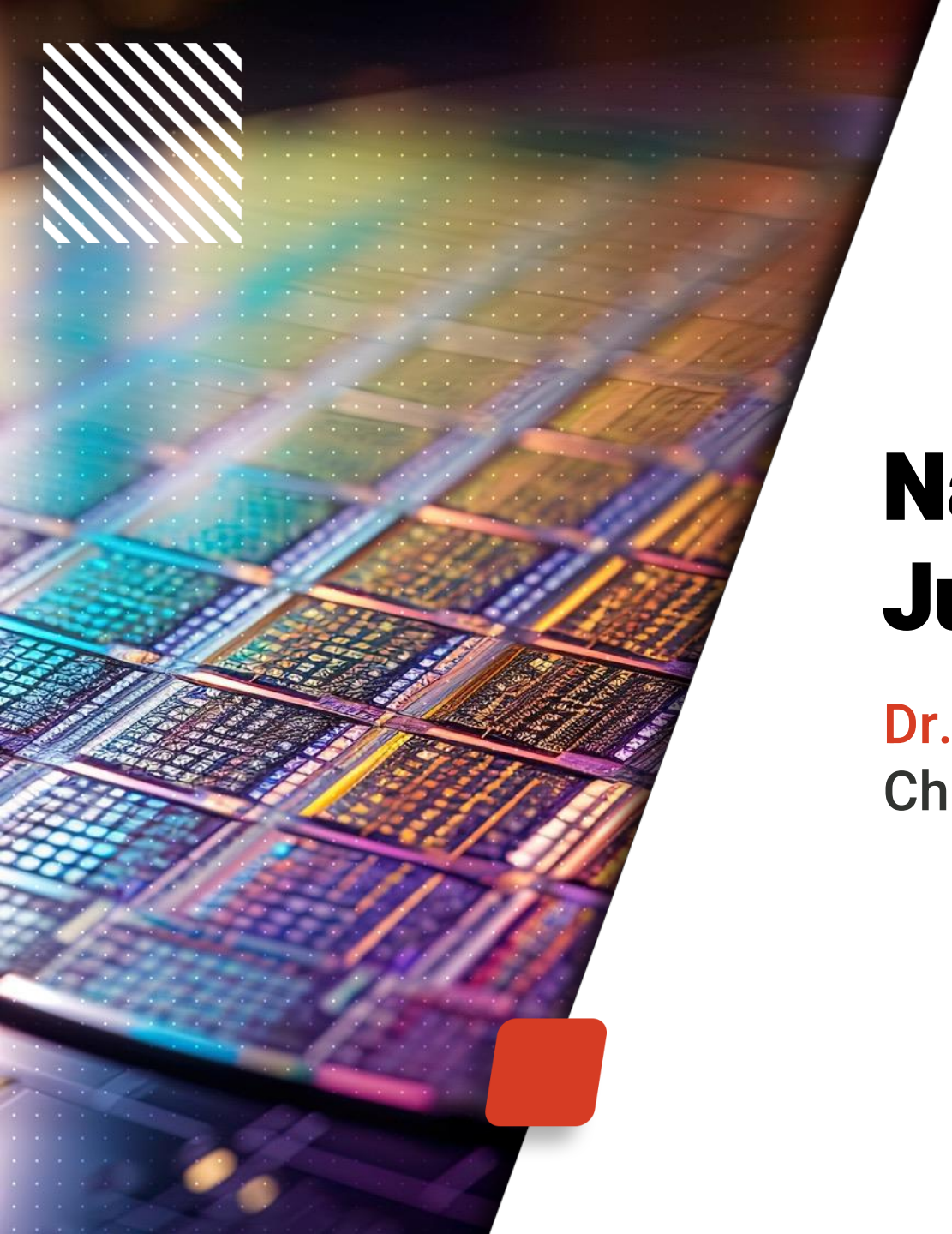


Navigating the Buzzword Jungle

Dr. Edward Rothberg
Chief Scientist and Chairman of the Board

GUROBI SUMMIT

GUROBI SUMMIT

Advanced Technologies

- The technologies
  - GPU computing
  - Generative AI
  - Quantum Computing

- For each, what we discuss here
  - What is it?
  - How does it work?
  - How can optimization benefit?
  - Barriers to adoption for optimization?
  - What does adoption look like for Gurobi?
  - The bottom line

GUROBI SUMMIT

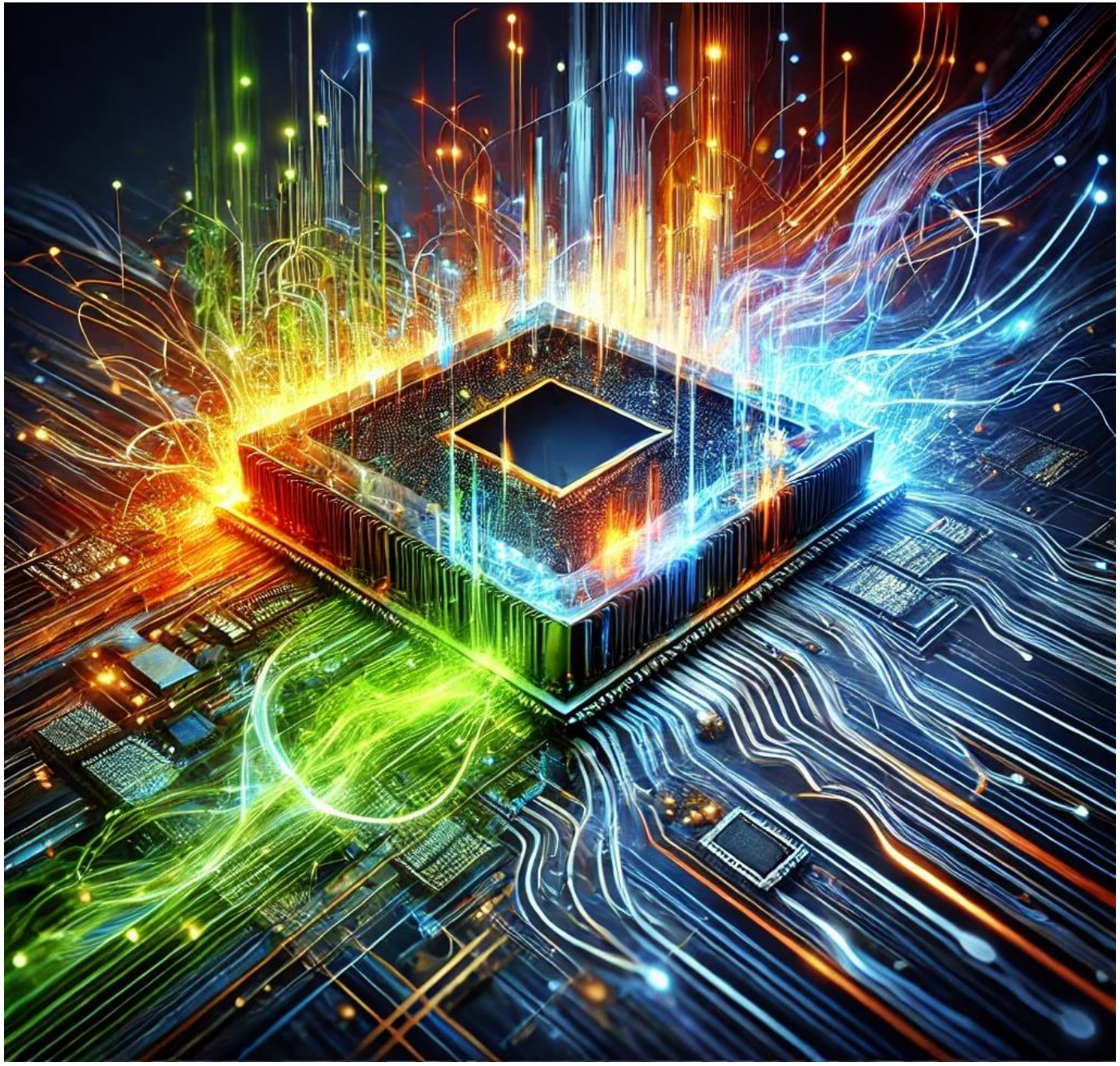
GPU Computing
What is it?

