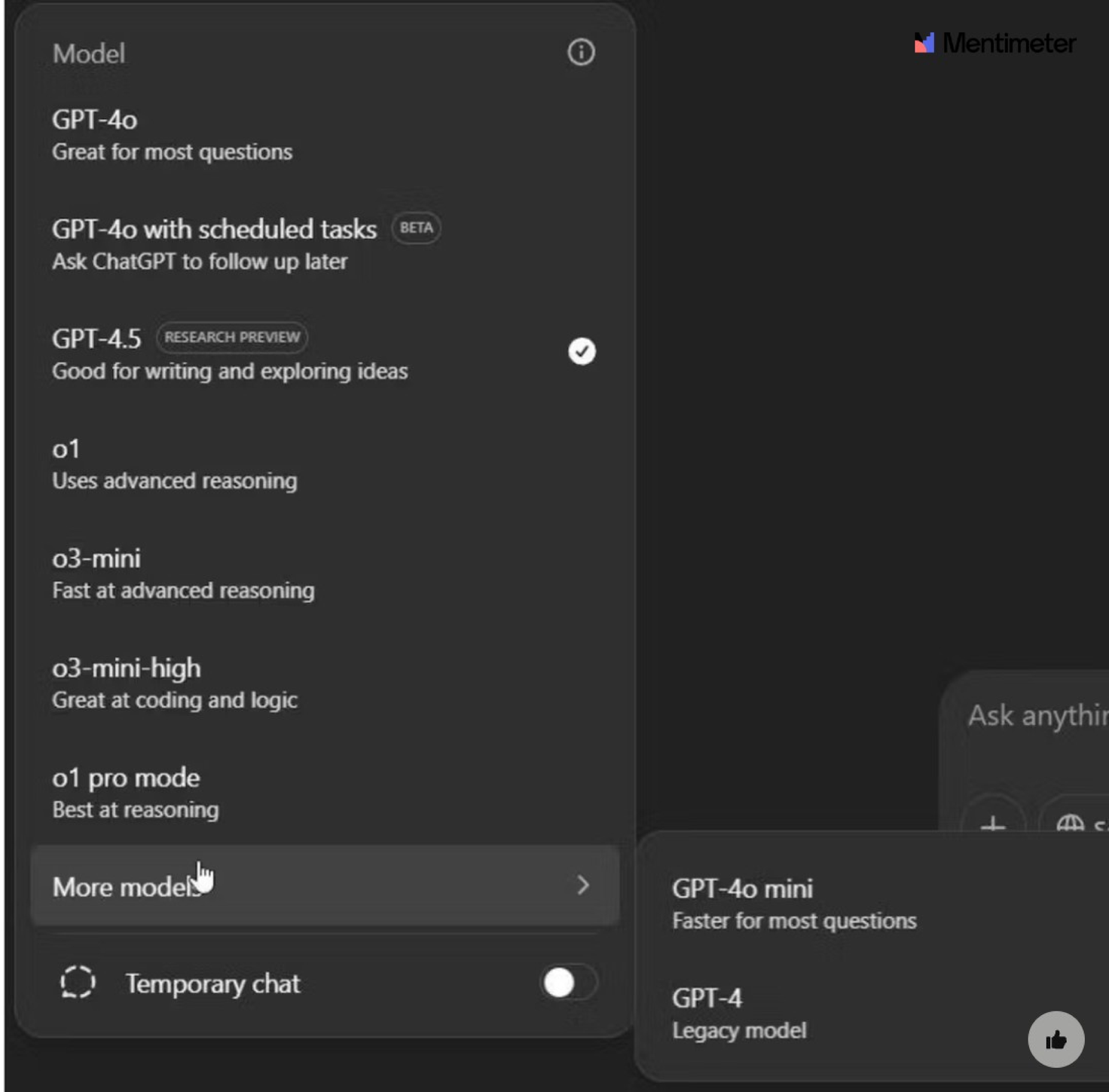
1) The model with the biggest number is mostly not the best

2) Mini means worse, except for the mini that is the second best

3) o1 pro beats o3-mini-high beats o1 beats o3-mini, naturally (unless you are doing creative work then GPT-4.5 probably beats o1 pro, maybe)

4) Also some of them can see images and some can use the web and some do search even when search is turned off and some of them can run code and some cannot.

[LinkedIn post van Ethan Mollick]



