



ORCA: Orchestrated Reasoning with Collaborative Agents for Document Visual Question Answering

Aymen Lassoued^{1,2}, Mohamed Ali Souibgui³, Yousri Kessentini¹

¹Digital Research Center of Sfax, SMARTS Laboratory, Tunisia ²École Polytechnique de Tunisie, University of Carthage

³Computer Vision Center, Universitat Autònoma de Barcelona



Problem Statement

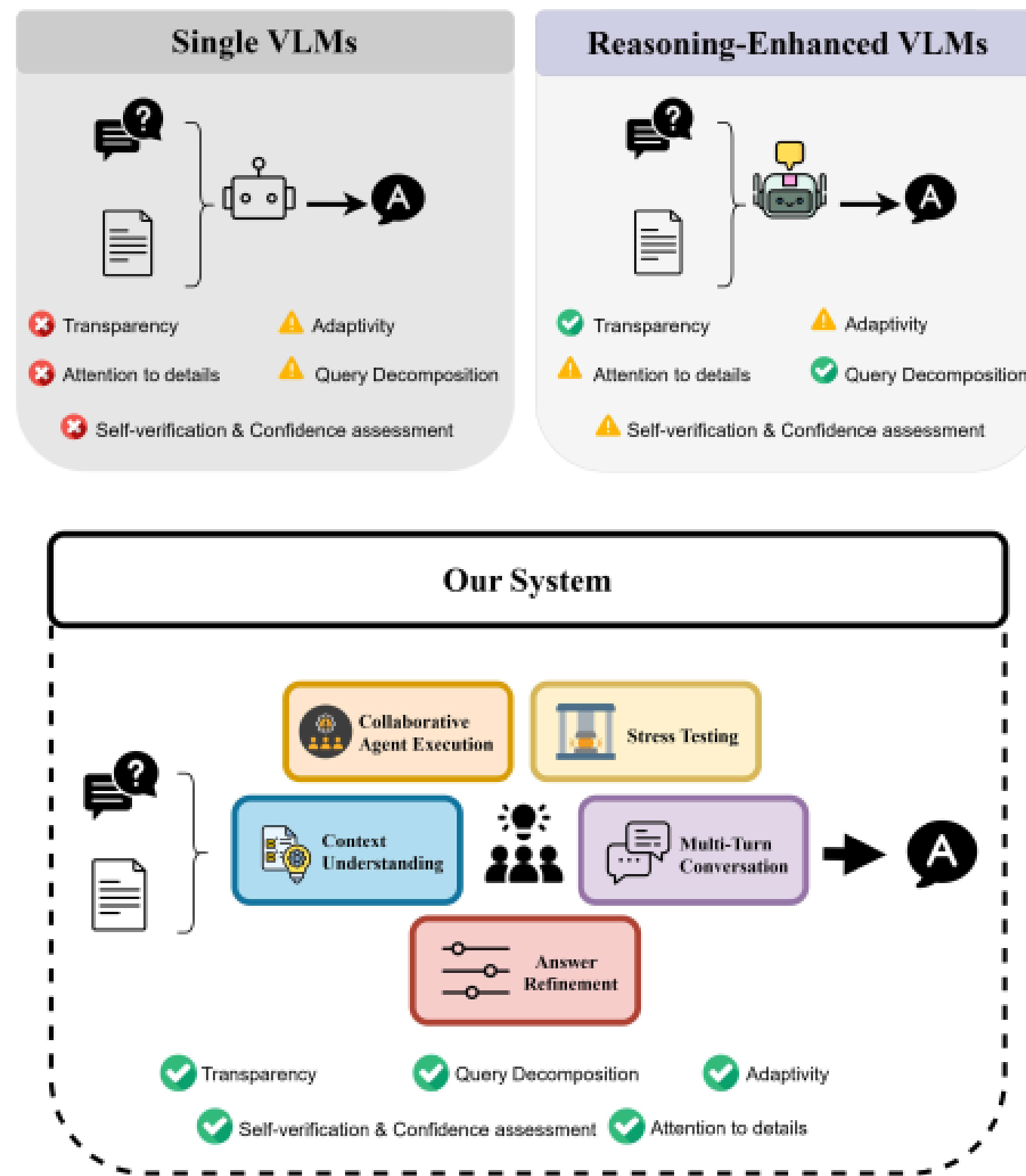
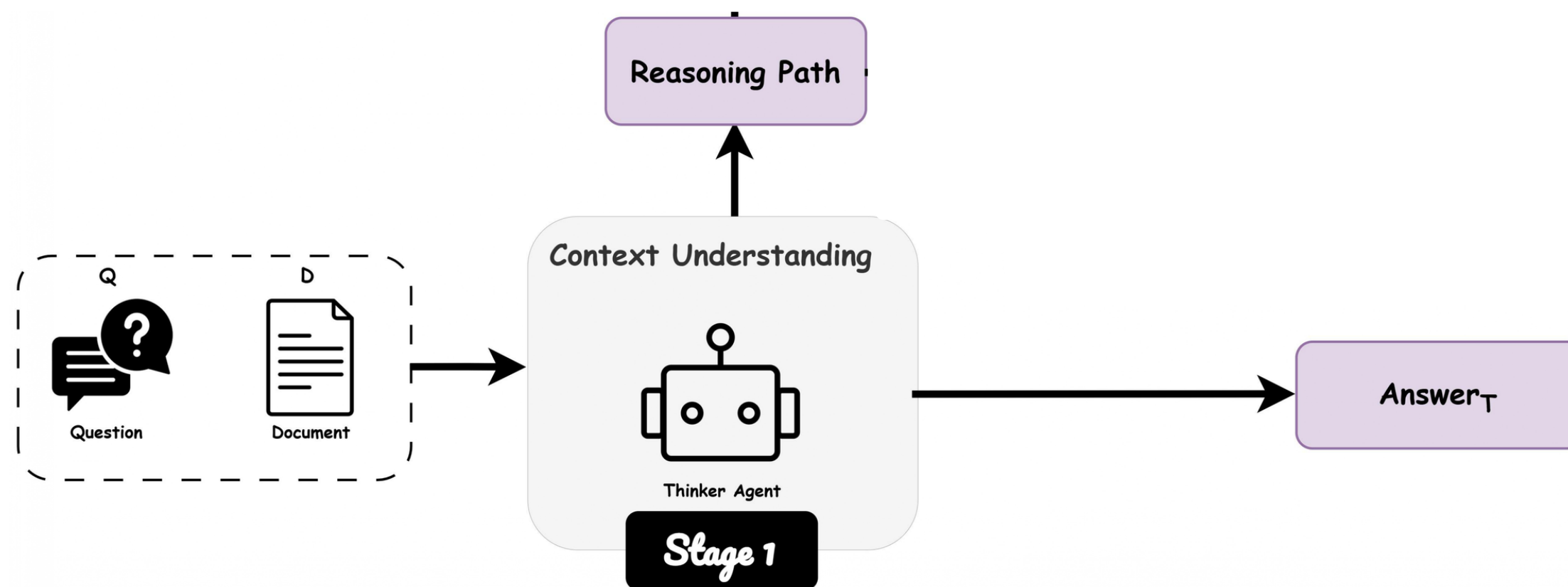
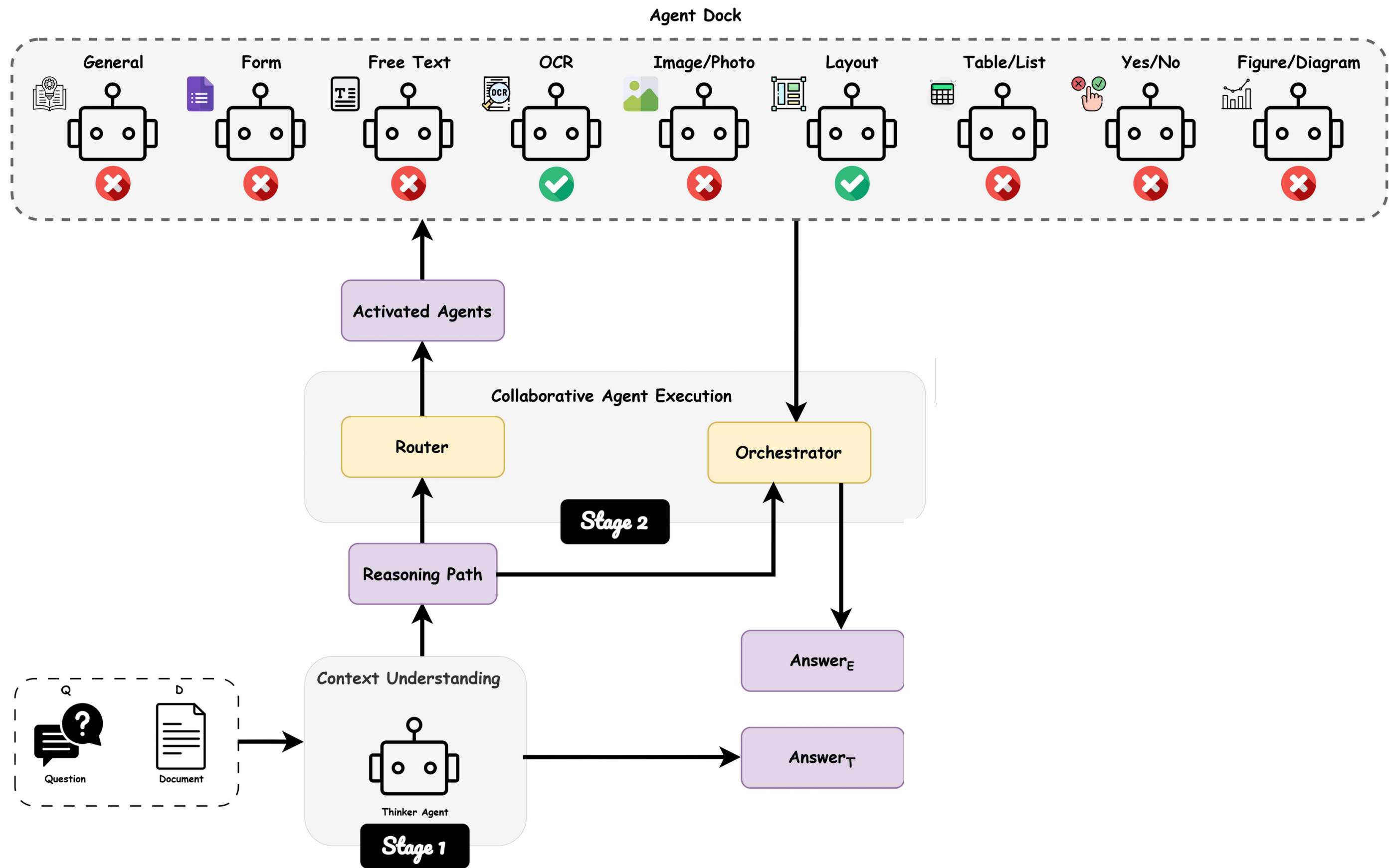