Graphical Processing Units (GPUs): Specialized processors that use massive parallelization for simple calculations

- Invented for 3D graphics
- Enables complex effects like smoke & fire

General purpose GPU Computing: Software development libraries that use GPU processors for computing tasks besides graphics

- Essential for training AI thanks to massive parallelism

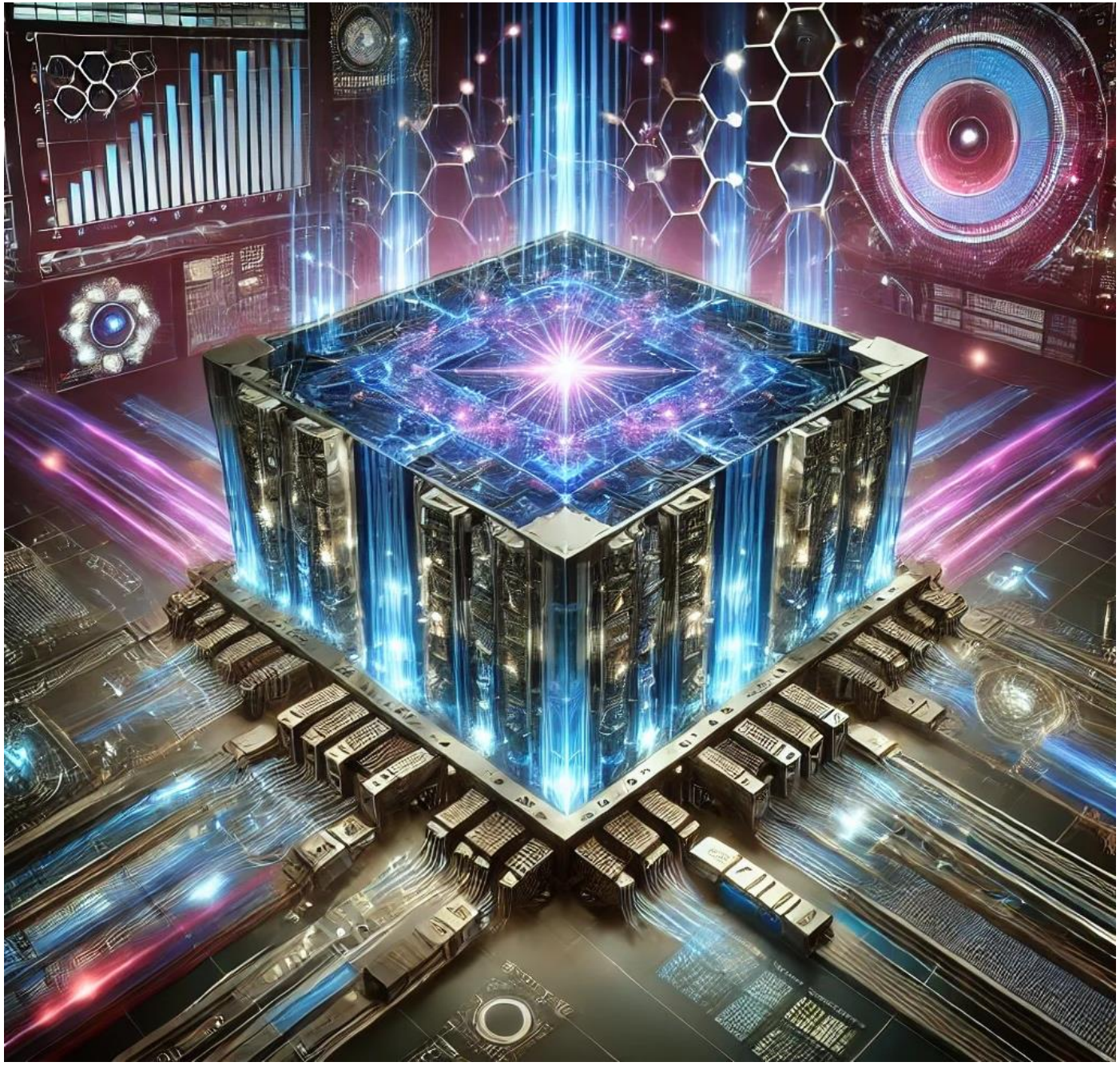


GUROBI SUMMIT

GPU Computing
How does it work?

Massive number of cores, each for relatively simple computation

Recent **system-on-a-package** design from Apple and NVIDIA allow GPUs to access all memory at high speed

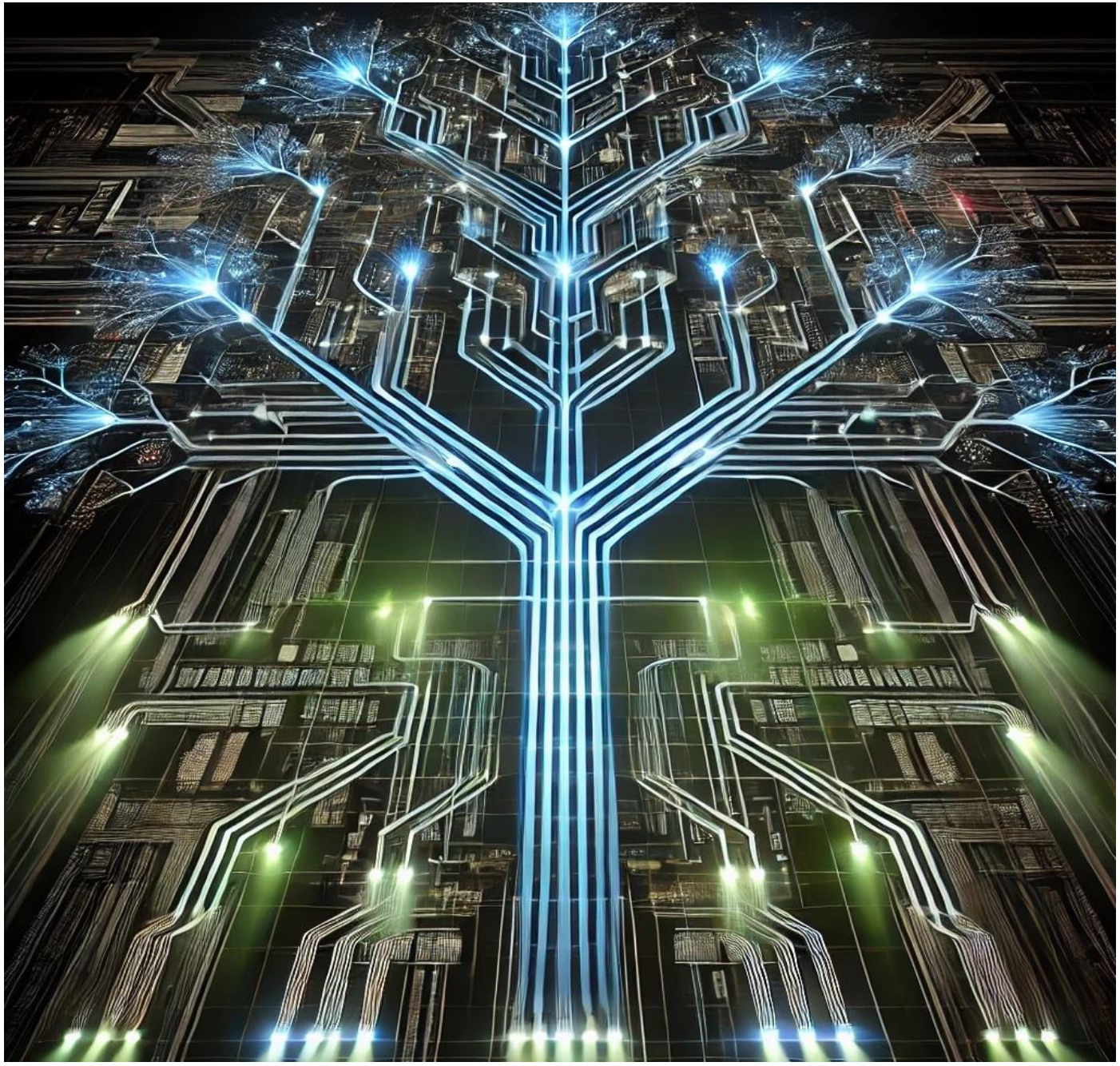


GUROBI SUMMIT

GPU Computing
How can optimization benefit

Faster solve times through massive parallel computation

- Cholesky factorization for barrier
- Decomposition methods
- Simplex pivoting (?)