AI landschap

Sommige taken zijn een makkie voor AI

Om vervolgens totaal te falen bij een taak van vergelijkbare moeilijkheidsgraad

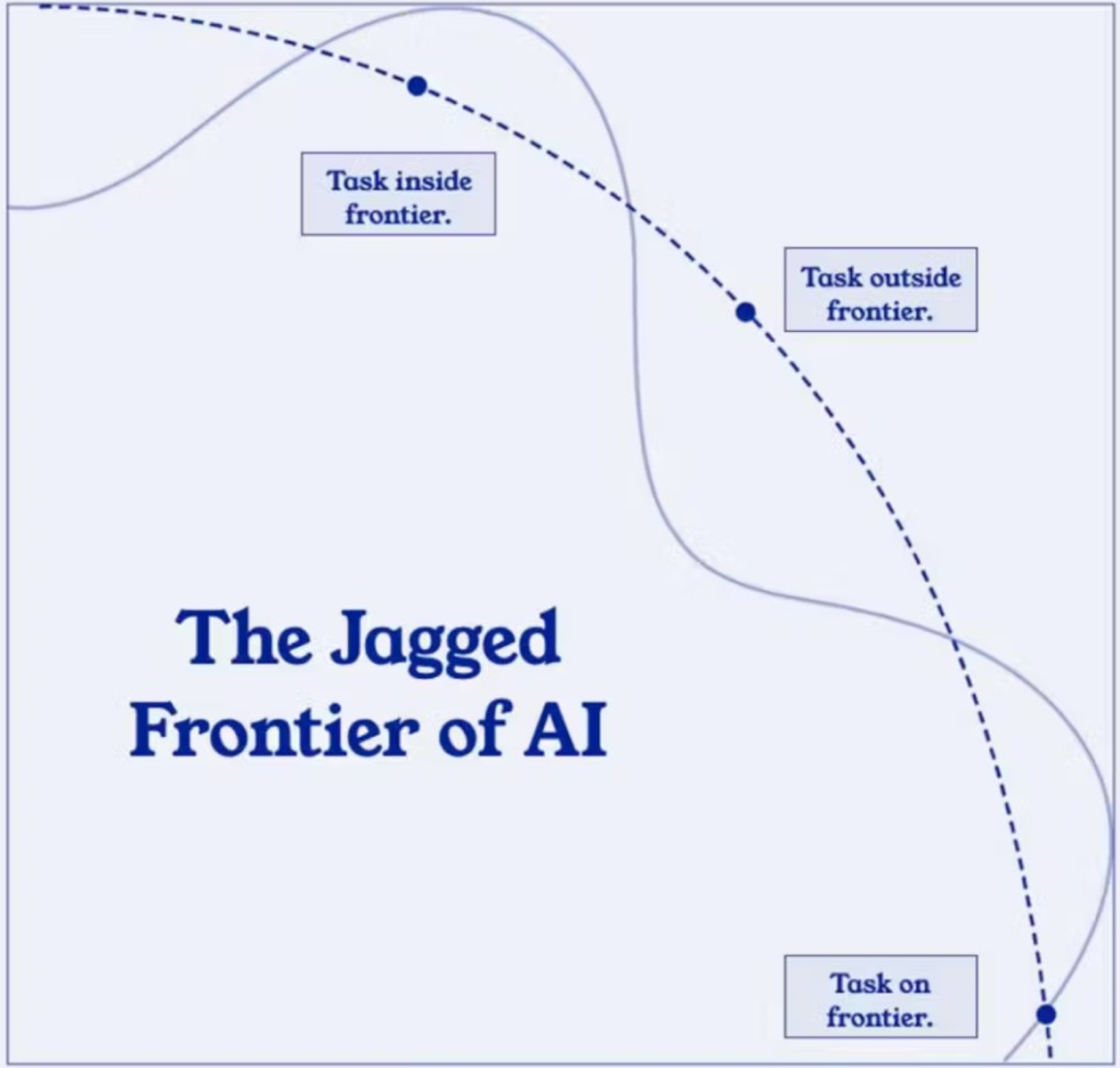
Bijvoorbeeld:

- “how many times does the letter ‘r’ occur in the word ‘strawberry’?

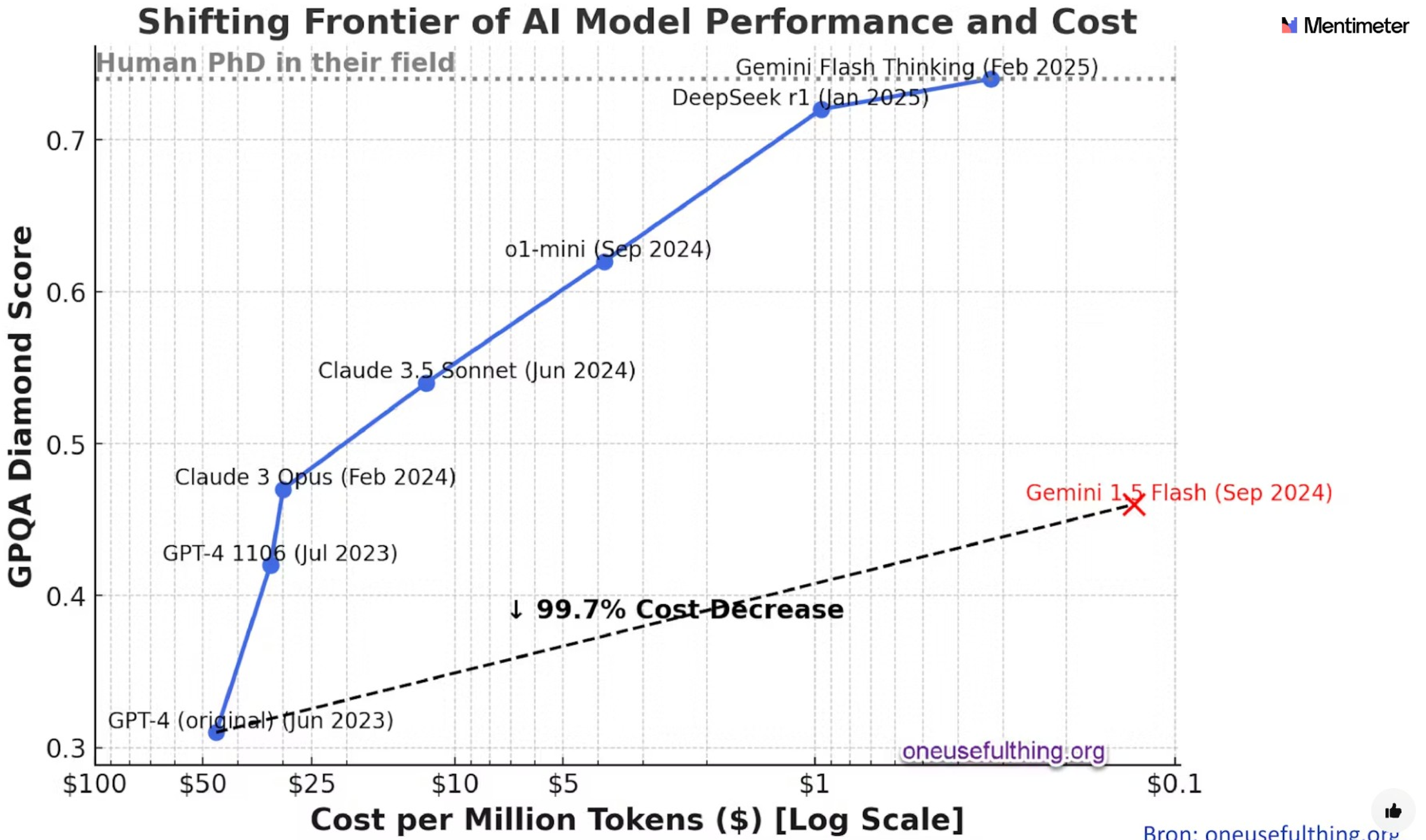
  Mening LLM: 2.
  Mijn mening: 3

- “Naar welke property uit de SAREF ontologie moet ik deze JSON key mappen? `Device_serial_number`”

