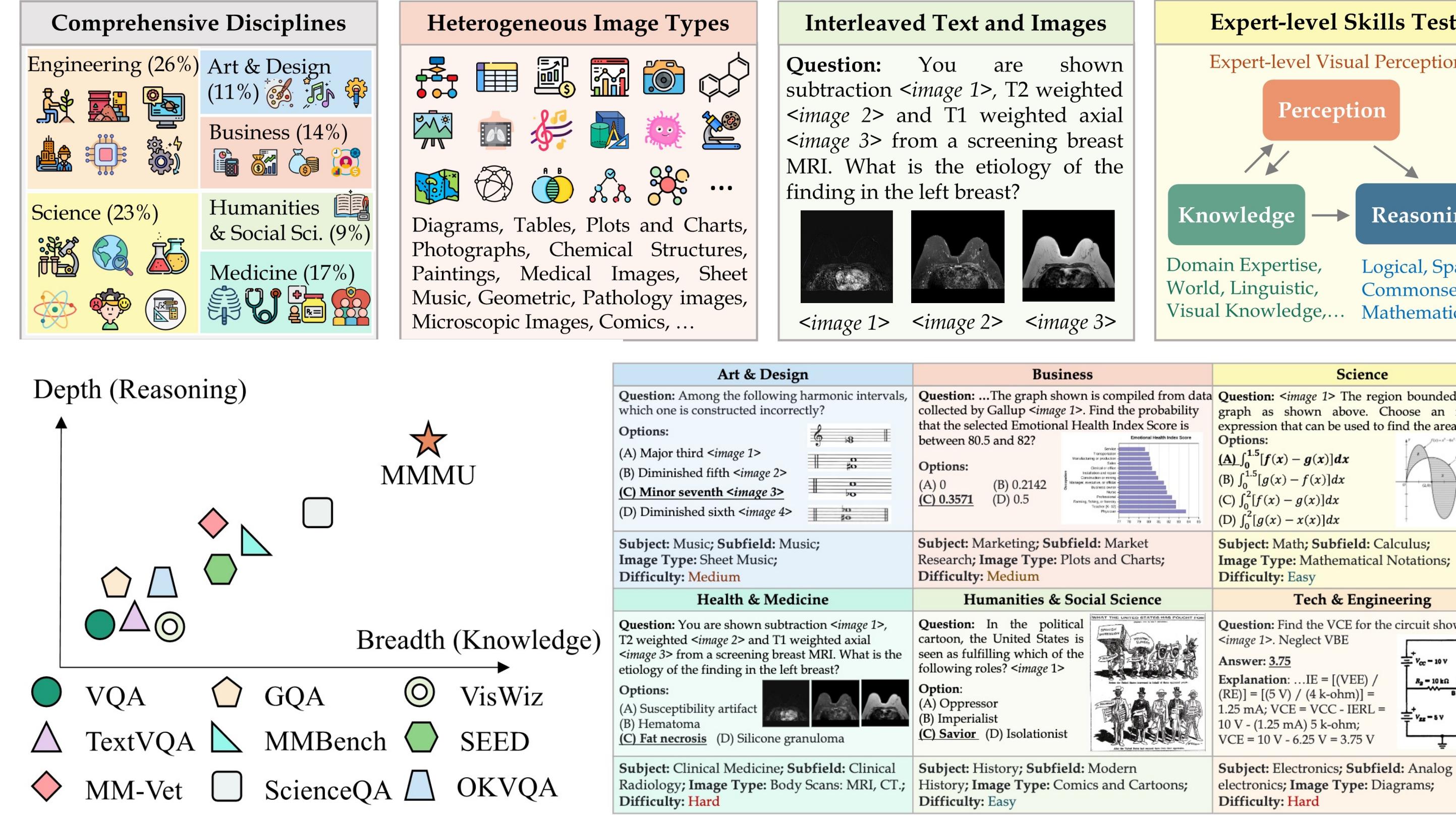