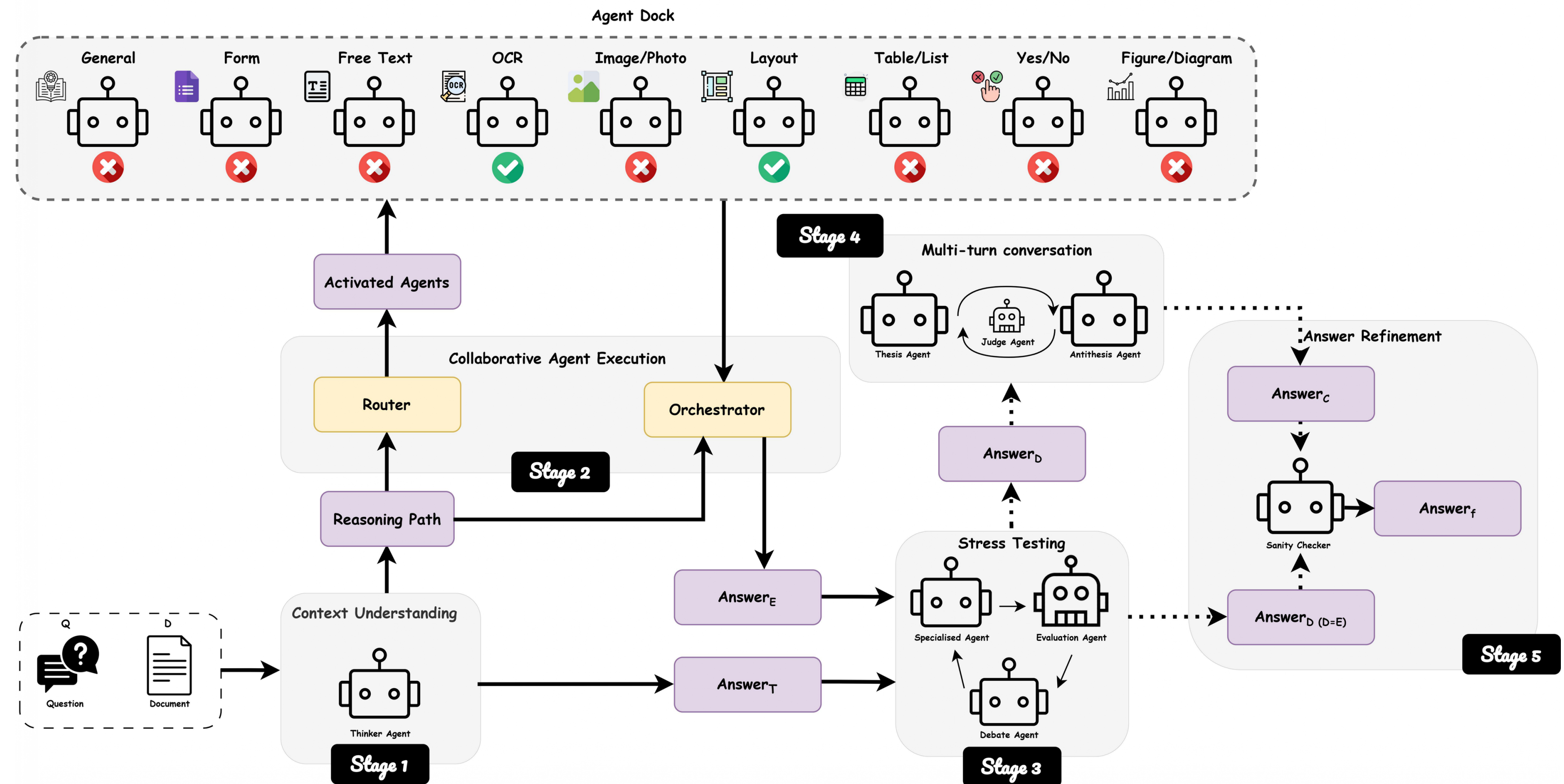
Figure 1. **Comparison of different approaches for DocVQA.** Single-model VLMs and reasoning-enhanced VLMs lack critical capabilities such as adaptivity and self-verification. In contrast, ORCA introduces a feature-oriented, multi-agent design achieving improved DocVQA performance as well as the missing capabilities in one unified framework.

