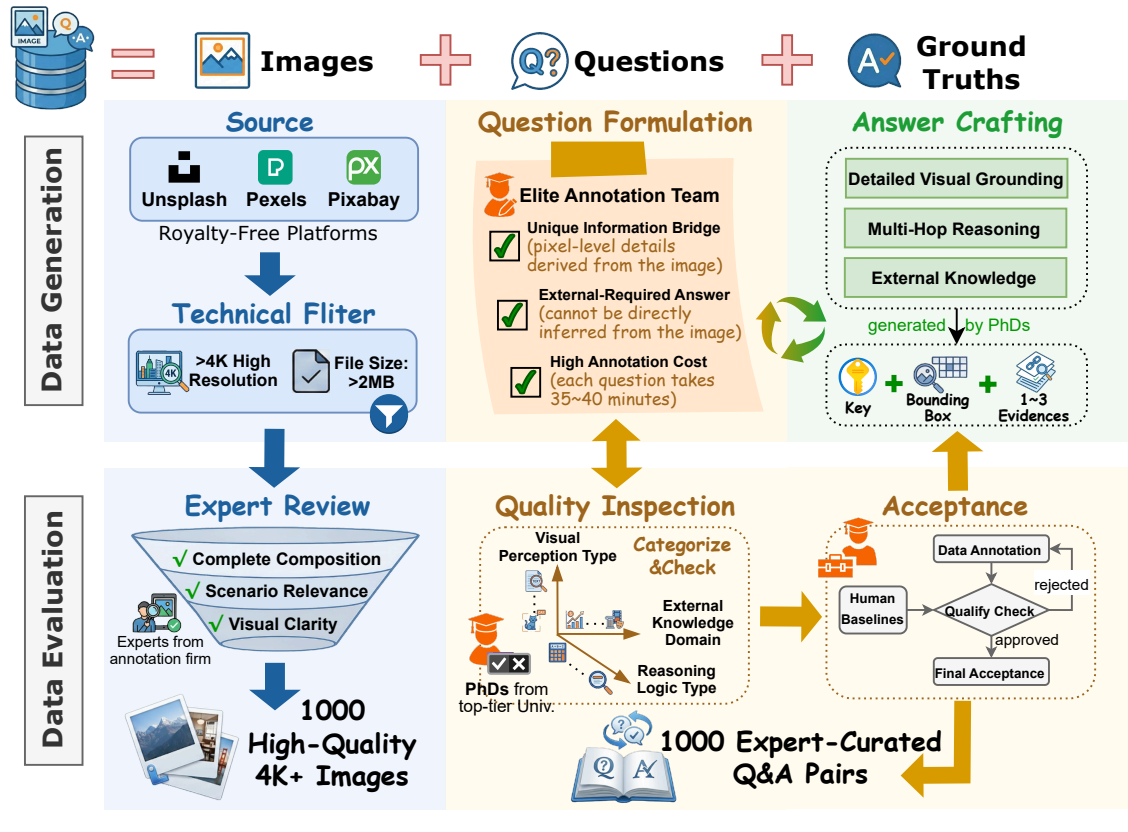
Images
Questions
Ground Truths
Data Generation
Source
Unsplash
Pexels
Pixabay
Royalty-Free Platforms
Technical Fliter
>4K High Resolution
File Size: >2MB
Question Formulation
Elite Annotation Team
Unique Information Bridge (pixel-level details derived from the image)
External-Required Answer (cannot be directly inferred from the image)
High Annotation Cost (each question takes 35~40 minutes)
Answer Crafting
Detailed Visual Grounding
Multi-Hop Reasoning
External Knowledge
generated by PhDs
Key
Bounding Box
1~3 Evidences
Data Evaluation
Expert Review
Complete Composition
Scenario Relevance
Visual Clarity
Experts from annotation firm
1000 High-Quality 4K+ Images
Quality Inspection
Visual Perception Type
Categorize &Check
External Knowledge Domain
Reasoning Logic Type
PhDs from top-tier Univ.
Acceptance
Data Annotation
Human Baselines
Qualify Check
rejected
approved
Final Acceptance
1000 Expert-Curated Q&A Pairs
Data Example
Q: What is the ISNI for the text on the object right of the woman-cover book?
A: ISNI-0000000121363255
WOMAN
RIGHT
Spatial Relationship
Direct Lookup
Entity Recognition
OCR & External Knowledge