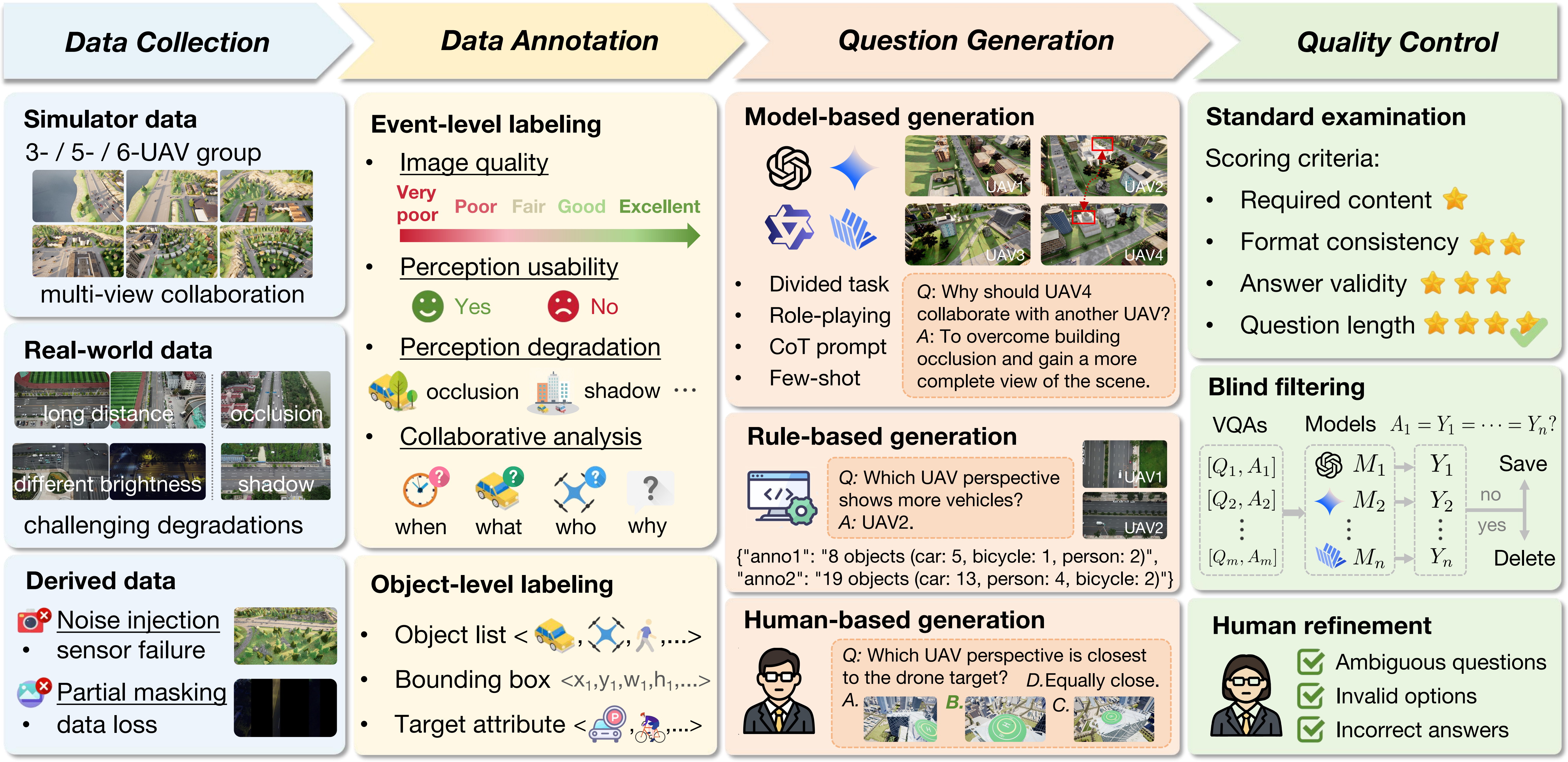

Data Collection

Simulator data

3- / 5- / 6-UAV group

multi-view collaboration

Real-world data

challenging degradations

Derived data

Noise injection
- sensor failure

Partial masking
- data loss

Data Annotation

Event-level labeling

- Image quality

- Perception usability

- Perception degradation

occlusion shadow ···

- Collaborative analysis

Object-level labeling

- Object list < , , ,...>
- Bounding box $<x_1, y_1, w_1, h_1, ...>$
- Target attribute < , ,...>

Question Generation

Model-based generation

- Divided task
- Role-playing
- CoT prompt
- Few-shot

Q: Why should UAV4 collaborate with another UAV? *A*: To overcome building occlusion and gain a more complete view of the scene.

Rule-based generation

Q: Which UAV perspective shows more vehicles? *A:* UAV2.

{"anno1": "8 objects (car: 5, bicycle: 1, person: 2)", "anno2": "19 objects (car: 13, person: 4, bicycle: 2)"}

Human-based generation

Q: Which UAV perspective is closest to the drone target? *D.* Equally close.
A. **B.** C.

Quality Control

Standard examination

Scoring criteria:

- Required content ★
- Format consistency ★★
- Answer validity ★★★
- Question length ★★★★

Blind filtering

VQAs Models $A_1 = Y_1 = \cdots = Y_n?$

$[Q_1, A_1]$ M_1 Y_1 Save
$[Q_2, A_2]$ M_2 Y_2 no
$\vdots$ $\vdots$ $\vdots$ yes
$[Q_m, A_m]$ M_n Y_n Delete

Human refinement

☑ Ambiguous questions
☑ Invalid options
☑ Incorrect answers