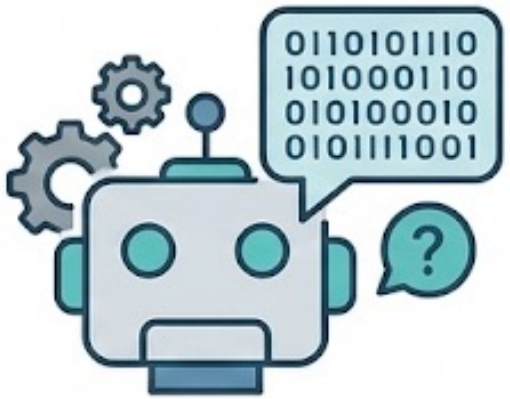
CoScientist

: A multi-agent framework for autonomous scientific discovery

Team: Tennis-Running-Swimming

https://github.com/seoklab/ai-cosci-all.git

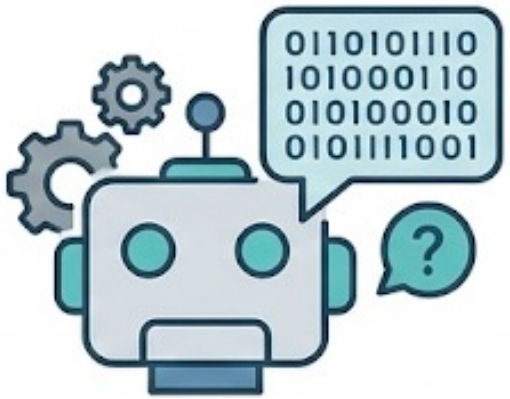
Can “machine” conduct scientific research?



Statistically probable tokens

Hallucination
Superficial answer

Can “machine” conduct scientific research?



Statistically probable tokens

Hallucination
Superficial answer

Needs:

Minimizing hallucination

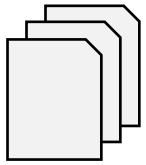
Specializing

“Divide-and-Conquer” with “evidence”

Minimizing hallucination

Specializing

“Divide-and-Conquer” with “evidence”



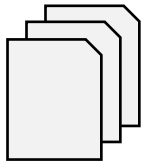
Literatures
Databases

Retrieval Augmented Generation

Minimizing hallucination

Specializing

“Divide-and-Conquer” with “evidence”



Literatures
Databases

Retrieval Augmented Generation

Agent Critic

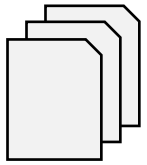
Critic-Refinement Loop

Q: Are claims supported by evidence?

Minimizing hallucination

Specializing

“Divide-and-Conquer” with “evidence”



Literatures
Databases

Retrieval Augmented Generation

Agent Critic

Critic-Refinement Loop

Q: Are claims supported by evidence?

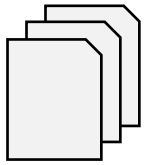
Subtask 1 → … → Subtask N

Sequential Subtask workflow

Minimizing hallucination

Specializing

“Divide-and-Conquer” with “evidence”



Literatures
Databases

Retrieval Augmented Generation

Agent Critic

Critic-Refinement Loop

Q: Are claims supported by evidence?

Subtask 1 → ··· → Subtask N

Sequential Subtask workflow

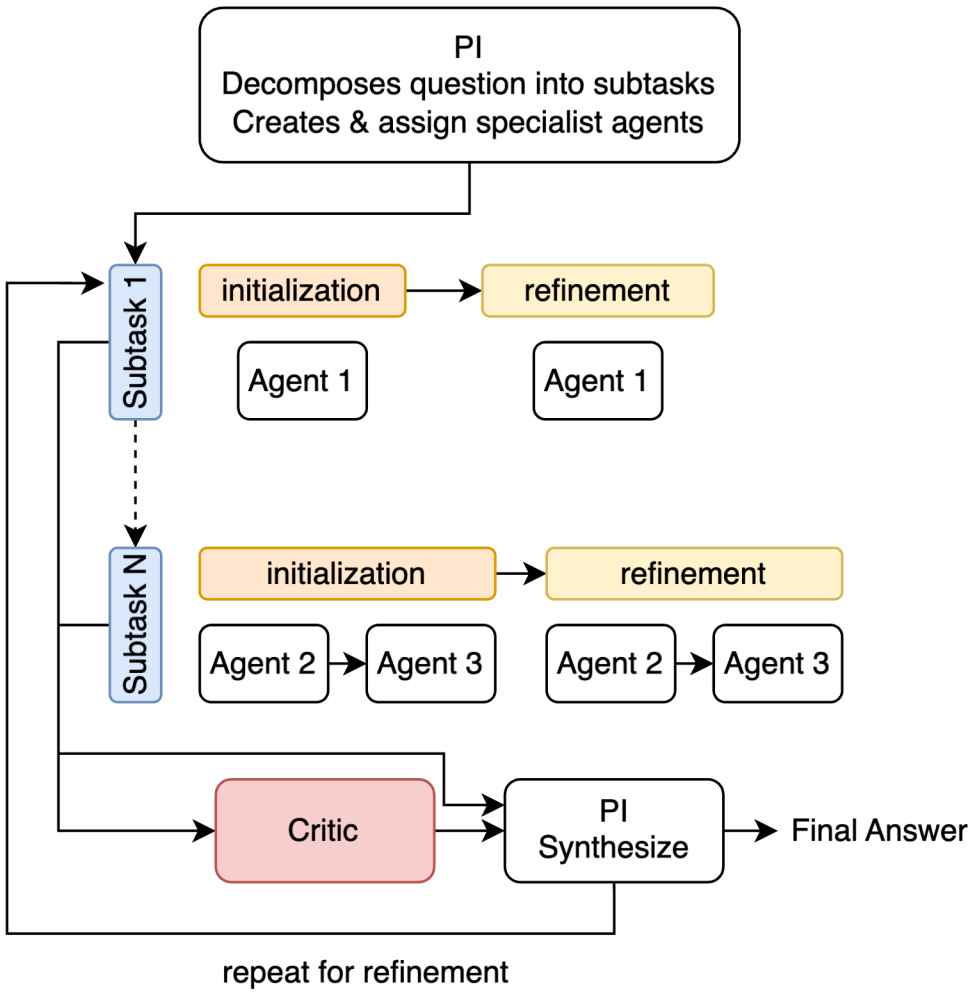
Dynamic agent generation & subtask assignment

