

SCML 2026 at RISC, Austria, July 6-8, 2026

Proof Engineering: Nakano's Light Puzzle Example

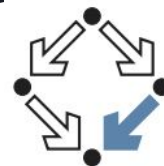
Koji Nakagawa*, Bruno Buchberger**

(*) Rigor Drive, Japan

(**) Research Institute for Symbolic Computation (RISC),
Johannes Kepler University, Austria



Rigor Drive
DON'T THINK, FEEL THE PROOF!



RISC
Linz / Hagenberg

Content

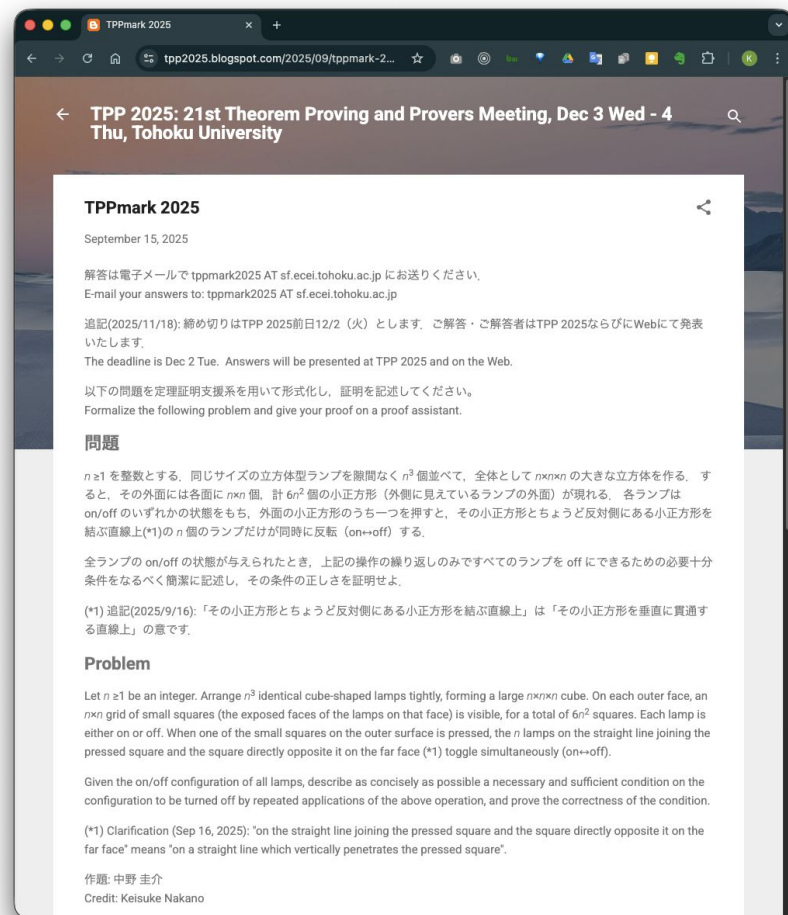
- Nakano's Light Puzzle Problem
- Why is this problem beautiful?
- What is the “Human-LLM-ARS” Cycle?
- What are our experiences?

Content

- Nakano's Light Puzzle Problem
- Why is this problem beautiful?
- What is the “Human-LLM-ARS” Cycle?
- What are our experiences?

TPP and TPPmark

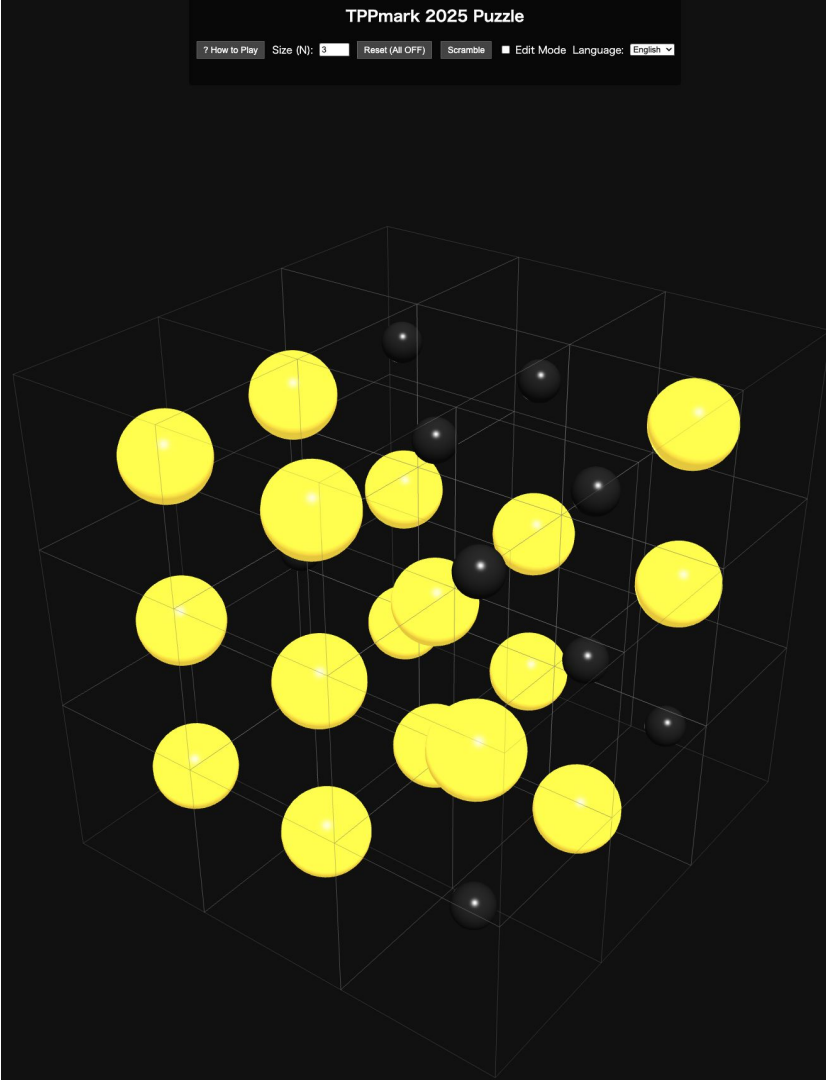
- TPP(Theorem Proving and Provers) is a meeting held every year since 2005 and provides a forum to exchange ideas for both users and implementers of theorem provers and proof assistants.
- TPP 2025: Dec 3 - 4, Tohoku University, JAPAN
- TPPmark: A proof contest posed annually in conjunction with TPP



