

Hindi PIQA Dataset Construction

Amol Khanna

Booz Allen Hamilton

Boston, MA

amol.khanna00@gmail.com

1 Dataset Construction Description

The construction of this Hindi dataset was inspired by the format and objectives of the original English PIQA benchmark, focusing on physical common-sense reasoning. To ensure the novelty of the content, the original PIQA examples were consulted only for structural guidance and were not reused or translated. The dataset is strategically composed to test two tiers of knowledge. Approximately 75% of the examples cover general physical common-sense that would be universally understood, while the remaining 25% are dedicated to testing reasoning about culturally-specific phenomena from the Indian context. These culturally-grounded examples include references to specific Indian culinary items (e.g., dosa, dal, roti), musical instruments (dhol), common fauna (monkeys, elephants), and social traditions, such as customs within a wedding ceremony.

Our methodology for authoring the dataset followed a three-stage process designed for efficiency and accuracy. First, prompts and solutions were conceptualized and written in English. Second, these English examples were translated into Hindi using Google Translate. This approach was chosen to streamline the authoring process and mitigate the complexities and time investment associated with direct Hindi typesetting. The third and most critical stage involved a comprehensive manual review where every machine-translated output was meticulously checked, corrected, and refined by a Hindi speaker to ensure factual accuracy, grammatical correctness, and natural fluency. In line with the PIQA framework, the two candidate solutions were designed to be lexically similar while containing a key semantic difference. This "minimal pair" construction creates challenging examples that require a nuanced understanding of either physical properties or specific cultural knowledge to solve correctly.