Emails from thenuancedperspective+10-day-agentic-ai-course@substack.com

Subject: Day 9: Real-World Agentic Systems

Body: View this post on the web at https://thenuancedperspective.substack.com/p/day-9-real-world-agentic-systems  
  
  
  
​Missed previous days’ posts? Find them here [ https://substack.com/redirect/5cf7f540-2ad7-4c1e-938d-761915709de2?j=eyJ1IjoiNWRsNHd4In0.9s-meWzGjhoUinHaqHT72VjjydZlHVP0pxRYVWyxmVQ ]  
  
So far, we’ve covered all the ingredients that make up an agent: tools, planning, RAG, memory, structure, and even coordination in multi-agent setups.  
  
But you might be thinking:  
  
“Where does all this actually show up in the real world?”  
  
Let’s walk through a few public-facing systems that exhibit agentic behavior — as far as we can tell.  
  
⚠️ A quick note:  
  
These aren’t open source. We don’t know their exact internals. What follows is an informed simplification based on how they behave externally — just enough to understand how the agentic stack might show up in practice.  
  
NotebookLM (Google): Agentic Search on Your Own Data  
  
Google’s NotebookLM [ https://substack.com/redirect/f14d5631-68d4-4530-93c2-e426aa1c69ee?j=eyJ1IjoiNWRsNHd4In0.9s-meWzGjhoUinHaqHT72VjjydZlHVP0pxRYVWyxmVQ ] acts like a personal research assistant. You upload your files, and it helps you work with them — summarizing, answering questions, even generating audio versions or study guides.  
  
Let’s focus on its core: Q&A over your content — essentially a scaled-up, personal RAG system.  
  
How it likely works:  
  
User uploads files (PDFs, notes, slides, etc.)  
  
Preprocessing  
  
The system stores them in a way that allows for retrieval later.  
  
User asks a question  
  
e.g., “What were the key insights from my Q2 strategy deck?”  
  
Planning  
  
Interprets task type — summary, Q&A, comparison?  
  
Identifies relevant docs or sections.  
  
RAG  
  
Retrieves only the most useful parts of your documents.  
  
LLM Generation  
  
Responds clearly and grounded in your actual content.  
  
Memory  
  
Short-term: Keeps track of the current conversation.  
  
Long-term: Likely minimal or none.  
  
Tools  
  
Might include file viewers, summarization modules, etc.  
  
What makes it agentic: It interprets user goals, searches across your data, and composes clear responses — rather than just giving static outputs.  
  
Perplexity: Agentic Search on the Open Web  
  
Perplexity [ https://substack.com/redirect/a578110c-d110-49bb-89f9-d29d67362963?j=eyJ1IjoiNWRsNHd4In0.9s-meWzGjhoUinHaqHT72VjjydZlHVP0pxRYVWyxmVQ ] aims to do what traditional search engines like Google don’t, give you a direct, answer-like response with sources, instead of a page of links.   
  
How it likely works:  
  
User asks a question  
  
e.g., “What’s the latest research on Alzheimer’s treatments?”  
  
Planning (implicit or explicit)  
  
Interprets the intent — What does "latest" mean? What counts as credible?  
  
Tool Use  
  
Makes search queries via web APIs (search tool).  
  
RAG  
  
Retrieves relevant content from web pages to feed into the model.  
  
LLM Response  
  
Synthesizes a coherent answer from the retrieved content, often with citations.  
  
Memory  
  
Short-term: Likely remembers session context.  
  
Long-term: May store preferences (e.g., “always use WSJ for news”).  
  
What makes it agentic: It fetches info, chooses what to use, and constructs an answer, all in a multi-step reasoning loop.  
  
DeepResearch (OpenAI): Deep Agentic Workflows  
  
DeepResearch [ https://substack.com/redirect/6883e86a-9922-4fae-8322-3c4b627feffe?j=eyJ1IjoiNWRsNHd4In0.9s-meWzGjhoUinHaqHT72VjjydZlHVP0pxRYVWyxmVQ ] is designed for open-ended, complex research tasks, things like market analysis, competitive landscapes, or technical deep dives.  
  
This is where the agent stack really stretches its legs.  
  
How it likely works:  
  
User asks a broad task  
  
e.g., “Analyze the generative AI landscape for education startups.”  
  
Planning  
  
Breaks it into subtasks: funding, trends, companies, risks  
  
Forms a plan or flow of actions — a mini execution graph  
  
Tools  
  
Likely integrates:  
  
Web search  
  
Document readers (e.g., PDFs)  
  
Data tools (e.g., spreadsheets, graphs)  
  
Report generation modules  
  
Agentic RAG  
  
This isn’t one-shot retrieval. It fetches, reflects, and re-fetches as the task evolves.  
  
Memory  
  
Episodic: Tracks which parts of the task are done  
  
Semantic: Stores key facts, names, insights  
  
Multi-step Reasoning  
  
This is the “agent loop” in action:  
  
Plan → retrieve → read → rethink → generate → refine → repeat  
  
What makes it agentic: Heavy planning, iterative tool use, and the ability to drive a task forward on its own.  
  