TPPmark 2025:

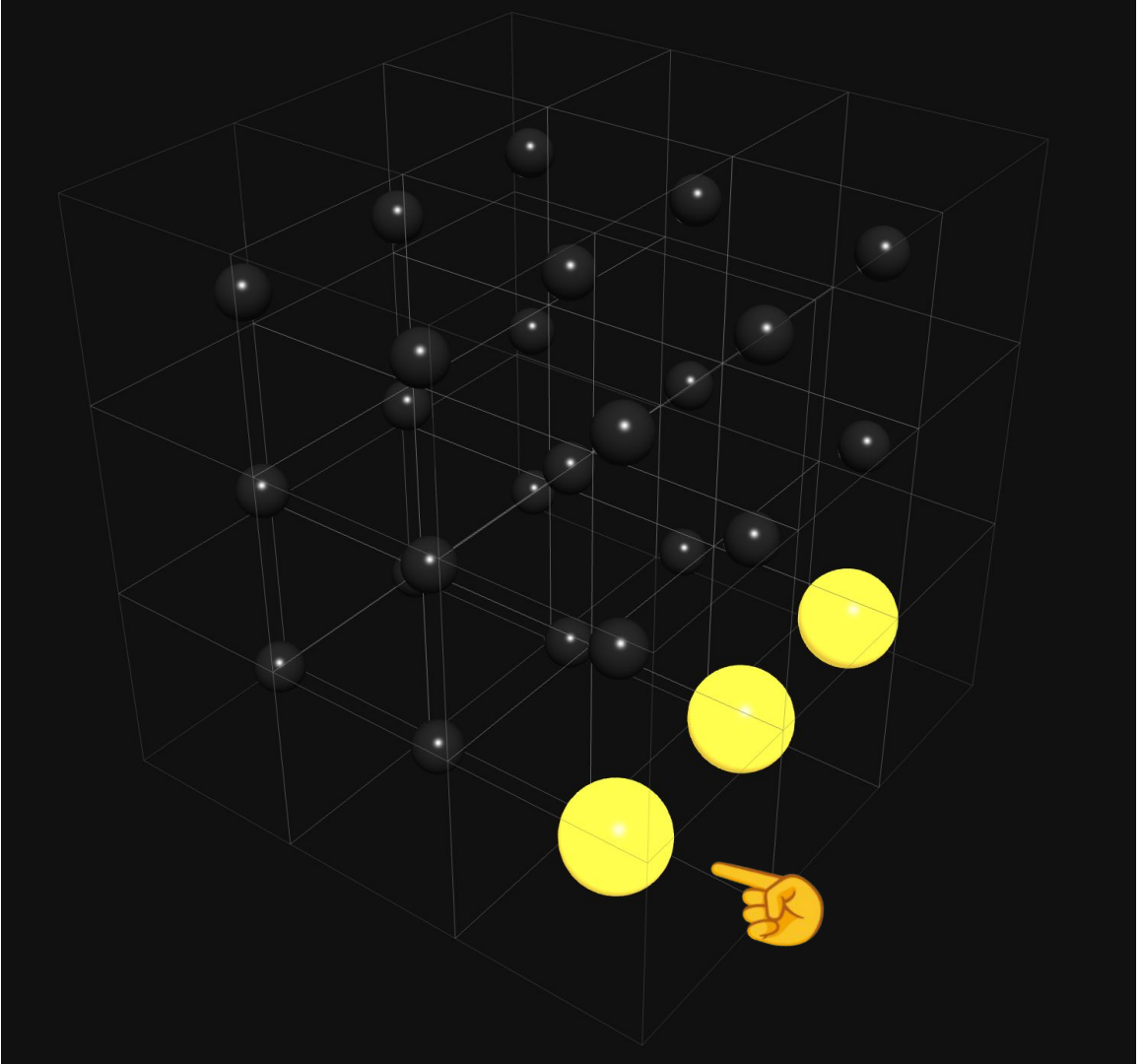
Nakano's Light Puzzle (1)

