

# Polish Subset of the Multilingual Physical Commonsense Reasoning Dataset

Michał Marcińczuk  
Sages / Warsaw, Poland  
m.marcinczuk@sages.com.pl

## 1 Dataset Description

The dataset comprises 103 manually created and verified examples for Polish. The list of topics covers mechanics, fluids, thermodynamics, phase changes, material strength, sound, light, and practical everyday physics experiments.

The average length of the prompt together with its answer is **24.3**. There are 54 examples that are 25 words or longer. Figure 1 illustrates the length distribution for the entire dataset.

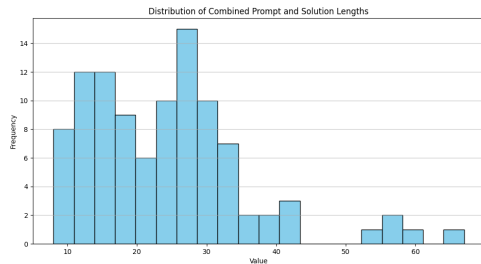


Figure 1: Distribution of prompt lengths including their corresponding solutions.

Examples of Polish-specific words:

- *bigosznica* is a colloquial Polish term for a pot or large cooking vessel used specifically for making bigos.
- *warzecha* is a colloquial or regional Polish term for a wooden cooking spoon.
- *pierogi* is are a traditional Polish dish consisting of dumplings made from unleavened dough and usually filled with various ingredients.
- *kropelka* (literally “little drop”) is a colloquial term for cyanoacrylate glue, commonly known in English as super glue.

## 2 Methodology

The methodology for constructing the Polish dataset consisted of two main stages. In the first stage, a list of physics topics was created that covered fundamental laws of physics, material properties, and principles governing the interactions between materials. For this purpose, relevant materials available online were collected. The list included, among other things, basic experiments that allow testing of these physical properties under home conditions.

During the second stage, the identified topics were used to generate illustrative examples, with one or several examples prepared for each topic. The process began with a true statement, from which a false counterpart was created by making as few changes as possible. Subsequent examples were extended versions of the basic example, in which some elements were replaced with alternative terms related to the task description. For some examples, twin examples were also created, i.e., examples containing the same set of answers, but with the correct answer changing due to a modification of the goal. For instance:

prompt	solution0	solution1	label
A stone thrown horizontally will move	in a straight line	along a parabola.	1
A stone thrown vertically will move	in a straight line	along a parabola.	0

## 3 Target LLMs for Polish

- **PLLuMM<sup>1</sup>** (Polish Large Language Model) is an open-source artificial intelligence initiative developed by a consortium of six Polish research institutions. Its primary goal is to create a Polish-language alternative to models

<sup>1</sup><https://huggingface.co/CYFRAGOVPL/PLLuM-8x7B-chat>

like ChatGPT, enhancing digital sovereignty and supporting innovation in public administration and business sectors. On-line chat: <https://pllum.org.pl/>

- **Bielik**<sup>2</sup> is an open-source Polish language model developed collaboratively by the Speak-Leash team and many volunteers. It is designed to understand and generate Polish text, and the project encourages community use, adaptation, and further development for various Polish language AI applications. On-line chat: <https://chat.bielik.ai/>
- **APT3**<sup>3</sup> (APT3-1B-Instruct-v1) is a Polish large language model developed by Azurro (with contributions from the Polish AI community). It's an instruction-tuned version of the APT3-1B model, trained on millions of Polish instructions. The model is released under the CC BY-NC 4.0 license, meaning it can be freely used for non-commercial purposes. It is openly available for download and experimentation on Hugging Face.

---

<sup>2</sup><https://huggingface.co/collections/speakleash/bielik-v3-family-681a47f877f72cae528bdab1>

<sup>3</sup><https://huggingface.co/collections/Azurro/apt3-66fa965b5eea43a116b1c545>