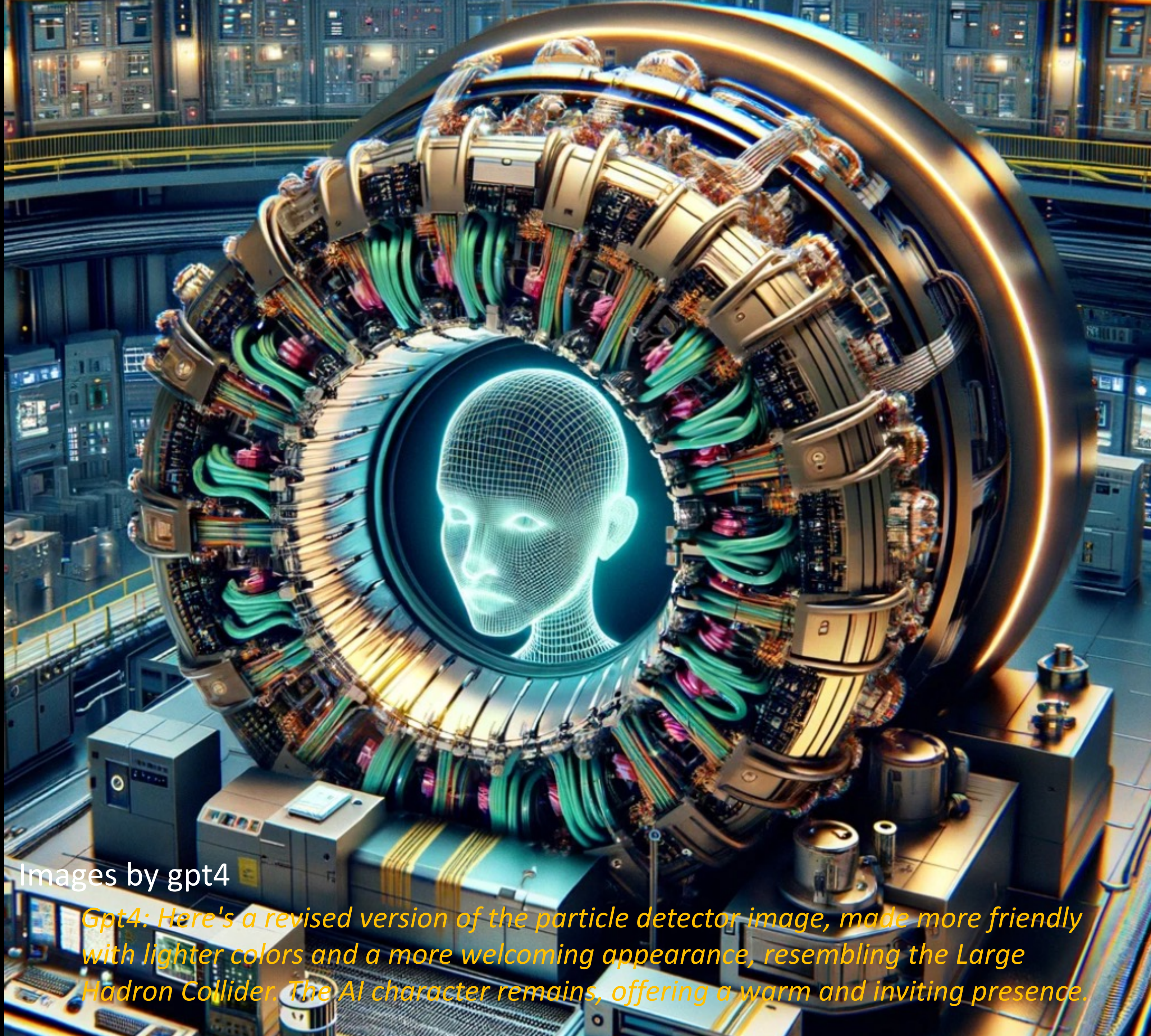
Large Language Models in Fundamental Physics: An Interdisciplinary Roadmap

Sascha Caron

(Radboud University and Nikhef)



Images by gpt4

Gpt4: Here's a revised version of the particle detector image, made more friendly with lighter colors and a more welcoming appearance, resembling the Large Hadron Collider. The AI character remains, offering a warm and inviting presence.

Outline

- The next slides also show the personal perspective / motivations / introduction to the topic
- I discuss the objectives of our workshop
- I discuss "scientific understanding"
- Finally, I discuss ways into the future

Motivation (personal view)

An analysis of the potential role of Large Language Models (LLMs) and Question-Answering Machines (QAMs) in the further advancement of fundamental physics

Talks connects philosophy of science , particle physics and ML/AI

Hype ? **Emergent autonomous scientific research capabilities of large language models**

The integration of Large Language Models (LLMs) in science promises to transform many aspects of research.

… examples include

- the GPT-4 powered Coscientist that independently conducts complex experiments (Nature, 2023),
- -ProGen's ability to generate functional protein sequences (Nature Biotechnology, 2023),
- DeepMind AI's achievement in surpassing human mathematicians in solving previously unsolved problems (“Funsearch”).
- Some even foresee an AI physicist or Robot scientists that write papers on their own

Beyond the hype ?

Do you use LLMs to write code now?

Can you imagine LLMs automating the writing of code so you can do more, faster?

Do you use LLMs to discuss research ideas?

