Dataset Creation Methodology

We developed a PIQA-style commonsense reasoning dataset in Nepali following a systematic approach:

Step 1: Topic Selection We compiled an extensive list of practical, everyday activities spanning household tasks, personal care, outdoor activities, crafts, sports, and recreational pursuits. These topics were selected to ensure broad coverage of common-sense scenarios that require practical reasoning.

Step 2: Initial Question Generation For each topic, we created PIQA-style questions in Nepali, initially focusing on crafting high-quality questions with correct answers. Each question presented a practical problem requiring commonsense reasoning to solve. We also took little help from AI, using ChatGPT, Gemini and Claude chat interfaces but not APIs.

Step 3: Quality Curation We conducted careful manual review of all generated examples, selecting 2-5 of the highest quality questions per topic based on clarity, practical relevance, and reasoning requirements.

Step 4: Distractor Creation For each selected question, we manually crafted plausible but incorrect answer choices, ensuring they were realistic enough to serve as effective distractors while maintaining clear distinction from the correct answers.

Step 5: Dataset Expansion and Final Review We expanded the dataset by approximately 200 manually curated data points across all topics, followed by a comprehensive final review to ensure consistency, quality, and linguistic accuracy throughout the entire dataset.