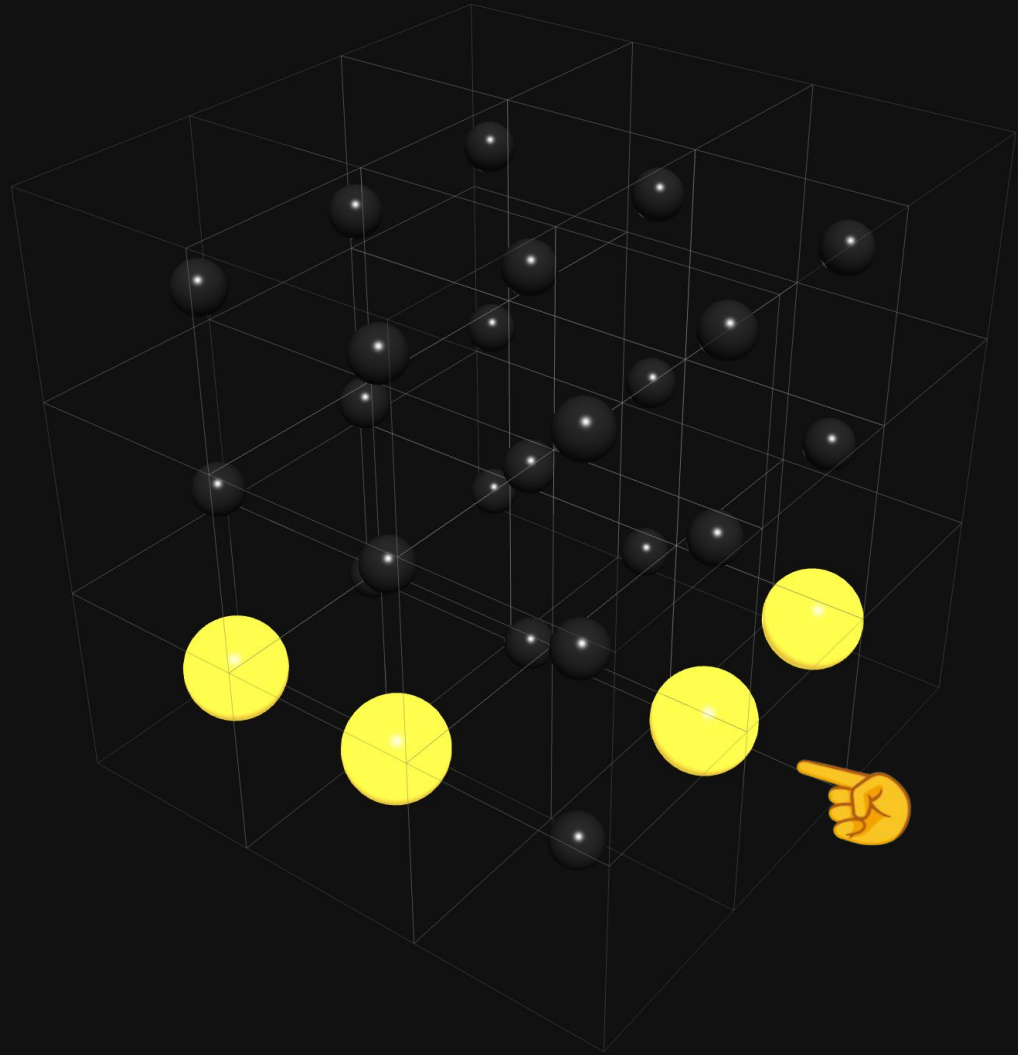
- Let $n \geq 1$ be an integer. Arrange n^3 identical cube-shaped lamps tightly, forming a large $n \times n \times n$ cube
- On each outer face, an $n \times n$ grid of small squares for a total of $6n^2$ squares
- Each lamp is either on or off.
- When one of the small squares on the outer surface is pressed, the n lamps on the straight line joining the pressed square and the square directly opposite it on the far face toggle simultaneously (on $\leftrightarrow$ off).