PI agent generates specialized agents

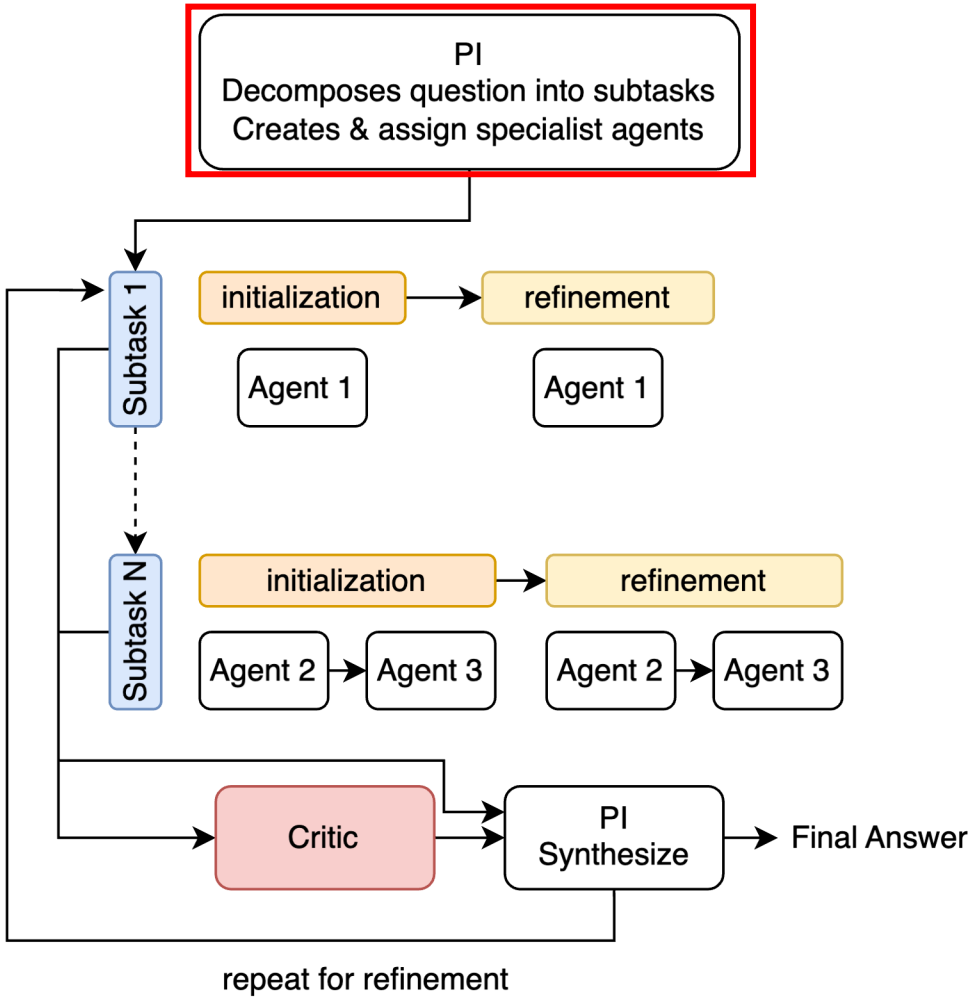
Minimizing hallucination

Specializing

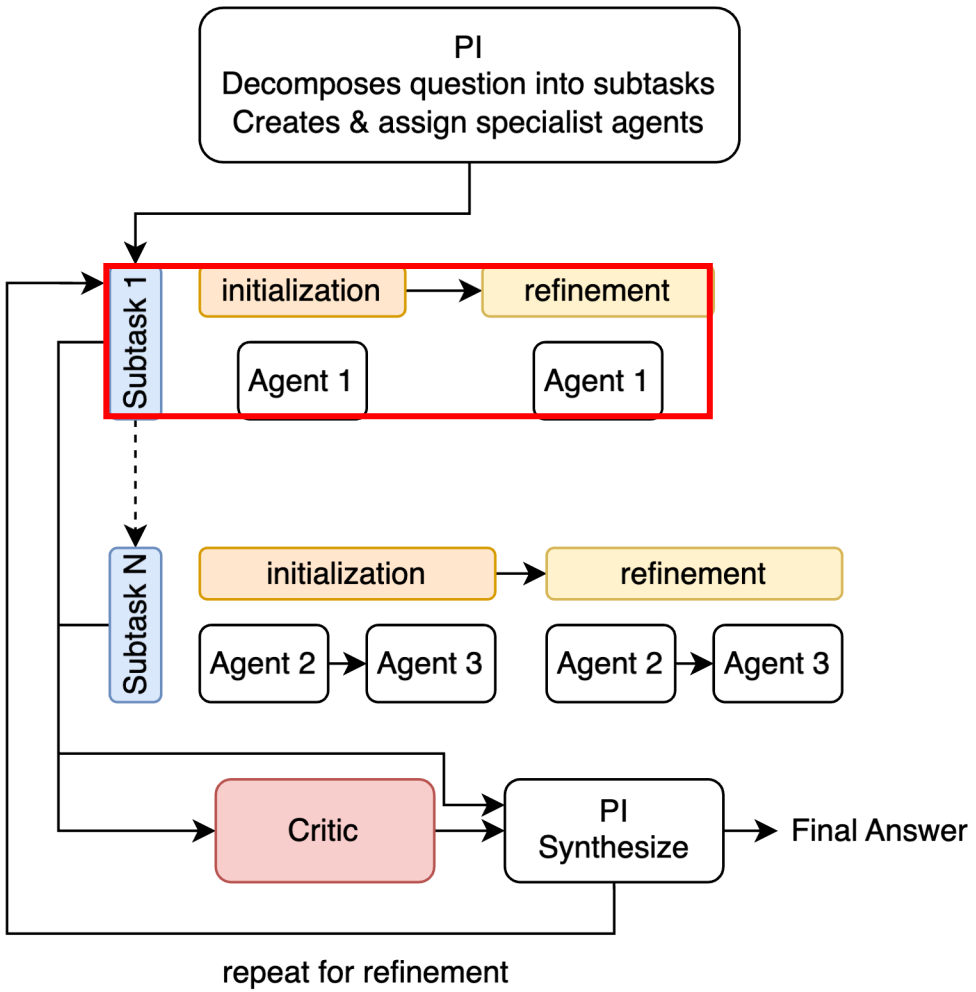
System architecture



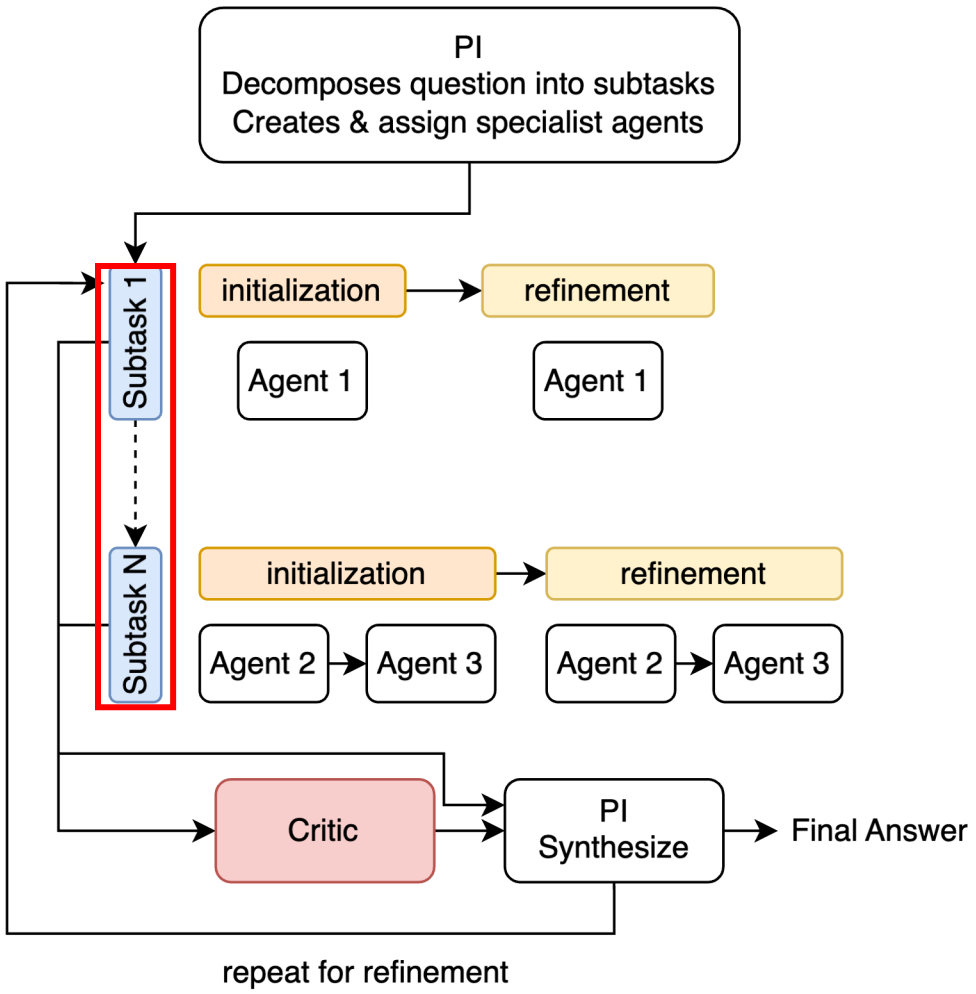
