



SlideAgent

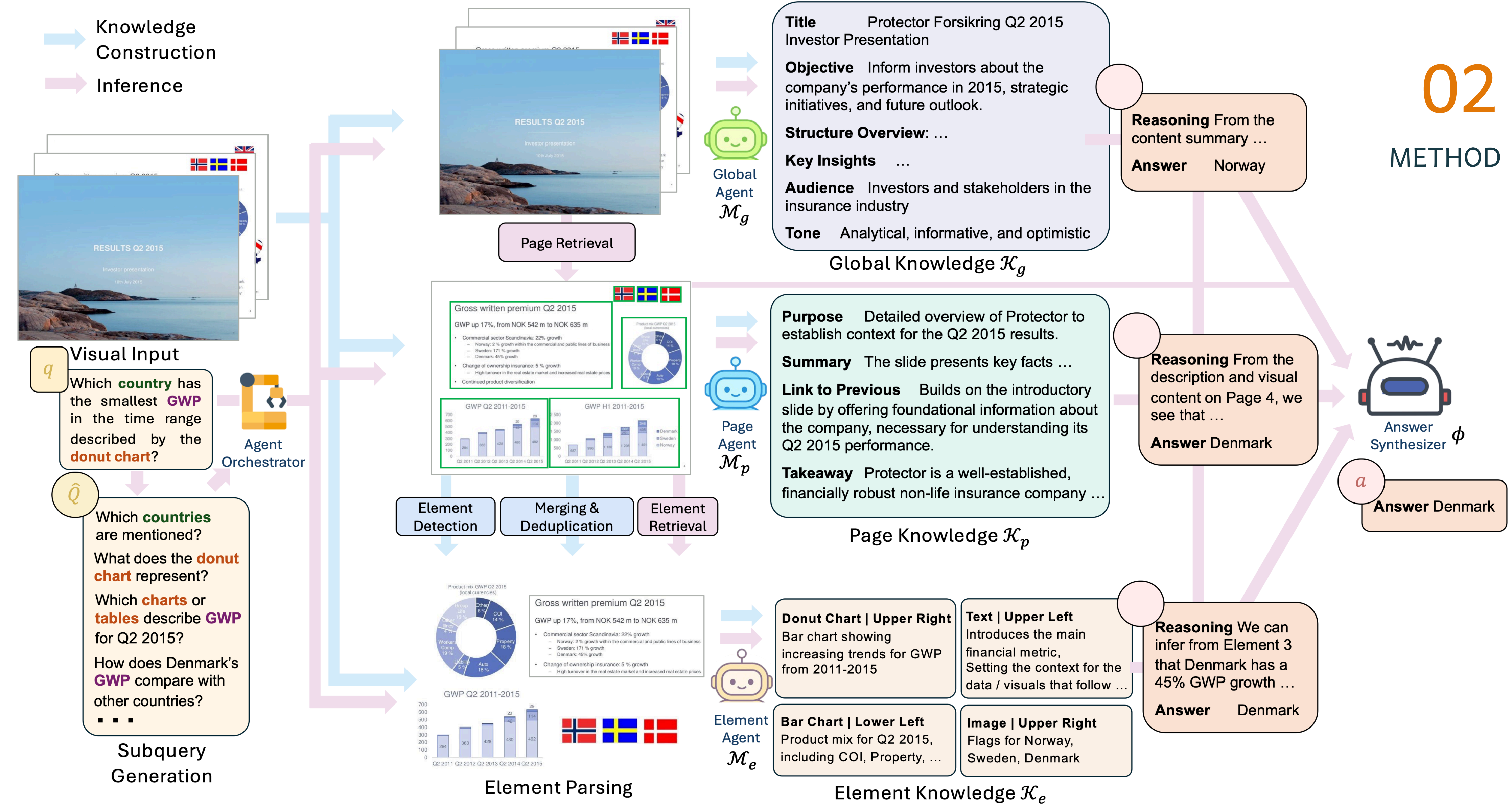
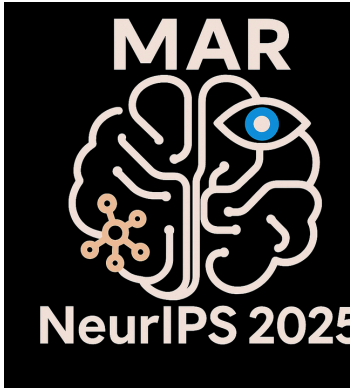