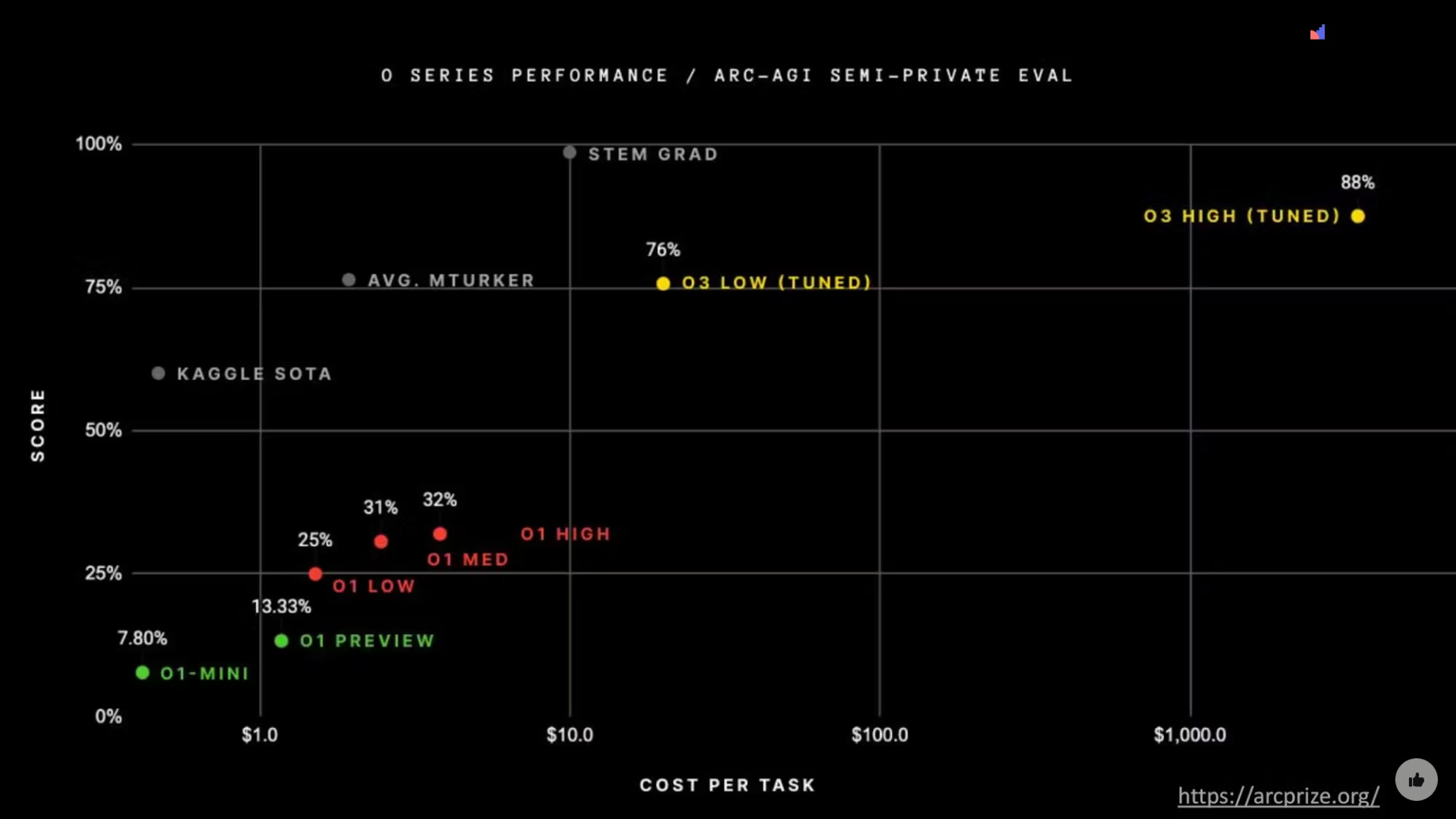
  LLM: daar is een extensie op SAREF voor nodig
  Juiste antwoord: `s4ener:serialNumber`