If you play around with the above agentic systems they feel like very different products but actually rely on the same foundational components we’ve been learning — tools, planning, memory, and RAG!  
  
Can you guess what level of agents they are? Based on what we covered back in Day 2 [ https://substack.com/redirect/4dd041cb-97bd-411b-a99d-297714227e02?j=eyJ1IjoiNWRsNHd4In0.9s-meWzGjhoUinHaqHT72VjjydZlHVP0pxRYVWyxmVQ ], here’s a rough guess (since none of these are open source, we can’t say for sure):  
  
NotebookLM looks like it sits between level 2 and level 3. It behaves more like a high-control workflow agent. It retrieves well, maybe does complex RAG behind the scenes, but it doesn’t make decisions on your behalf. You’re still directing it step by step.  
  
Perplexity leans towards level 3 and maybe even level 4. You give it a query and it plans what to search, how to organize sources, and how to answer. It has more decision-making power baked in. You’re delegating the task, not just asking for help.  
  
DeepResearch feels like a strong level 4 system. It takes a high-level task, breaks it down, searches, reasons, and builds out a full report. It reflects, re-asks, and loops. You’re not guiding it through each step, it acts much more independently.  
  
If you’re curious, try them out. They all have free versions. Watch for how much control you still have and how much the system is deciding on its own. It’s a good way to build that instinct for agent design.  
  
Tomorrow, at the same time (for one last time), we’ll wrap things up, summarise what we’ve learned, go over some best practices, and take a quick look at where agentic AI is headed.  
  
-Aish [ https://substack.com/redirect/bf3f7fc8-d76e-4d2c-8def-5defe378afb1?j=eyJ1IjoiNWRsNHd4In0.9s-meWzGjhoUinHaqHT72VjjydZlHVP0pxRYVWyxmVQ ]​  
  
----------------------------------------------------------------------------------------------------  
  
All the frameworks and abstractions you've seen above are original, developed through hands-on experience building over 100 real-world AI applications across enterprise settings. You won’t find them packaged like this anywhere else online.  
  
If this approach resonates with you, and you're serious about learning how to design agents that actually work in production, not just hype demos, check out our top-rated 6-week course here [ https://substack.com/redirect/04250d3b-c061-4143-8d1f-4e5408fbe342?j=eyJ1IjoiNWRsNHd4In0.9s-meWzGjhoUinHaqHT72VjjydZlHVP0pxRYVWyxmVQ ]!  
  
Use the code 10DAY for a limited time 10% off!  
  
It's built for all backgrounds, whether you're a Product Manager, Architect, Director, C-suite leader, or someone exploring agentic AI with intent. Let’s move beyond the buzzwords and into systems that solve real problems!  
  
The Nuanced Perspective is a reader-supported publication. To receive new posts and support my work, consider becoming a free or paid subscriber.  
  
  
  
Unsubscribe https://substack..GzqtWPp3L45aTzn7AtGQvfIsUgsBP1yO07hRKyj\_ZWc?

End of email

Subject: Day 10: AI Agent Lessons and What’s Ahead

Body: View this post on the web at https://thenuancedperspective.substack.com/p/day-10-ai-agent-lessons-and-whats  
  
  
  
Missed previous days’ posts? Find them here [ https://substack.com/redirect/36a3d590-2ea1-4312-82ec-f225501b34d4?j=eyJ1IjoiNWRsNHd4In0.9s-meWzGjhoUinHaqHT72VjjydZlHVP0pxRYVWyxmVQ ]  
  
We’ve spent 9 days getting under the hood of agentic AI. Today’s the last day, and we’re bringing it all together.  
  
A Quick Recap  
  
Here’s what we covered:  
  
What agents are (Day 1): not just chatbots that generate text, but systems that can decide and act.  
  
Types of agents (Day 2): from tightly controlled workflow agents to fully autonomous ones, all depending on how much decision-making you hand over.  
  
Tools and RAG (Days 3–4): the bread and butter of agent action and knowledge grounding.  
  
MCP (Day 5): a clean way to structure everything an agent needs, tools, memory, prior messages, in one payload.  
  
Planning and reasoning models (Day 6): why plain LLMs aren’t enough for complex decisions, and how newer models are built for multi-step tasks.  
  
Memory (Day 7): short-term vs. long-term memory, what to store, how to retrieve, and why it matters for continuity.  
  
Multi-agent systems (Day 8): orchestration, peer-to-peer collaboration, and the messiness of coordination.  
  
Real-world systems (Day 9): how Perplexity, NotebookLM, and DeepResearch likely use these patterns in different ways.  
  
We’ve covered the moving parts that show up in real-world systems. But all of it falls apart if you’re not thinking about two things: observability and evaluation.  
  