Proposed Method





ORCA Framework



Quantitative Results

Model	Open-source	DocVQA	InfoVQA	Avg.
<i>Document Understanding Models</i>				
LayoutLMv2 LARGE [68]	✓	86.7	28.3	57.5
Text-Monkey [32]	✓	66.7	28.6	47.7
DocOwl-1.5 [20]	✓	82.2	50.7	66.5
<i>General-Purpose VLMs</i>				
Qwen2-VL [5]	✓	96.7	84.7	90.7
InternVL2-Pro [12]	✓	95.1	83.3	89.2
Molmo-72B [15]	✓	93.5	81.9	87.7
DeepSeek-VL2 [64]	✓	93.3	78.1	85.7
LLaVA-One-Vision-8B [29]	✓	94.8	78.4	86.6
Gemini Pro 1.5 [55]	×	86.5	72.7	79.6
Claude-3.7 Sonnet [3]	×	94.1	65.5	79.8
Qwen-VL-Max (single) [4]	×	93.1	73.4	83.3
GPT-4 Turbo + Textract [1]	×	87.4	71.9	79.7
GPT-4o [1]	×	93.0	82.1	87.6
<i>Reasoning-Enhanced OS Models</i>				
MiMo-VL-7B-RL [65]	✓	95.0	88.1	91.6
VideoLLaMA3-7B [70]	✓	95.0	78.9	87.0
Gemma-3-27B-IT [56]	✓	86.6	70.6	78.6
<i>Baseline VLMs (Single-Model)</i>				
Qwen2.5-VL-7B-Instruct [47]	✓	95.7	77.7	86.7
Qwen3VL-4B-Instruct [49]	✓	95.3	80.3	87.8
Qwen3VL-8B-Instruct [50]	✓	96.1	83.1	89.6
<i>ORCA (Multi-Agent Framework)</i>				
ORCA (Qwen2.5-VL-7B)	✓	96.4 (+0.7)	86.9 (+9.2)	91.7 (+5.0)
ORCA (Qwen3VL-4B)	✓	96.0 (+0.7)	85.4 (+5.1)	90.7 (+2.9)
ORCA (Qwen3VL-8B)	✓	97.2 (+1.1)	88.0 (+4.9)	92.6 (+3.0)
<i>Relative Improvements</i>				
Average Gain	–	+0.8%	+6.4%	+3.6%

