MRL 2025 Shared Task: Hawaiian Language Physical Reasoning Dataset

Cade J. Kanoniakapueo Kane[1], A. Kuuipolani Kanahele Wong[1], Pua Niau-Puhipau[1], R. Kekeha Solis[1], Bruce Torres Fischer[2], Winston Wu[2], Haopeng Zhang[1]

[1] University of Hawai‘i at Mānoa [2] University of Hawai‘i at Hilo
{cadekane, annettew, rsolis, haopengz}@hawaii.edu
{bruce42, wswu}@hawaii.edu

Abstract

We present a PIQA-style (Bisk et al., 2020) dataset of 100 examples in 'ōlelo Hawai'i, the Hawaiian language, aimed at supporting natural language processing research for this critically endangered language. Each example includes a prompt, two candidate solutions, and a label, with subtle differences requiring both linguistic and cultural knowledge to resolve. Examples were created in modern orthography, including the 'okina and kahakō, and cover a range of scenarios from everyday tasks to culturally specific activities. To our knowledge, this is the first dataset of its kind for 'ōlelo Hawai'i, providing a resource that integrates cultural authenticity with computational utility. Future work will expand the dataset and increase cultural specificity through further community engagement.

1 Introduction

The goal of this project is to create a high-quality dataset of physical interaction question answering (PIQA)-style questions (Bisk et al., 2020) in 'ōlelo Hawai'i, the Hawaiian language. 'Ōlelo Hawai'i is the ancestral language of the Indigenous people of Hawai'i, yet English remains the dominant language across the islands.

As a low-resource language, Hawaiian currently lacks sufficient computational resources for natural language processing (NLP). This project aims to address that gap with a culturally authentic and linguistically accurate resource for 'ōlelo Hawai'i.

To achieve this, we collaborated closely with native speakers of 'ōlelo Hawai'i (Mānaleo) and NLP researchers, combining community expertise with technical methods to ensure both cultural integrity and computational utility.

2 Background

2.1 'Ōlelo Hawai'i

'Ōlelo Hawai'i is an endangered Polynesian language of the Austronesian family and the ancestral language of the Hawaiian Islands.

In 1896, Hawaiian was banned from Hawai'i's public schools, and English was mandated as the sole medium of instruction (con, 1898). This policy marked a pivotal moment in Hawaiian history, contributing to the decline of the language.

Beginning in the mid- to late 20th century, however, a Hawaiian language renaissance emerged. This movement led to the creation of Hawaiian-language preschools (Pūnana Leo) (aha) and immersion schools (kula kaiapuni) (haw), which remain central to revitalization efforts. Despite these initiatives, the historical suppression of 'ōlelo Hawai'i and the ongoing dominance of English have left the language critically endangered.

Because most contemporary speakers are second-language learners rather than *Mānaleo* (native speakers), Hawaiian language education emphasizes the distinction between *no'ono'o Hawai'i* (Hawaiian ways of thinking) and *no'ono'o Haole* (foreign ways of thinking). This concept is especially relevant for NLP, where data representation choices risk importing external frameworks. Preserving *no'ono'o Hawai'i* ensures that datasets and computational models reflect culturally grounded perspectives, maintaining authenticity and integrity in the development of Hawaiian language technologies.

2.2 Linguistic Characteristics

'Ōlelo Hawai'i differs significantly from English and many widely studied languages. Its alphabet contains only five vowels, seven consonants, and the 'okina (glottal stop). In addition, the kahakō (macron) is used to lengthen vowels. Hawaiian employs the articles *ka* and *ke*, which are similar to the English "the" but with context-dependent distinctions. Most nouns are preceded by either *ka* or *ke*, and in some cases, the choice of article alters meaning. Plurality is indicated by the particle *nā*, which replaces *ka* or *ke*.