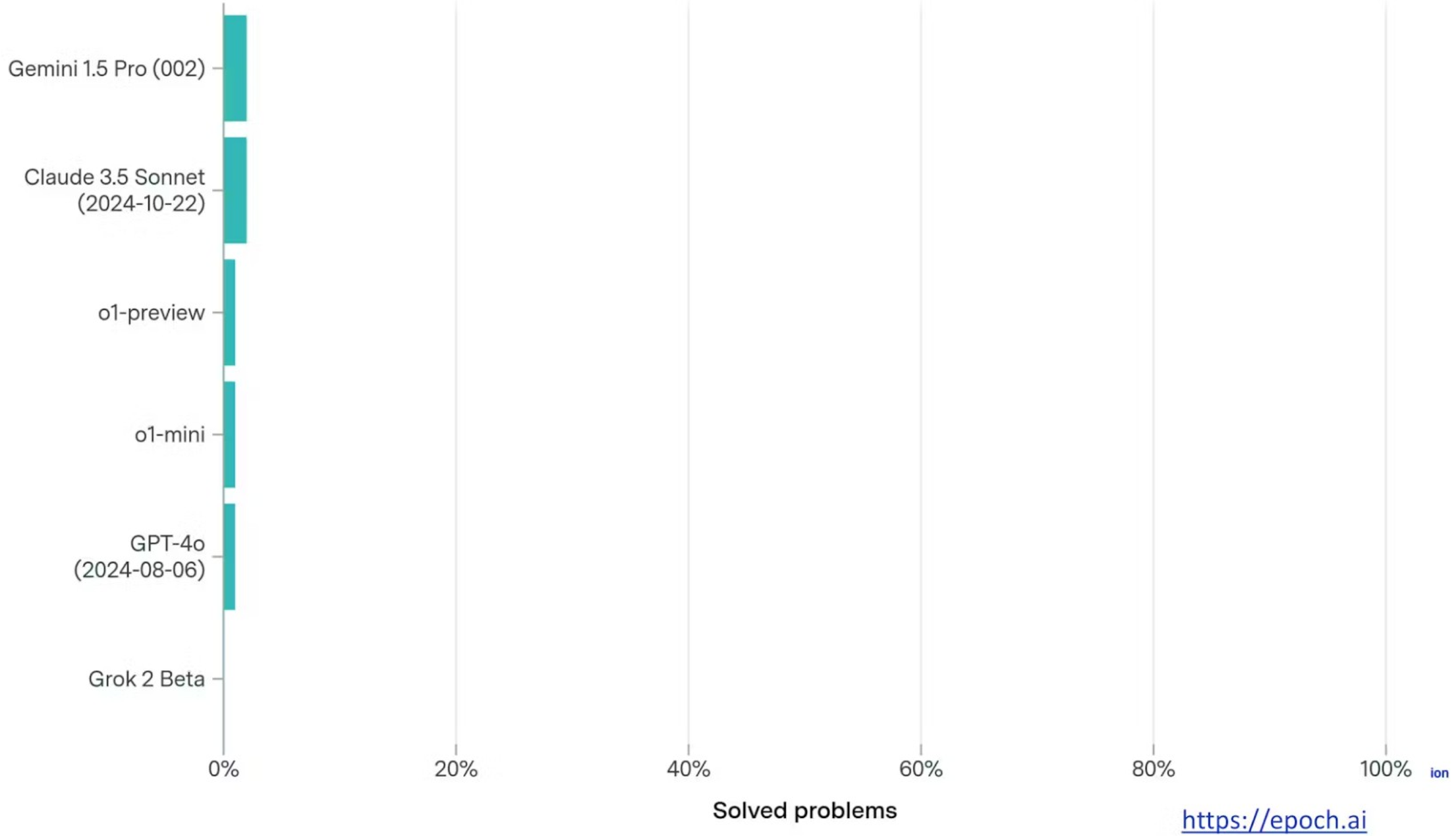
Dell'Acqua, Fabrizio, et al. "Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality." *Harvard Business School Technology & Operations Mgt. Unit Working Paper* 24-013 (2023).



Bron: oneusefulthing.org


O SERIES PERFORMANCE / ARC-AGI SEMI-PRIVATE EVAL
100%
75%
50%
25%
0%
SCORE
STEM GRAD
88%
O3 HIGH (TUNED)
76%
AVG. MTURKER
O3 LOW (TUNED)
KAGGLE SOTA
31%
32%
25%
O1 HIGH
O1 MED
O1 LOW
13.33%
7.80%
O1 PREVIEW
O1-MINI
$1.0
$10.0
$100.0
$1,000.0
COST PER TASK
https://arcprize.org/

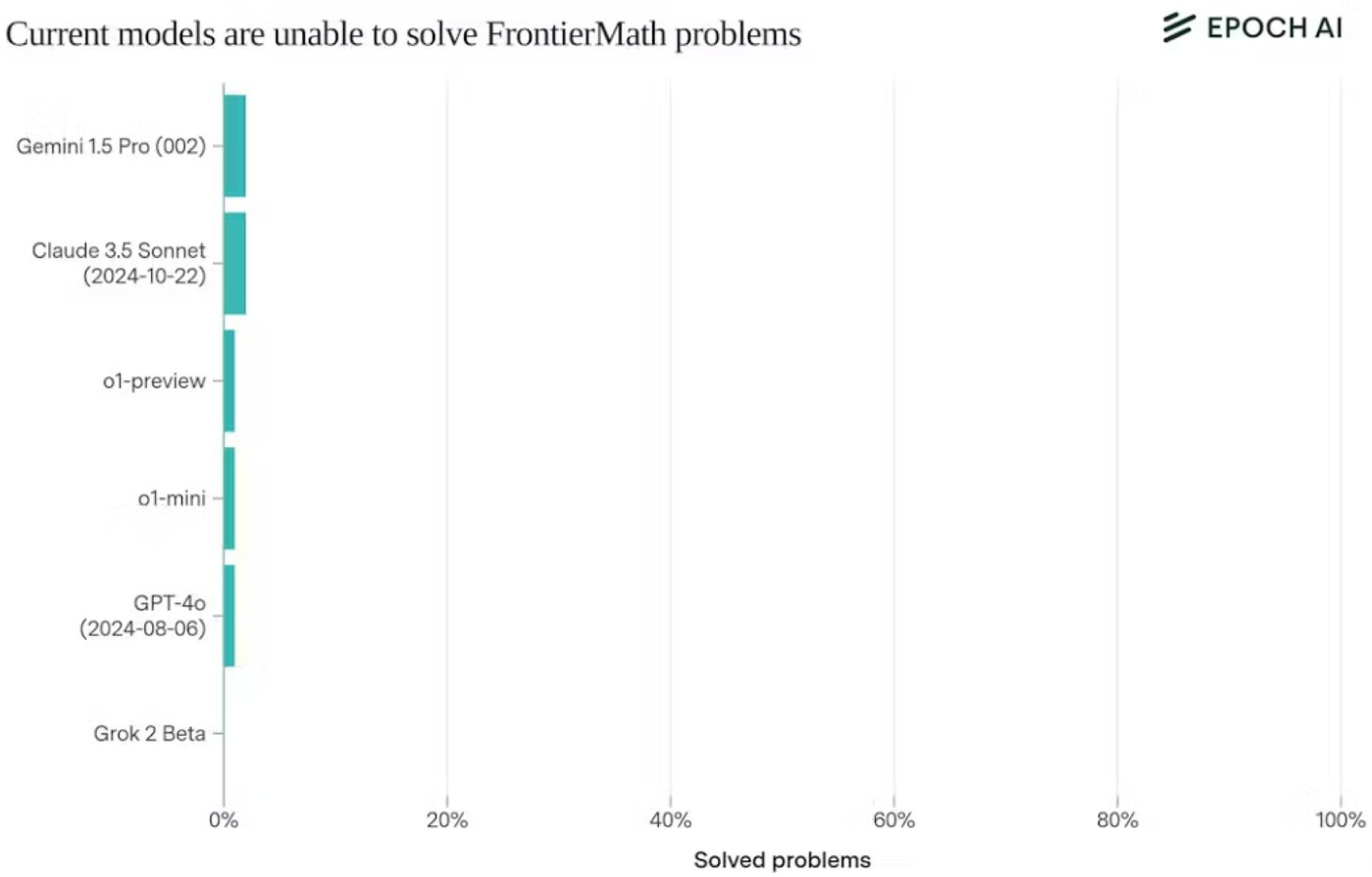
Current models are unable to solve FrontierMath problems