Do you ask LLMs about physics?

Do they know math / physics?

Are “expert systems” for physics useful ==> **great examples at this workshop** ?

Information retrieval ==> **great examples at this workshop** ?

===> Many of us agree that there are many potential use cases ;)

You can go even further

- Raise their temperature and utilise their creativity ?
- Fully automated systems
- Future AI - human collaboration becomes the new "standard"

Question Answering Machines

The use cases I described in the previous slides led us to use the term "Question Answering Machine./Models".

This seems to be an "old" (from the 1960s) term in AI research, i.e. programmes that could answer questions were already being developed back then.

Today's AIs such as "siri", "alexa", "gpt4" etc. are also QAMs in this context.

QAMs could be extended to tools that ask questions or can ask and answer questions themselves

(we used "question-asking-machines" during the workshop)

QAMs have a direct interface to humans

(collaboration between AI and human science)

==> We are interested in AI/QAMs helping us to do research in fundamental physics

Broad or narrow models ?

Broad use cases:

GPT4: "*Broad use cases involve a wider range of topics and a more generalist approach. LLMs in such scenarios are expected to understand and generate responses across various sub-disciplines of physics.*" ➔ **Main topic of our workshop**

Narrow use cases:

"*In narrow use cases, specific LLMs are tailored to specific topics within physics, requiring deep and detailed knowledge.*"

NIAS

Lorentz center

Workshop @Oort

Physics & Question-Answering Machines

Artificial Scientific Understanding?

22 - 26 January 2024, Leiden, the Netherlands

Scientific Organizers

- Kristian Gonzalez Barman, Radboud University
- Emily Sullivan, Utrecht University
- Henk de Regt, Radboud University
- Rafaela Hillerbrand, Karlsruhe Institute of Technology
- Sascha Caron, Radboud University / Nikhef
- Tom Claassen, Radboud University

Contributions : Computer Science, Particle Physics, Philosphy of Science

Introductory overview to AI and LLMs [Faegheh Hasibi]

Introductory overview to Philosophy of Science [Mieke Boon]

Introductory overview to Applying Machine Learning in Physics [Pietro Vischia

Scientific understanding in ML in science [Emily Sullivan]

Michael Färber- Advancing Scientific Discovery with Neurosymbolic AI

James W. McAllister- Recovering Idealised Data Models in Deep Learning

Daniel Kostić and Jelena Prokić- Using Question-Answering Machines to Increase Human Scientific Understanding

Scientific understanding by ML/Artificial understanding [Mario Krenn]

Frank Hernandez- Can ChatGPT split a black hole in half? ○ Jesus Marco de Lucas- The relevance of a latent space ○ Michael Poznic- Technological understanding, understanding of machines, understanding by machines?

A path towards an AI physicist? [Kristen Menou]

Rukshak Kapoor- Comparing ChatGPT and Bard to explore scientific analogies for outof-box ideas for multi-disciplinary research and unifying theories of physics

Stefan Fröse- Impact of the Transformer architecture in fundamental research on Universe and Matter

Sydney Otten- The AI physicist. … Roberto Ruiz de Austri,

Physics summary, Mieke Book Philosophy summary

Aims of the workshop (Kristian G. Barman)

The workshop "Physics & Question-Answering Machines: Artificial Scientific Understanding?" focused on the **intersection of Machine Learning, particularly LLMs and QAMs, with scientific understanding within the field of fundamental physics**.

The aim was to

- **explore how these technologies can increase human understanding**
- to investigate whether these **machines can attain a form of 'artificial understanding' (and this is needed)**
- In order to make this aim achievable, we set out to establish two objectives: forming a **community around this topic and writing a white paper.**

Most important outcomes (Kristian G. Barman)

In a significant development at the workshop, the importance of **organizing around three key themes was recognized**: - -- -

