Error Analysis

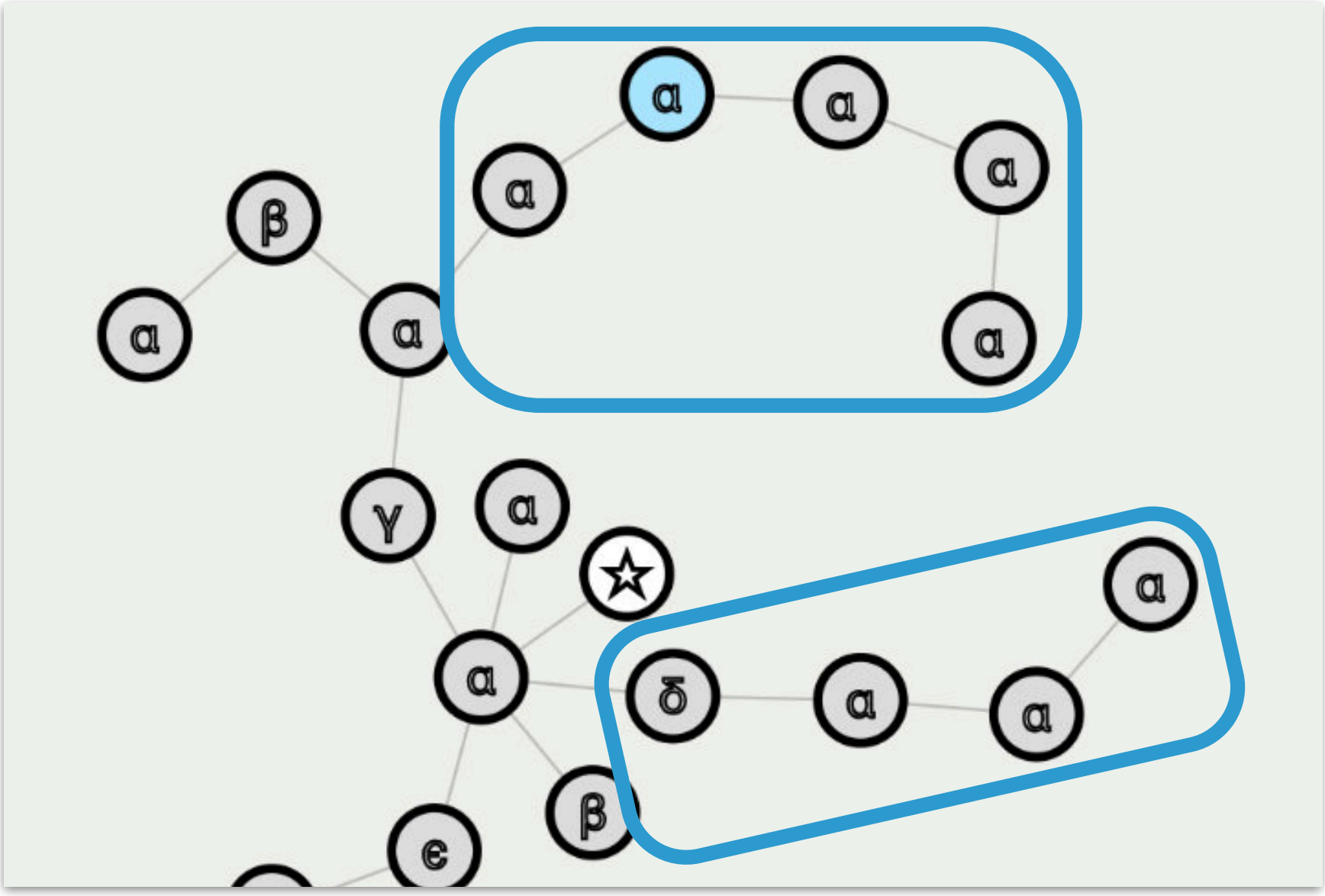
Your LLMs have commitment issues

- We manually investigated the logs of ReDel runs and found two main failure cases

- **Overcommitment**: An agent attempts to complete an overly-complex task itself.

- **Undercommitment**: The agent performs no work and re-delegates the task it was given.

System	FOQA		TP		WA	
	OC	UC	OC	UC	OC	UC
RD (4o)	22.7	11.3	41.1	0.5	31.3	44.8
RD (3.5-t)	40.8	1.1	96.7	0	54.6	17.7


α
α
α
α
α
β
α
α
γ
α
☆
α
δ
α
α
α
β
ε

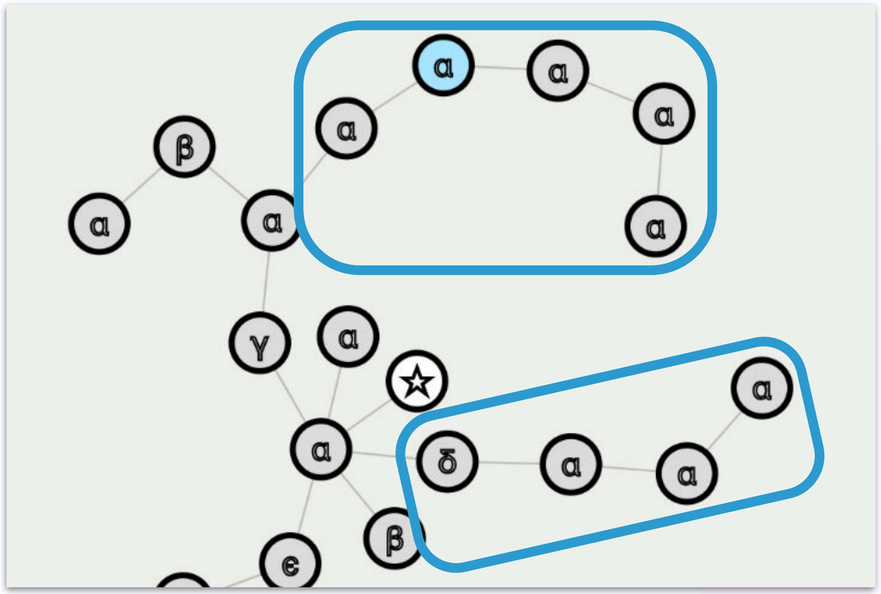
31

Error Analysis

Your LLMs have commitment issues

- We manually investigated the logs of ReDel runs and found two main failure cases
- **Overcommitment**: An agent attempts to complete an overly-complex task itself.
- **Undercommitment**: The agent performs no work and re-delegates the task it was given.

System	FOQA		TP		WA	
	OC	UC	OC	UC	OC	UC
RD (4o)	22.7	11.3	41.1	0.5	31.3	44.8
RD (3.5-t)	40.8	1.1	96.7	0	54.6	17.7



Future Research Directions