https://epoch.ai

AI landschap

o3



https://epoch.ai

LLMs: NEXT PUBLIC RELEASES IN 2025

ESTIMATES AS OF FEBRUARY 2025

Mentimeter

Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec
	OpenAI GPT-4.5 (Orion)	OpenAI GPT-5 (GPT + o3)	OpenAI o4
	xAI Grok-3	xAI Grok-Video	xAI Grok-4
	Anthropic Reasoning model	Anthropic Claude 4	
	Meta AI Llama 4		Meta AI Llama 5
		Google DeepMind Gemini 3	Google DeepMind Gemma 3
		Microsoft Phi-5	Baidu ERNIE 5

Estimates only, selected highlights only. Full models table available at https://lifearchitect.ai/models-table/ Alan D. Thompson. February 2025. https://lifearchitect.ai/

Mentimeter

Take aways

1. om een goed idee te krijgen van de potentie voor je use case moet je vlieguren maken

2. de *jagged frontier* is daarnaast continu in beweging

3. Het ene model is het andere niet;

4. (en de wereld van AI is veel groter dan LLMs)

Mentimeter

Agenda

1. Introductie
2. Peiling (mentimeter)
3. AI landschap
4. Mapping experiment
5. Potentieel van AI voor BOMOS (mentimeter)
6. Q&A en afsluiting

Mentimeter

Vertaling experiment: de dataset

- Interconnect OfficeGraph
  - Github
  - Paper: van der Weerdt, Roderick, et al. "OfficeGraph: A Knowledge Graph of Office Building IoT Measurements." European Semantic Web Conference. Cham: Springer Nature Switzerland, 2024.
  - 444 devices of 19 types
  - 14 million measurements
  - 11 months
  - 90 million RDF triples
- SAREF ontologie


```
{ "data_battery_level":0.34,
"data_building":"urn:Building:SmartThings:d563d271-4855-4728-8367-cbfb205cf3da",
"data_room":"urn:Room:SmartThings:38f802d2-bb3c-4bf4-ba28-23c5ae3a64d0",
"data_status":0,
...}
```

Mentimeter

Vertaling experiment: 3 niveaus

- Simple:
  - Input: JSON variable name
  - Output: ontology property identifier
- Complicated:
  - Input: JSON variable name
  - Output: RDF triple(s) following ontology
- Hard:
  - Input: JSON object
  - Output: RDF triple(s) following ontology

```
In:     "data_battery_level"
Out:     saref:hasValue
```

```
In:     "data_battery_level"
Out:    <meas_uri> rdf:type saref:Measurement .
        <meas_uri> saref:hasValue <input> .
        <device> saref:isMeasuredIn <uom> .
```

Mentimeter

Vertaling experiment: voorbeeld JSON bericht


```
{
  "data_battery_level": 0.34,
  "data_building": "urn:Building:SmartThings:d563d271-4855-4728-8367-cbfb205cf3da",
  "data_room": "urn:Room:SmartThings:38f802d2-bb3c-4bf4-ba28-23c5ae3a64d0",
  "data_status": 0,
  "device_link_type_description": "ZigBee",
  "device_manufacturer_description": "Samsung",
  "device_model_description": "SmartSense Button",
  "device_parent_serial_number": "urn:Device:SmartThings:4c73905f-b0a9-46d6-99df-11fd0a45498d",
  "device_product_id": 117,
  "device_serial_number": "urn:Device:SmartThings:c6e8db8f-6f27-4482-979e-2b790628f902",
  "device_type_description": "SmartThings",
  "message_timestamp": "2022-03-01T23:00:08.6609426Z",
  "message_type_description": "Heartbeat",
  "message_type_id": 2
}
```

Semantic Treehouse

SPECIFICATIONS VALIDATOR | ISSUES GROUPS | DOCUMENTATION

BERT HEUVEL

Mentimeter

Sensor Device JSON data

search

1...1 device_link_type_description
1...1 data_building
1...1 data_room
1...1 data_battery_level
1...1 device_manufacturer_description
1...1 device_model_description
1...1 device_parent_serial_number
1...1 device_product_id
1...1 device_serial_number
1...1 device_type_description
1...1 message_timestamp
1...1 message_type_description
1...1 message_type_id

AI mapping suggestion

saref:hasValue

A relationship defining the value of a certain property, e.g., energy or power. Note that, even if numeric values are expected to enable reasoning, measurement values could use other datatypes.

Justification: `saref:hasValue` is used for `data_battery_level` because it defines the value of a certain property, such as the battery level.

ACCEPT SUGGESTION

SAREF

v 1.3

saref:hasModel
s4bldg:isContainedIn
saref:hasTemperature
saref:hasState
saref:hasModel
saref:hasValue
saref:hasTemperature
s4bldg:isContainedIn
saref:hasModel
saref:hasValue
saref:hasTemperature
s4bldg:isContainedIn
saref:hasModel
saref:hasValue
saref:hasTemperature
s4bldg:isContainedIn
saref:hasModel
saref:hasValue
saref:hasTemperature
s4bldg:isContainedIn
saref:hasModel
saref:hasValue
saref:hasTemperature
s4bldg:isContainedIn
saref:hasValue
saref:hasModel
saref:hasTimestamp

Mentimeter

Vertaling experiment: methode

- Edit distance
  - Levenshtein distance
  - Longest common subsequence
- Statistical
  - Word2vec
- LLM based mapping
  - JSON variable to ontology property identifier
  - JSON variable to RDF triple(s)
  - JSON object to RDF triples