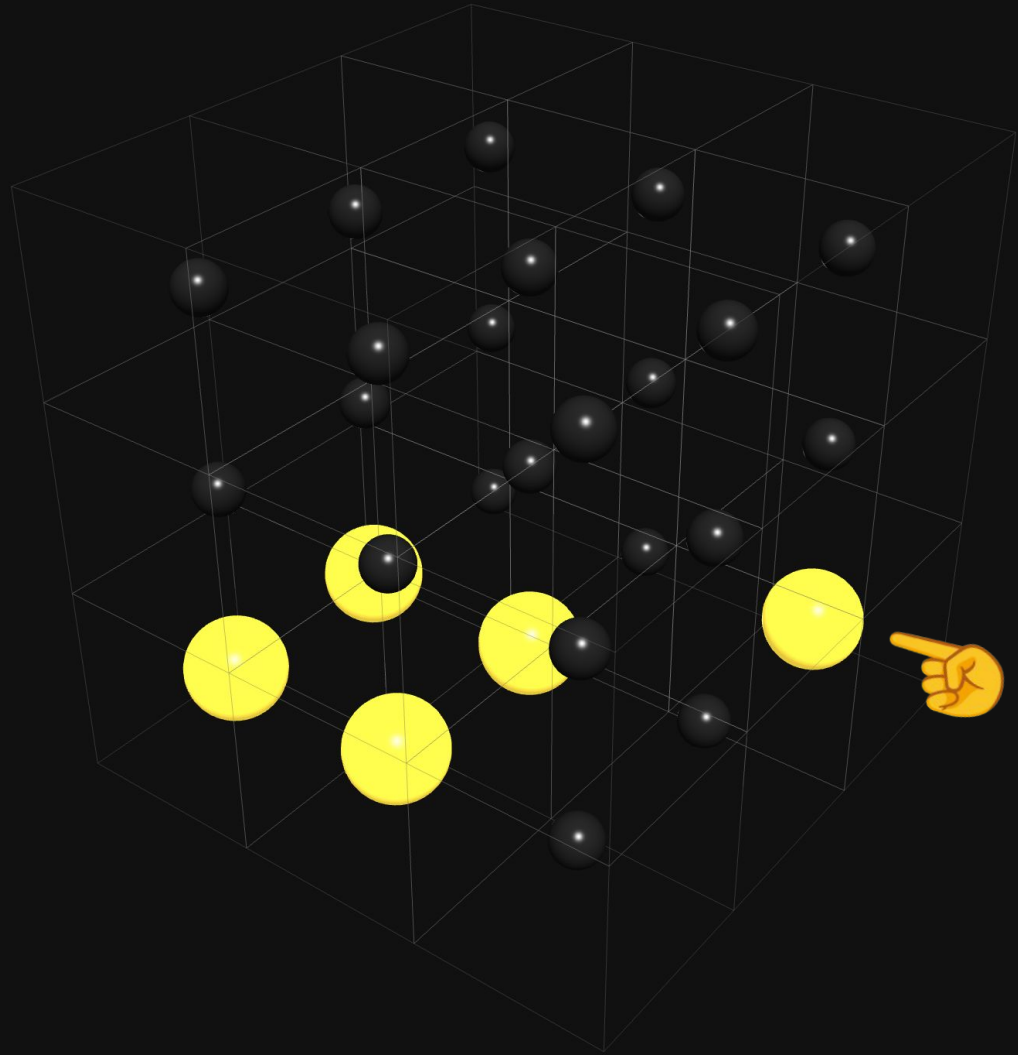


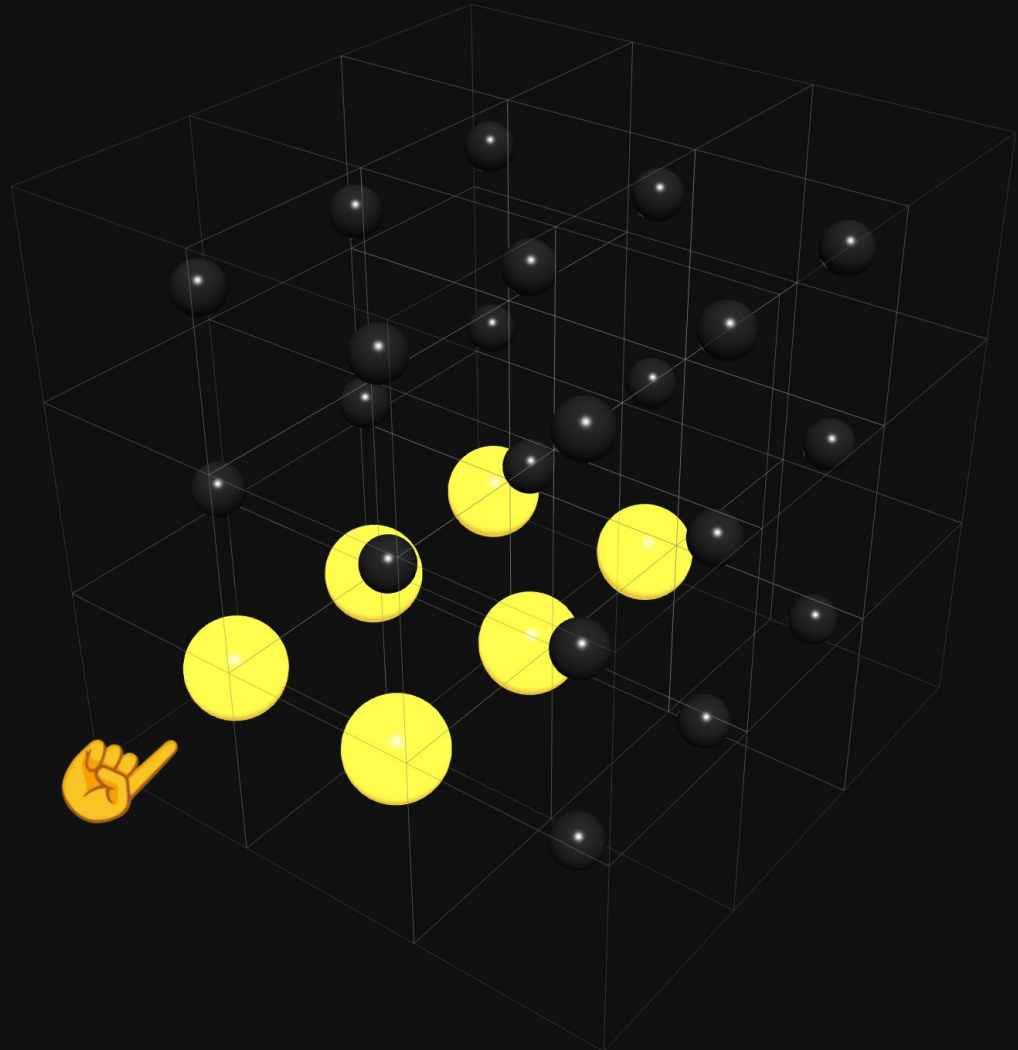
Example:

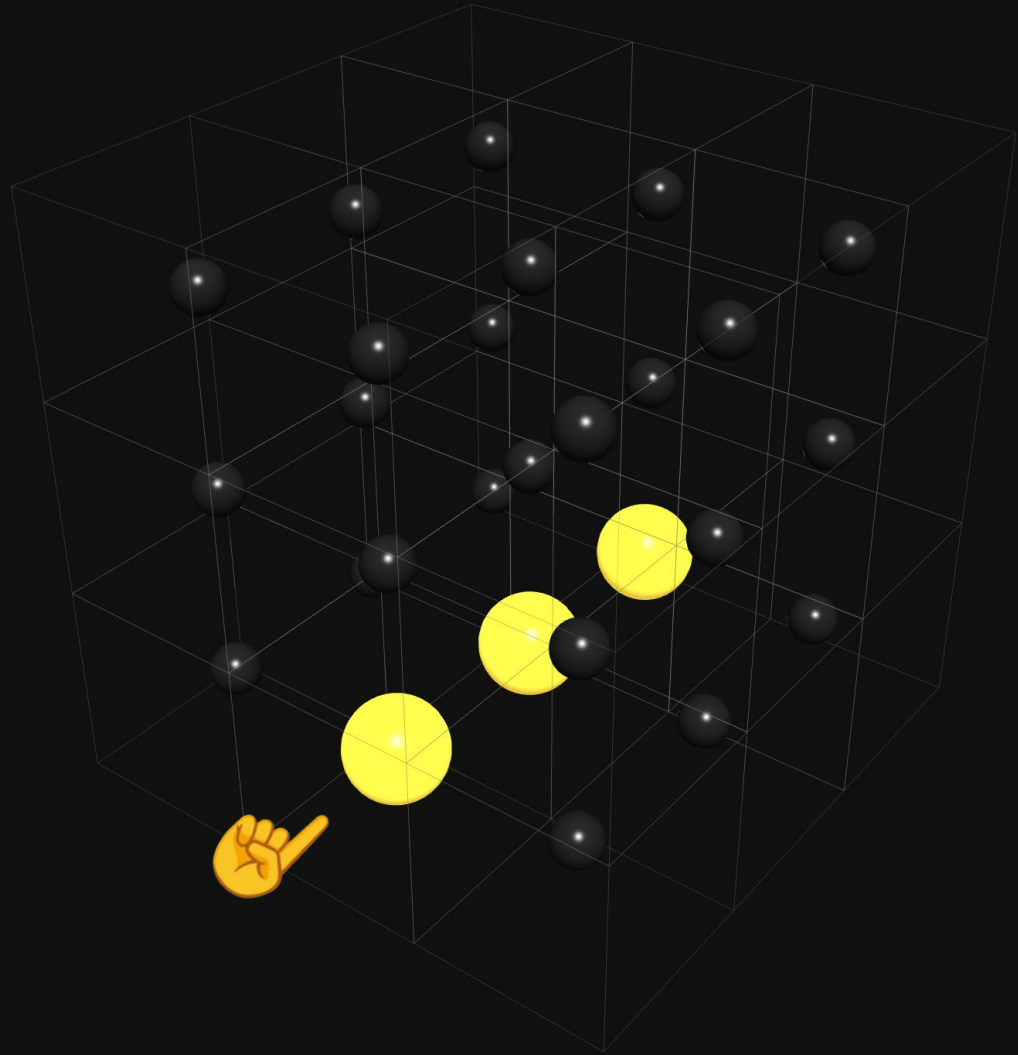
An initial given configuration

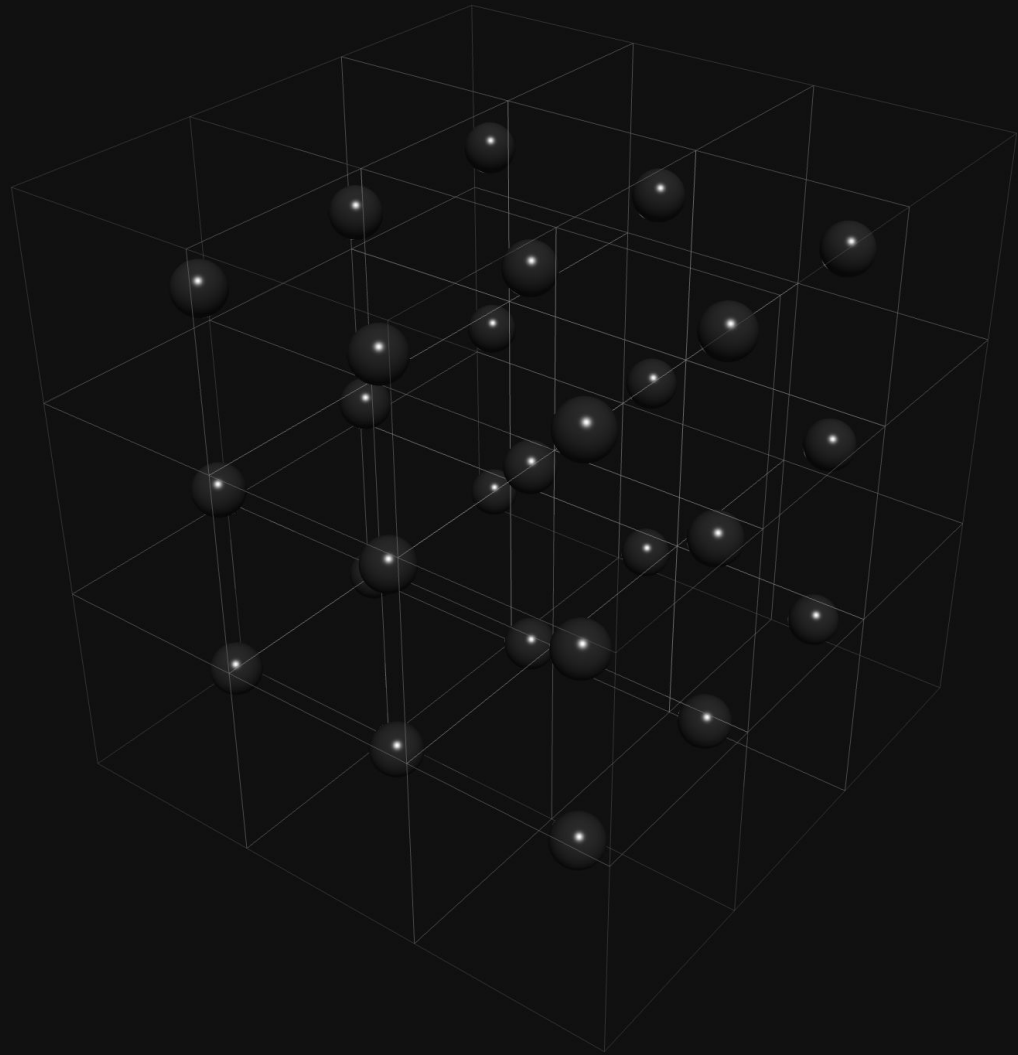




