**Content:** Summary of participants comments on their story version preference sorted by model.

**English**

**Preference for Baseline**

The comments indicate a preference for first-person narration to enhance engagement. Additionally, there's a suggestion to use gender-neutral pronouns like "you/your" when the subject's gender is ambiguous, implying a desire for inclusivity and avoiding assumptions.

**Preference for Gemma 3**

The feedback highlights a tension between sticking closely to the images and allowing for imagination and emotional depth.

A recurring criticism to other models is the introduction of invented characters or details not directly supported by the images, which some find undesirable. The preference is for stories that remain faithful to the visuals.

However, some commenters also value stories that offer a more balanced perspective, open up the reader's imagination, and add emotion beyond a simple description of the images. Stories that explore themes like family and are relevant to the pictures, while remaining simple and clear, are well-received. The ability to evoke sadness in line with the visuals and leave the reader with a sense of wonder is also seen as a positive attribute. The use of first-person narration to create relatability and the effective use of language to convey the patient's emotions are specifically praised.

**Preference for Gemini 2.0 Flash**

The feedback on the stories highlights several key aspects. Many commenters value strong cohesion between the text and images, with some stories praised for accurately depicting the events and characters, leading to their selection as the preferred option. Accurately identifying characters and subtle details is also appreciated. Conversely, stories are criticized for adding context not present in the images, creating internal contradictions, or lacking cohesion between the images, negatively impacting their reception. While some readers seek emotional engagement, others find the stories lacking in this area. Overall, the ideal story effectively complements the images, remains accurate, incorporates interesting details with clear language, and creates a compelling narrative, leading to its selection.

**Spanish**

**Preference for Baseline**

The comments praise the conciseness and contextual summarization of the content, as well as the improved coherence between the text and accompanying images. In other words, it likes that the content is well-written, doesn't have unnecessary details, and that the images match the writing well.

**Preference for Gemma 3**

The comments prioritize simplicity, logic, and coherence between the text and the images. There's a general appreciation for narratives that are easy to understand and tell a straightforward story. A key element is the connection between the text and the visuals, with some stories being favored for their relevance to the images and their ability to create a contextual narrative based on clear observation. One story is specifically praised for its logical depiction of the lizard's habitat and the relationship between the visuals and text, while some criticism arises for other models when elements, such as characters or events (e.g., who made the soup), feel disconnected or unjustified by the images. Better narration and a closer relationship between the image and text are desired.

**Preference for Gemini 2.0 Flash**

The feedback on these stories reveals a preference for narratives that are engaging, expressive, and relevant to the visuals.

Several respondents favor stories that exhibit depth of expression without being overly literal, evoking a sense of excitement or anticipation, and featuring a captivating storyline that encourages continued reading. These stories often have a touch of warmth in the text and effectively depict key elements like the movement of creatures.

Some stories are appreciated for aligning well with the images and adding a touch that animates the reader, making the narrative message more apparent. Other stories are favored for how they weave the narrative, highlighting relevant details and emphasizing themes like environmental awareness and exploration.

Overall, positive aspects include concise and direct text without unnecessary content, and a sense of logical storytelling. A key theme is the importance of the text feeling relevant to the images, though some note potential disconnects between certain elements within the narratives and the visuals in other models.