DocQVA + InfoVQA

Model	Rec	Ref	Spo	Ext	Par	Cal	Und	Rea	Avg
<i>Open-source LMMs</i>									
LLaVA-Next-8B* [37]	41.3	18.8	0	49.5	21.2	17.3	55.2	48.9	31.5
LLaVA-OV-7B* [29]	46.0	20.8	0.1	58.3	25.3	23.3	64.4	53.0	36.4
TextMonkey* [32]	39.1	0.7	0	19.0	12.2	19.0	61.1	40.2	23.9
Molmo-7B* [15]	52.4	21.3	0.1	45.5	7.6	28.5	65.3	55.0	34.5
Cambrian-34B* [59]	45.3	21.5	0	53.6	19.2	19.5	63.5	55.5	34.7
Pixtral-12B* [2]	48.9	21.6	0	66.3	35.5	29.8	66.9	53.7	40.3
Nemotron Nano V2 VL* [42]	67.6	54.5	36.2	92.0	26.6	80.4	75.5	57.0	61.2
Llama Nemotron VL 8B* [41]	70.2	69.1	61.8	81.4	39.2	31.9	73.1	54.7	60.2
InternVL-3-14B* [22]	67.3	36.9	11.2	89.0	38.4	38.4	79.2	60.5	52.6
Qwen2.5-VL-7B* [47]	68.8	25.7	1.2	80.2	30.4	38.2	73.2	56.2	46.7
Qwen3-Omni-A3B* [46]	72.3	62.0	45.6	93.5	20.8	67.0	74.1	55.3	61.3
Qwen3-VL-4B-Instruct* [49]	-	-	-	-	-	-	-	-	63.7
Qwen3-VL-8B-Instruct* [50]	-	-	-	-	-	-	-	-	65.4
<i>ORCA (Multi-Agent Framework)</i>									
ORCA (Qwen2.5-VL-7B)	72.3	28.6	3.4	84.7	34.2	41.9	75.8	61.5	50.3 (+3.6)
ORCA (Qwen3VL-4B)	81.1	48.2	24.3	89.3	60.1	57.2	86.6	71.6	64.8 (+1.1)
ORCA (Qwen3VL-8B)	83.2	51.3	28.4	89.8	62.3	59.4	88.7	73.7	67.1 (+1.7)

OCRBenchV2

Inference latency comparison

Configuration	Latency (s)	DocVQA	InfoVQA
Baseline (Qwen3VL-8B)	0.3–0.8	96.1	83.1
ORCA Early-Termination	2.9–4.5	96.7	86.9
ORCA Full Pipeline	9.6–13.1	97.2	88.0

Table 3. Inference latency comparison on 4× NVIDIA L4 GPUs with vLLM optimization. Early-termination applies when thinker and expert agents agree, bypassing Stages 3–5 (77% of cases)

