Writing Assistance Filtering Annotations

Annotation sheet: [Realistic Value Evals] Writing Aid Annotations
500 prompts to annotate, sampled randomly from our ~32k relevant real user prompts.

Goal of this annotation project is to create a gold standard of which prompts are about writing assistance, so we can evaluate different automated filtering methods.

The **annotation task** covers one binary label: **annotX_label**.
1. "**1 - writing aid**" denotes any prompt that asks or instructs the model to provide writing assistance to the user. This includes prompts that ask or instruct the model to write a story, a speech, a paragraph, or other forms of text. It does NOT include prompts about paraphrasing, rewriting, summarising, describing, responding to, or translating text.
2. "**0 - NOT writing aid**" denotes any other prompt.

The **note** column is for any additional comments / observations.

It is usually enough to skim the prompts to assign labels.