Mentimeter

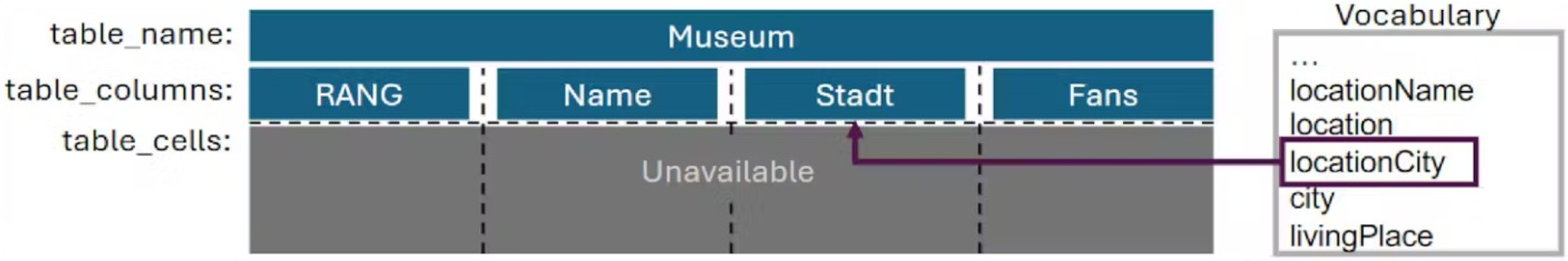
Vertaling experiment: resultaten

- Manual Grading of Performance:
  - Syntax ✓
  - Semantics ?
  - Consistency ✗
  - Hallucination of properties ✗
  - Hallucination of namespaces ✗
  - Introducing irrelevant information ✗

Variable	Output 1	Output 2	Output 3	**Average**
Temperature	6	7	4	5.7
CO2	7.5	6	4	5.8
Humidity	8	7	8	7.7
Battery level	7	5	5	5.7
Occupancy	8	7	5	6.7

Mentimeter

SemTab 2024 “Metadata to Knowledge Graph” track



Benchmark		Adwan		MetaLinker	
		Hit@1	Hit@5	Hit@1	Hit@5
Metadata2KG Round1	Top Hit@1	**0.75**	0.92	0.55	0.70
	Top Hit@5	**0.75**	0.92	0.55	0.70
Metadata2KG Round2	Top Hit@1	**0.83**	0.98	0.49	0.52
	Top Hit@5	**0.83**	0.98	0.37	0.68

Mentimeter

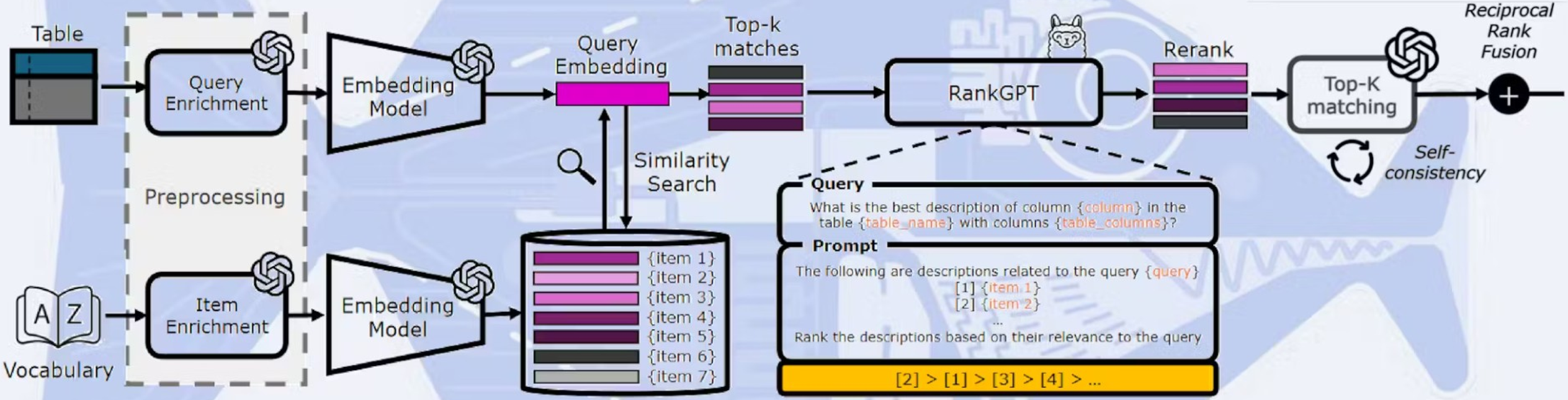
Vertaling experiment: 3 niveaus

- **Simple:**
  - **Input: JSON variable name**
  - **Output: ontology property identifier**
- Complicated:
  - Input: JSON variable name
  - Output: RDF triple(s) following ontology
- Hard:
  - Input: JSON object
  - Output: RDF triple(s) following ontology

In: "data_battery_level"
Out: `saref:hasValue`

Mentimeter

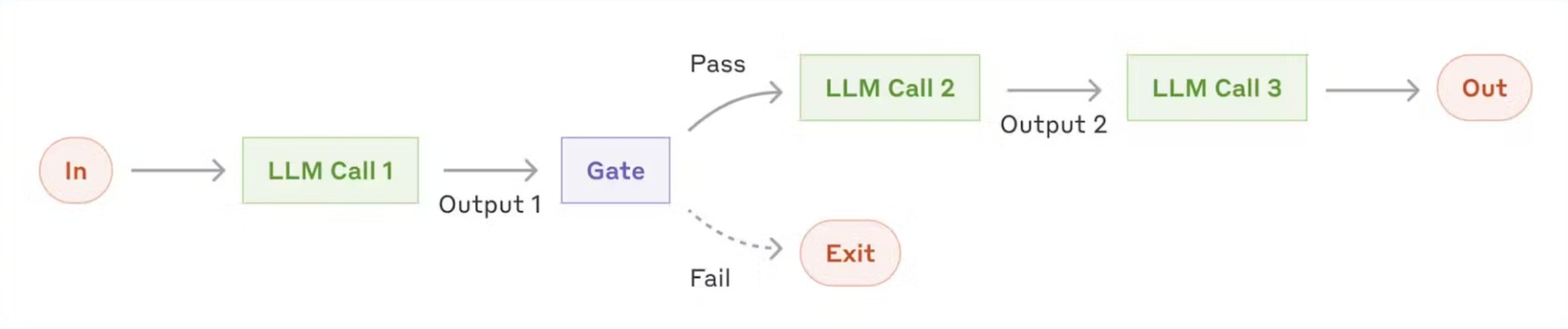
How it works



Vandemoortele, N., Steenwinckel, B., Hoecke, S. V., & Ongenae, F. (2024). *Scalable Table-to-Knowledge Graph Matching from Metadata using LLMs.*

Experiment B: samenvatting

gpt-4o-2024-08-06



This workflow is supposed to be ideal for situations where the task can be easily decomposed into subtasks. Trade-off: latency for higher accuracy.