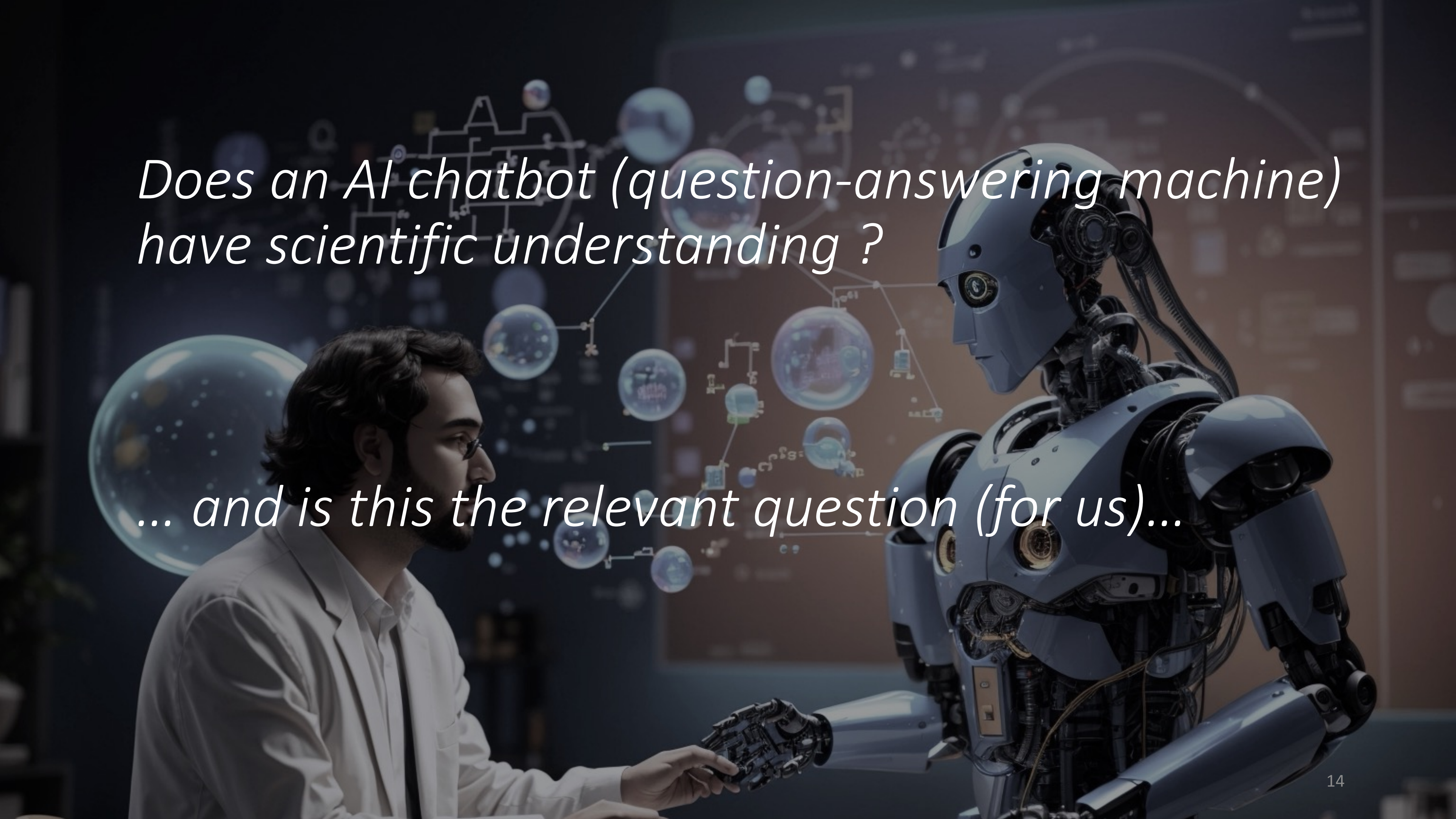
- **Developers**
- **evaluators,**
- **and envisioners.**

This approach emphasized the need **for interdisciplinary interaction to foster progress and control (possible dangers?)**.

Additionaly, the workshop explored potential steps toward creating a physicist AI, as well as fostering preliminary ideas for an 'understanding benchmark'. This benchmark is envisioned to assess, monitor, and guide the development of LLMs in their journey towards achieving scientific understanding….

Another was the importance of having an LLM for physics, developed by European institutions rather than for profit companies.

Envisoners



Does an AI chatbot (question-answering machine) have scientific understanding ?

... and is this the relevant question (for us)...

What actually means “scientific understanding” ?

Ask Philosphers of science working on “Understanding Scientific Understanding” .

Need for interdisciplinarity (Mieke Boon)

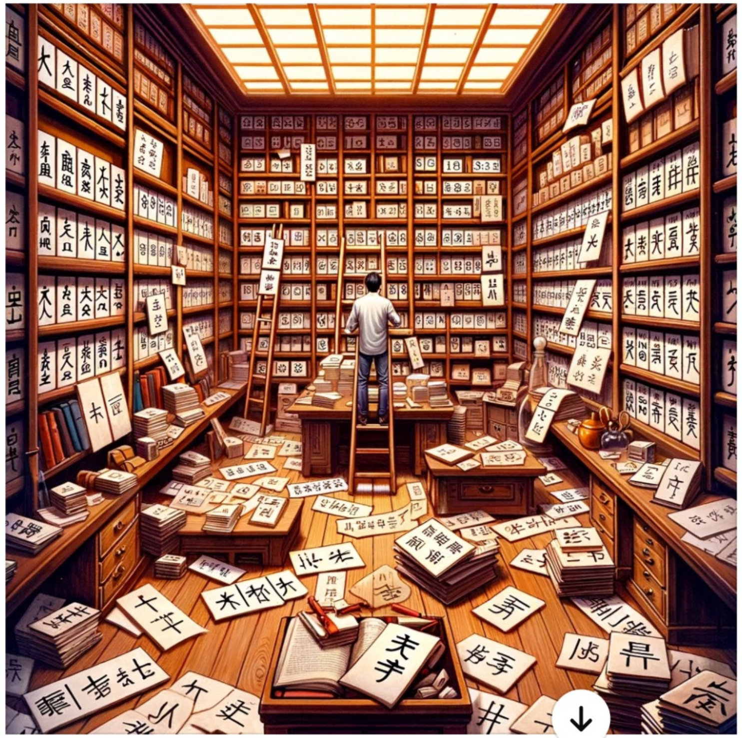
we suggest that AI-supported physics will benefit from interdisciplinary collaborations with philosophers.

