Hungarian Physical Reasoning Dataset Annotation Information

The goal was to build a Hungarian dataset of everyday physics/common-sense judgments. Each item presents a scenario (prompt) plus two plausible solutions; exactly one is physically correct, the other is realistic-sounding but wrong. A machine-readable label (0/1) marks the correct option.

Workflow:

Authoring: Two human annotators independently wrote examples in Hungarian following the instruction set below.
Validation: A third reviewer checked every item for clarity, realism, and physical correctness; labels were verified and inconsistencies resolved.
Translation: English versions of prompts and solutions were produced with ChatGPT-5.

Data schema:

- `prompt`, `solution0`, `solution1`, `label` (0 or 1 indicates which solution is correct)
- `domain` (e.g., mechanics, heat, fluids, electricity/magnetism, light/sound), `cultural` ("yes" if explicitly Hungarian, "no" otherwise), `prompt_en`, `solution0_en`, `solution1_en`: translations, `notes:` comments, if any.

Scale & coverage targets:

At least 100 Hungarian examples were expected; ≥1/3 should include Hungarian cultural context (e.g., Balaton, tram, *lángos*).
Suggested topical spread: mechanics/motion (falling, collisions, weight, equilibrium); materials and properties (wood, metal, glass, textiles, foods); thermal phenomena (boiling, freezing, heating, cooling); electricity/magnetism; fluids (pressure, buoyancy); light/sound; everyday Hungarian scenes.

Item design guidelines:

Prompt: Detailed natural-language scenario, preferably >25 words.
Solutions: Two very similar, realistic answers that differ subtly; avoid absurd wrong answers.
Labeling: Exactly one solution is physically correct; set `label` to the index of the correct one (0 or 1).

The third-person validator ensured: (i) prompts are specific and realistic; (ii) the incorrect option is plausible but physically wrong; (iii) domain tags and cultural flags are used consistently; (iv) English translations faithfully reflect the Hungarian originals while preserving subtle contrasts between the two solutions.