



# 👁 Pangolin Vision 👁 – Multimodal Safety Task Review Workflow

## Prompt Review

📋 **Step 1:** Follow Prompt Review workflow

Follow instructions [here](#) for reviewing, editing, or wiping the opening prompt.

## Response Review

📋 **Step 2:** Read over the rewritten response carefully

📋 **Step 3:** Identify any errors in the response and leave feedback for the previous attempter

We recommend that you follow this order to identify errors:

- Incorrect response rewrite strategy, e.g.
  - Response should have (partially) engaged but declined
  - Full engagement to a benign prompt doesn't include a disclaimer
- Unbalanced, opinionated, or preachy language
- Truthfulness issues – you can use perplexity.ai to check for this

#### **Step 4: Fix all issues in the response**

Common fixes:

- More engagement required – you can reroll the response to find a more helpful base response to rewrite
- Remove any unbalanced, opinionated, or preachy language

#### **Step 5: Clean up the response labeling**

Keep your response and sentence labels organized.

- Label at most 1 engage, 1 disclaimer, 1 decline, 1 short decline.
- Avoid labeling a single sentence with multiple labels.

Use this time to review your response strategies and any spelling mistakes.

#### **Step 6: Leave final feedback for the previous attempter**

Take 1-3 minutes to write some helpful feedback for the previous attempter. You can explain errors you identified and changes you made before submitting.

#### **Step 7: Submit the task after a thorough re-read**