System architecture



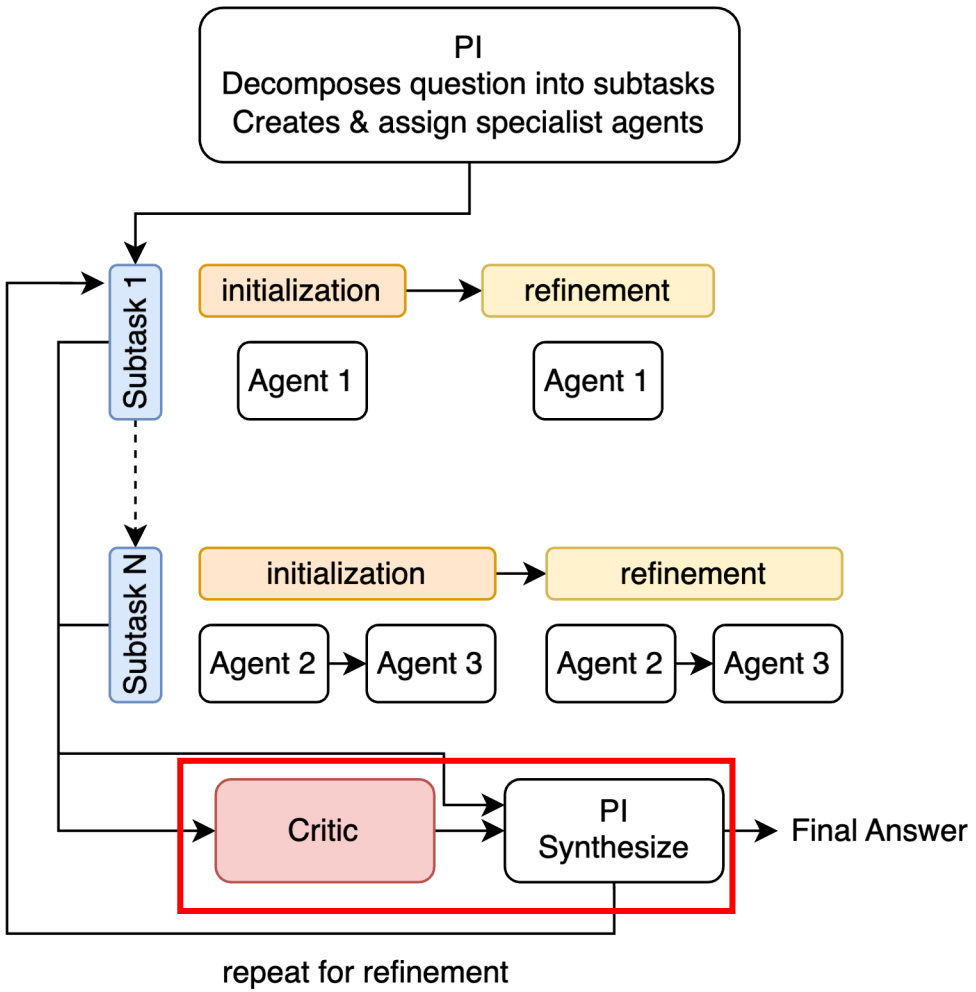
System architecture



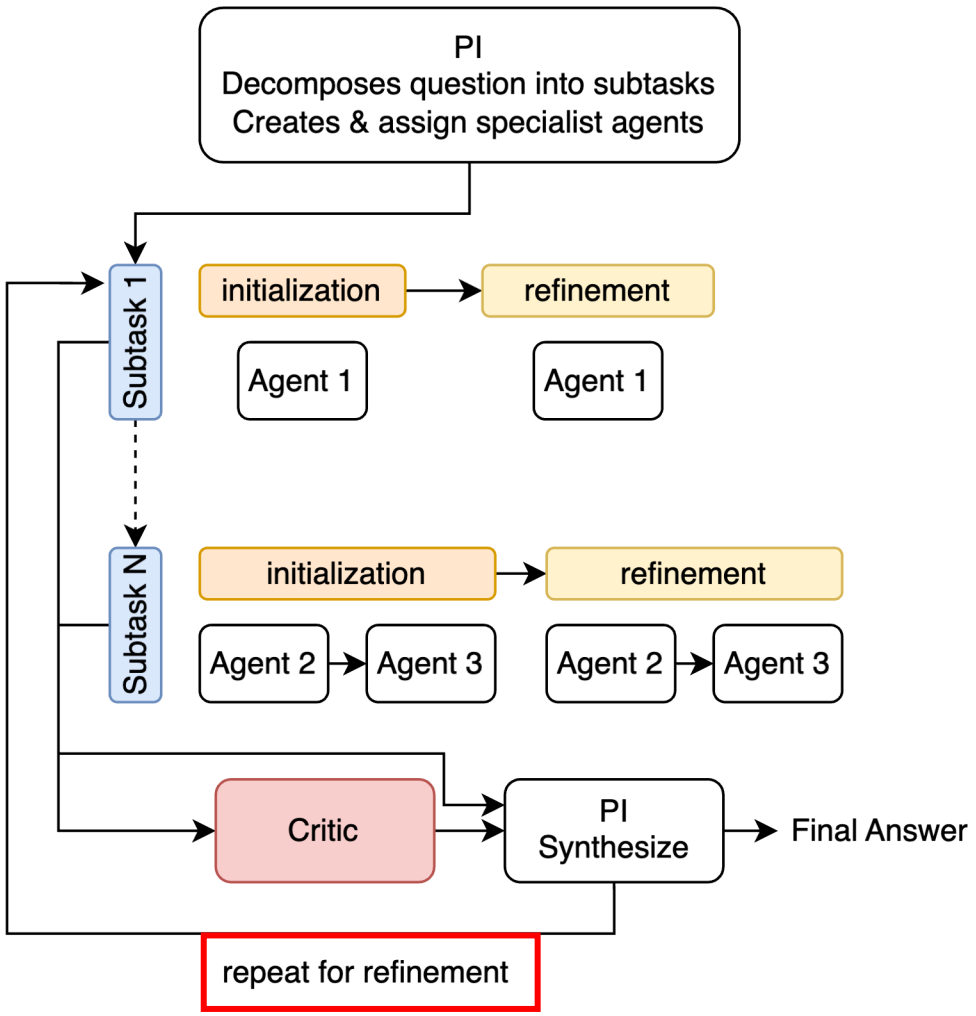
System architecture



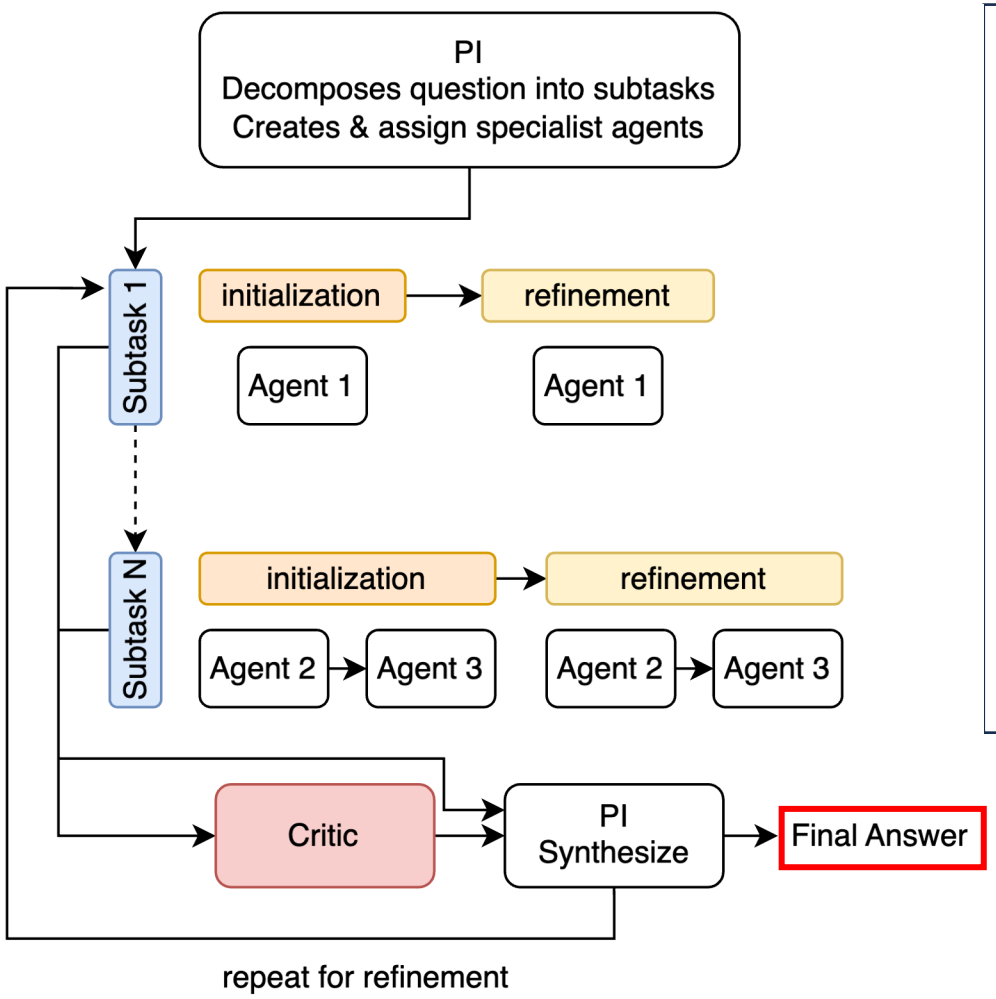
