

1. Benchmark Asset Construction

Topic-Diverse



Raw Human
Slides (30k+)



Filter &
Exact Text



GPT-4o
Annotation



Wikipedia Content
Compilation

Purpose-Diverse



Real-world Scenarios
(6 domains)



Functional
Task collection



Standardized Source
Documents &
Instructions (N=189)



Gold
Quiz Bank

Format-Agnostic
Unification



PPTX



HTML



Image

2. Slide Generation System Evaluation

Branch A (Content Fidelity)



Extraction-Answer
Two-Stage
"Open-Book" Exam

Branch B (Computational Aesthetics)



Harmony



Engagement

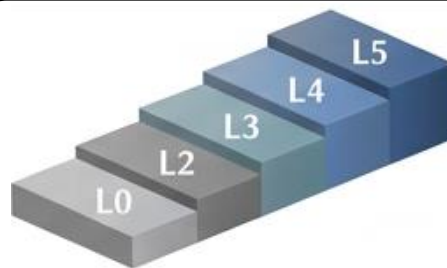
'Aa

Usability



Visual Rhythm

Branch C (Editability Intelligence)



L0 Static

L1 Patchwork

L2 Vector

L3 Structural

L4 Parametric

L5 Cinematic

Final
Report &
Metrics

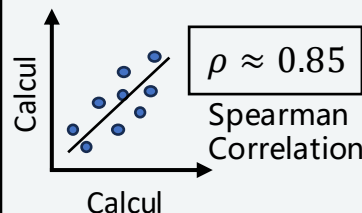


&
V-E
Metrics

3. Reliability Validation



Human
Preferences



Between automated
Metrics and human
preferences