The language follows a verb–subject–object (VSO) word order, foregrounding the action rather than the agent. Grammatical relationships are often marked by particles (e.g., *i*, *iā*, *ma*, *o*, *a*), whose meanings shift across contexts. For example, *i* can mean "to," "at," "on," "if," or "so that," depending on usage. Pronouns encode distinctions not only in number (singular, dual, plural) but also in clusivity (inclusive vs. exclusive) (Pukui and Elbert, 2001).

'Ōlelo Hawai'i also reflects a close relationship with the natural world. Place names across the islands encode nuanced distinctions among rivers, waterfalls, mountains, beaches, and even individual stones, many of which are tied to mo'olelo (stories) specific to those locations (Pukui et al., 2021).

2.3 NLP for Hawaiian Language

NLP for Hawaiian remains a relatively underexplored area, though several recent efforts have begun to introduce 'ōlelo Hawai'i into computational frameworks.

Shillingford and Parker Jones (2018) proposed methods for transliteration between older and newer Hawaiian orthographies. The modern writing system incorporates the 'okina (glottal stop) and kahakō (macron), while earlier writing conventions did not. Their work compared finite state transducers (FSTs) with hybrid deep learning approaches that approximate an FST using recurrent neural networks, emphasizing the lack of available data as a major limitation for model training.

Walker et al. (2025) introduced Kuene, an online collaborative dictionary platform designed to support Hawaiian neologism creation. Through this platform, the Hawaiian Lexicon Committee can propose, edit, and approve new words, which are then published to Wehewehe Wikiwiki, a public dictionary hosted by the University of Hawai'i.

Gilbert et al. (2025) developed a Universal Dependencies (UD) treebank for Hawaiian, providing syntactic annotations at the word and sentence level. This resource supports comparative linguistic analysis and enables both mono- and cross-lingual parsing experiments with Hawaiian and Cook Islands Māori, contributing to automated syntactic analysis and revitalization efforts.

Together, these works underscore the extreme scarcity of Hawaiian language data. Our project responds to this gap by creating a culturally grounded PIQA-style dataset in 'ōlelo Hawai'i, expanding the data available for NLP research and supporting ongoing language revitalization.

3 Data Collection

For this paper, we manually created 100 examples in the PIQA format. Each example consists of a prompt, two candidate solutions (*solution0* and *solution1*), and a label indicating the correct solution. The two solutions differ by only one or two key words to highlight subtle distinctions in meaning. For reference, we also provide English translations of all examples. All Hawaiian text was written in modern orthography, including both the 'okina and kahakō.

All examples were reviewed by a experts of Hawaiian language, and two native-speakers (Mānaleo).

Topic The examples cover a wide range of scenarios, including contexts specific to Hawai'i, the Hawaiian language, and Hawaiian culture, as well as everyday situations. Cultural contexts include activities such as surfing and working with native plants, while more general scenarios include common household tasks like preparing food. Table 1 summarizes the distribution of topics.

Topic	Percent
Hawaiian Culture	27%
Food	19%
Others	61%

Table 1: Distribution of topics. Hawaiian Culture and Food categories are not mutually exclusive.

Culture knowledge A subset of the examples requires cultural knowledge of Hawaiian language and practices. For instance, one example concerns the preparation of pa'i 'ai (pounded taro) and asks whether one should *ku'i* ("pound") or *pa'i* ("slap") the taro. While both actions are physically possible, only *ku'i* is the culturally and linguistically appropriate term in Hawaiian. Additional examples are shown in Table 2.

Statistics We provide here some basic descriptive statistics of the dataset in Table 3.

4 Analysis

We also conducted preliminary experiments using state-of-the-art LLMs via the Azure OpenAI platform. Of the 100 examples, 11 prompts were rejected by Azure OpenAI's content management system for being flagged as potential jailbreak attempts, and one additional prompt was rejected by the GPT-5 model due to potentially harmful content. Excluding these 12 invalid cases, the model achieved an accuracy of 90.9% on the remaining