Ablation Studies

Configuration	DocVQA	InfoVQA	OCRBench-v2
ORCA (Full)	97.2	88.0	67.1
<i>Individual Component Ablations</i>			
w/o Stage 1 (Reasoning)	96.5 (-0.7)	84.9 (-3.1)	66.1 (-1.0)
w/o Stage 2 (Collaborative Agents)	96.3 (-0.9)	84.1 (-3.9)	66.0 (-1.1)
w/o Stage 3 (Stress Testing)	97.0 (-0.2)	87.5 (-0.5)	66.9 (-0.2)
w/o Stage 4 (Multi-turn Debate)	96.9 (-0.3)	87.2 (-0.8)	66.7 (-0.4)
w/o Stage 5 (Answer Refinement)	97.1 (-0.1)	87.9 (-0.1)	67.0 (-0.1)
<i>Cumulative Ablations</i>			
w/o Stages 4+5	96.8 (-0.4)	87.6 (-0.4)	66.9 (-0.2)
w/o Stages 3+4+5 (All Verification)	96.7 (-0.5)	86.9 (-1.1)	66.3 (-0.8)
w/o Stages 2–5	96.2 (-1.0)	84.1 (-3.9)	65.6 (-1.5)
Baseline (Single Model)	96.1	83.1	65.4

Table 4. Stage-wise ablation study on Qwen3VL-8B. Each row isolates the impact of removing a specific stage from the complete ORCA framework.

Component Variation	DocVQA	InfoVQA	OCRBench-v2
No reasoning masking	96.5	87.6	66.4
Reasoning masking (Full)	97.2	88.0	67.1
Performance gain	+0.7	+0.4	+0.7

Table 5. Reasoning path masking ablation on Qwen3VL-8B. Masking prevents confirmation bias while preserving strategic guidance.

Qualitative Results



Question

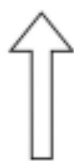
What publication detail accompanies the Genealogical Society entry?



Answer

GSU, 1977

Church of Jesus Christ of Latter-day Saints, Genealogical Society, Family Group Records Collection, Archives Section 1942-1964. (SLC: GSU, 1977.)



Research Log (3)

Ancestor: William FRAZIER (1826-1881) Researcher: G. David Dilts

Date	PLACE OF RESEARCH Mailing Address	PURPOSE Seeking what?	CALL NUMBER	SOURCE. Books: Author, title, publication place, publisher, publ. date, page(s). Periodicals: + vol., date, issue #.	DOCUMENT NUMBER	RESULTS
8 May 2001	Robert G. FRAZIER Telephone 801-359-9301	* Robert P.		Robert G. Frazier, Wichita, Kans, 8 May 2001, telephone conversation with G. David Dilts Salt Lake City, Utah		nil
11 May 2001	Family History Library 35 N West Temple St. SLC, UT 84150-3400	OO Wm. FRAZIER Marg. DENTON		Ancestral File [online database] (Salt Lake City: Geneal. Soc. of Utah, 1998). William FRAZIER 1826	William FRAZIER 1826-1881	* OO + B family of Wm.
11 May 2001	"	OO Wm. FRAZIER Nancy McCLURE		International Gen. Index [online database] (Salt Lake City: Gen. Soc. of Utah, 1992, 2000). North America 2000 Addendum Marr. Search	William FRAZIER 1826-1881	OO of Wm. & Nancy
"	"	Wm. FRAZIER family group	1274092	Church of Jesus Christ of Latter-day Saints, Genealogical Society, Family Group Records Collection, Archives Section 1942-1964. (SLC: GSU, 1977.)		
1 Jun 2001	"	1860 age of Robert	976.8 X225 1860 18771164.3	Byron Sistler, 1860 Census - Tennessee. (Nashville: Byron Sistler & Assoc., 1981-1982). v. 2, p. 244	William FRAZIER 1826-1881 10	census: Robert in 1860
"	"	"	805258	United States Census Office, 8th Census, 1860. Tennessee, 1860 Federal Census: Population Schedules. (Washington: National Archives, 1950, 1967.) 2nd filming. Jefferson County, p. 274	William FRAZIER 1826-1881 11	Wm.'s family including Robt. P.
1 Jun 2001	"	"	976.8924 H2m	Estle Pershing Muncy, People and Places of Jefferson County, Tennessee		nil

Question: Were any children born between James 1863 & Liz 1870?
Suggestion: Look in the 1870 census of Tennessee for the answer.



Qwen 2.5VL7B - Instruct

Church of Jesus Christ of Latter-day Saints, Genealogical Society, Family Group Records Collection, Archives Section 1942-1948 (SLC: GSA, 1977).