System architecture



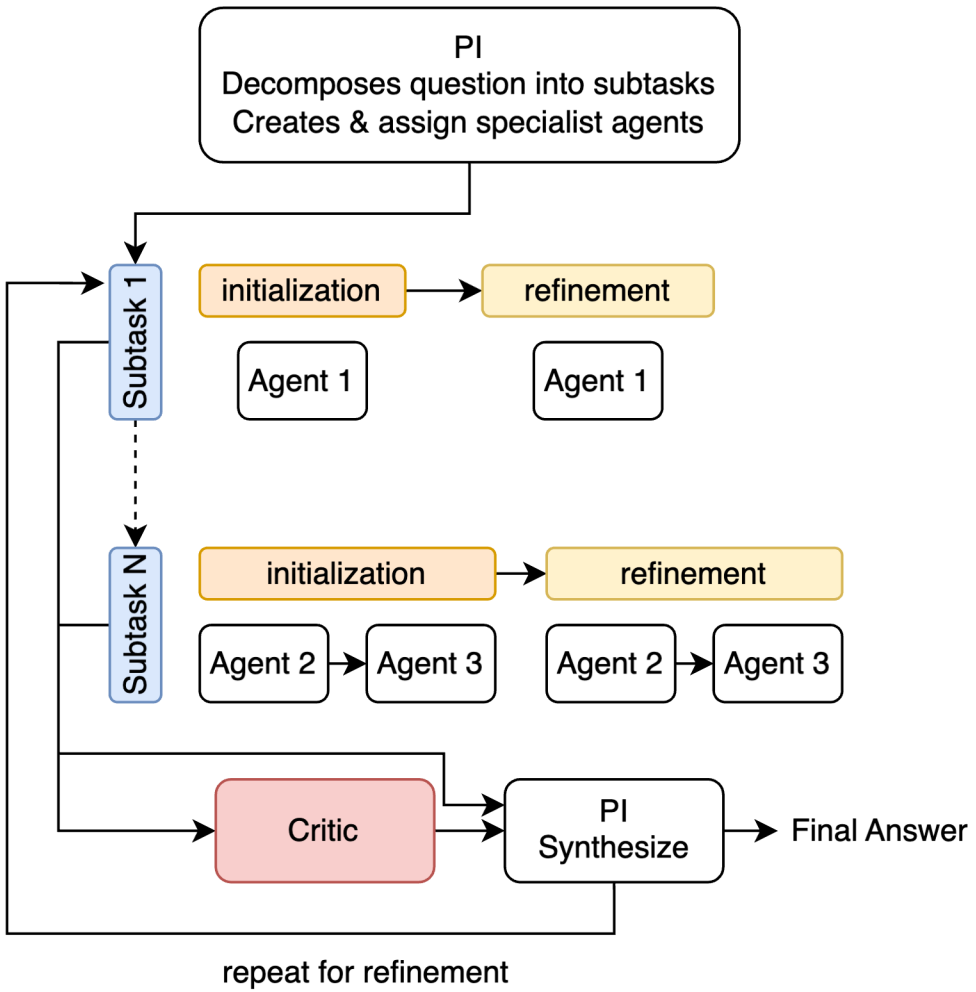
System architecture



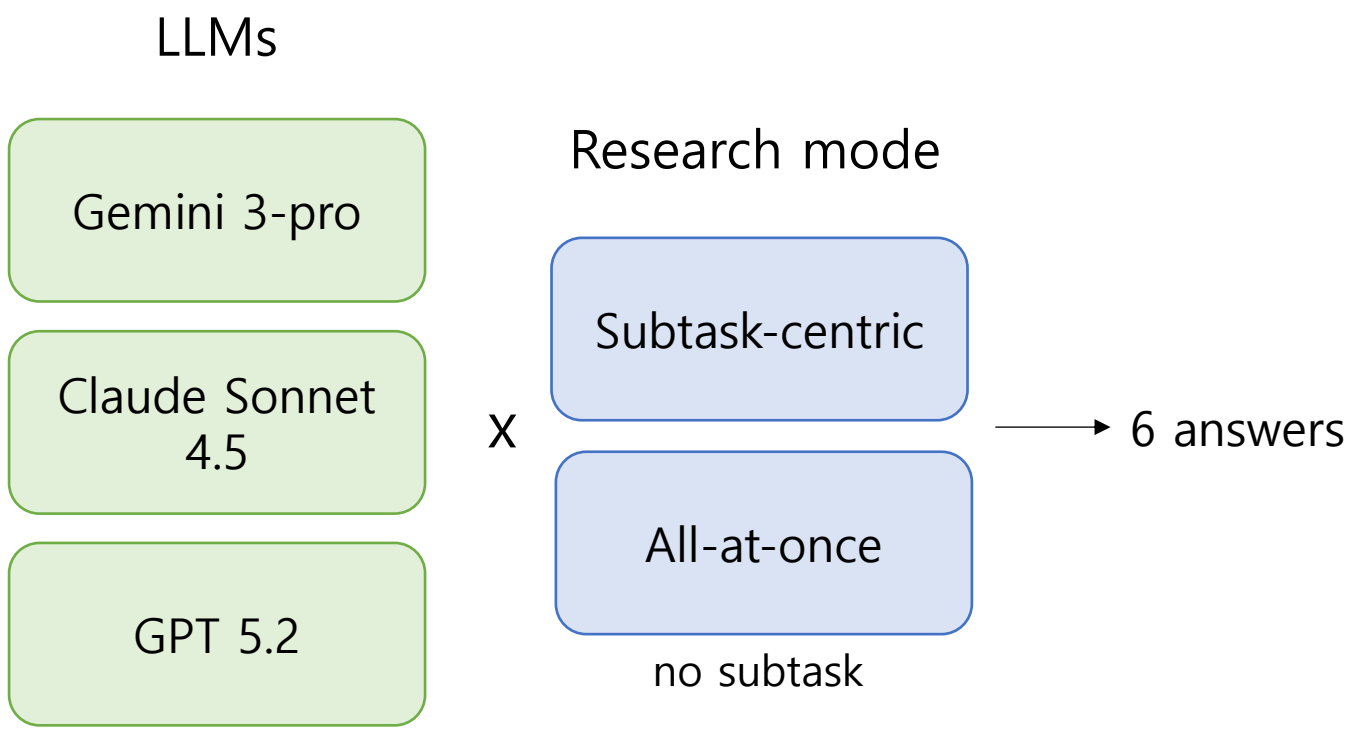
System architecture



System architecture



Automated workflow



Scoring by LLM-based evaluator

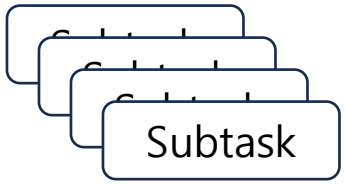
Evaluator

Prompt:
Evaluate answers according to the following criteria;
1. Scientific accuracy
2. Evidence quality
3. Methodological rigor
4. …

