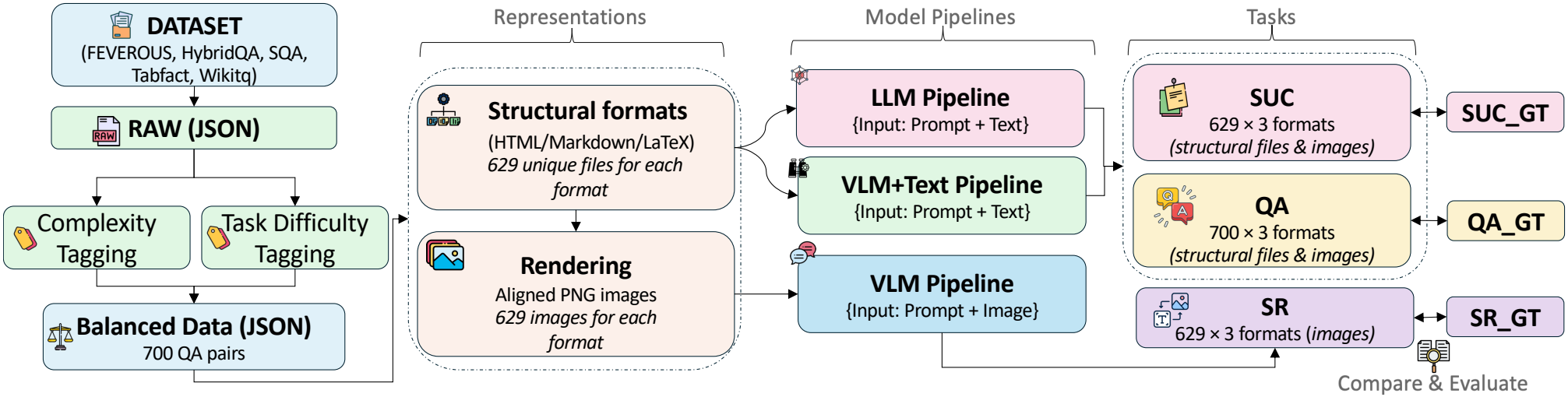
DATASET
(FEVEROUS, HybridQA, SQA, Tabfact, Wikitq)
RAW (JSON)
Complexity Tagging
Task Difficulty Tagging
Balanced Data (JSON)
700 QA pairs
Representations
Structural formats
(HTML/Markdown/LaTeX)
629 unique files for each format
Rendering
Aligned PNG images
629 images for each format
Model Pipelines
LLM Pipeline
{Input: Prompt + Text}
VLM+Text Pipeline
{Input: Prompt + Text}
VLM Pipeline
{Input: Prompt + Image}
Tasks
SUC
629 × 3 formats
(structural files & images)
QA
700 × 3 formats
(structural files & images)
SR
629 × 3 formats (images)
SUC_GT
QA_GT
SR_GT
Compare & Evaluate