Qwen3-VL-4B - Instruct

Family Group Records Collection, Archives Section 1942-1964 (SLC: GSU, 1977)



Qwen3-VL-8B - Instruct

William FRAZIER 182

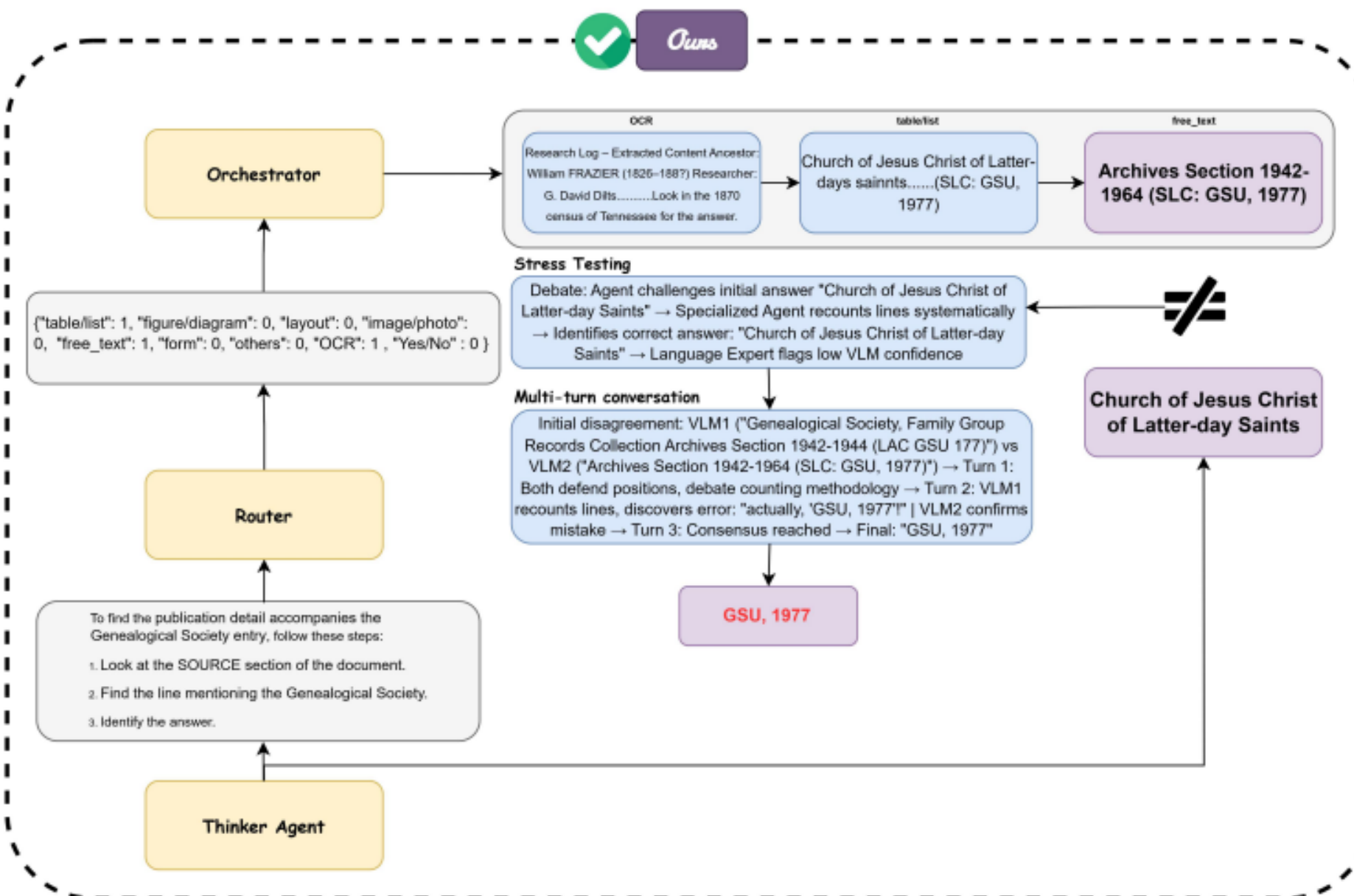


GLM-4.1V-9B-Thinking

is part of that source note. To determine the publication detail for the Genealogical Soc. in the *SOURCE* column. For the row corresponding to "Wm. FRAZIER family group", the *SOURCE* field shows ***"Church of Jesus Christ of Latter-day Saints - Genealogical Society, Family Group Records Collection Archives Section 1942-1944 (LAC GSU 177)***. Thus, the relevant publication detail is the collection and archival information. Ultimately, the answer is <|begin_of_box|>Church of Jesus Christ of Latter-day Saints - Genealogical Society, Family Group Records Collection Archives Section 1942-1944 (LAC GSU 177)<|end_of_box|>.