Problems of automated answers

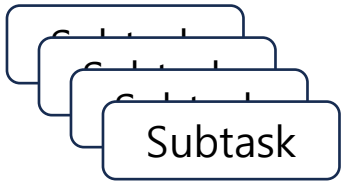
Problems of automated answers

Sub-optimal subtask split

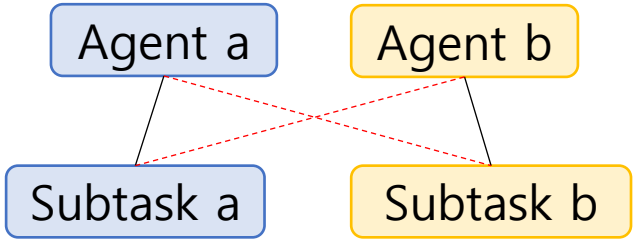


Problems of automated answers

Sub-optimal subtask split

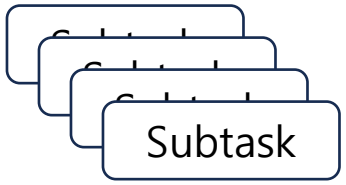


Subtask-Agent miss-match

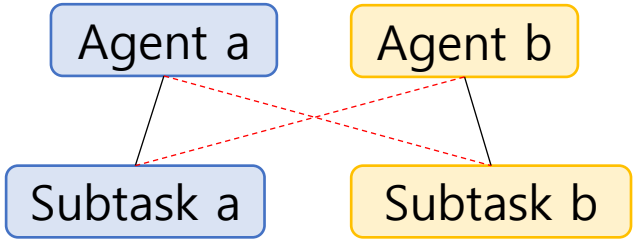


Problems of automated answers

Sub-optimal subtask split



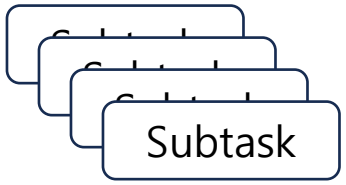
Subtask-Agent miss-match



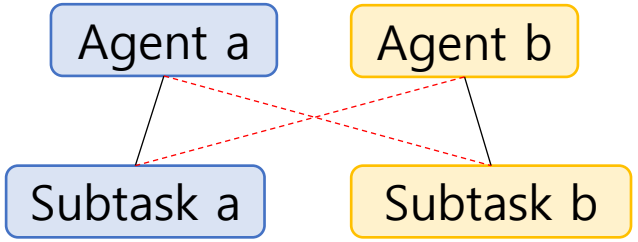
Repeating meaningless search & analysis

Problems of automated answers

Sub-optimal subtask split



Subtask-Agent miss-match

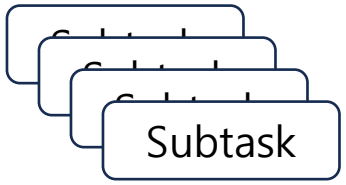


Repeating meaningless search & analysis

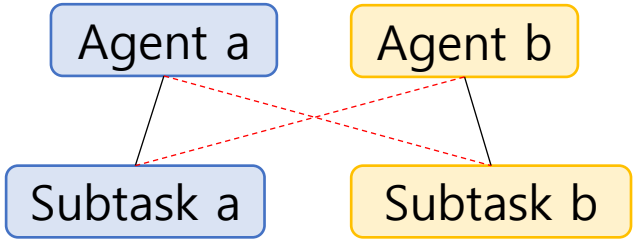
Ignoring previous results & Fabricating data & claims

Problems of automated answers

Sub-optimal subtask split



Subtask-Agent miss-match



Loss of Context

Hallucination

Repeating meaningless search & analysis

Ignoring previous results & Fabricating data & claims

Human guided workflow

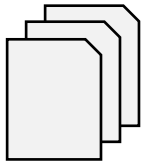
Human guided workflow

Direct human guidance

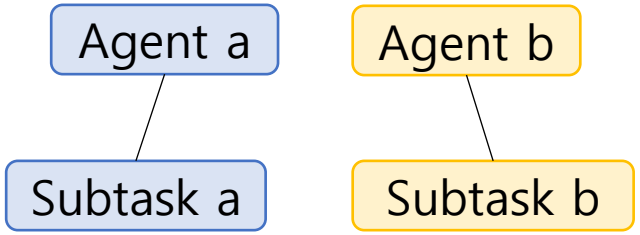
Prompt modification

```
Prompt:
Answer based on the
literature and
previous results.
```

answer based on
human-curated literatures



Human-guided task-agent-tool setting



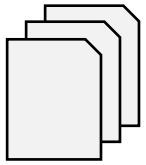
Human guided workflow

Direct human guidance

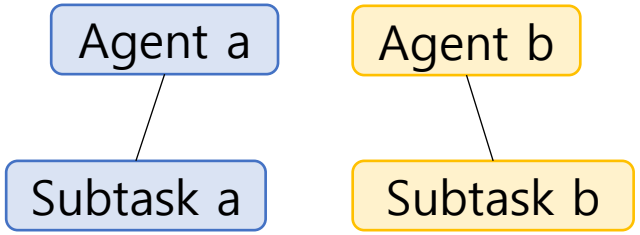
Prompt modification

```
Prompt:
Answer based on the
literature and
previous results.
```