Hierarchical Agentic Framework for Multi-Page Slide Deck Understanding

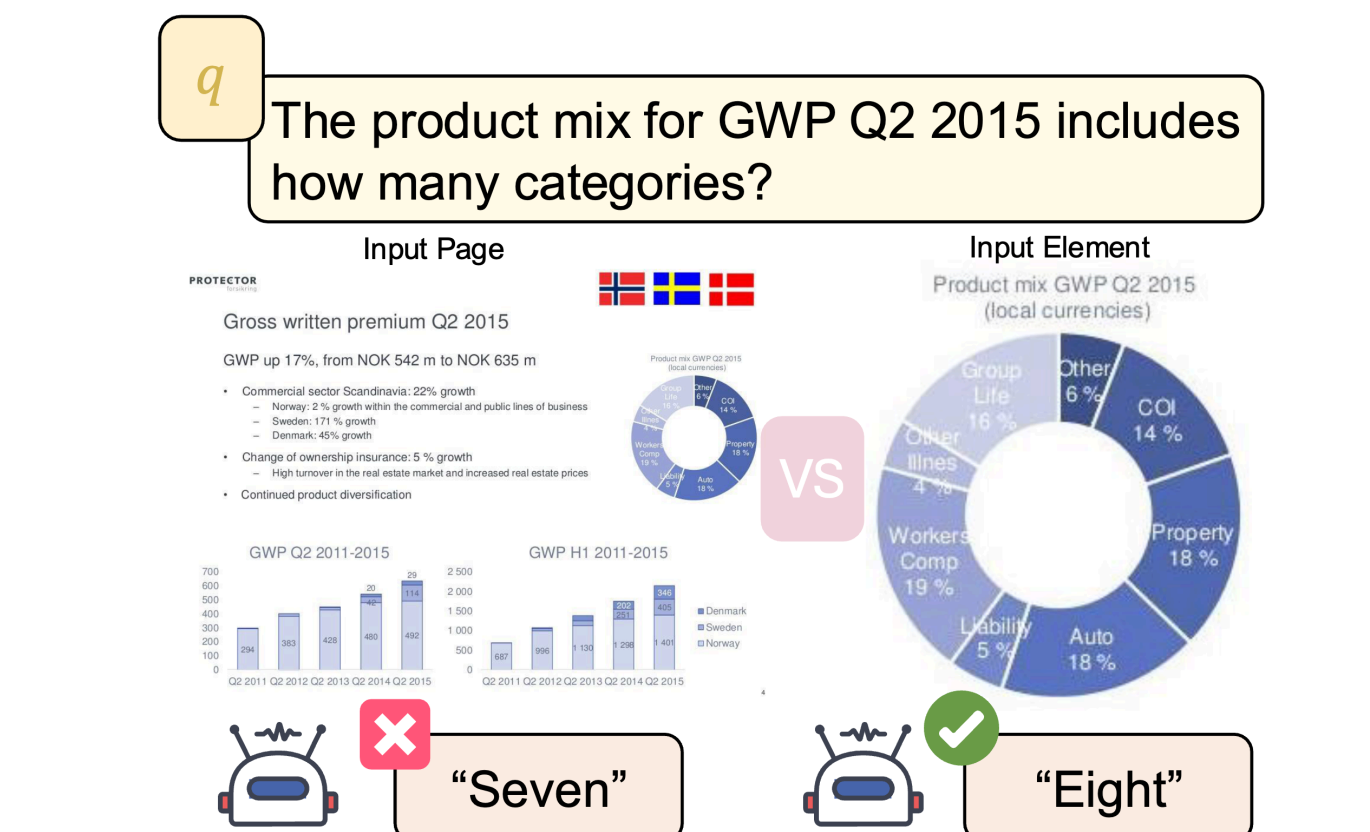
Yiqiao Jin¹, Rachneet Kaur², Zhen Zeng²,
Sumitra Ganesh², Srijan Kumar¹

