**[Anthropic](https://www.linkedin.com/company/anthropicresearch/" \t "_self)** 𝗷𝘂𝘀𝘁 𝗿𝗲𝗹𝗲𝗮𝘀𝗲𝗱 𝗮 𝗕𝗶𝗯𝗹𝗲 𝗳𝗼𝗿 𝗔𝗜 𝗔𝗴𝗲𝗻𝘁 𝗗𝗲𝘃𝗲𝗹𝗼𝗽𝗲𝗿𝘀  
  
It talks about how they developed multi-agent research systems -   
  
1. 𝗕𝗲𝘆𝗼𝗻𝗱 𝗦𝗶𝗻𝗴𝗹𝗲 𝗔𝗴𝗲𝗻𝘁𝘀  
  
➜ Single agents hit context and time walls on complex research tasks, where bigger models aren't the solution—orchestrated coordination is  
  
➜ When one Claude model stalls on broad queries like "list every S&P 500 IT board member," parallel subagents with separate context windows can explore different angles simultaneously  
  
2. 𝗧𝗼𝗸𝗲𝗻 𝗘𝗰𝗼𝗻𝗼𝗺𝗶𝗰𝘀  
  
➜ Multi-agent systems consume ~15× more tokens than standard chat, but token volume alone explained 80% of performance variance on BrowseComp evaluations  
  
➜ The trade-off is clear: spend more tokens to unlock capabilities that single agents fundamentally cannot achieve  
  
3. 𝗢𝗿𝗰𝗵𝗲𝘀𝘁𝗿𝗮𝘁𝗼𝗿 𝗣𝗮𝘁𝘁𝗲𝗿𝗻  
  
➜ A lead agent analyzes queries, develops strategy, and spawns specialized subagents with focused objectives  
  
➜ Each subagent operates independently with its own tools and context window, then returns distilled insights, preventing chaos of unstructured multi-agent interactions  
  
4. 𝗜𝗻𝘁𝗲𝗹𝗹𝗶𝗴𝗲𝗻𝘁 𝗖𝗼𝗺𝗽𝗿𝗲𝘀𝘀𝗶𝗼𝗻  
  
➜ Traditional RAG uses static chunk retrieval while multi-agent research uses adaptive, multi-step search  
  
➜ Subagents act as intelligent filters—exploring different aspects in parallel and compressing vast information into key insights for the lead agent  
  
5. 𝗖𝗼𝗼𝗿𝗱𝗶𝗻𝗮𝘁𝗶𝗼𝗻 𝗖𝗼𝗺𝗽𝗹𝗲𝘅𝗶𝘁𝘆  
  
➜ Coordination complexity grows rapidly, with early agents making errors like spawning 50 subagents for simple queries or scouring endlessly for nonexistent sources  
  
➜ Success requires embedding clear heuristics: delegate with explicit objectives and output formats, scale effort by complexity (1 agent for facts, 10+ for open research), and start broad then narrow based on findings  
  
6. 𝗣𝗿𝗼𝗱𝘂𝗰𝘁𝗶𝗼𝗻 𝗥𝗲𝗹𝗶𝗮𝗯𝗶𝗹𝗶𝘁𝘆  
  
➜ Agents are stateful and non-deterministic—small changes cascade into large behavioral shifts that can't be fixed with simple restarts  
  
➜ Solutions include checkpointing progress, full execution tracing, rainbow deployments for live updates, and robust retry logic  
  
7. 𝗢𝘂𝘁𝗰𝗼𝗺𝗲 𝗘𝘃𝗮𝗹𝘂𝗮𝘁𝗶𝗼𝗻  
  
➜ Unlike traditional software with fixed paths, agents take different valid routes to the same goal, making standard evaluation methods inadequate  
  
➜ Effective evaluation uses LLM judges for scalable assessment plus human reviewers for edge cases, focusing on whether agents achieved the right outcomes through reasonable processes  
  
8. 𝗔𝘀𝘆𝗻𝗰𝗵𝗿𝗼𝗻𝗼𝘂𝘀 𝗙𝘂𝘁𝘂𝗿𝗲  
  
➜ Current systems run subagents synchronously—the lead waits for all to complete, which simplifies coordination but limits parallelism  
  
➜ Asynchronous execution where agents spawn new agents mid-task promises major performance gains, but introduces challenges in state consistency and result merging  
  
P.S. Check out my profile for more resources on AI Agents 👋