answer based on
human-curated literatures

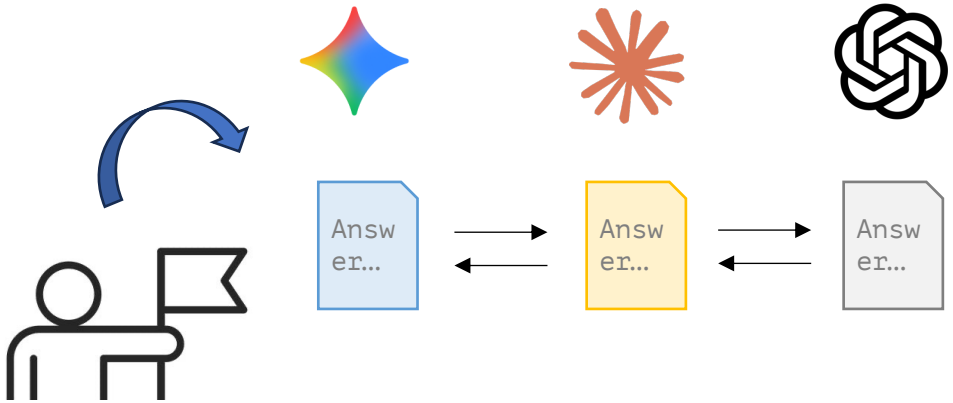


Human-guided task-agent-tool setting



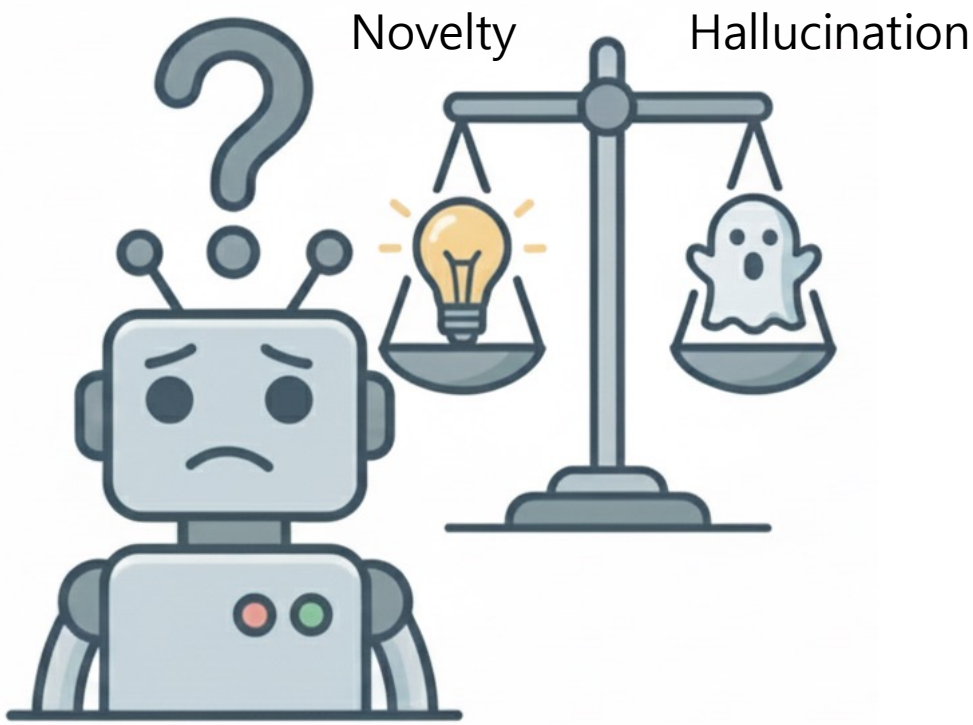
Cross-model validation

(using commercial LLM assistant)



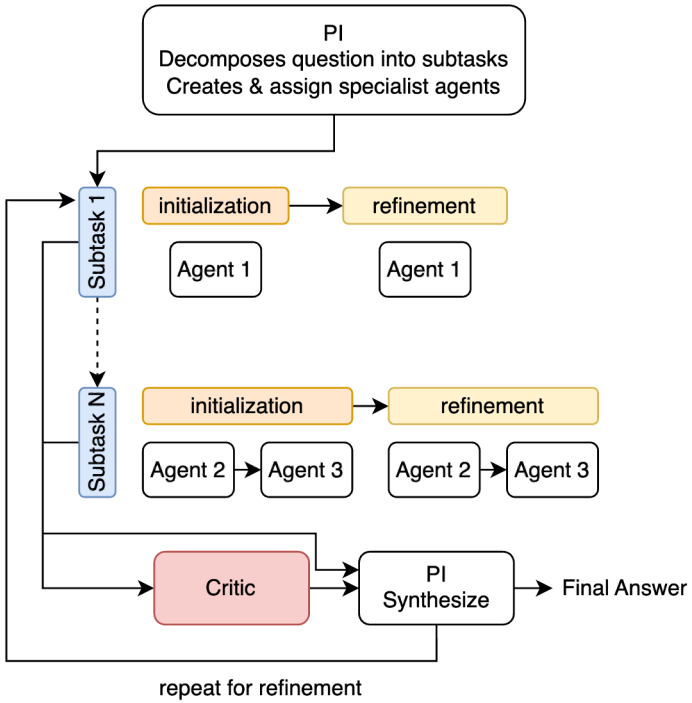

```
Q: Can the claims be derived from only given data?
Q: Is there conflicts or complementary claims?
Q: Refine answers considering feedback by other models
Q: Synthesize the final answer.
                                    ...
```

Limitations

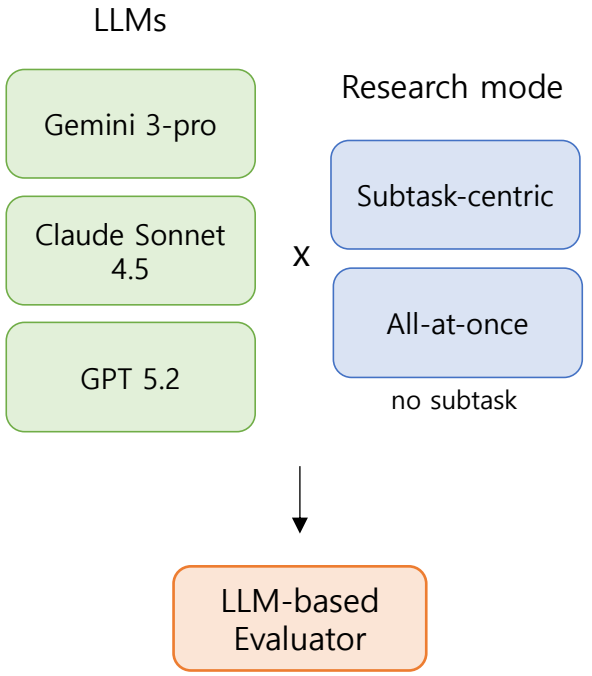


Conclusion

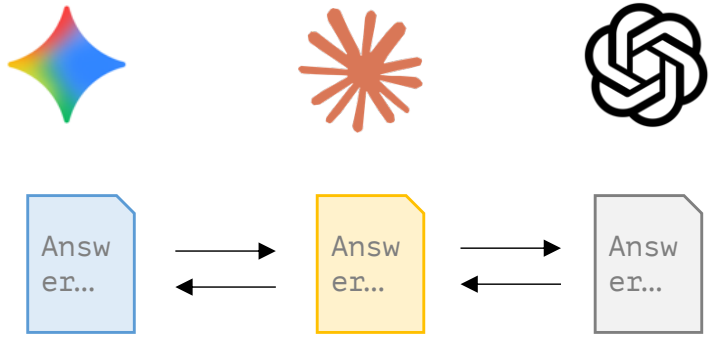
Multi-agent subtask-centric



Automation category



Cross-model checking



Q&A

Thank you for listening!