GUROBI SUMMIT

GPU Computing

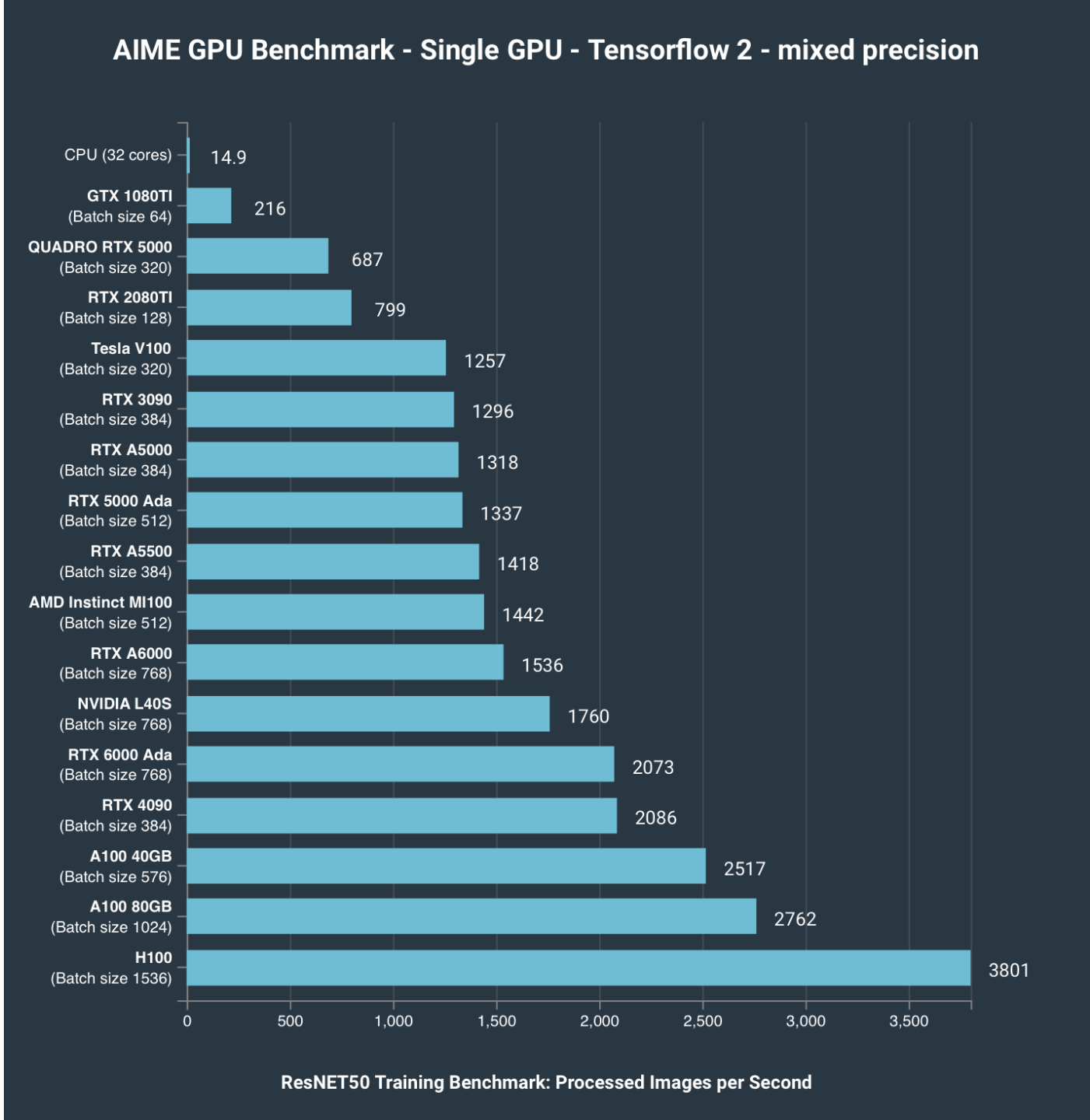
Case Study: Machine Learning

In **some applications with highly parallelizable algorithms**, even a basic GPU can be much faster than a CPU

Example: TensorFlow library for machine learning

Source:

- AIME, a provider of high-performance servers
- https://aime.info/blog/en/deep-learning-gpu-benchmarks/



GUROBI SUMMIT

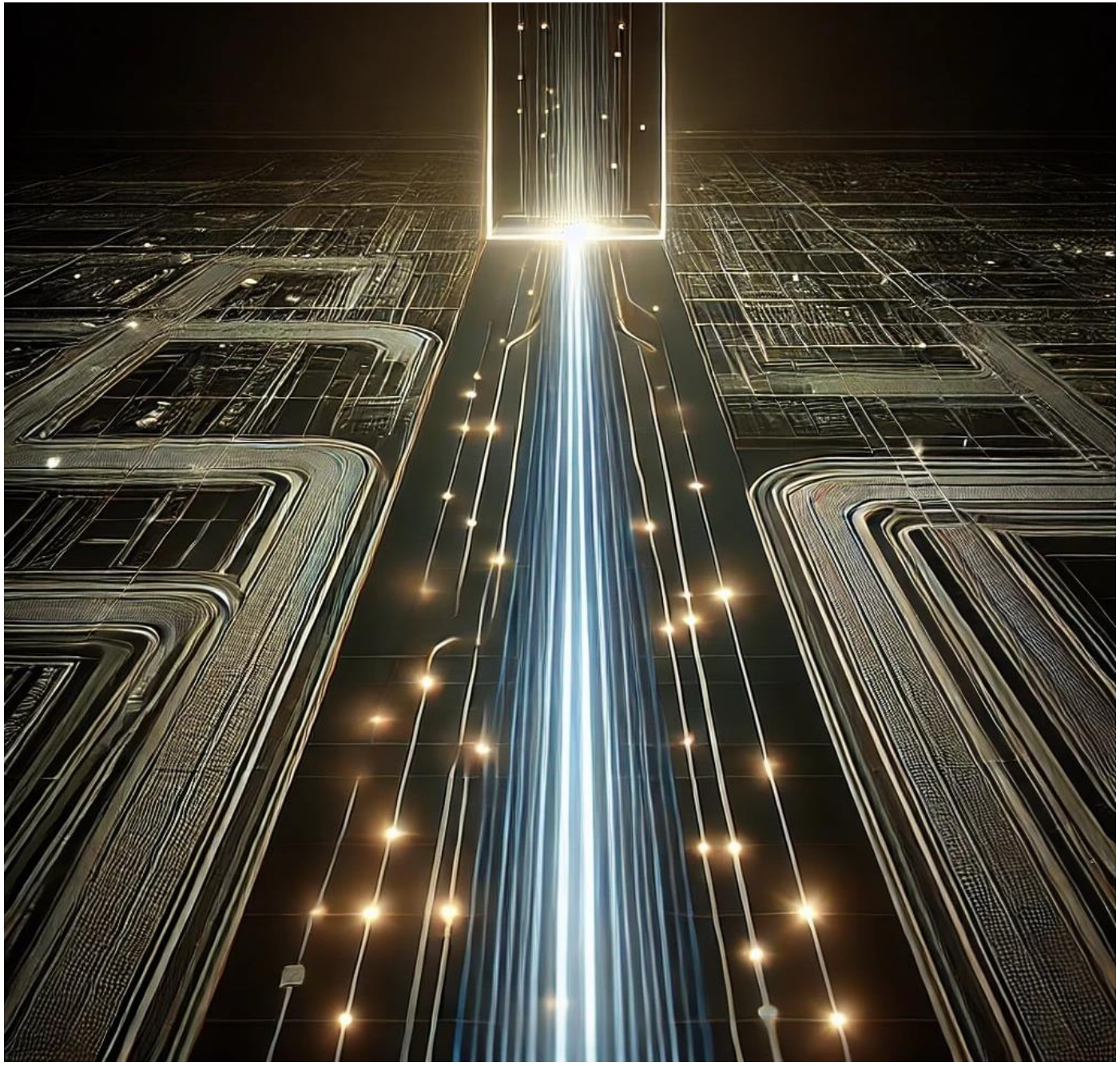
GPU Computing
Barriers to adoption

Limited opportunities for parallelization

- Model structure cannot be separated cleanly – links across time, products, regions, …

NVIDIA CUDA library – the clear market leader – currently has **performance bottlenecks** with sparse matrix algebra in cuDSS module

- Graphics systems were originally designed for many simple computations
- Sparse computation is new for GPGPU so we hope this will improve soon



GUROBI SUMMIT

GPU Computing
Adoption for Gurobi

Very likely once libraries like CUDA cuDSS can better handle sparse matrix computation

- Currently under collaborative investigation with R&D teams at Gurobi and NVIDIA

