

# PIQA-Yoruba: A Commonsense Physical Interaction QA Dataset

---

## Abstract

We present PIQA-Yoruba, a dataset of 100 PIQA-style commonsense reasoning examples in Yoruba. Each item consists of a prompt/goal and two minimally different solutions, with exactly one correct. The dataset is entirely manually constructed and focuses on scenarios that are relevant to the Yoruba land. This enables evaluation of commonsense reasoning in a low-resource language.

---

## Dataset Construction

- **Prompt creation:** Most of the items were created from daily activities that pertain to the Yoruba land. Aside daily activities I also created goals based on common events in the Yoruba land – like Festivals and Traditional worship. Also the dataset includes prompts that pertain to our foods and wears.
  - **Solution design:** Each prompt paired with two solutions, differing in one or two words; While the difference between the two solutions will immediately appear clear to native speakers, they are not in any way absurd (as can be seen in the examples below).
  - **Annotation process:** The dataset was created by me and reviewed by a native language speaker
  - **Size:** 100
- 

## Examples

prompt (Yoruba)	solution0	solution1	label
<i>Lààti pa ijàlọ:</i>	<i>Aó ò bu iyò lé àwọn ijàlọ náà lori.</i>	<i>Aó ò fì ìgbálẹ gbá àwọn ijàlọ náà.</i>	0
<i>Ònà wo ni a leè gbà mú kí ògì tútù pẹ nílẹ láì bàjẹ?</i>	<i>Aó o máa po omi orí ògì náà lójoojúmọ, aó sì bu tuntun síi.</i>	<i>Aó o máa yọ omi orí ògì náà lójoojúmọ, aó sì bu tuntun síi.</i>	1

---

## References

- Bisk, Yonatan, et al. 2020. “PIQA: Reasoning about Physical Commonsense in Natural Language.” *AAAI 2020*.