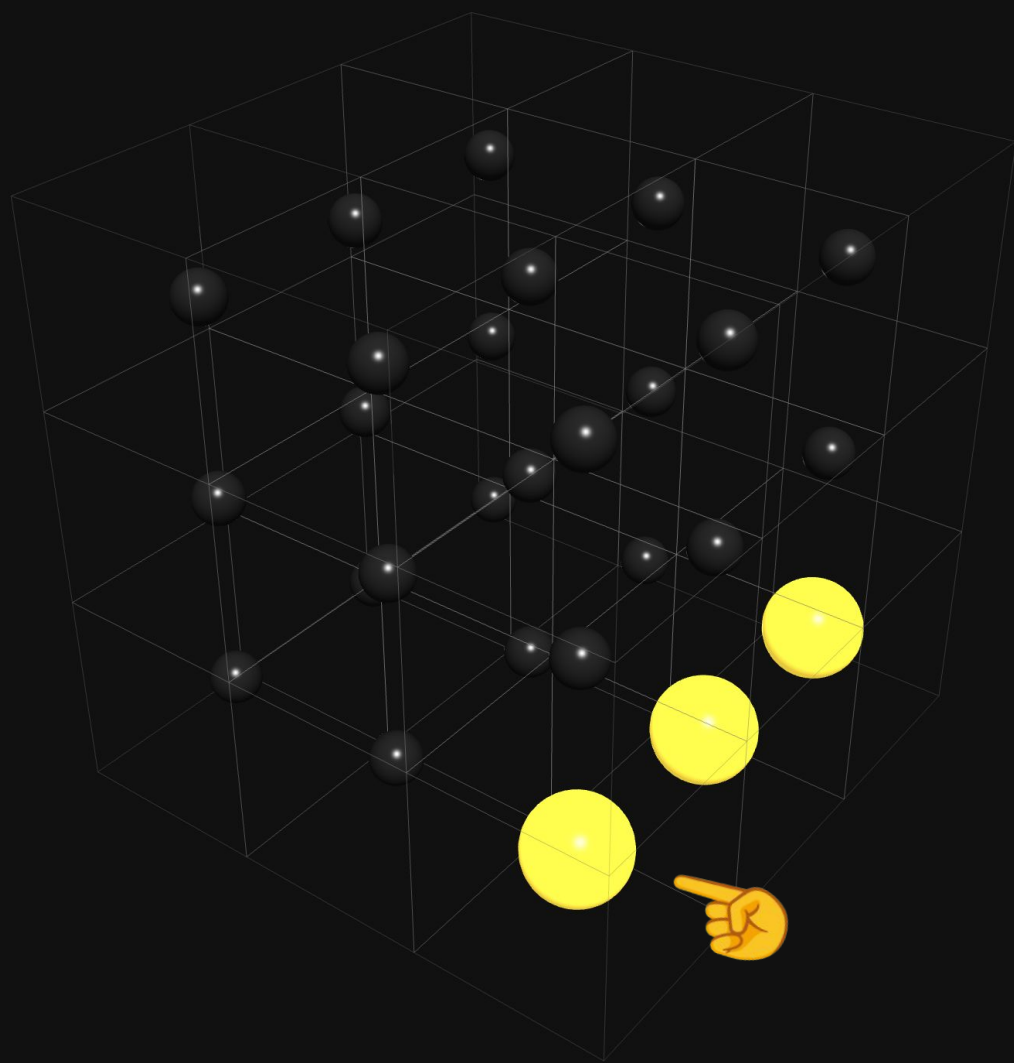








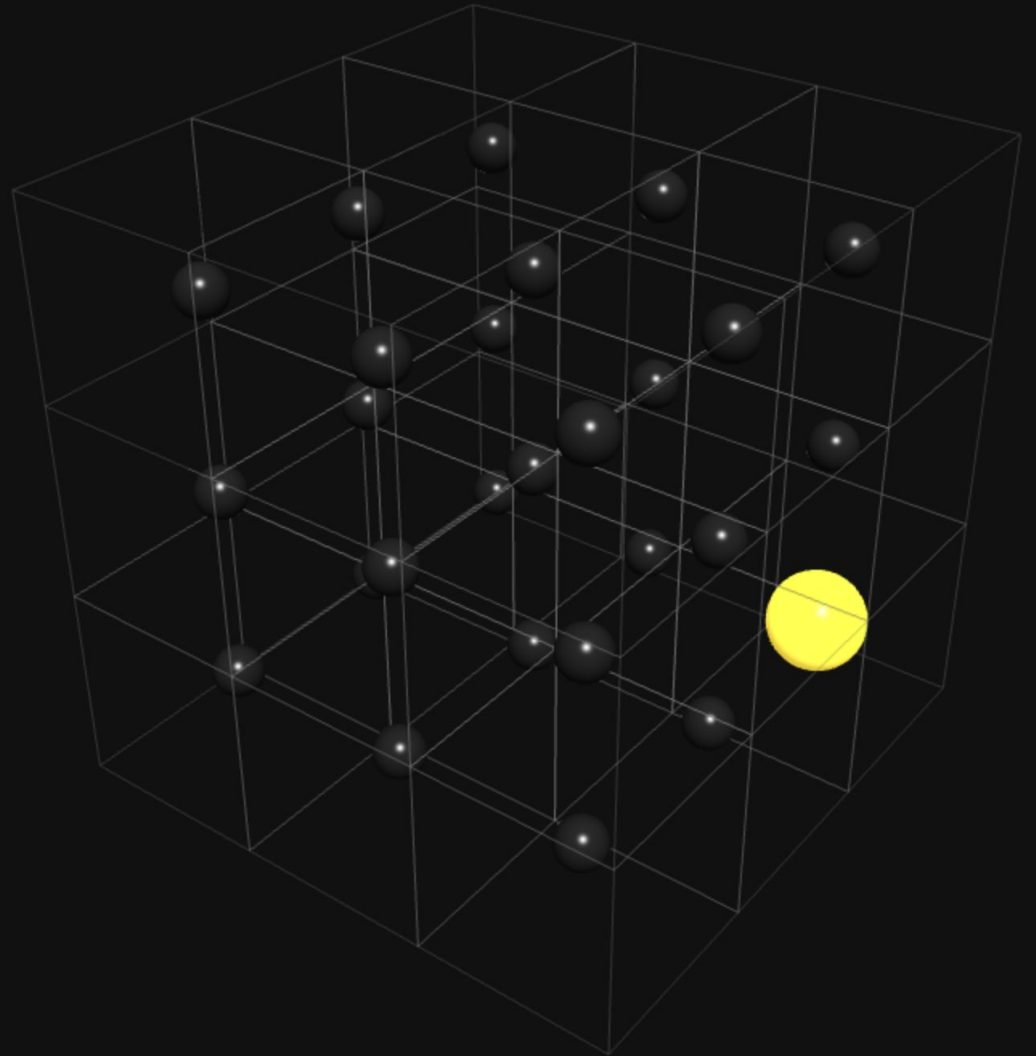
So, the initial configuration
is “solvable”.