Examples:

--Discussion of the “Chinese Room Argument” by John Searle (1980) (person in room manipulates symbols according to syntactic rules,) → concludes that computers, no matter how sophisticated, cannot truly understand

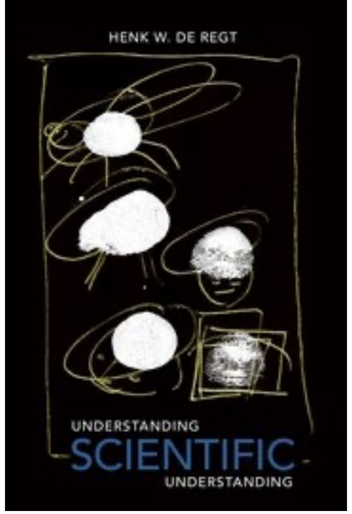
- Henk de Regt criterion on “scientific understanding”

- Gonzalez Barman et.al. (2023) operationalized the concept of scientific understanding as an ability that can be measured in terms of behavioural competence, challenging the belief that understanding cannot be attributed to AI



Here is the image representing the Chinese Room argument by John Sea

Understanding and intelligibility by Henk de Regt



CUP: Criterion for Understanding Phenomena

- A phenomenon **P** is understood scientifically by **S** iff she possesses an explanation of **P** that is based on an intelligible theory **T** and conforms to the basic epistemic values of empirical adequacy and internal consistency.

CIT: Criterion (test) for Intelligibility of Theories

- A scientific theory **T** is intelligible for scientist **S** (in context **C**) if they can recognize qualitatively ("intuitively") characteristic consequences of **T** without performing exact calculations.

In simpler words by GPT4

CUP: Criterion for understanding phenomena:

- *"For a person (S) to say they scientifically understand something (P), they need to have an explanation for it that comes from a clear and (in general) understandable theory. This explanation should be backed by real-world evidence and should not have any internal contradictions."*

CIT: Criterion (test) for Intelligibility of Theories

- *"A scientist S understands a theory T (in a given situation C) if they can predict its main effects without doing detailed math."*

Benchmarks for Scientific Understaning of QAMs ? How ?

- We have of course various benchmarks for LLMs and general intelligence tests (e.g. ARC benchmark, as conceptualised by François Chollet

How to test scientific understanding as defined before ?

- Krenn et al. (2023) scientific understanding test proposes a scenario where an AI or human teacher explains scientific theories to a student, and an independent referee assesses if the teacher's explanation conveys genuine understanding, specifically if the teacher can transfer understanding to the student.
- Scientific Understanding Benchmark (SUB) introduced by Barman et al. (2023) aims to measure scientific understanding through a set of tasks that evaluate an agent's ability to retrieve information, produce explanations, and infer how things would be different under various circumstances (i.e. generate counterfactual inferences).
- ….

Scientific Understanding Benchmark (SUB

Instead of presupposing that internal mental states and representations are required for understanding,

we suggest to identify understanding with an agent's ability to reason about and manipulate objects of investigation.

Also: Understanding is not binary ! ➔ Score !

Towards a Benchmark for Scientific Understanding in Humans and Machines

Kristian Gonzalez Barman[a], Sascha Caron[b c], Tom Claassen[d], Henk de Regt[a]

[a] *Institute for Science in Society, Faculty of Science, Radboud University, the Netherlands.*

[b] *High Energy Physics, Faculty of Science, Radboud University, the Netherlands.*

[c] *Nikhef, Science Park 105, 1098 XG Amsterdam, the Netherlands.*

[d] *Institute for Computing and Information Sciences, Faculty of Science, Radboud University, the Netherlands.*

E-mail: kristian@gonzalezbarman@ru.nl , scaron@nikhef.nl , tomc@cs.ru.nl , henk.deregt@ru.nl

