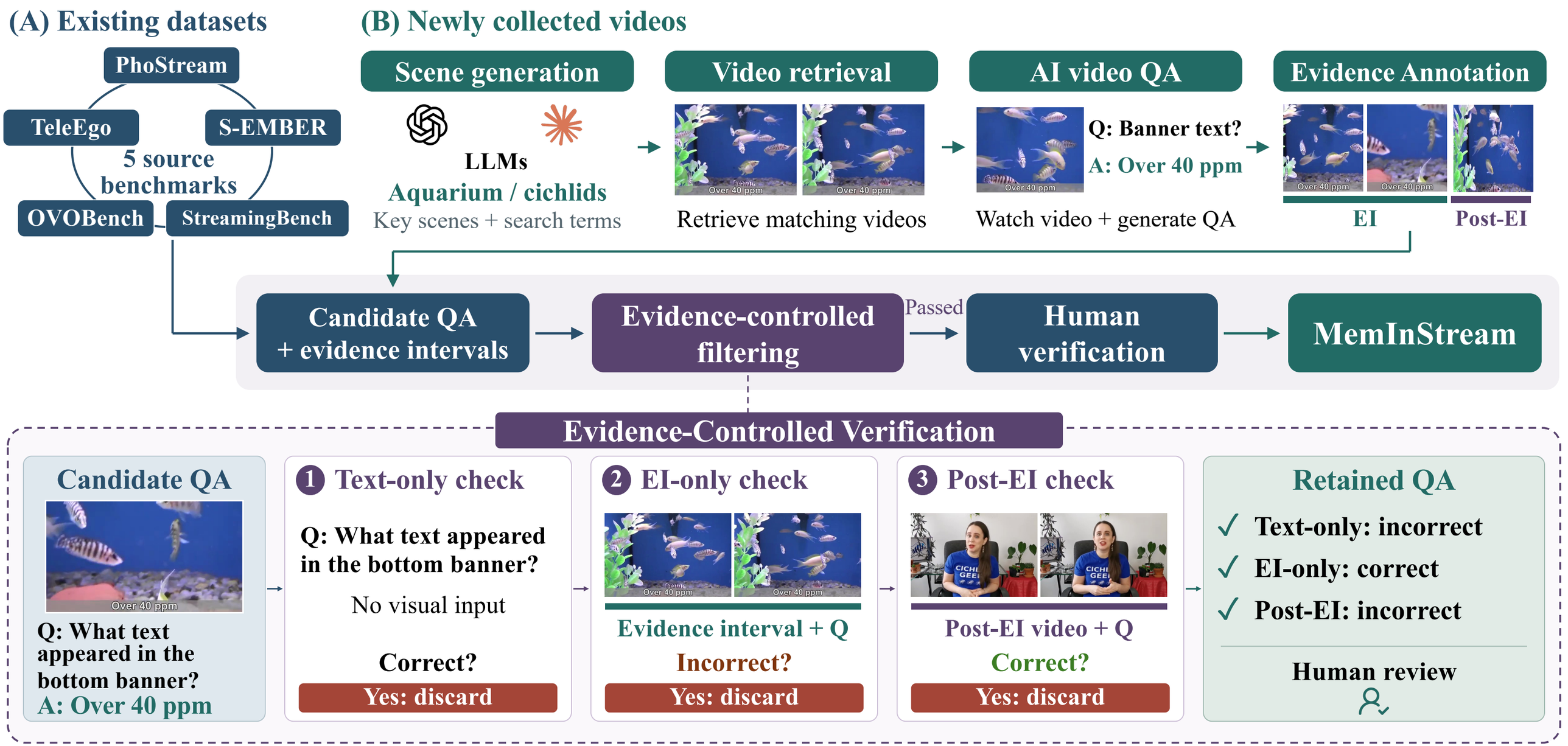

(A) Existing datasets
PhoStream
TeleEgo
S-EMBER
5 source benchmarks
OVOBench
StreamingBench
(B) Newly collected videos
Scene generation
LLMs
Aquarium / cichlids
Key scenes + search terms
Video retrieval
Over 40 ppm
Retrieve matching videos
AI video QA
Q: Banner text?
A: Over 40 ppm
Watch video + generate QA
Evidence Annotation
EI
Post-EI
Candidate QA + evidence intervals
Evidence-controlled filtering
Passed
Human verification
MemInStream
Evidence-Controlled Verification
Candidate QA
Q: What text appeared in the bottom banner?
A: Over 40 ppm
1 Text-only check
Q: What text appeared in the bottom banner?
No visual input
Correct?
Yes: discard
2 EI-only check
Evidence interval + Q
Incorrect?
Yes: discard
3 Post-EI check
Post-EI video + Q
Correct?
Yes: discard
Retained QA
Text-only: incorrect
EI-only: correct
Post-EI: incorrect
Human review