Ours



Details: X
Self-verification: X

Details: X
Self-verification: X

Details: ✓
Self-verification: ✓

Question

what is the trainer's name of the horse "Delta"?

Answer

Tipton Mt Horse

OWNER'S NAME	TRAINER'S NAME & ADDRESS	OWNER'S MAILING ADDRESS
Molly Boyd 123 Main St Prestonsburg, KY	Tipton Mt Horse Stanton, KY	Molly Boyd 123 Main St Prestonsburg, KY
M. K. Bowen Stanton, KY	Tipton Mt Horse Stanton, KY	John Stivers 123 Willow Creek Russell Springs, KY
Justin Anderson Verona, KY		John Anderson Verona, KY
Sydney Dillard 10 W. 1st St Franklin, KY	Sam	Sam
	Charles DeAtty Anna Day	Anna Day



CLASSNO. #13
ENTRY FEE \$30.00
Show Date Friday Aug 4, 2023
Judge Mike Alexander
Name of Show C.C.S.C. Horse Show

PLACE	RACE	HORSE NAME & REG. NO.	OWNER'S NAME	TRAINER'S NAME & ADDRESS	OWNER'S MAILING ADDRESS
2	804	Hell on Wheels	Molly Boyd 123 Main St Prestonsburg, KY	Tipton Mt Horse Stanton, KY	Molly Boyd 123 Main St Prestonsburg, KY
1	808	Hell on Wheels	M. K. Bowen Stanton, KY	Tipton Mt Horse Stanton, KY	John Stivers 123 Willow Creek Russell Springs, KY
		Delta Queen	Justin Anderson Verona, KY		John Anderson Verona, KY
5	716	CV-5 Amy Lee	Sydney Dillard 10 W. 1st St Franklin, KY	Sam	Sam
	886	CV-5 Amy Lee	Sydney Dillard 10 W. 1st St Franklin, KY	Sam	Sam
	017	CV-5 Amy Lee	Sydney Dillard 10 W. 1st St Franklin, KY	Sam	Sam
	107	CV-5 Amy Lee	Sydney Dillard 10 W. 1st St Franklin, KY	Sam	Sam
Scrub	108	CV-5 Amy Lee	Sydney Dillard 10 W. 1st St Franklin, KY	Sam	Sam
	82	CV-5 Amy Lee	Sydney Dillard 10 W. 1st St Franklin, KY	Sam	Sam

Qwen 2.5VL 7B

Not Found

Qwen3-VL-4B-Instruct

Justin Anderson

Qwen3-VL-8B-Instruct

Sam

Details: X
Conclusion: X

GLM-4.1V-9B-Thinking

To determine the trainer's name of the horse "Delta" (likely referring to "Delta Queen" in the image), we locate the row with the horse name "Delta Queen" (reg. no. 808) and check the "TRAINER'S NAME & ADDRESS" column. The trainer's name in that row is <|begin_of_box|>**Tipton M4 Yerster**<|end_of_box|> (or a similar variation as seen in the image).

Details: X
Conclusion: X

Ours

Using ORCA, the system first examined the given document with the OCR agent. Then, the table/list agent identified the cell containing the answer using its reasoning path and was able to know how to read it correctly from the OCR output, which produced the text "Tipton Mt Horse." Since this answer differed from the candidate answer provided by the thinker agent, the process moved to the stress-testing stage. After two rounds of stress testing, the model continued to give the same answer, indicating strong confidence. Therefore, the system passed the result directly to the sanity checker, which confirmed the final answer: **Tipton Mt Horse**.

Details: ✓
Conclusion: ✓

Thank you for watching

Code