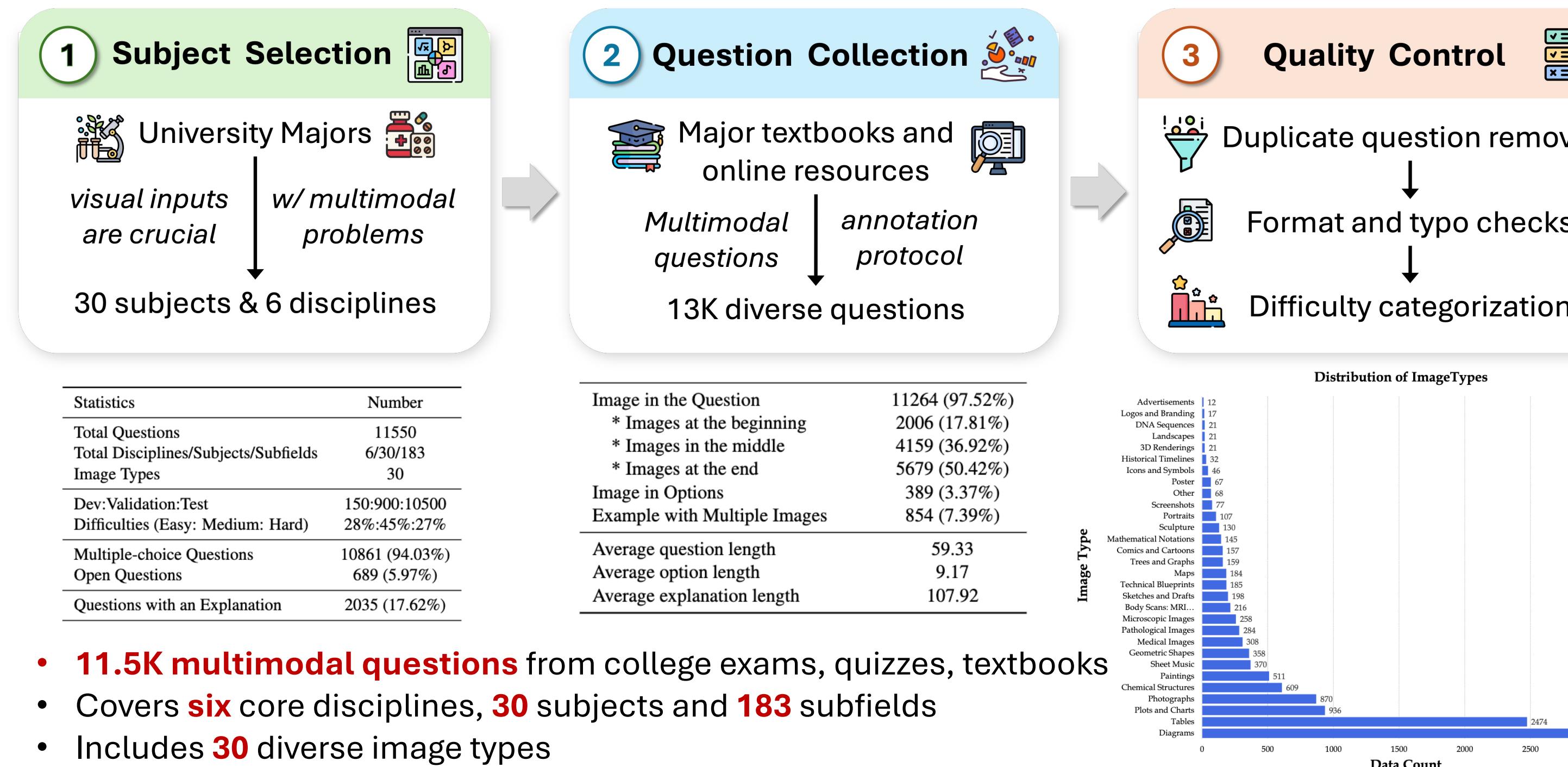