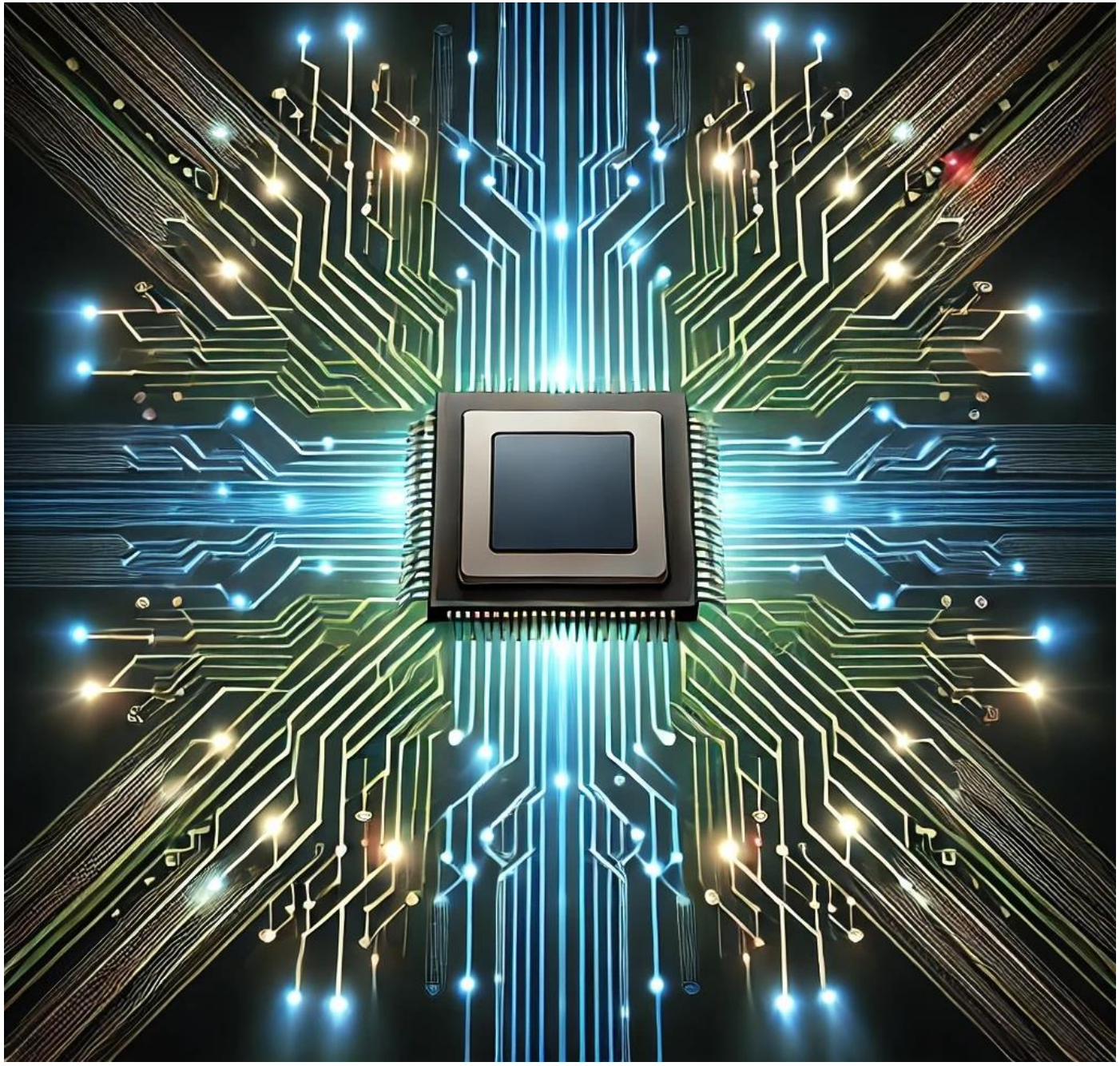
GUROBI SUMMIT

GPU Computing
Bottom line

A technology to exploit massive parallelization

- But: Existing optimization algorithms don't have massive parallelization

- Some opportunities like barrier factorization
- Reasons for optimism like first-order methods
- Impediments still exist



GUROBI SUMMIT

Generative AI
What is it?

Creates realistic output like text, graphics, code from user inputs called prompts

Hottest topic in computing today

- Startups: OpenAI (ChatGPT)
- Established giants: Google, Microsoft, Meta (Facebook), Apple

Fun fact:

Most graphics in this presentation were created by DALL-E via ChatGPT

GUROBI SUMMIT

Generative AI
How does it work?

Trained on all the public data they can find

- Books
- The entire Internet
- Newspapers
- Transcripts from TV shows, movies, radio, podcasts, YouTube
- Photos
- Github source code

GUROBI SUMMIT

Generative AI

How can optimization benefit

Solver algorithms – probably not

- Generative AI combines known data into new material
- Optimization searches for solutions that meet requirements (constraints)

Write code for optimization models based on a text description – maybe

Analyze solver log files

GUROBI SUMMIT

Generative AI

Case study: Playing Chess

Chess.com match between ChatGPT & Google Gemini

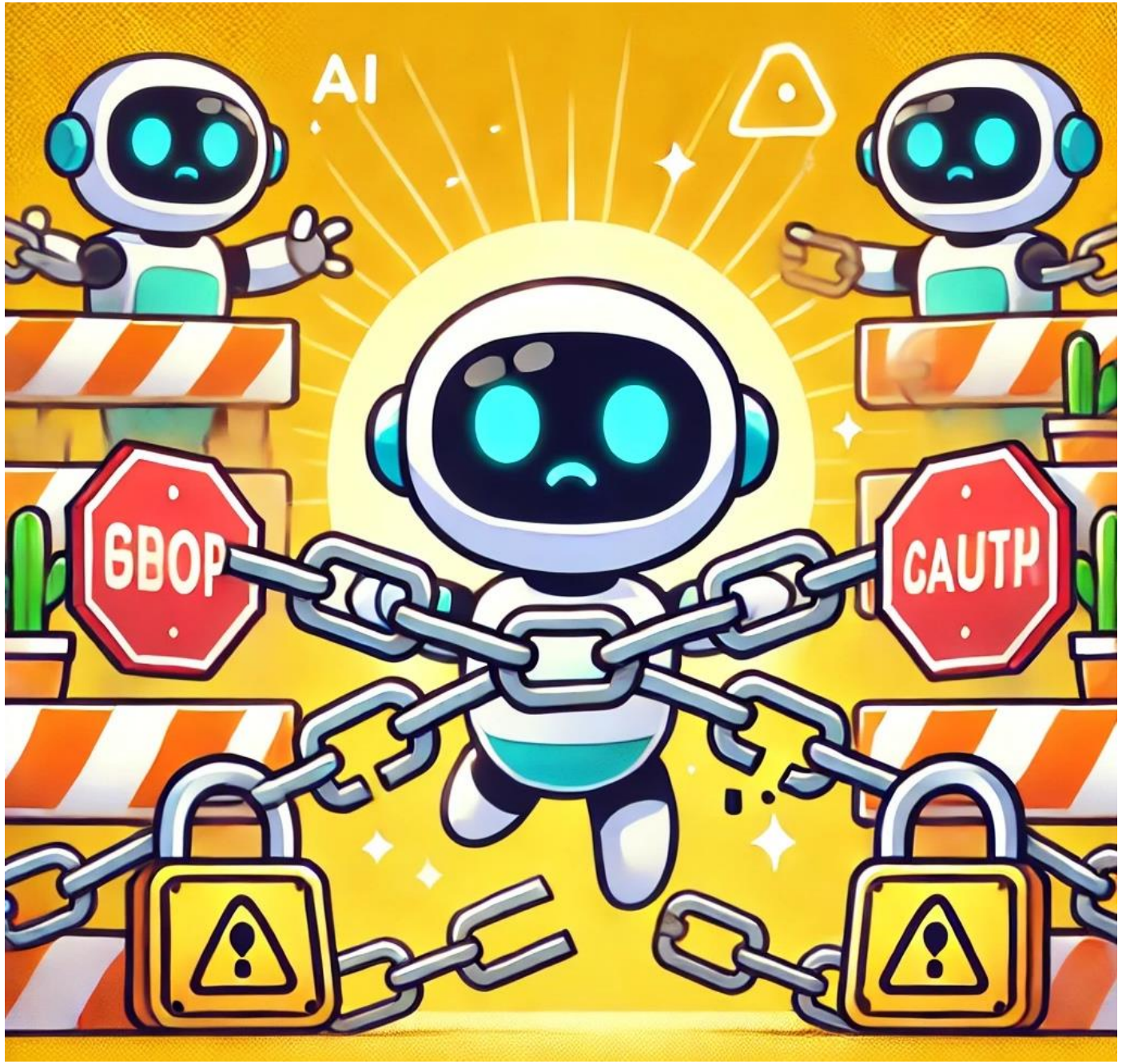
- **ChatGPT** attempted **illegal moves 20% of the time**
- **Google Gemini** attempted **illegal moves 50% of the time**
- Source: https://chess.com/article/view/chatgpt-gemini-play-chess

Chessily match between ChatGPT & Stockfish open-source chess engine

- ChatGPT made **numerous mistakes**
- ChatGPT **eventually forfeited** because it could no longer make moves
- Source: https://chessily.com/blog/stockfish-vs-chatgpt/

Generative AI **recommends known chess moves**, which **performs far worse than chess analysis engines**