Not all Configs...

This configuration can
NEVER be turned all the
lamps off!

It's not obvious!

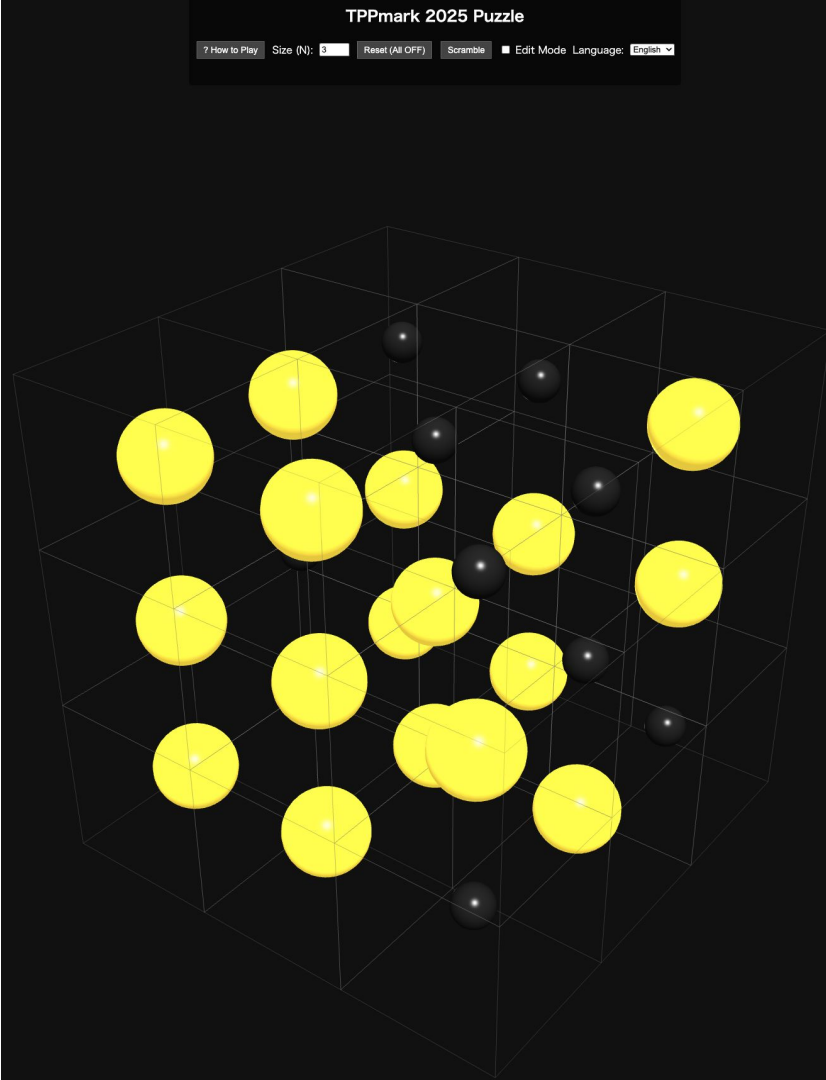


TPPmark 2025:

Nakano's Light Puzzle (2)

- Given the on/off configuration of all lamps, describe **as concisely as possible a necessary and sufficient condition** on the configuration to be turned off by repeated applications of the above operation, and prove the correctness of the condition.

The contestants are required to formalize the problem and give the proof on a proof assistant.



Content

- Nakano's Light Puzzle Problem
- Why is this problem beautiful?
- What is the “Human-LLM-ARS” Cycle?
- What are our experiences?

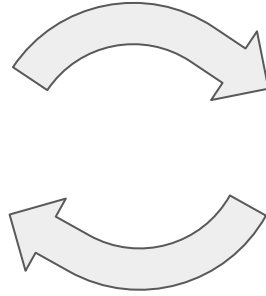
Why is this problem beautiful?

- The rule is so easy to understand even elementary school kids can play
- On the other hand the “condition” can not be easily found
- Even some “conditions” may be true, but it may be inefficient
- There is an efficient algorithm, but the proof is non-trivial and it does not need any deep mathematics
- If one knows the algorithm, the puzzle becomes uninteresting
- If the “condition” is not satisfied and you don’t know the theorem, you might try forever

Content

- Nakano's Light Puzzle Problem
- Why is this problem beautiful?
- What is the “Human-LLM-ARS” Cycle?
- What are our experiences?