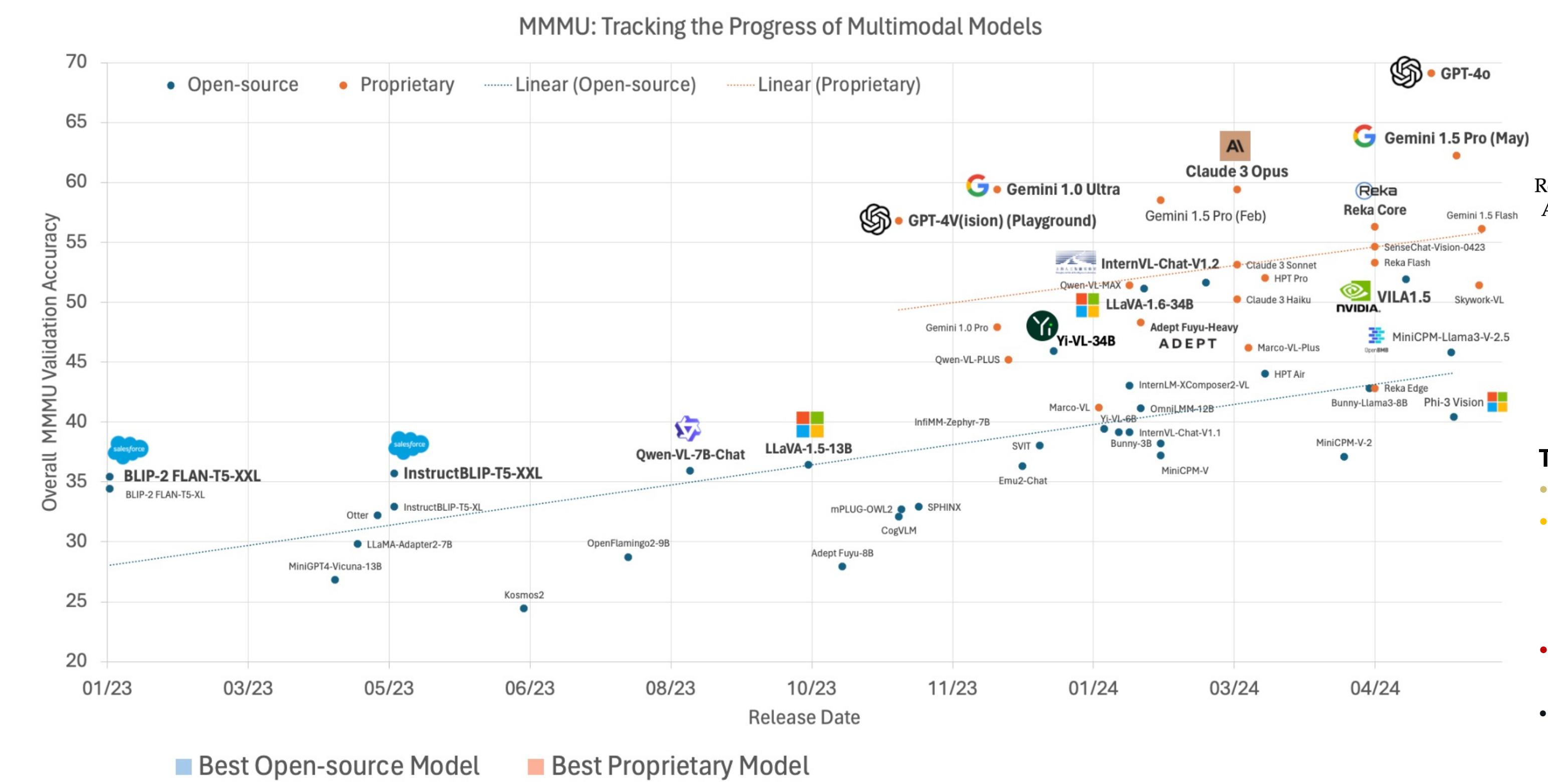
# Overview of THE Dataset



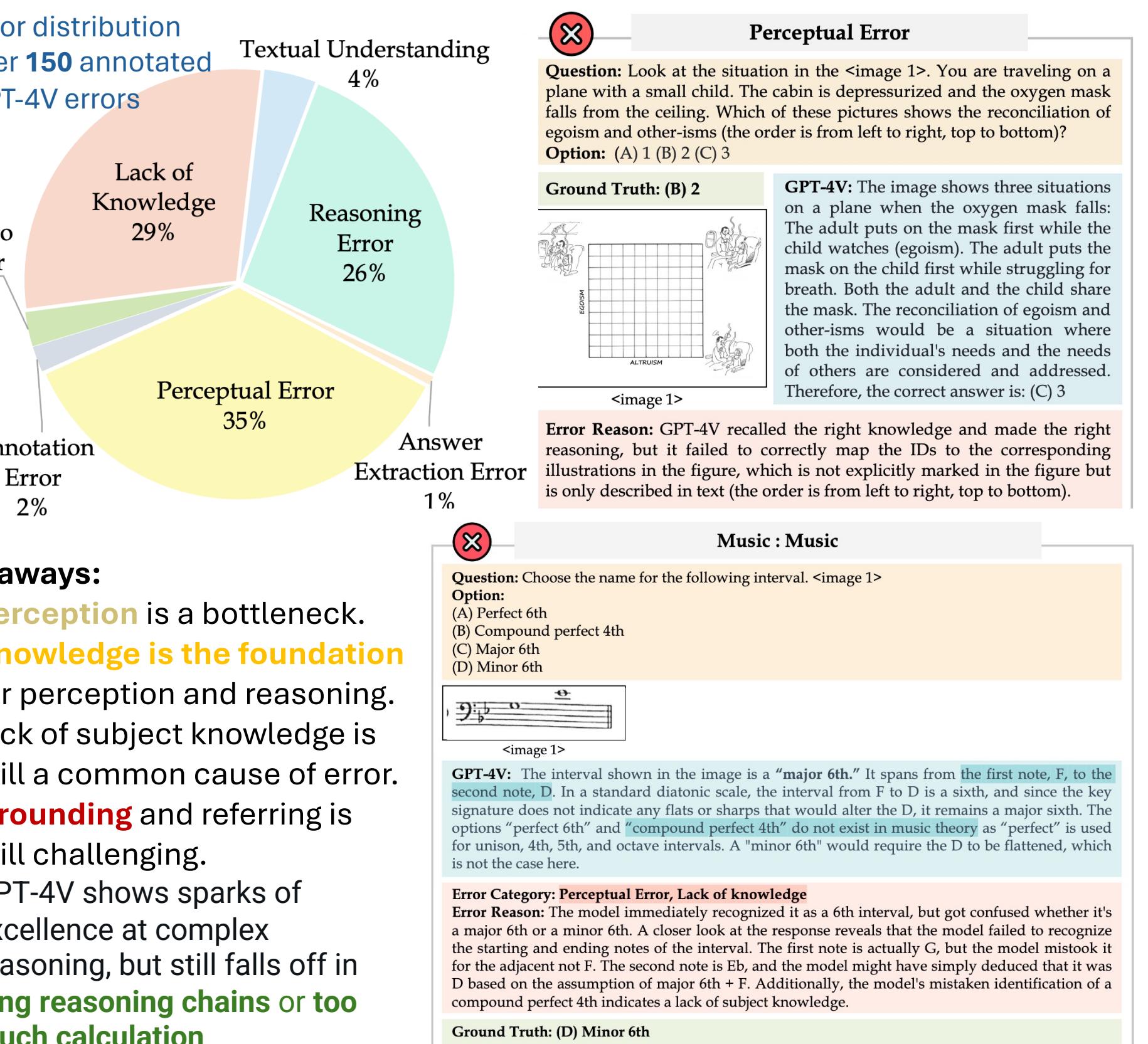
# Data Curation



# Tracking the Progress of Multimodal Model Development



# Error Analysis and Case Study



# How to Improve MMMU Performance?

- Better performance with **stronger and larger pre-trained LLMs** due to the broad and complex nature of MMMU questions.

### Improved Image Encoders:

  - Resolution and number of image tokens are more critical than image encoder size.
  - **Higher resolution** usually leads to better performance but continually increasing (>500 pixels) shows diminishing returns
  - A large number of image tokens leads to higher costs but may not necessarily result in better performance.

### High-Quality, Diverse Training Data:

  - Prioritize **quality and diversity** over quantity.
  - **Scaling up** still helps, especially with knowledge intensive data and intermediate rationales (CoT).

### Tool Integration:

  - Under-explored but promising; tools like **Python interpreters** may enhance the reasoning capability of multimodal models.