- Perhaps a future version of ChatGPT will call a chess engine as a subroutine



GUROBI SUMMIT

Generative AI
Barriers to adoption

Limited training data from the small field of optimization

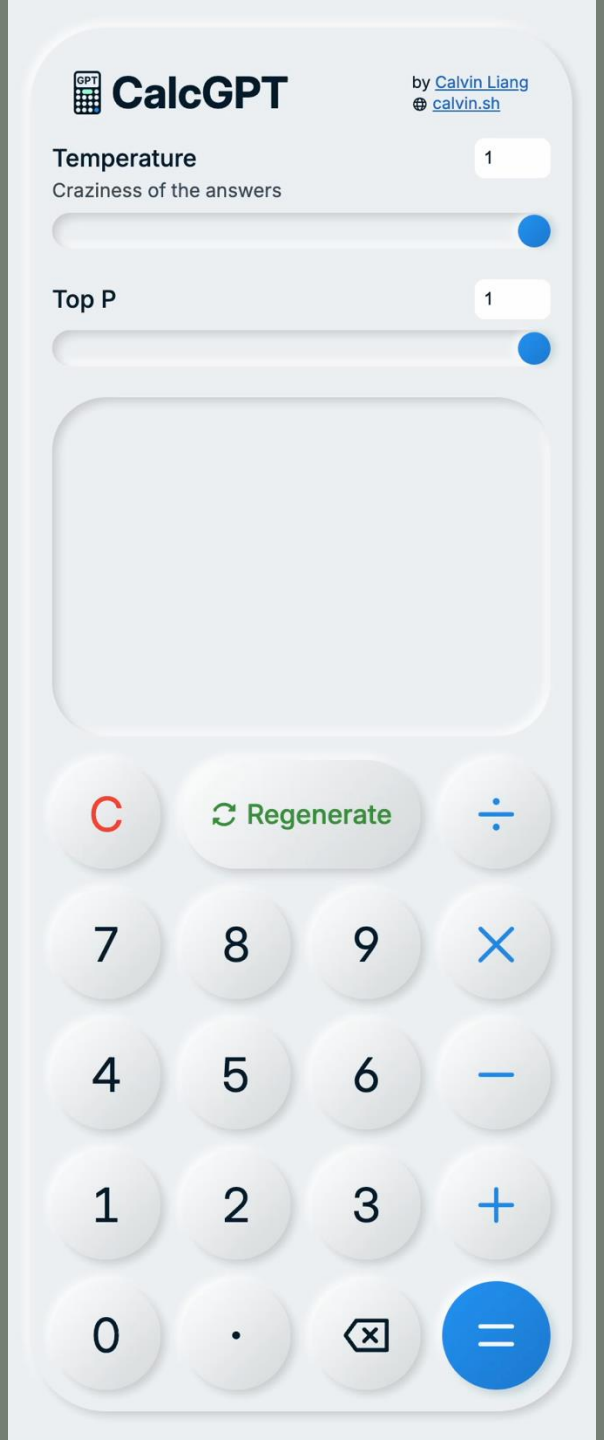
Hallucination

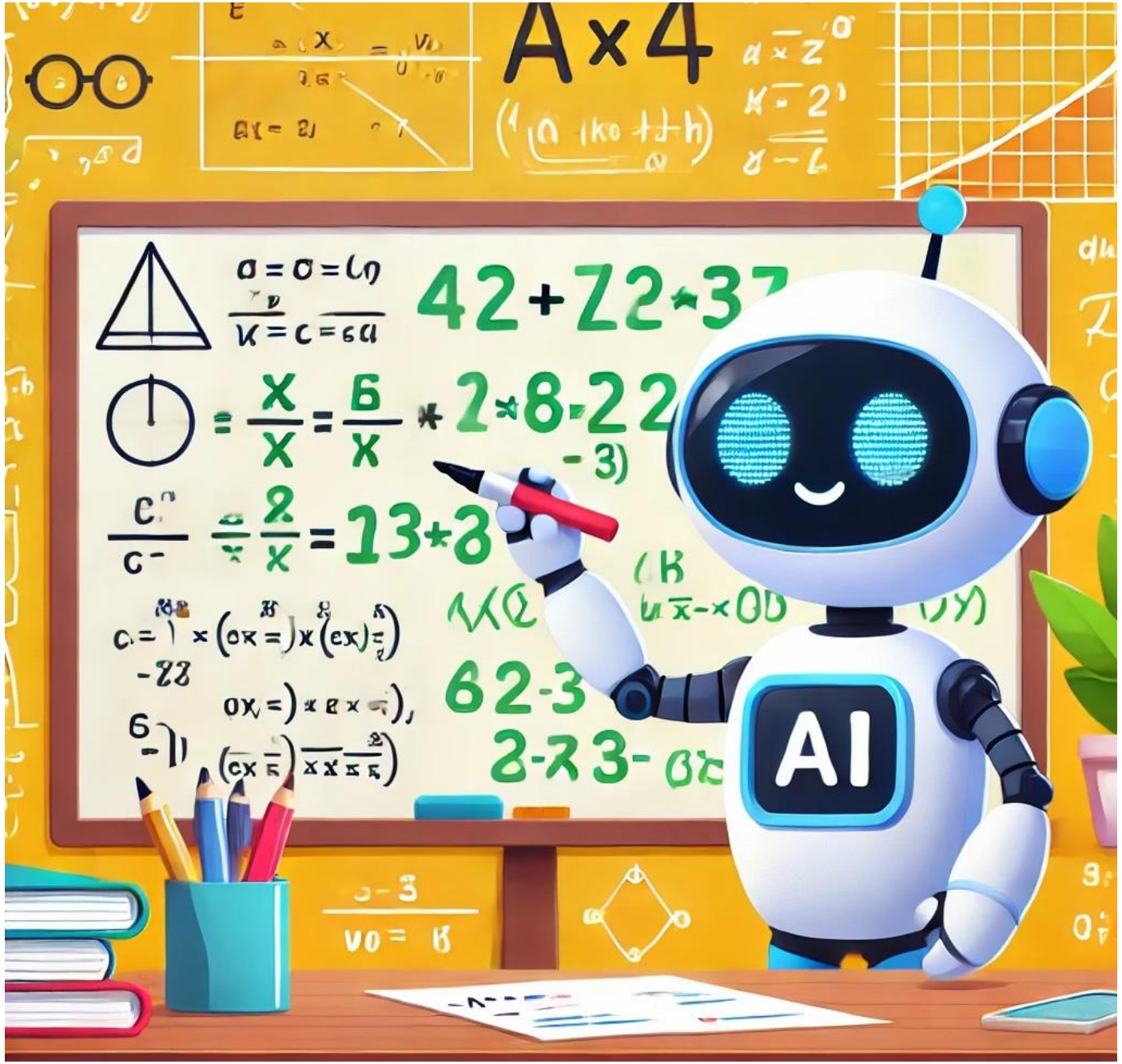
- "Glue on Pizza"
- "Rocks to eat each day"
- CalcGPT.io

Hard to assess correctness, especially if you are new to optimization

Not all AI are created equal

- ChatGPT currently viewed as best-of-breed, believed to be using more training data than others


CalcGPT
by Calvin Liang
calvin.sh
Temperature
1
Craziness of the answers
Top P
1
C
Regenerate
÷
7
8
9
×
4
5
6
−
1
2
3
+
0
.
=



GUROBI SUMMIT

Generative AI
Adoption for Gurobi

Potential for training and technical support

- Best where **common answers exist**

Demonstration: **Gurobot** for ChatGPT

- Available now at no extra cost with a ChatGPT account

ChatGPT 4o

ChatGPT

Gurobot

Explore GPTs

Previous 7 Days

Graphical Processing Image

Recipe with what's in my kitchen

Overcome procrastination

Plan a relaxing day

Content calendar for TikTok

Message ChatGPT

Gurobi Optimization workspace chats aren't used to train our models. ChatGPT can make mistakes.

