

# Bosnian PIQA-Style Dataset: Physical Commonsense with Cultural Context

Contributors: Murat Kolic, kolicm@icloud.com

October 6, 2025

## 1 Overview

We present a manually constructed PIQA-style dataset in Bosnian (Latin script, Ijekavian standard), designed to evaluate **physical commonsense reasoning** and everyday procedural knowledge with explicit **Bosnian cultural grounding**. Each item consists of a natural-language prompt describing a physical task or situation and two candidate solutions that are surface-similar; exactly one is correct. The task format mirrors common PIQA evaluation: `[prompt, solution0, solution1]` with a binary `label` (0 or 1) indicating the correct solution.

## 2 Motivation and Language/Dialect

Resources for South Slavic languages covering physical commonsense remain scarce. This dataset targets Standard Bosnian while incorporating regionally salient vocabulary and scenarios (e.g., *burek*, *somun*, *džezva*, *ajvar*, *hurmašice*, *roštilj*, *sehara*, *ćilim*, *slava*, *Bajram*, *planinarenje*, *mine warning signs*). These allow assessment of models on culturally familiar tasks and safety practices common in Bosnia and Herzegovina (BiH).

## 3 Data Construction

- **Authoring:** Items were written by native Bosnian speakers following PIQA principles: realistic prompts, physically grounded alternatives, and unambiguous correctness.
- **Physical reasoning:** Correct answers align with basic physics, materials, and safety (e.g., heat expansion of glass, oil fires, electrical safety, food hygiene).
- **Surface similarity:** The two solutions are phrased to be as similar as possible while differing in the crucial physical choice (e.g., hot vs. cold syrup for baklava; short vs. long proofing for dough).
- **Cultural coverage:** Cooking (*ćevapi*, *kahva* in *džezva*, *burek* and *jufka*, *ajvar*, *baklava*), household and workshop tasks, mountain/outdoor scenarios (Bjelašnica, Trebević, Neretva/Una), religious and social customs (mosque etiquette, *Bajram* greetings, Orthodox *slava*).
- **Plausible distractors:** Incorrect solutions are plausible but physically wrong or unsafe, avoiding absurd options to keep the benchmark challenging.

## 4 Dataset Format

We release a .TSV file with the following columns:

- `prompt` – Bosnian natural-language description of a scenario (often multi-sentence).
- `solution0`, `solution1` – two surface-similar candidates.
- `label` – 0 or 1, the index of the correct solution.

**Note:** Translated PIQA items can be included but must be flagged externally and do not count toward the 100-item minimum. The included file contains only original items crafted in Bosnian.

## 5 Size and Characteristics

The dataset currently contains **100+** items (see file). Prompts and answers emphasize realistic Bosnian contexts. Many items exceed 25 words across prompt+solutions to reduce superficial cues and encourage genuine physical reasoning.

## 6 Quality Control and Native Review

All items are authored to be native-sounding Bosnian. Items involving safety (gas leaks, electricity, mines) were crafted conservatively to avoid harmful advice; nonetheless, they should *not* substitute professional guidance.

## 7 Reproducibility Guidelines

To recreate a comparable dataset in Bosnian:

1. Define 6–8 domains common in BiH (traditional food prep, home maintenance, winter driving, outdoor safety, first aid, workshop tasks, religious/social customs, mountain activities).
2. For each domain, draft 15–20 scenarios grounded in physics (heat, pressure, friction, materials, hygiene).
3. For every prompt, write two minimal-contrast solutions; ensure one is unambiguously correct given physical principles and local practice.
4. Avoid absurd distractors; prefer “tempting but wrong” steps (e.g., pouring water on an oil fire).
5. Balance labels (0/1) and vary linguistic patterns to avoid position bias.
6. Run native review: fix ambiguity, grammar, and culture-specific details.

## 8 Ethical and Safety Considerations

Some items mention potentially dangerous situations (e.g., explosive remnants of war, fire, electricity). We phrase correct answers with standard safety-first recommendations. This dataset is an evaluation resource; users must not treat it as a safety manual.

## 9 Availability

The dataset is provided as `bosnian_piqa_dataset.tsv` alongside this paper (L<sup>A</sup>T<sub>E</sub>X source: `bosnian_piqa_paper.tex`).

## 10 Acknowledgments

We thank community members from BiH who advised on cultural details. Upon acceptance to the NeurIPS 2026 D&B Track, contributors may opt in to authorship and help perform minor modifications or additional manual curation if requested by organizers.