

# SCCT Annotation Guidelines

Hello, dear annotator, and thank you for your time!

Here is a brief background for the task and the specific annotation guidelines. This task is a part of the DharmaBench paper and evaluation benchmark for Sanskrit and Tibetan.

# Introduction

## Task: Classifying canonical literature (SCCT)

**Essence & Relevancy:** This task classifies Tibetan canonical literature (allochthonous) into two primary categories: “scriptures” and “non-scriptures.” Its primary objectives are to (a) evaluate LLM performance in canonical literature classification into these two broad categories, and (b) illuminate the evolutionary processes of works that are considered scriptures by the tradition (allochthonous texts) as well as para-canonical works through (works whose provenance is debated) linguistic analysis. The task aims at identifying distinctive features characteristic of each category, including compositional styles, syntax, vocabulary, and other linguistic markers. In particular, such an analysis should help reveal recurring stylistic patterns and compositional strategies that a scripture creator possibly employed to achieve two critical goals: canonical acceptance and popular success.

**Real-world relevancy:** Its broader implication is that by mapping these linguistic patterns, the task provides insights into how textual authority was constructed and maintained within Indian and Tibetan literary traditions.

## Example

## Annotation guidelines

1. Read the sample of the text carefully.
  2. If the sample contains any material that is not part of the original text, such as table of contents, editorial notes, translator's remarks and so forth, reject the sample.
  3. If the sample is too short (i.e., shorter than two full sentences), reject the sample.
  4. If there are any conversion errors in the Tibetan (due to erroneous input in the original material, for example), if the error is minor, correct it. If there are many conversion errors that cannot be resolved easily, reject the sample.

## Edge cases and common mistakes

1. In the case of ambiguity, please consult the team and the supervisors.
    - a. If you can't reach an agreement, discard the sample.

# Something is off?

Reach out to us.

Thank you very much!