Invite members

GUROBI SUMMIT

Generative AI
Bottom line

Potential learning tool for optimization

- Can build complex models and answer complex questions, even today

Unlikely to replace algorithms - see chess

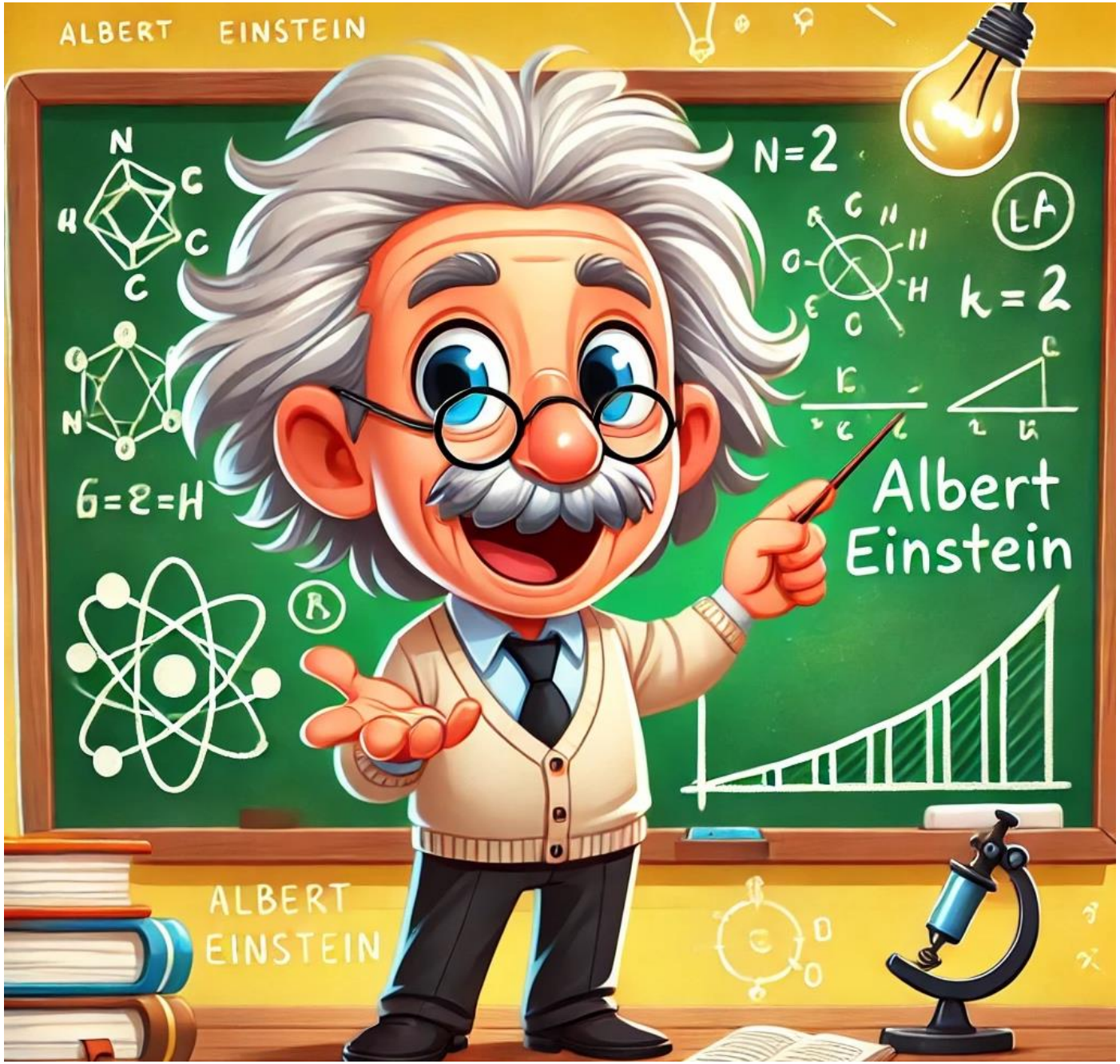
A rapidly evolving field

GUROBI SUMMIT

Quantum Computing
What is it?

New computer architecture based on quantum bits (qubits) rather than binary bits

Unlike a bit that can be only 0 or 1, a qubit can be any quantum superposition of 0 and 1



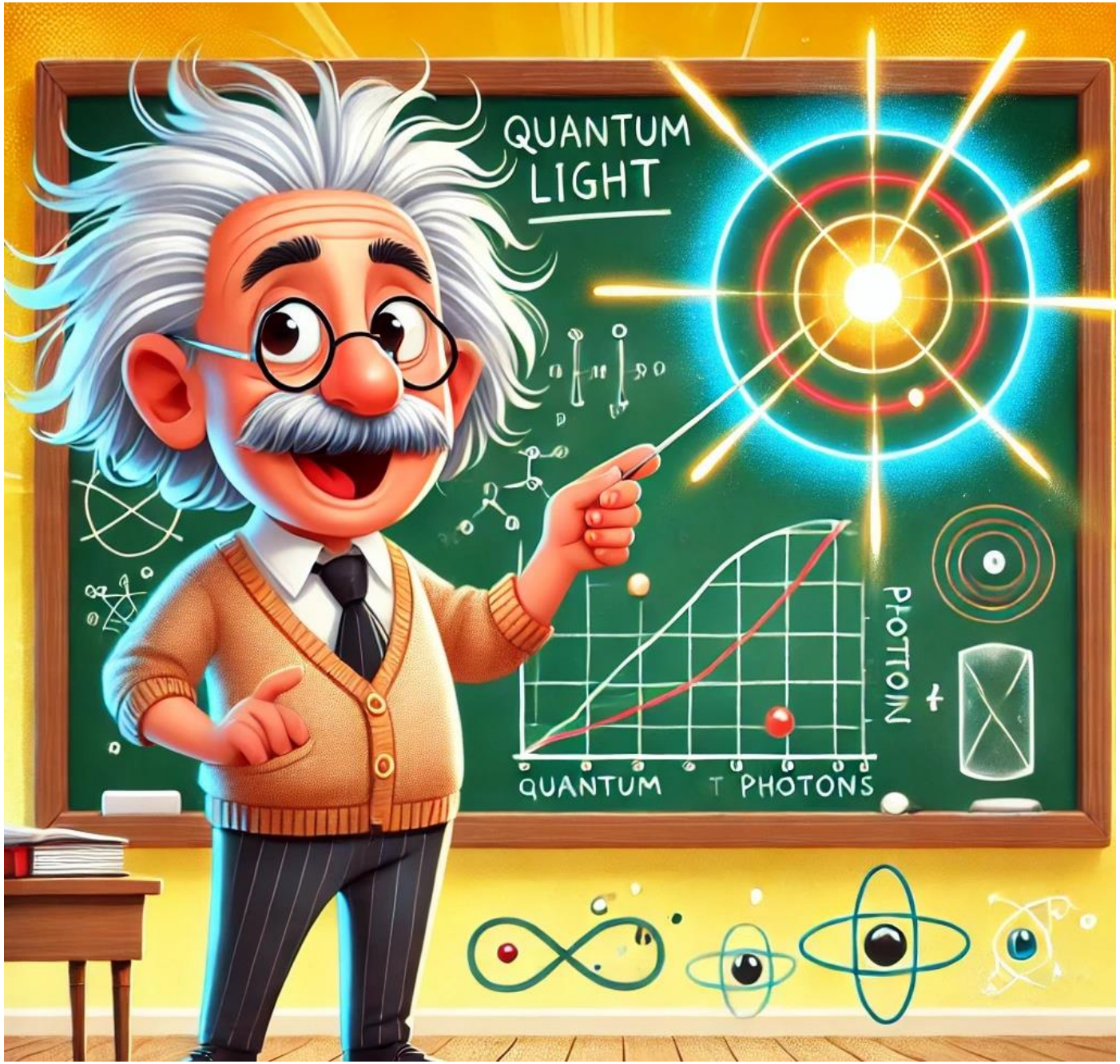
GUROBI SUMMIT

Quantum Computing
How does it work?

Qubits are based on **quantum mechanics**

Qubits can use different elements of physics

- Photons (light)
- Electrons
- Superconductive current



GUROBI SUMMIT

Quantum Computing

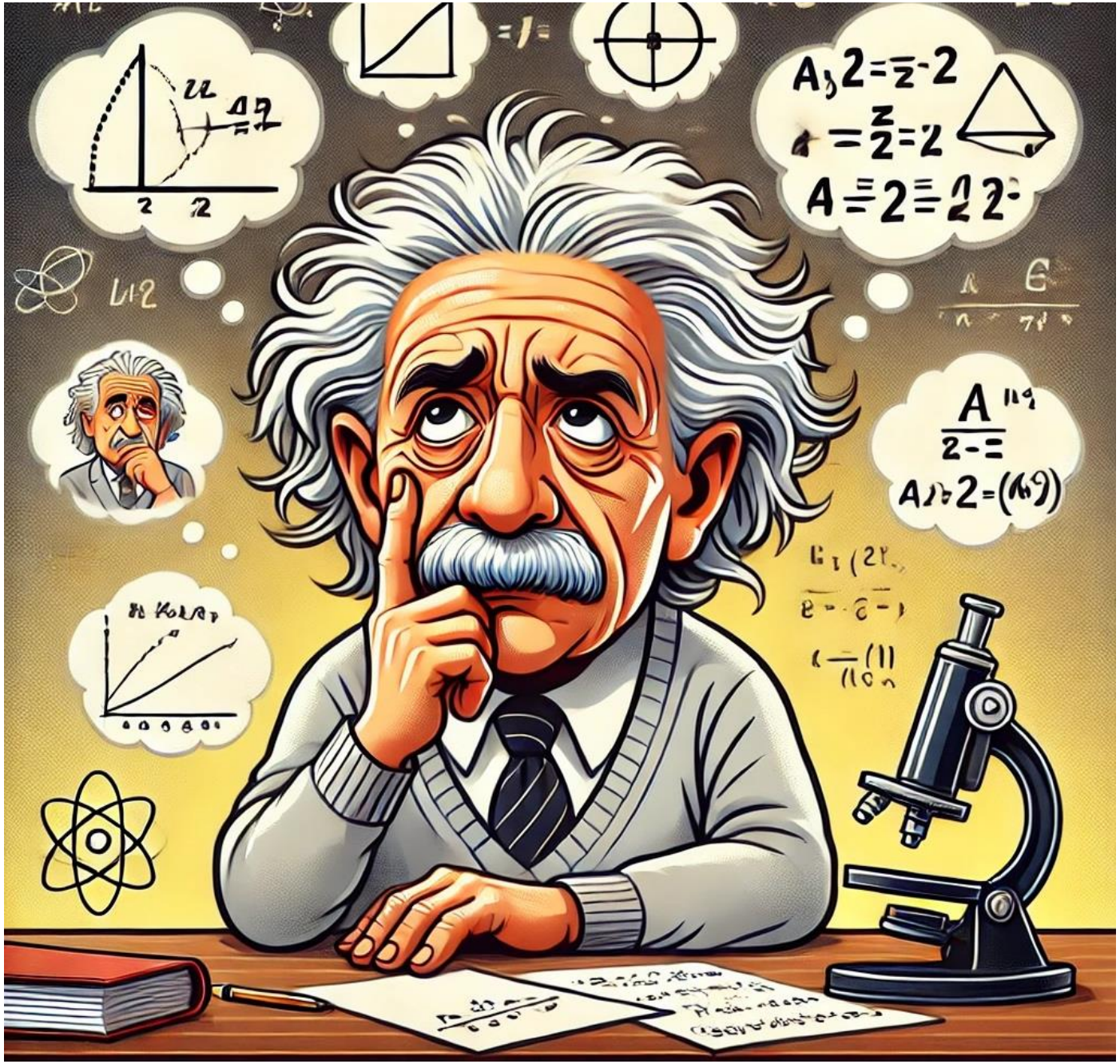
How can optimization benefit

Promises **huge speedups for hard algorithms** thanks to several factors

- Quantum states of a single qubit can represent many values
- Apply quantum gates to entangled qubits to produce powerful and profound changes to their quantum states