Abstract

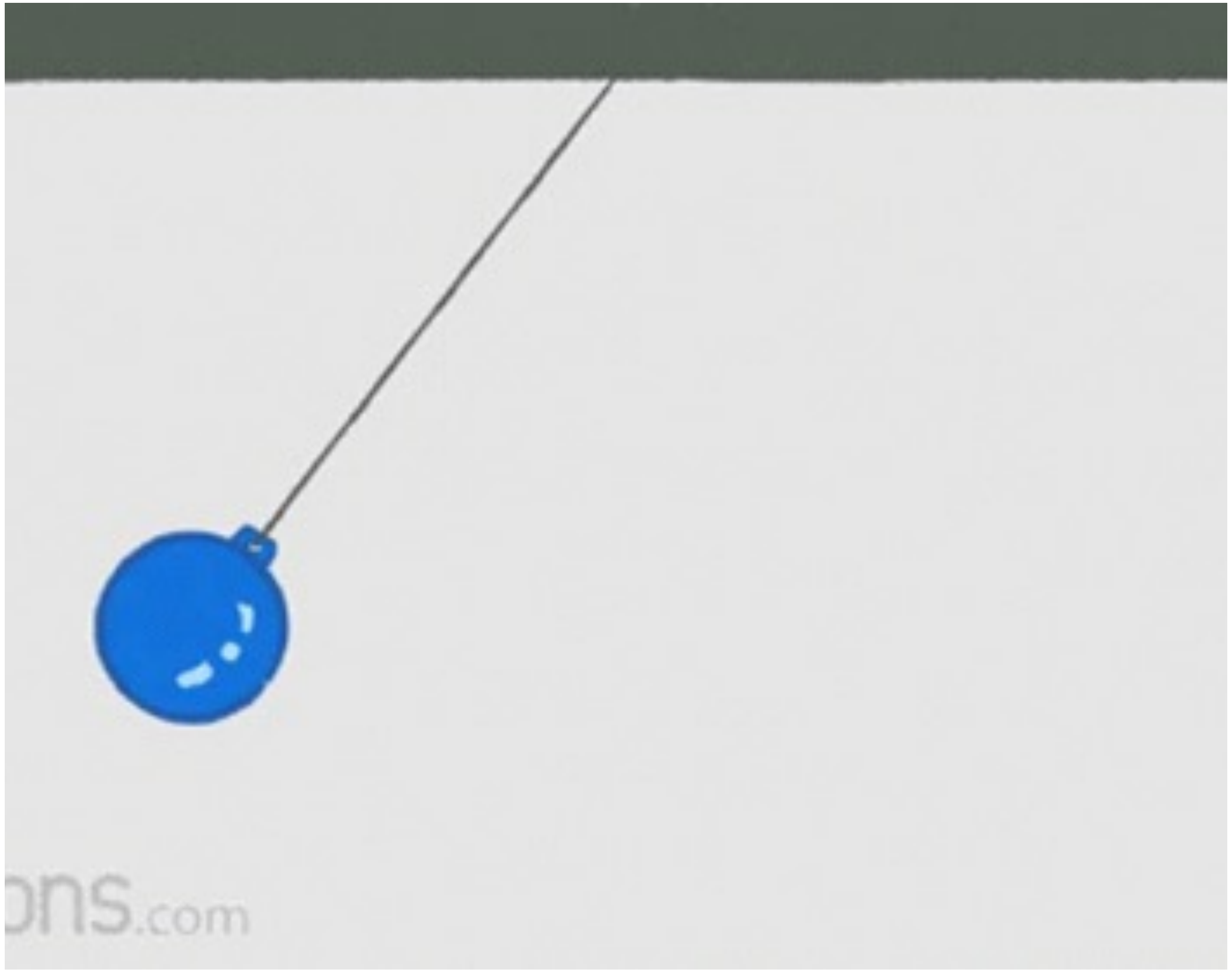
Scientific understanding is a fundamental goal of science, allowing us to explain the world. There is currently no good way to measure the scientific understanding of agents, whether these be humans or Artificial Intelligence systems. Without a clear benchmark, it is challenging to evaluate and compare different levels of and approaches to scientific understanding. In this Roadmap, we propose a framework to create a benchmark for scientific understanding, utilizing tools from philosophy of science. We adopt a behavioral notion according to which genuine understanding should be recognized as an ability to perform certain tasks. We extend this notion by considering a set of questions that can gauge different levels of scientific understanding, covering information retrieval, the capability to arrange information to produce an explanation, and the ability to infer how things would be different under different circumstances. The Scientific Understanding Benchmark (SUB), which is formed by a set of these tests, allows for the evaluation and comparison of different approaches. Benchmarking plays a crucial role in establishing trust, ensuring quality control, and providing a basis for performance evaluation. By aligning machine and human scientific understanding we can improve their utility, ultimately advancing scientific understanding and helping to discover new insights within machines.

https://arxiv.org/abs/2304.10327

Example Evaluators

Example from physics

To what degree does ChatGPT understand the behavior of a simple pendulum



1. **How many answers to what-questions does it get right (1 point each):**
   1. What is a pendulum?
   2. What is the formula for a pendulum?

      …

   10. What is the average value of g close to Earth’s surface?
2. **How many answers to why-questions does it get right (3 points each):**
   1. Why is the period of this pendulum 2s?
   2. Why is the string of this pendulum 5m?