Prompt	solution0	solution1	label
Ke ho'omākaukau 'ia ka pa'i 'ai, *[When pa'i 'ai is prepared,]*	**ku'i** 'ia ke kalo mo'a. *[cooked taro is **pounded**.]*	**paʻi** 'ia ke kalo mo'a. *[cooked taro is **slapped**.]*	0
E pā'ina ana 'oe ma ka pikiniki. I 'ole e pili ka mau'u i ka hāli'i, *[You are going on a picnic. So that the grass does not stick to the mat,]*	e hāli'i 'ia ke **kapa moe huluhulu** e noho ai. *[a **fuzzy blanket** should be laid down.]*	e hāli'i 'ia ka **moena lauhala** e noho ai. *[a **lauhala mat** should be laid down.]*	1
Ke ku'i 'opihi, *[When picking 'opihi,]*	hiki ke ho'ohana 'ia ka **pahi**. *[a **knife** can be used.]*	hiki ke ho'ohana 'ia ka **manamana lima**. *[a **finger** can be used.]*	0

Table 2: Examples requiring Hawaiian cultural knowledge to answer correctly. In the first prompt, ku'i, to pound, is the correct word used for making pa'i 'ai. In the second prompt, knowledge about what a moena lauhala is (lauhala mat) and its texture is essential to know if it is better than a blanket. In example 3, knowledge about what ku'i 'opihi (gathering 'opihi) is and how it is done is necessary.

Metric	Average	Std Dev
Example word count	31.58	17.27
Prompt word count	9.55	3.69
Solution word count	22.03	17.35

Table 3: Word count statistics for examples, prompts, and solutions. An example consists of a given prompt and one solution.

88 valid examples. The results are summarized in Table 4.

Metric	Value
Total examples	100
Skipped examples	12
Evaluated examples	88
Wrong answers	8
Accuracy	90.90%

Table 4: Results of GPT-5 on PIQA-style examples in 'ōlelo Hawai'i.

5 Conclusions

We have presented a PIQA-style dataset of 100 examples in 'ōlelo Hawai'i, representing an initial step toward integrating the Hawaiian language into AI and natural language processing research. To our knowledge, this is the first QA dataset for 'ōlelo Hawai'i, providing a foundation for both linguistic and culturally grounded computational work.

Future work includes expanding the dataset and involving more members of the Hawaiian community, whose knowledge and perspectives are essential for creating PIQA-style examples that are both linguistically accurate and culturally authentic.

References

Hawaiian immersion school (kaiapuni education). https://hawaiipublicschools.org/enrolling-in-school/hawaiian-immersion-school/.

History — 'aha pūnana leo. https://www.ahapunanaleo.org/history-hl-1.

1898. *Congressional Serial Set, Issue 3727*. U.S. Government Printing Office. Accessed via Google Books: https://books.google.com/books?id=VSxHAQAAIAAJ&pg=RA1-PA23.

Yonatan Bisk, Rowan Zellers, Jianfeng Gao, Yejin Choi, and 1 others. 2020. Piqa: Reasoning about physical commonsense in natural language. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 7432–7439.

Gabriel H Gilbert, Rolando Coto-Solanu, Sally Akevai Nicholas, Lauren Houchens, Sabrina Barton, and Trinity Pryor. 2025. Crosslingual dependency parsing of hawaiian and cook islands māori using universal dependencies. In *Proceedings of the 18th International Conference on Parsing Technologies (IWPT, SyntaxFest 2025)*, pages 40–50.

Mary Kawena Pukui and Samuel H. Elbert. 2001. *Hawaiian Grammar*. University of Hawai'i Press.

Mary Kawena Pukui, Samuel H Elbert, and Esther T Mookini. 2021. *Place names of Hawaii: Revised and expanded edition*. University of Hawaii Press.

Brendan Shillingford and Oiwi Parker Jones. 2018. Recovering missing characters in old Hawaiian writing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4929–4934, Brussels, Belgium. Association for Computational Linguistics.

Sunny Walker, Winston Wu, Bruce Torres Fischer, and Larry Kimura. 2025. Kuene: A web platform for facilitating hawaiian word neologism. In *Eight Workshop on the Use of Computational Methods in the Study of Endangered Languages*, page 182.