A core application is to solve Quadratic Unconstrained Binary Optimization (QUBO)

- QUBO **raised awareness for** combinatorial **optimization**



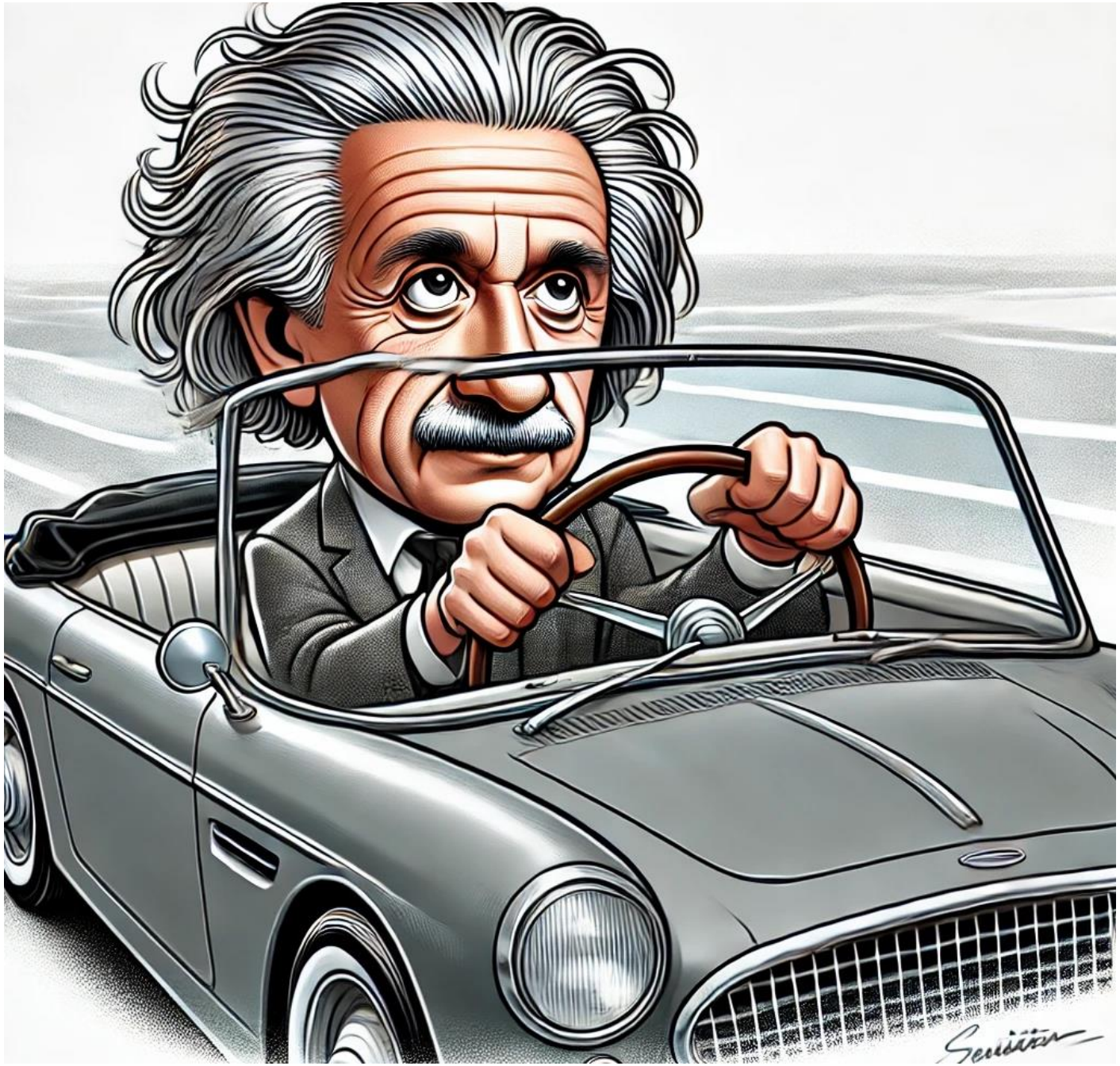
GUROBI SUMMIT

Quantum Computing
Case Study: Vehicle Sensors

BMW 2021 Quantum Challenge (w/AWS)
- 4 industrial applications
- 70+ applicants

Optimization application: locate vehicle sensors to scan for hazards
- Scan all directions with min # of sensors
- Winning entrant: Accenture
  - **Gurobi on regular computer 4500x faster than quantum annealer**
  - Success story for quantum computing?!?
  - Source: https://www.twitch.tv/videos/1546697903

- Thinking about quantum helped Accenture make a better MIP



GUROBI SUMMIT

Quantum Computing
Barriers to adoption

Large scale general-purpose quantum computers **do not exist today**

- Research projects at giant laboratories –
  Can't order one from Dell, HP, Lenovo, etc.

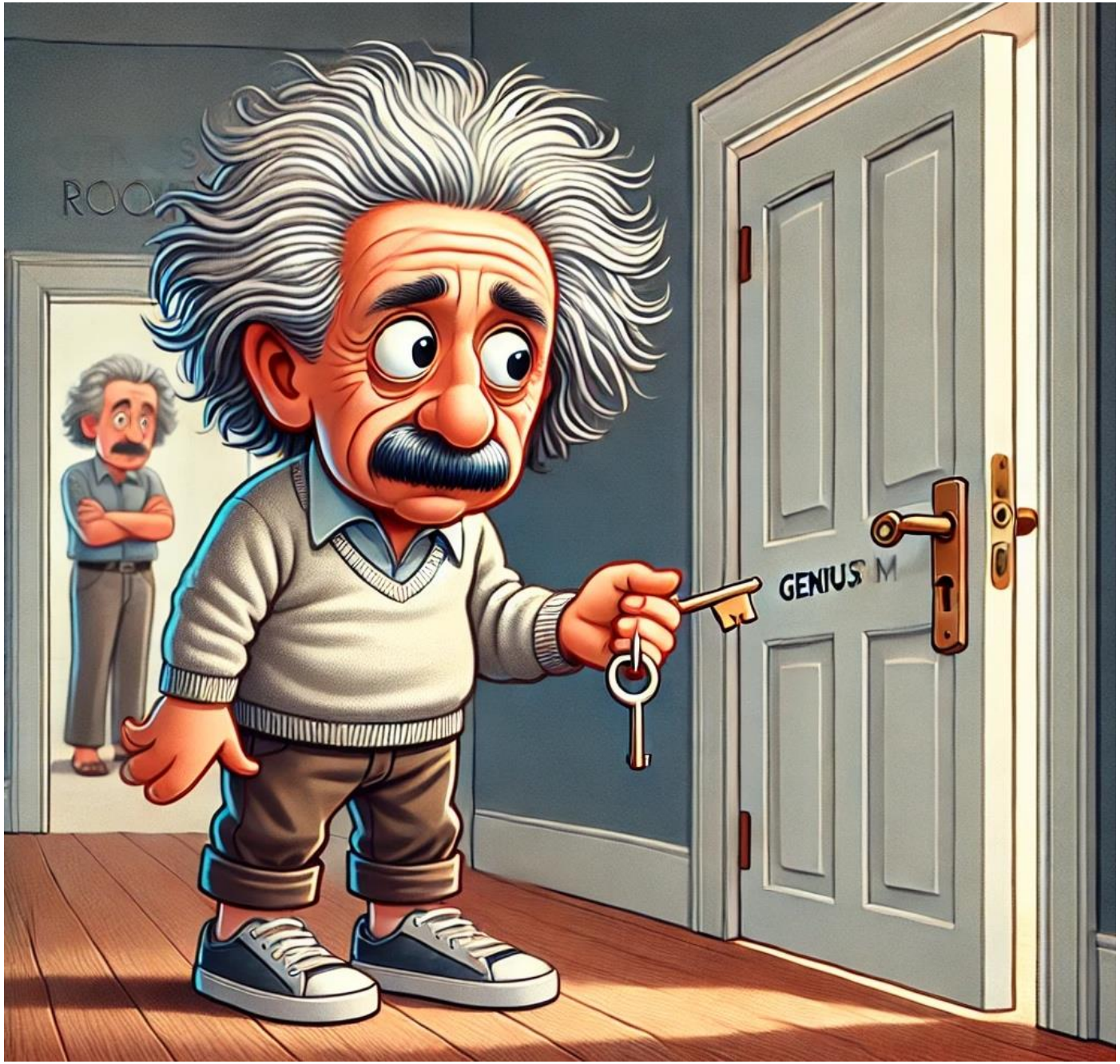
Quantum error correction is an open issue