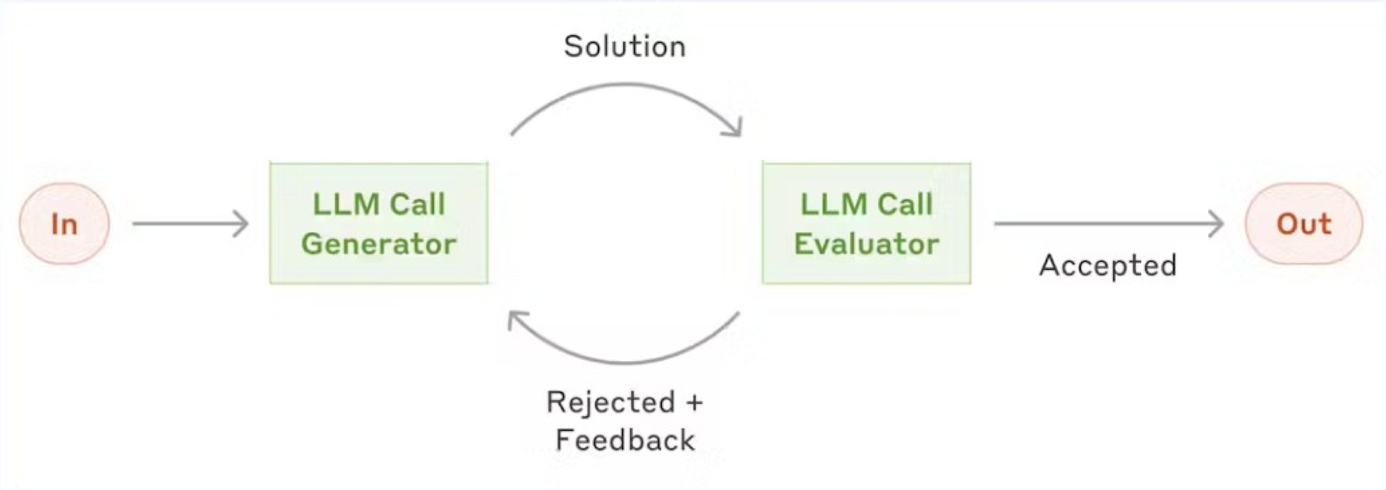
Results

- Accuracy: ~40%
- Speed: ~23m

Conclusion: Bad trade-off! Diagnosis: LLM is hampered – not assisted - by similarity search step
Not a problem with the architecture! Just the waý it's implemented

Experiment C: evaluator-optimizer workflow

`Gemini-2.0-flash-exp` (generator) & `gemini-2.0-flash-thinking-exp-01-21` (evaluator)



Experiment	Generator speed	Evaluator speed	Generator accuracy	Evaluator accuracy
#1	15s	46s	71%	85%
#2	16s	27s	50%	85%
#3	20	27s	57%	64%

- A single error in every experiment that seems particularly easy to fix with semantic search
- Other errors are soft-errors: user is not misled, no hallucinations.

Conclusion: good to very good starting point for users.

Experiment D: Claude with Citations

`claude-3-5-sonnet-20241022`

Here's an example:

5. Timestamp Information:

The `message_timestamp` aligns with SAREF core's hasTimestamp property, which represents "a relationship stating the timestamp of a measurement"

```
###  https://saref.etsi.org/core/hasTimestamp
saref:hasTimestamp rdf:type owl:DatatypeProperty ;
                   rdfs:comment "A relationship stating the timestamp of an entity (e.g. a measurement).
```

LLM generated text

Citation (guaranteed verbatim)