¹Georgia Institute of Technology,
²J.P. Morgan AI Research

J.P.Morgan



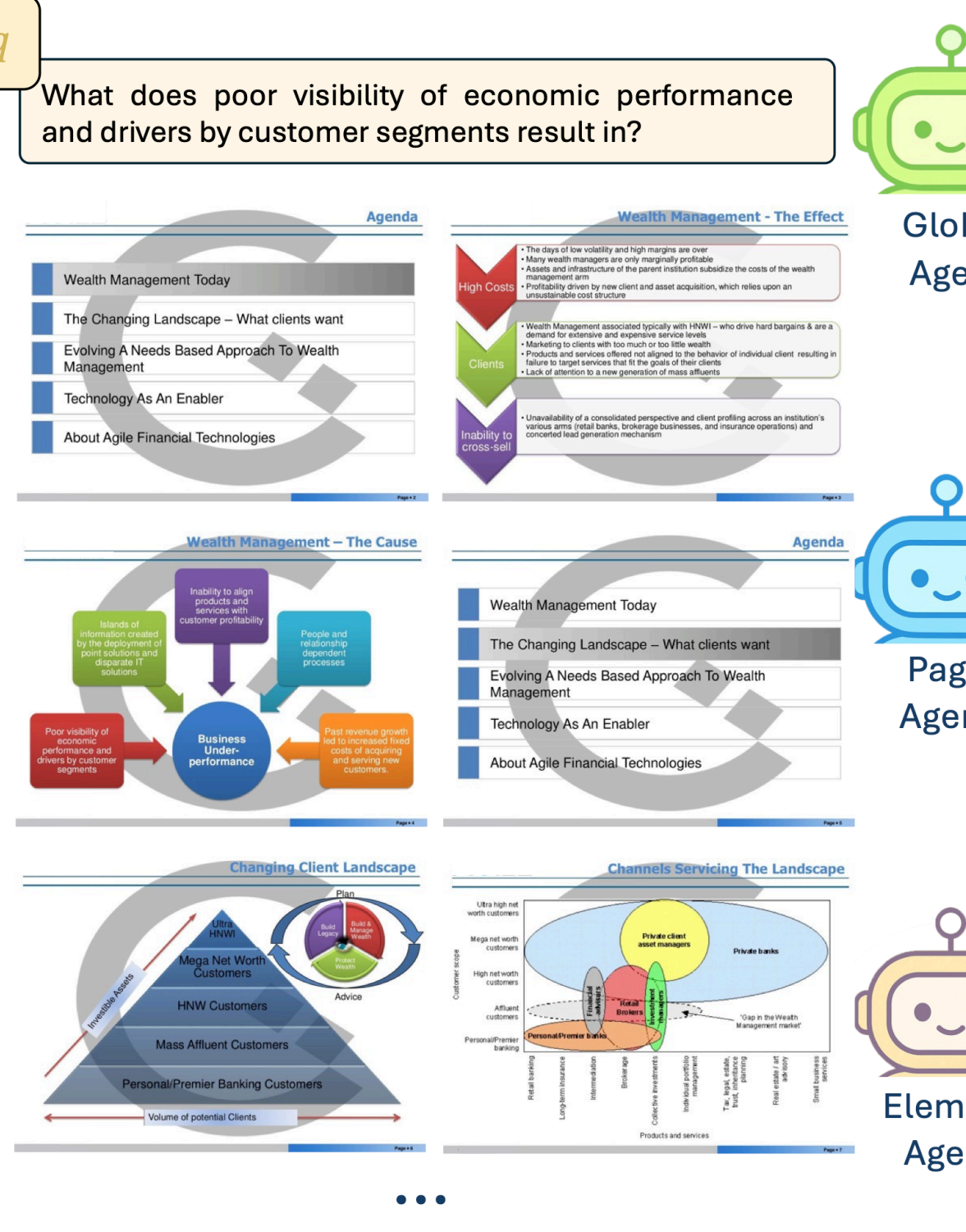
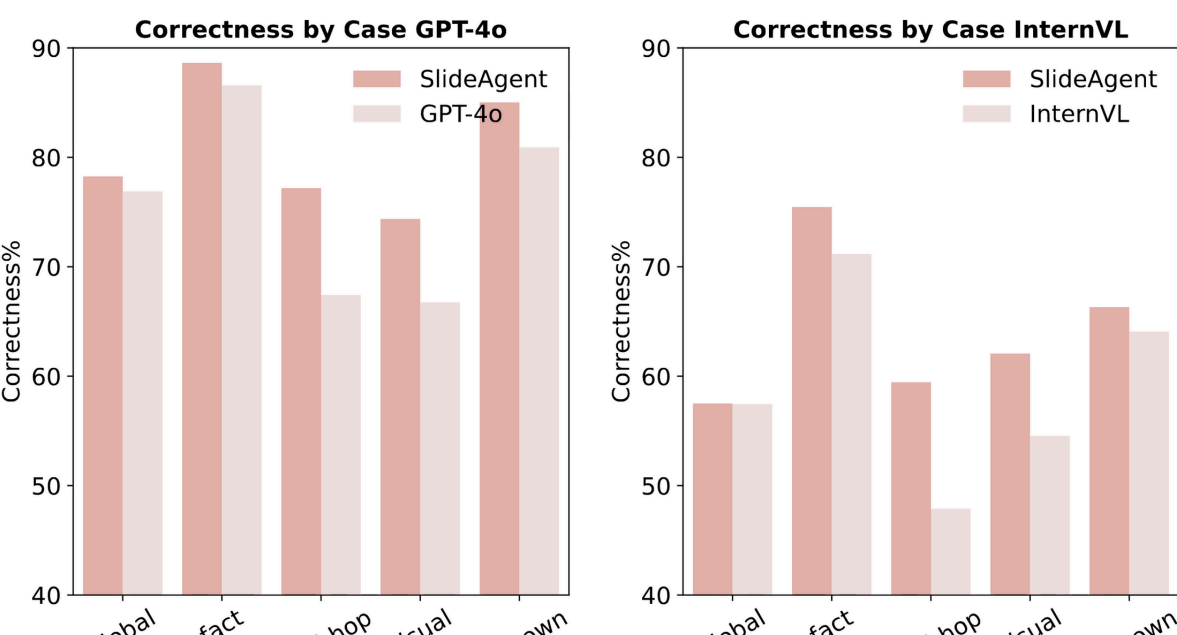
01 Infographics are structured visuals designed to convey complex information.