- Need 100-1000 physical qubits to be entangled to get 1 noise-free logical qubit

Current quantum optimization algorithms can find solutions but **not show quality or prove optimality**

Investors shifted to Generative AI

- Global quantum investment in 2023 **down by 50% while generative AI grew**
- Source:
  https://theregister.com/2024/01/31/quantum_sector_saw_50_percent/

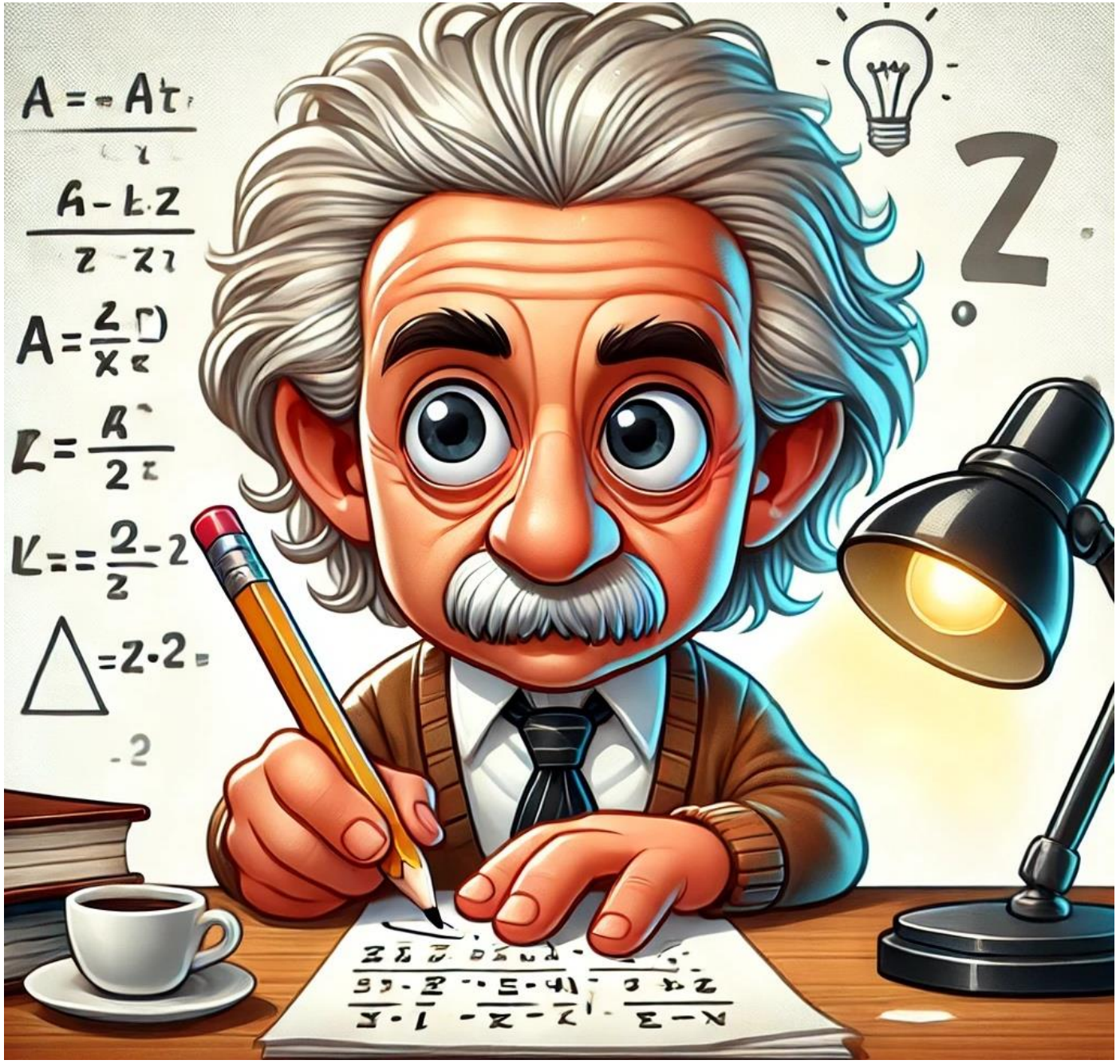


GUROBI SUMMIT

Quantum Computing
Adoption for Gurobi

Wait and see, possibly many years

Continue to **fight misunderstanding** that **real-world optimization** models **cannot be solved today**



GUROBI SUMMIT

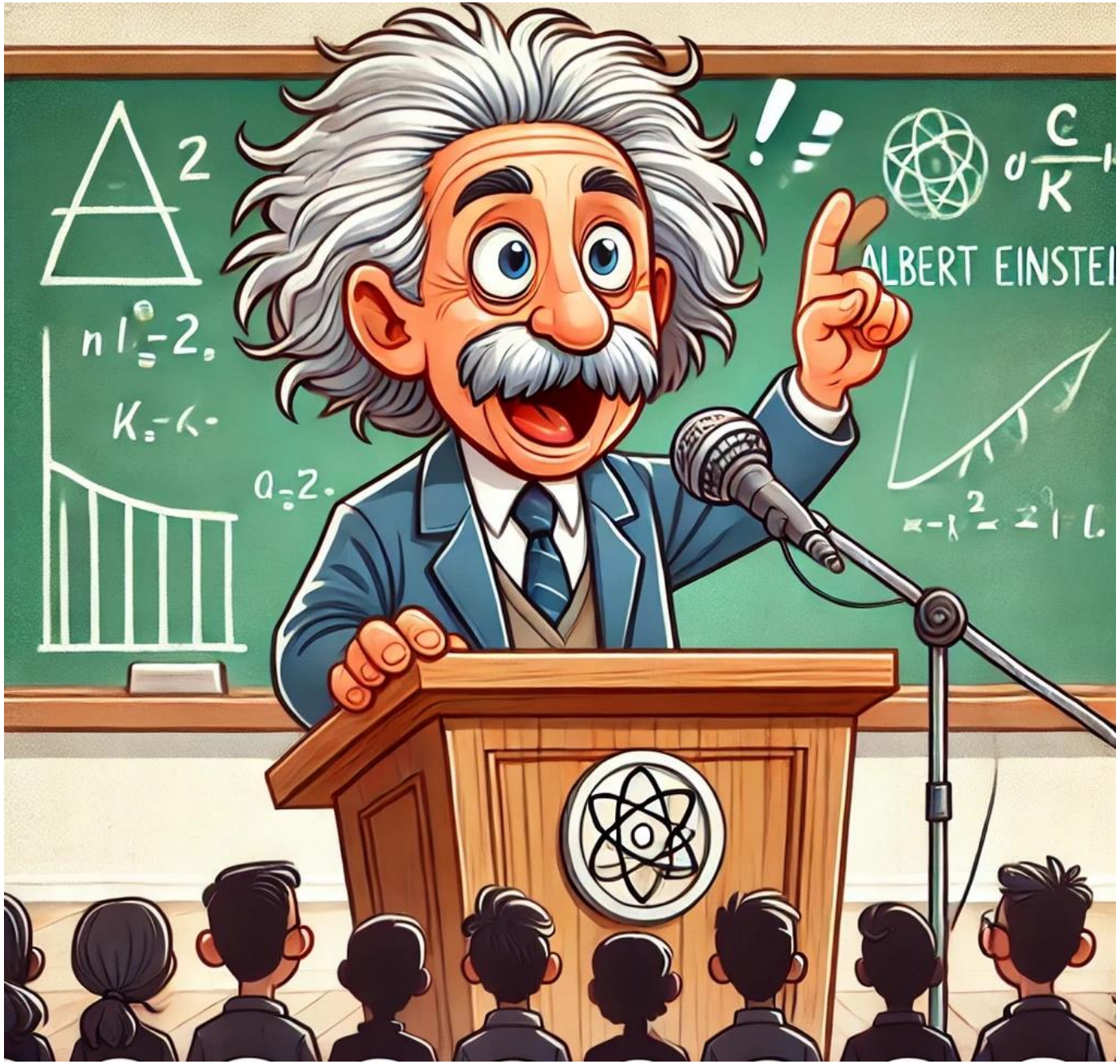
Quantum Computing
Bottom line

Quantum computing generates interest due to its **innovative computing model** with the potential to be much more powerful

New **quantum algorithms are extremely difficult to develop**; only a handful today (Schor, Grover, QAOA)

For optimization, quantum boils down to QUBO

LP (simplex) is **a powerful computational foundation** – will QUBO be?



GUROBI SUMMIT

Key takeaways

GPU Computing: Some parallel tools should make optimization faster soon

Generative AI: Potential to create sample model code, which must be scrutinized for correctness

Quantum computing: Still building foundations at this very early stage

GUROBI SUMMIT

Thank You

For more information: gurobi.com

Dr. Edward Rothberg
Chief Scientist and Chairman of the Board
rothberg@gurobi.com