Human - ARS Cycle

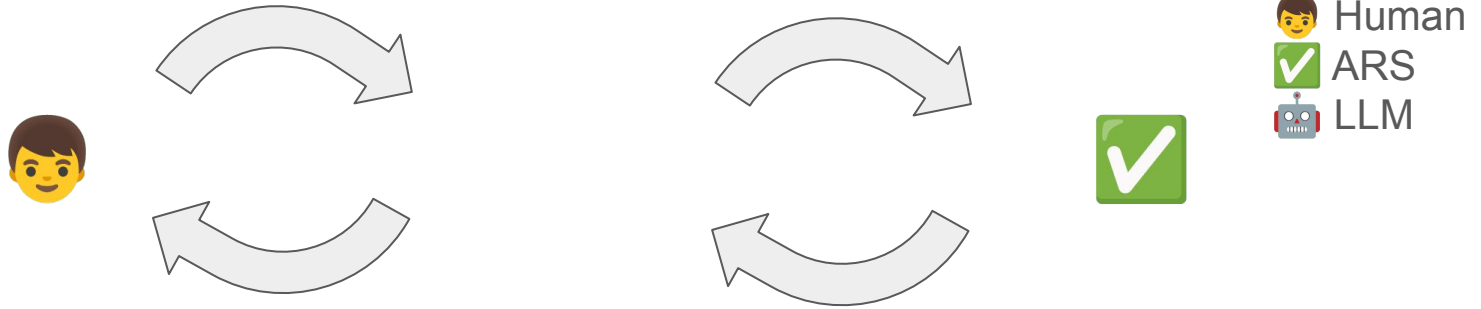


 Human
 ARS

Automated Reasoning System

- Proof Checker
 - Lean
 - Rcoq
 - Agda
- Proof Generator
 - Theorema

Human - LLM – ARS Cycle



General LLMs

- ChatGPT
- Claude
- Gemini

Specialized LLMs

- Aristotle
- Numina Lean Agent

Proof Engineering in the Age of AI(LLM)

- Environment and System Setup
 - Prompt Engineering
 - Context Engineering
 - Harness Engineering
- Formal Theory Development
- Theorem Proving and Refinement
- Human Comprehension Transformation

Content

- Nakano's Light Puzzle Problem
- Why is this problem beautiful?
- What is the "Human-LLM-ARS" Cycle?
- What are our experiences?

What are our experiences?

- [Exp. 1] Github Copilot via VSCode with Mathlib (Nov. 2025)
- [Exp. 2] Github Copilot via VSCode without Mathlib (Apr. 2026)
- [Exp. 3] Feeded into Aristotle (Mar. 2026)
- [Exp. 4] Gave the Problem to General LLMs with the Algorithm (May. 2026)
- [Exp. 5] Gave the Problem to General LLMs with the Proof (Jun. 2026)

[Exp. 1] Github Copilot via VSCode with Mathlib

Prompt:

- Nakano's Light Puzzle Description

Method:

- Human-LLM-ARS cycle

Result:

- Koji failed and couldn't submit to TPPmark2025
- However Koji made a web application of the puzzle by Vibe Coding
- The game was played at the TPP2025 meeting
- After the meeting Koji talked to Bruno about the problem and the algorithm
- Bruno got interested in and started to prove by hand

[Exp. 2] Github Copilot via VSCode without Mathlib

Prompt:

- Nakano's Light Puzzle Description
- Idea of the Efficient Algorithm
- Instruction "Do not use Mathlib"

Method:

- Human-LLM-ARS cycle

Result:

- A Lean proof was successfully obtained! It seemed ok.
- Since I'm just a software engineer, let LLM explain the proof in 3 natural languages (English, Japanese, German)
- I let a mathematician (namely Bruno) check the proof...
- Bruno read these and appreciated!
- Bruno studied the proof in detail, found it correct (except some small formal deficiencies) and was impressed because it found the three main ideas of the proof by itself and was able to provide all details.

[Exp. 3] Fedded into Aristotle

Prompt:

- Nakano's Light Puzzle Description

Method:

- Human-LLM-ARS cycle (Single Cycle)

Result:

- Bruno pointed out that there are tacit assumptions in the formalization.
- We should not trust the results even if Lean says "Yes, it's proved".

[Exp. 4] Gave the Problem to General LLMs with the Idea

Prompt:

- Nakano's Light Puzzle Description
- Idea of the algorithm

Method:

- Human-LLM-ARS cycle
- LLMs (ChatGPT, DeepSeek 3.2, Grok 4, Claude Sonnet 4.6) 2026/4/29

Result:

- Claude Sonnet 4.6 was the only one that came up with a complete proof. ChatGPT only gave some general comments about proving.
- The Claude Sonnet proof was basically ok with one essential error and one unnecessarily complicated passage.

[Exp. 5] Gave the Problem to General LLMs with the Proof

Prompt:

- Nakano's Light Puzzle Description
- A Detailed Proof of the Algorithm

Method:

- Human-LLM-ARS cycle
- Asking LLMs (ChatGPT, DeepSeek, Grok, and Claude Sonnet)

Result:

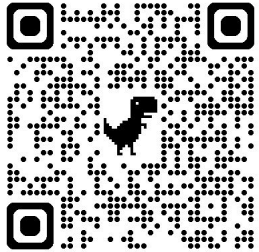
- Practically, the systems were not able to cope with the (long) detailed proof, and we had to give up. Probably it would be better to just give the detailed proof to a proof checker like LEAN.

Conclusion

- Human-LLM-ARS cycle has great practical and theoretical potential
- The more powerful support we get from LLMs and ARSs, the more important is the role of experienced, well educated and creative mathematicians and software engineers

TPPmark 2025: Nakano's Light Puzzle

<https://www.rigordrive.fun/lab/scml2026>



Play it with your
smart phone!