What’s Still Hard  
  
Observability  
  
Observability means tracking what your agent is actually doing — at every step. You’ll want:  
  
Logs of tool calls, decisions, retries  
  
Metrics to spot bottlenecks in latency and cost  
  
Visibility into when things go off-rail  
  
Step-wise traceability for debugging  
  
Tools like Comet Opik [ https://substack.com/redirect/944ce36b-81cd-415f-942a-79c8ced6c49a?j=eyJ1IjoiNWRsNHd4In0.9s-meWzGjhoUinHaqHT72VjjydZlHVP0pxRYVWyxmVQ ] help with this. You need to design observability into the system from day one, especially for high-autonomy agents.   
  
Evaluation  
  
Agents are non-deterministic. So you need continuous evaluation, not just manual testing. At a minimum, track:  
  
Goal or task completion rates  
  
Tool call success and failure  
  
RAG quality and hallucination metrics  
  
Model overthinking or inefficiency  
  
Latency and token usage at each step  
  
Evaluation is how you understand and improve your system.   
  
This is also where most teams get stuck: doing vibe checks instead of real evals, and ending up in PoC purgatory.  
  
Think of evals and observability as your testing pipeline — the agentic equivalent of software QA. The above metrics aren’t exhaustive; what you track will depend heavily on the specifics of your use case.  
  
Where Things Are Headed in Agentic AI  
  
This space is still early, but here’s where things are clearly moving.  
  
1. Protocols > Prompts  
  
Image Source: Reuven’s LinkedIn post [ https://substack.com/redirect/d42751d1-beb8-4bcc-92e8-8b68716b659f?j=eyJ1IjoiNWRsNHd4In0.9s-meWzGjhoUinHaqHT72VjjydZlHVP0pxRYVWyxmVQ ]  
  
As systems get more complex, we’ll move away from handcrafted prompts and toward shared standards. MCP (Model Context Protocol) standardizes how we package up structured context (tools, memory, RAG, and prior instructions) into the model.  
  
A2A, released by Google just a few days back, focuses on agent-to-agent communication, allowing agents across platforms to collaborate with a shared schema.  
  
Expect cleaner abstractions going forward, though it’ll take time before any of these become as standard as HTTP.  
  
2. Hybrid Reasoning Models  
  
Reasoning models are evolving. The future isn’t using one type of model for everything, but having models that know when to plan and when to act fast.  
  
We’re already seeing this with Claude 3.7 and similar models. The goal is to balance intelligence with efficiency — without overthinking everything.  
  
3. Better Memory Systems  
  
Memory today is mostly duct-taped in. Going forward, expect tighter, smarter, and more context-aware memory integration, memory that knows what to recall, when, and why. Think memory scoped to tasks, sessions, or personas, and easier to manage.  
  
4. Tool Ecosystem Maturity  
  
Right now, everyone’s building custom tools and wrappers. But over time, expect a more robust ecosystem, trusted, plug-and-play APIs, better abstraction layers, and shared security practices.  
  
The same way microservices matured in traditional software, tools will mature in the agentic stack too.  
  
A Final Word  
  
If you’ve followed along, you’ve probably noticed the theme:  
  
We didn’t start with architecture. We started with problems.  
  
That’s the real mindset shift. Don’t chase agents for the hype. Build them when they make solving a problem easier, faster, or smarter.  
  
Start simple. Measure everything. Scale when needed.  
  
Agent-first thinking breaks. Problem-first thinking scales.  
  
Thanks for reading, sharing, replying, and thinking along over these 10 days. Thanks again for following along. If you take away one thing from this series — let it be this: problem first, always.  
  
Onwards.  
  
– Aish [ https://substack.com/redirect/ac07be42-bc20-408e-b0ad-6cca10f1c270?j=eyJ1IjoiNWRsNHd4In0.9s-meWzGjhoUinHaqHT72VjjydZlHVP0pxRYVWyxmVQ ]  
  
If this was helpful, feel free to forward it to someone looking to learn in this space.  
  
And if you’d like to go deeper, our [ https://substack.com/redirect/463617c8-7f03-4384-84f7-3f0efe75864a?j=eyJ1IjoiNWRsNHd4In0.9s-meWzGjhoUinHaqHT72VjjydZlHVP0pxRYVWyxmVQ ]full 6-week course covers system design [ https://substack.com/redirect/463617c8-7f03-4384-84f7-3f0efe75864a?j=eyJ1IjoiNWRsNHd4In0.9s-meWzGjhoUinHaqHT72VjjydZlHVP0pxRYVWyxmVQ ], applied agentic concepts, and real evaluation workflows, the kind that support production-grade applications.  
  
The course is built for everyone, whether you’re a Product Manager, Architect, Director, C-suite leader, or someone seriously exploring agentic AI.  
  
P.S. Sharing two of our favorite cohort reviews below. You can find 50+ more on the course site. We’re the only AI course on the platform with a full 5-star rating.  
  
If you’re planning to enroll, use code: 10DAY for 10% off.  
  
The Nuanced Perspective is a reader-supported publication. To receive new posts and support my work, consider becoming a free or paid subscriber.  
  
  
  
Unsubscribe https://substack..22iML2r5I0pRELkNuw5R8xtsjTUFTGFwSKd832uklG0?

End of email