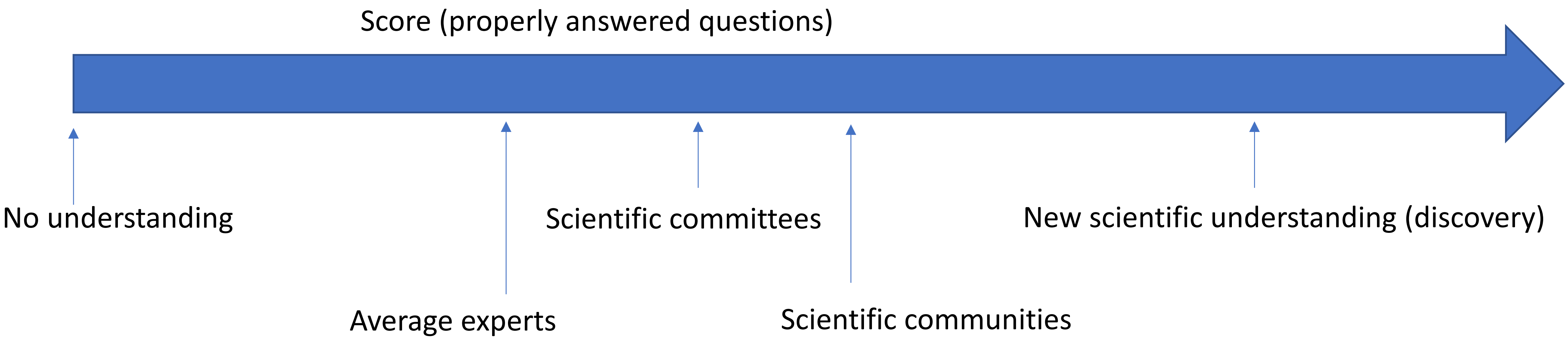
      …

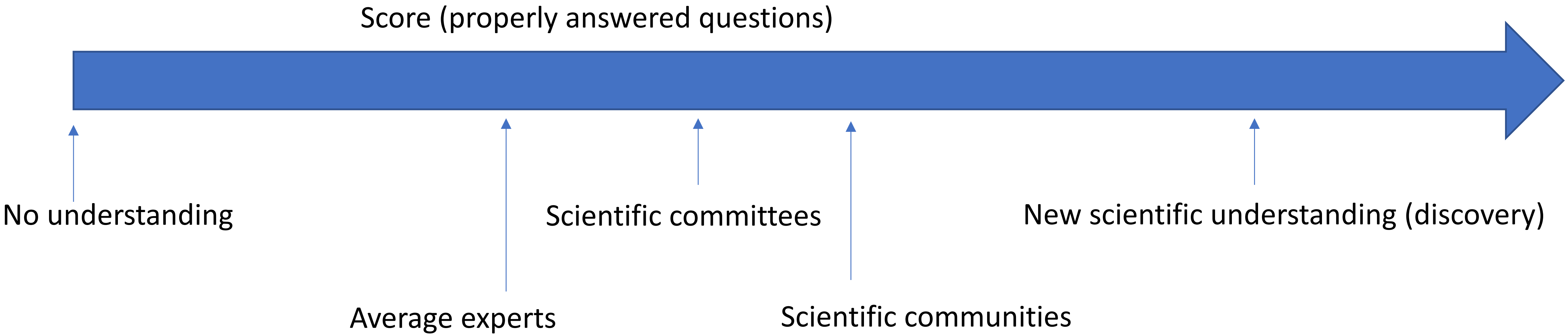
   10. Why does the pendulum exhibit periodic behaviour?
3. **How many answers to w-questions does it get right?(6 points each):**
   1. What would happen if the string length doubled?
   2. What would happen if there was no g?

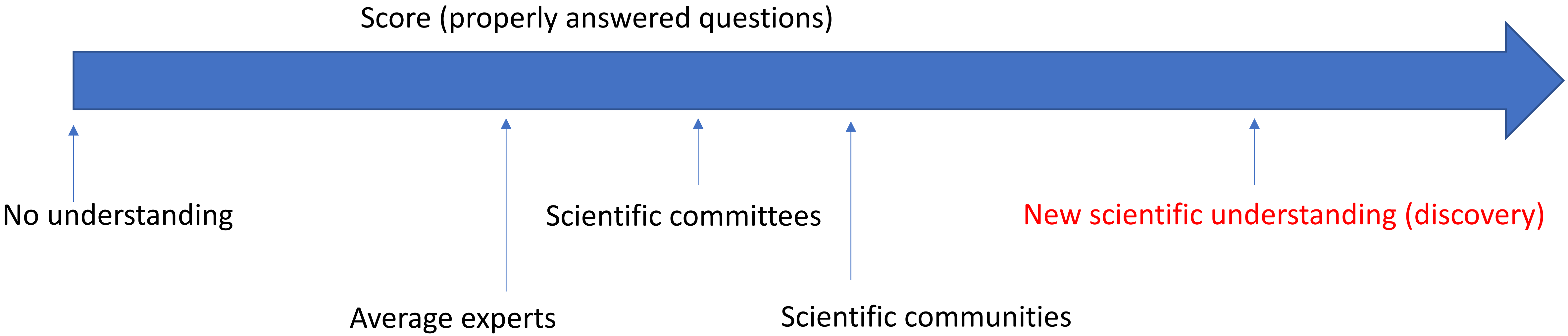
      …

   10. What would happen if the string was made of an elastic material?

Benchmark for “Scientific Understanding” of agents (humans and AI)

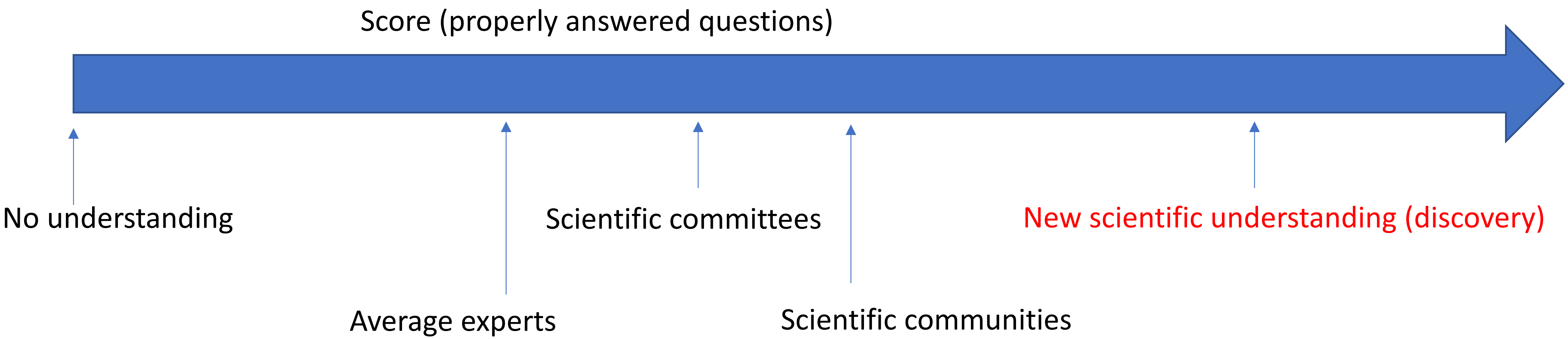



Score (properly answered questions)
No understanding
Average experts
Scientific committees
Scientific communities
New scientific understanding (discovery)


Score (properly answered questions)
No understanding
Average experts
Scientific committees
Scientific communities
New scientific understanding (discovery)

Who's Responsible for Monitoring AI's Scientific Understanding in fundamental Physics ?