- C1: Scalable Fine-Grained Reasoning**
Low-resolution visual encoders miss small fonts, icons, footnotes.
- C2: Domain-Specific Visual Semantics**
Identify relevant slides from large slide decks.
- C3: Metadata-Free Integration**



03 RESULTS



Model	SlideVQA			TechSlides			FinSlides		
	Overall	Num	F1	Overall	Num	F1	Overall	Num	F1
Multimodal LLMs (Type 1)									
Gemini 2.0	75.0	71.3	79.8	50.4	67.6	41.6	70.8	70.6	77.8
Gemini 2.5	83.8	78.3	91.8	51.1	71.4	41.2	76.2	75.8	100.0
Gemini 2.5-lite	71.2	60.8	87.0	47.3	58.1	41.9	57.0	56.6	68.3
Claude 4.1	78.4	74.3	82.3	61.0	81.4	52.3	56.5	54.8	73.3
Claude 3.5	62.5	68.3	54.6	52.5	80.2	39.5	48.5	49.5	29.6
GPT-4o	77.0	72.1	84.0	63.4	78.3	53.9	80.0	80.8	62.1
Multimodal RAG (Type 2) and Agentic Methods (Type 3) with GPT-4o as base models.									
COLPALI	78.8	73.7	83.4	64.1	73.2	54.5	80.9	81.5	62.7
VisRAG	78.2	73.1	85.4	64.7	72.6	54.7	79.2	81.1	75.8
VDocRAG	80.0	75.0	87.8	67.0	80.5	57.0	83.5	83.8	64.2
ViDoRAG	81.1	76.4	88.1	68.7	78.2	59.4	82.2	83.3	65.1
SlideAgent	84.9	80.4	90.5	70.9	82.5	66.2	85.5	85.9	79.6
Impr.	+7.9	+8.3	+6.5	+7.5	+4.2	+12.3	+5.5	+5.0	+17.5

