

# THCT 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:** To classify Tibetan canonical literature (allochthonous) by genre and theme as represented in the Tibetan Canon's subsections.

**Goal:** Relevant for evaluating LLM performance in identifying Buddhist-specific genres and topics, and seeks to determine the overall thematic content of complete works or individual sections.

**Background:** This task aims to classify Tibetan canonical literature (allochthonous) by genre and theme as represented in the Tibetan Canon's subsections. The classification is relevant for evaluating LLM performance in identifying Buddhist-specific genres and topics, and seeks to determine the overall thematic content of complete works or individual sections. Additionally, this task could ultimately help address content-based questions by examining the coherence (or incoherence) of philosophical ideas within texts as potential indicators of the religio-philosophical school affiliation of their creators and the texts' evolution and intellectual context.

## Example

- Text: "मृद्गायुक्तं विद्यायाम् अनुवादं विद्यायाम्" - Class: Treatises on Sanskrit
  - Text: "प्रत्येकं विद्यायाम् अनुवादं विद्यायाम्" - Class: Treatises on Abhidharma
  - Text: "प्रत्येकं विद्यायाम् अनुवादं विद्यायाम्" - Class: Epistles

## Data collection guidelines

1. Go over texts from the defined sources (rKTS and ACIP).
  2. Create separate folders for each section of the Tibetan canons and add materials.
  3. Share the folder containing these materials with the responsible RUNI data scientist.

# Annotation guidelines

## Annotation platform

We will be using the [Label Studio](#) app. We will upload the collected CSV files and set up the annotation project.

## Preparation

1. Log in to our Label Studio organization here.
2. Let us know, and we will approve your user and assign you to the respective project.

## Annotation guidelines

1. Read the sentence/paragraph thoroughly.
2. Ensure the label remains intact - it is pre-annotated with the source's metadata.
3. Annotation Speed vs. Accuracy Tradeoff: Emphasize **accuracy over speed** if necessary.

## Something is Off?

Reach out to us!

If you encounter technical problems with the folder setup, missing texts, or conceptual ambiguities, please contact the project leads.

Thank you very much for your contribution!