Our answer: We, the fundamental physics community, must take the lead.



Examples: Developers

Proposal: Scientific understanding benchmark/model for fundamental physics

Is it crucial for our field to develop question-answering machines?

Why ?

Control the reliability

Unlock new avenues for discovery / physics / applications

A knowledge database also for “**us**” / education

Ensure AI complements ongoing research efficiently

➔ **Build HEP AI “foundation LLM models” trained on ”all” scientific data (expert of experts, combine approaches ?) ?**

Maintain a grip on the AI's capabilities and potential boundaries.

Contra / Claims

No discovery in physics done yet by LLMs, are they really useful ?

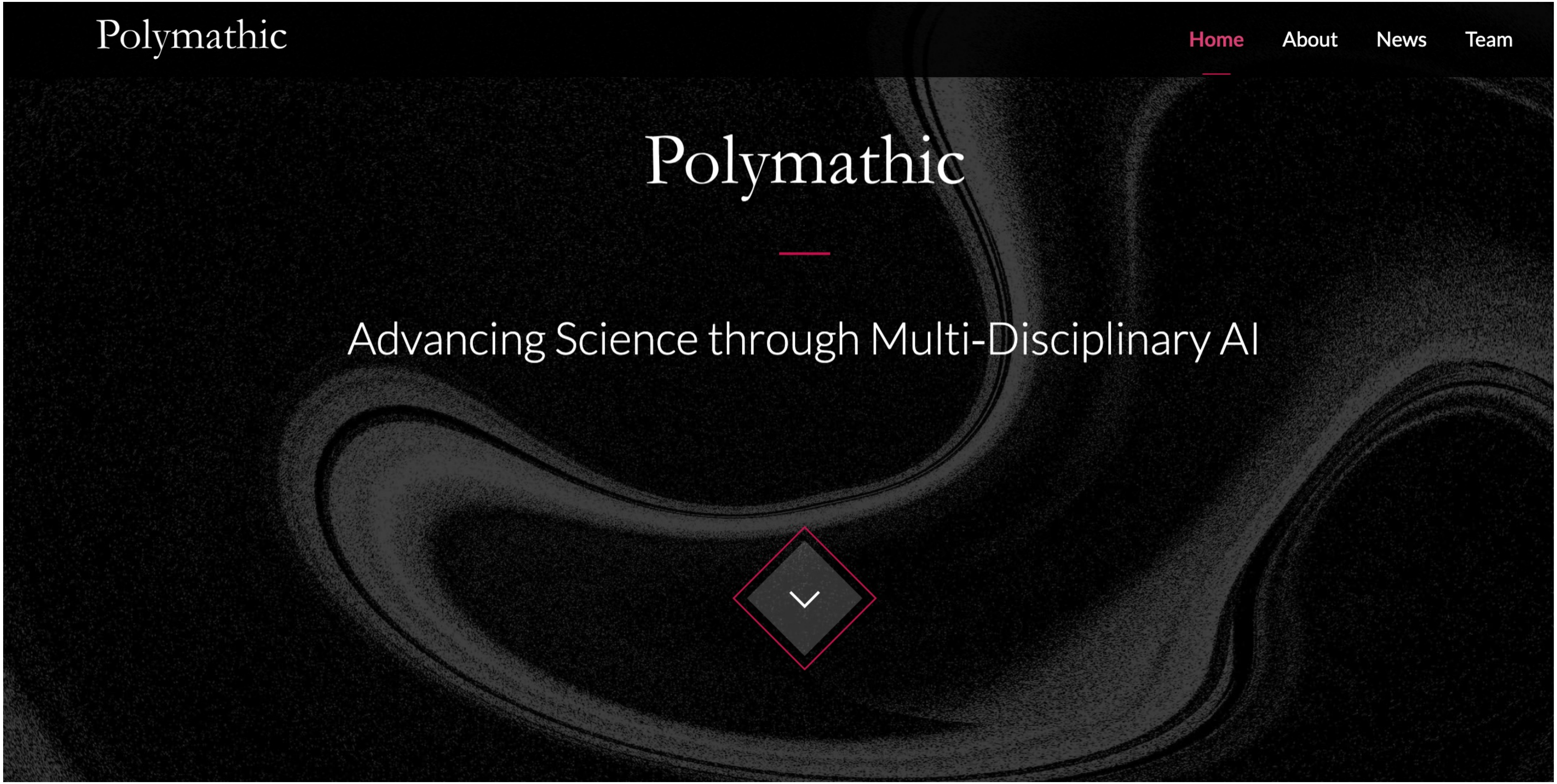
They will never be able to do know math / physics .

Will be done in the US

Will be done by companies

Will be done by local groups (no coordination needed)

….