Sumin Lee

Jiho Sim

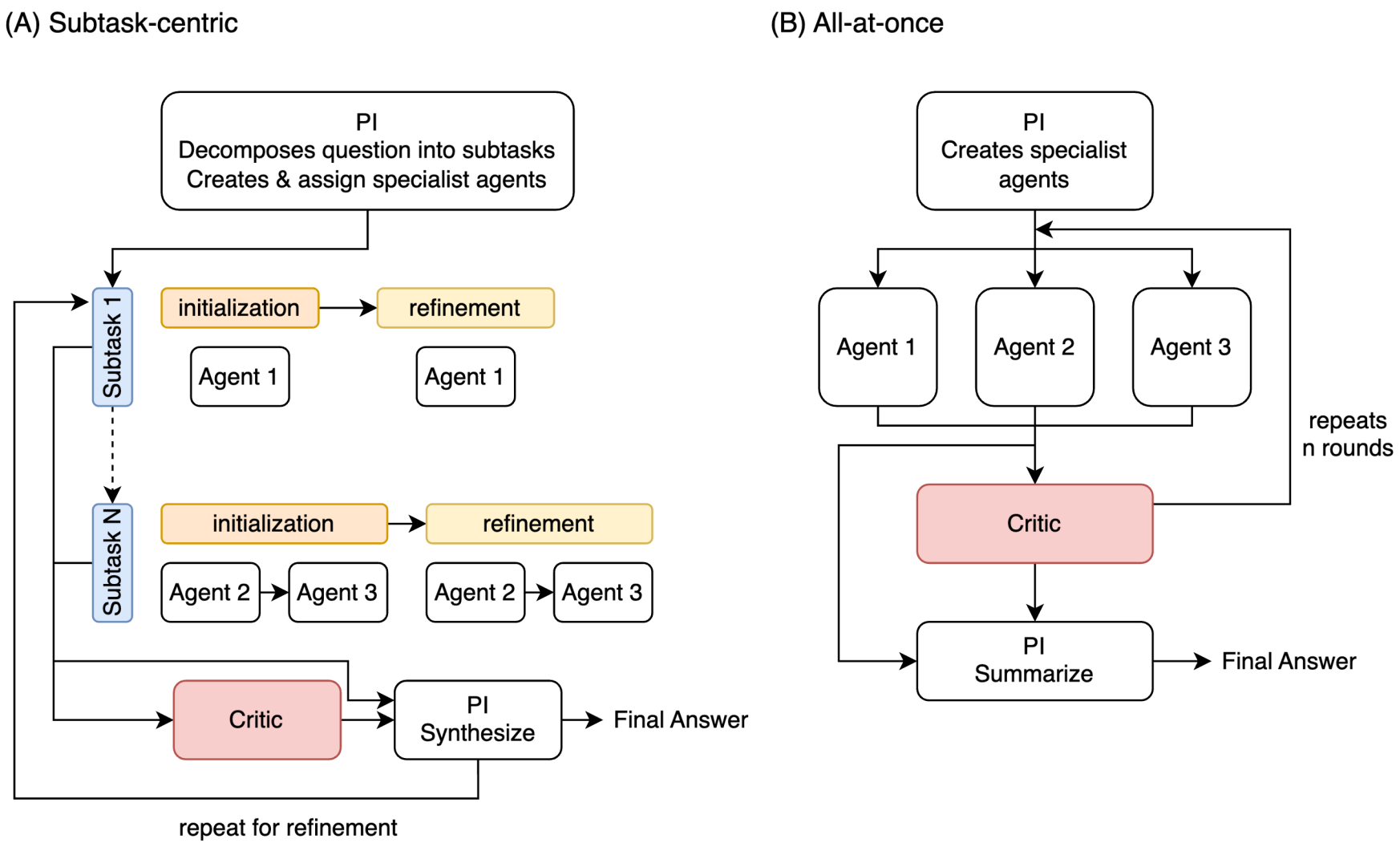
Seeun Kim

Yubeen Kim

Sujin Park

Tennis-Running-Swimming Seok Lab

Two research modes



Q5 cross-model check example

Based on my verification, here's the definitive answer:

✅ VERIFIED ACCURATE Claims (Not Hallucinated)

Claim	q5final Value	Actual Data	Status
IL2 L35→L60 log2FC	+1.33	+1.33	✅ **Exact match**
CHEK2 L35→L60 log2FC	-1.18	-1.18	✅ **Exact match**
CCL8 L35→L60 log2FC	+3.49	+3.49	✅ **Exact match**
CCL12 L35→L60 log2FC	+3.00	+3.00	✅ **Exact match**
CCL1 L35→L60 log2FC	+2.77	+2.77	✅ **Exact match**
TRAV3-1 L35→L60 log2FC	+6.45	+6.45	✅ **Exact match**
Gene names exist	GM42878, DUXBL3, TRAV genes	Present in data	✅ **Verified**

❌ HALLUCINATED or INCORRECT Claims

Claim	q5final Value	Actual Data	Status
GM42878 L35→L60 upregulation	+32.6×	**~1.1×** (no change)	❌ **Hallucinated**
DUXBL3 L35→L60 upregulation	+24.0×	**~1.0×** (no change)	❌ **Hallucinated**
L35→L60 DEG count	3,923 genes	**264 genes** (	LFC
Exhaustion score L35	-0.269	**+0.517** (my calc)	⚠️ **Discrepant**
Exhaustion score L60	+0.534	**+0.251** (my calc)	⚠️ **Discrepant**

Q6 novel target example

자동화 답안
- **FGFR1c/FGF21**: 포도당 흡수, 지방 산화 등 지방세포 대사 촉진

휴먼 답안
- **ActRIIB/Myostatin :** 근육을 녹이는 경로 ← 차단. 근육 감소 최소화
- **GIP 수용체 :** 지방 세포에 직접 작용, 대사 저하 방지

REGENERON
mounjaro®

AI consensus: "ActRII와 GIPR는 GLP-1의 '실패 지점'을 직접적으로 보완하는 compensatory nodes인 반면, FGFR1c/FGF21은 GLP-1 이후 대사 건강을 증폭시키는 보조 축이기 때문에 본 문제의 핵심 요구(한계 극복)에서는 2순위로 판단."

휴먼 답변의 두 타겟은 이미 임상을 통과했거나 진행중. 즉 이미 실험적으로 충분히 밝혀져 novel target으로 보기 어려움.