Results

- Accuracy: ~42%
- Speed: ~55s

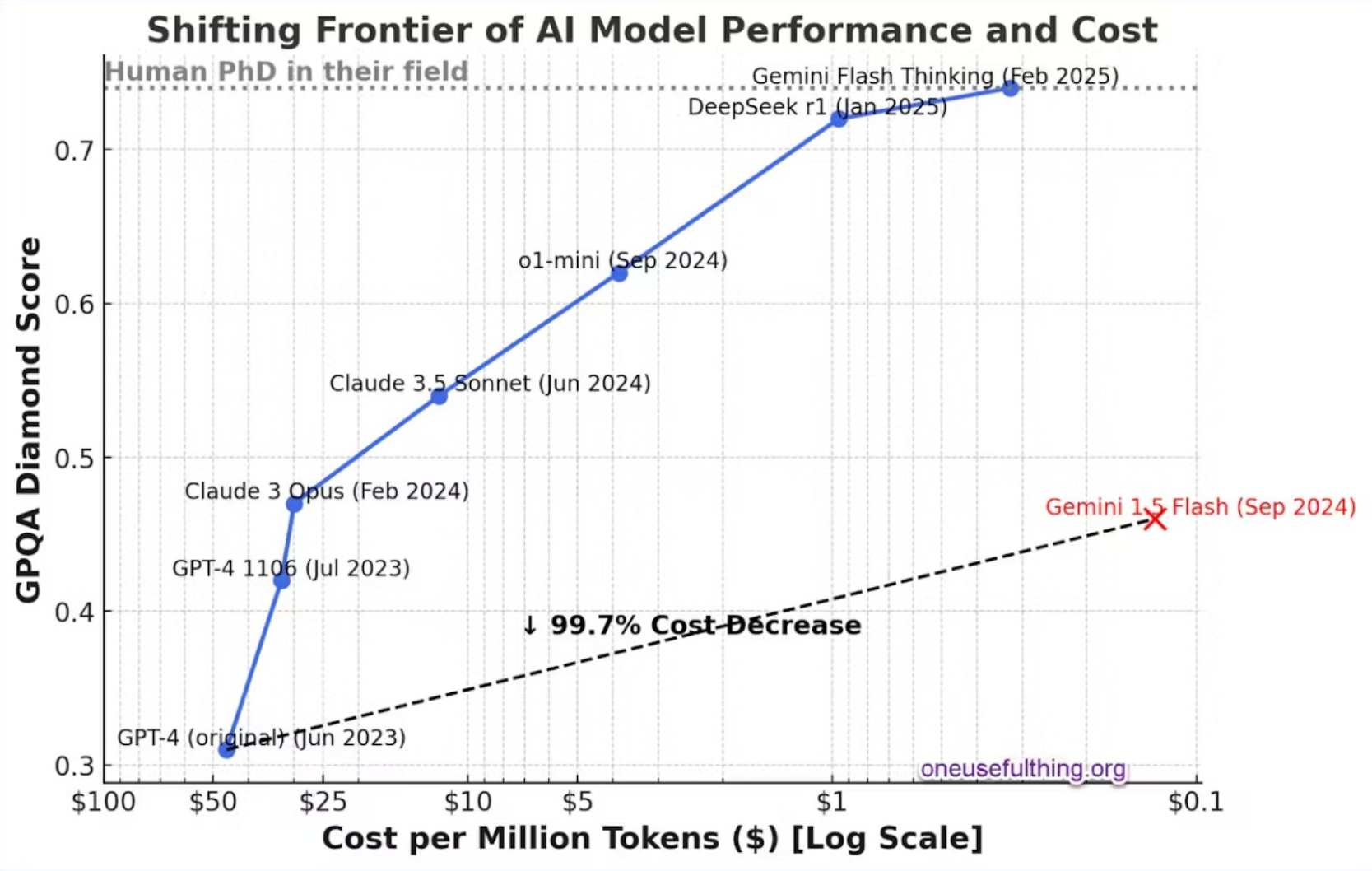
On two occasions, cited source text didn't make any sense i.r.t. the claim being made

Conclusion: has potential but not mature enough yet.

Mentimeter

Take aways

- Betere resultaten komen door de shifting frontier →
- Huidige LLMs zijn waarschijnlijk goed genoeg voor chatbot-oplossingen die gebruikers assisteren bij mappings
- Er zijn nog veel technieken om een dergelijke applicatie robuuster te maken
  - Fine tuning
  - Tools (function calling)
  - Verschillende workflows
    - Human in the loop
    - Evaluator-optimizer
    - Parallelization



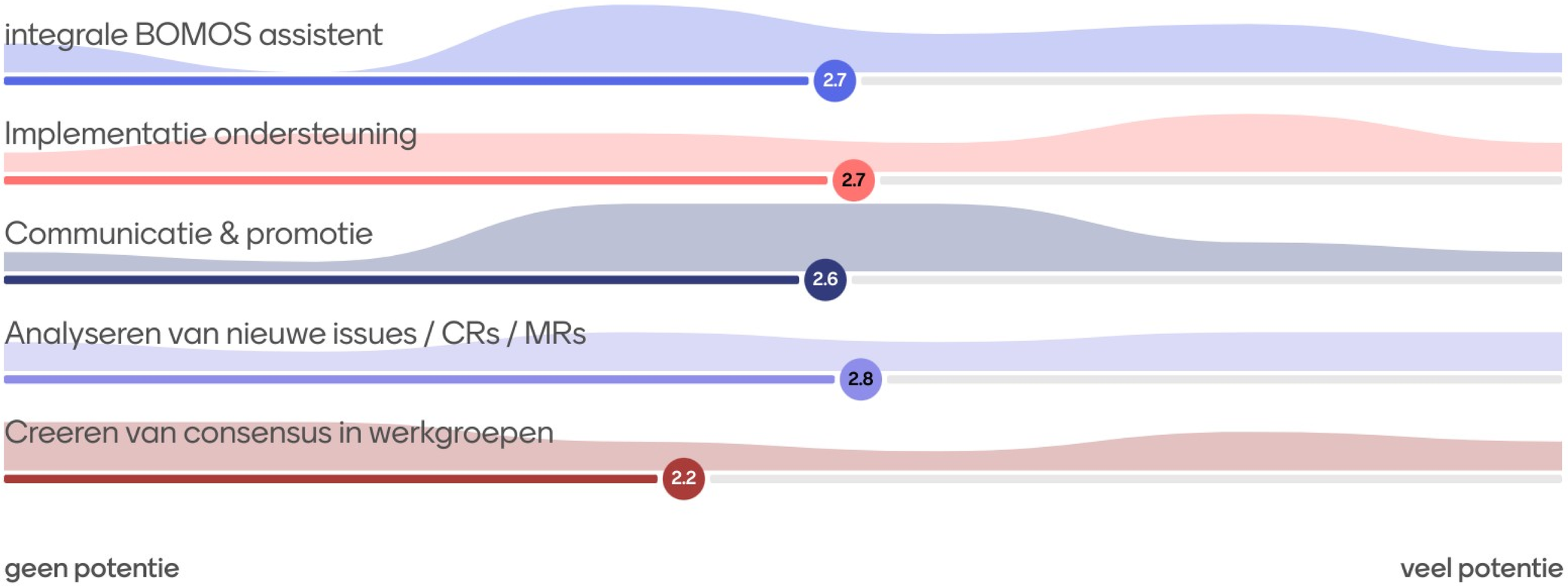
Mentimeter

Agenda

1. Introductie
2. Peiling (mentimeter)
3. AI landschap
4. Mapping experiment
5. Potentieel van AI voor BOMOS (mentimeter)
6. Q&A en afsluiting

Mentimeter

BOMOS & AI use cases



23

Mag ik je benaderen als we voor BOMOS & AI iets gaan organiseren?

Ja	Liever niet	Dat weet ik nog niet
0%	0%	0%

4

33

Mentimeter

Let’s talk

TNO innovation for life