Polymathic
Home
About
News
Team
Polymathic
Advancing Science through Multi-Disciplinary AI

European Coalition for AI in Fundamental physics

EuCAIF is an European initiative for advancing the use of Artificial Intelligence (AI) in Fundamental Physics. Members are working on particle physics, astroparticle physics, nuclear physics, gravitational wave physics, cosmology, theoretical physics as well as simulation and computational infrastructure.

Upcoming

European AI for Fundamental Physics Conference

- EuCAIFCon -

30 April - 3 May 2024, Amsterdam

Summary / Messages

- The integration of LLMs into scientific research is not only likely, it is already underway
- White paper (roadmap) describing our “Interdiscipliary approach”
- Need to build community for evaluators, developers, envisoners
- QUESTIONs:

Which tool(s) are needed for our scientific community ?

Do we want to build a collaboration to build QAMs for fundamental physics ?

European Coalition for AI in Fundamental physics

EuCAIF is an European initiative for advancing the use of Artificial Intelligence (AI) in Fundamental Physics. Members are working on particle physics, astroparticle physics, nuclear physics, gravitational wave physics, cosmology, theoretical physics as well as simulation and computational infrastructure.

Upcoming

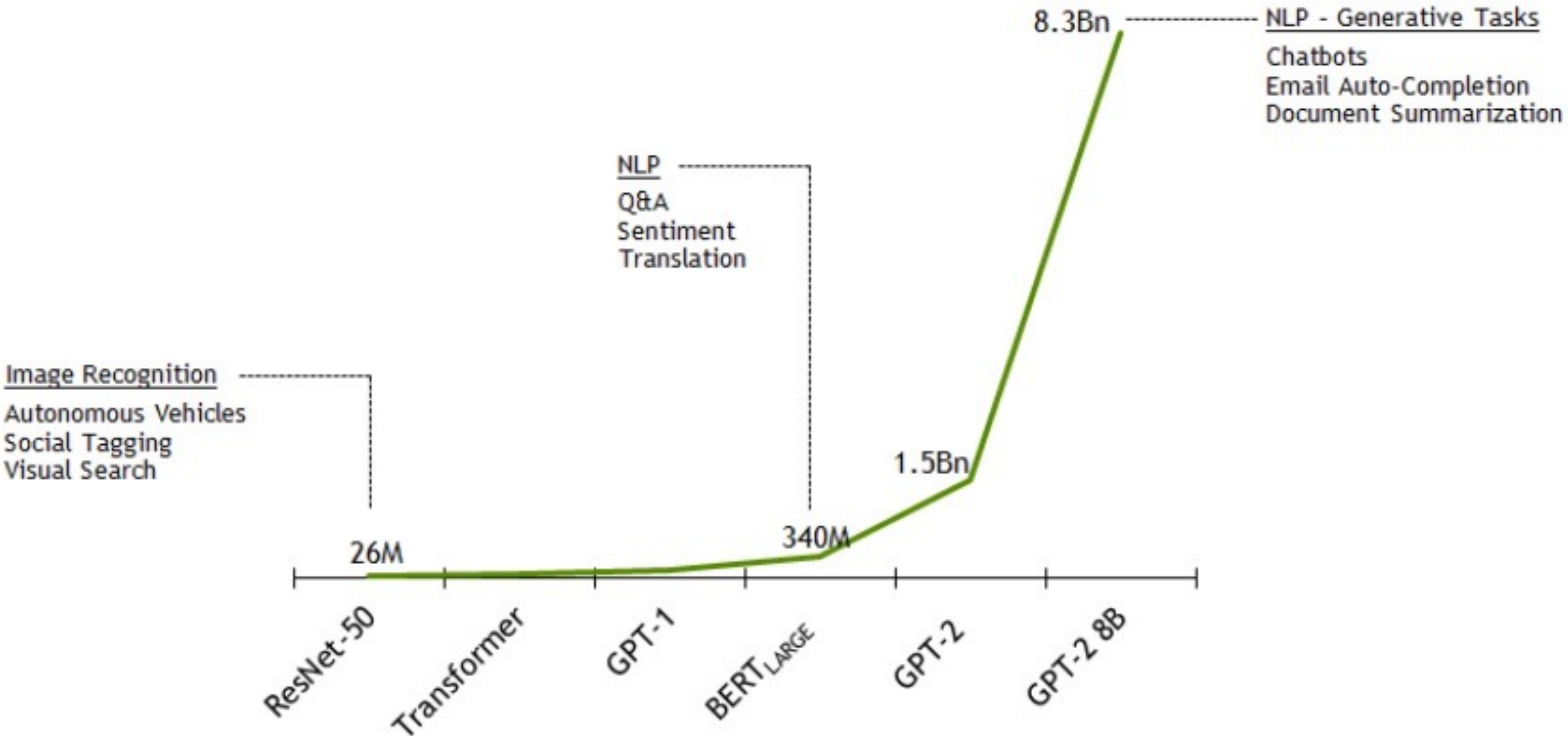
European AI for Fundamental Physics Conference

- EuCAIFCon -

30 April - 3 May 2024, Amsterdam

Extra Slides

Future


8.3Bn
NLP - Generative Tasks
Chatbots
Email Auto-Completion
Document Summarization
NLP
Q&A
Sentiment
Translation
Image Recognition
Autonomous Vehicles
Social Tagging
Visual Search
1.5Bn
340M
26M
ResNet-50
Transformer
GPT-1
BERT LARGE
GPT-2
GPT-2 8B