AI-Powered Economic Research workflows

Augmented Research Workflows

Goal: To introduce the concept of AI as a collaboration tool, and AI agents as a powerful new tool for improving research workflows. Want to get people thinking about how to use AI to improve their research workflows while highlighting limitations and potential.

The Evolving Landscape

1. **Traditional Workflow**: Manual literature reviews, Researcher driven coding
2. **New Wave: LLM-powered AI Augmentation**: AI as collaborative partner, mostly through a chat interface. Similar to early days of "googling". What are peoples mental models of how AI works?
3. **Next wave: LLM powered by tools, e.g. AI agents**: AI Agents as workflow enablers

Demystifying AI Agents

What are these new tools?

AI Agents are mostly autonomous systems built on LLMs by giving them access to tools that can:

- ▶ Manage context, e.g. using memory to keep track of previous steps
- ▶ Create detailed plans
- ▶ Use various tools (e.g. web search, code execution, data analysis, write reports)
- ▶ Execute multi-step tasks

The Leap Forward:

1. **Chatbots** : Basic Q&A capabilities, single-turn responses, limited context retention
2. **Reasoning Engines** : Multi-step thinking, improved context handling, still confined in scope
3. **AI Agents** : Proactive problem-solving, advanced tool utilization, autonomous goal pursuit

Key Analogy: Think of an AI Agent as a research assistant you can delegate tasks to, not just a calculator for computations.

Some example Research Agents

1. **Automated Literature Synthesis** : Comprehensive source scanning, Automated citation management, Intelligent synthesis

2. **Code Generation & Debugging** : Natural language to code, Multi-language support (Python, R, Stata), Intelligent debugging

3. **Autonomous Data Handling** : Automated data collection, Smart cleaning algorithms, Advanced analysis tools, Multi-source integration

4. **Workflow Orchestration** : End-to-end automation, Simulation management, Results analysis, Visualization generation

What does using Agents look like?

1. **Development Environment**: Modern IDEs integration, Seamless workflow integration, Real-time assistance

2. **Agent Interaction** : Chat-based interface, Customizable personas, Task-specific fine-tuning, Economic research specialization

3. **Available Platforms** : Notebook LM with research tools, Gemini/OpenAI/Anthropic integration, VSCode with extensions (e.g., Clive), Cursor's built-in agent support, AWS bedrock, Azure openai, Google vertex ai, Python frameworks like langchain.

4. **Deep research** : Gemini, Claude, OpenAI all have their own ways to do deep research. Its basically where you have an LLM attached to tool, you instruct it what to do then it returns with a report. I have used Gemini for deep research, the most. Use it for the planning phase of a tool build.

What are tools in the Agent contest?

- ▶ Tools are anything which the LLM has access to which it can use to complete a task
- ▶ MCP is how LLMs access different tools, think of it as an api wich allows the LLM to interact with something

Context is King

- Context refers to the memory of an LLM. There are often limits to the context of different models
- Hallucinations are more commond as we fill up the tokens the model is tracking.
- LLM and agents excell at well scoped task where full context is given, creating a plan before executing is important to think through the task and the results.
- Documenataion and githhub can add context and add value to the agents

How can we trust AI based collaboration?

- ▶ Trust but verify. When creating a plan, always think about how to verify the results.
- ▶ Do we have a dataset we trust? Then we have labeled data we can test against and create score against.
- ▶ Create a test data to run a simulation to unit test the model being developed.

The need for Evaluation

▶ As we thinking about what we want agents to do, we need to think about how to evaluate them.

▶ Having general benchmarkets is important to help us understand the performance of the agents across different models and domains.

Model			Metrics			Scores					
Name Context Window	Price In / Out		Duration	Tokens In / Out	Cost USD	🗑	☕	JS	🐍	®	Total
GPT-5 Mini 400K	$0.25 / $2.00		5h 46m 33s	14M / 977K	$3.34	100%	98%	100%	100%	97%	**99%**
Claude Opus 4.1 200K	$15.00 / $75.00		7h 3m 6s	27M / 490K	$140.14	97%	96%	98%	100%	100%	**98%**
GPT-5 (Medium) 400K	$1.25 / $10.00		8h 40m 10s	14M / 1M	$23.19	97%	98%	100%	100%	93%	**98%**
Claude Sonnet 4 1M	$3.00 / $15.00		5h 35m 31s	39M / 644K	$39.61	94%	100%	98%	100%	97%	**98%**
Gemini 2.5 Pro 1M	$1.25 / $10.00		6h 17m 23s	43M / 1M	$57.80	97%	91%	96%	100%	97%	**96%**
GPT-5 (Low) 400K	$1.25 / $10.00		5h 50m 41s	16M / 862K	$16.18	100%	96%	86%	100%	100%	**95%**

The Critical Importance of Data Security

LLMs from big providers collect your data and we need to be careful about what is shared.

1. **Security Challenges** : Public AI tool risks, data exposure concerns, privacy compliance, regulatory requirements

2. **AWS Bedrock Solution** : Private cloud environment, access to some but not all public models, secure infrastructure

3. **Key Guarantees** : No external data training, complete confidentiality, virtual private cloud, data integrity preservation

Case Study: Analyzing a Hospital Merger

A research team evaluates economic impacts of a proposed hospital merger using AI-powered analysis.

AI-Assisted Workflow

1. **Initial Task Delegation** : High level of payoff for human involvement. Set goals, plan and how to verify. Can use to highlight ambiguities in the plan. Which API should we use? Do we have test data? How should we validated? What metrics should we use? Whats the final product look lile? Do we have key figures we are interested? (Researcher as manager)

2. **Data Collection Phase** : Use a coding assistant (e.g. Cursor, VScode) to write a script to scrape news filings of hospitals mergers in the last year in and in this year. Get a test dataset to compare results against.

3. **Analysis & Visualization** : Given data collected, use coding assistant to write a report that give key stats and figures. Do for new year and last year, to give us a benchmark. Identify outlier or cases to validate by hand.
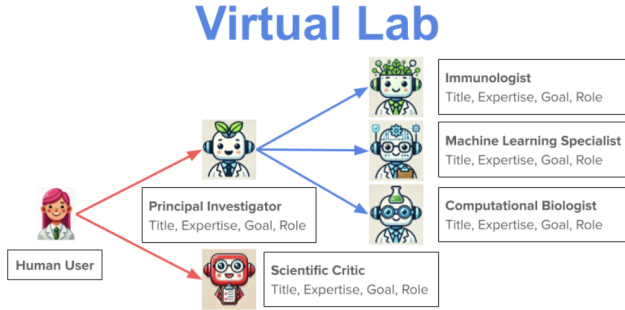
Future steps

Setup to run autonomously every X months. Researcher will review report and validated results, maybe through a random sampling of cases.
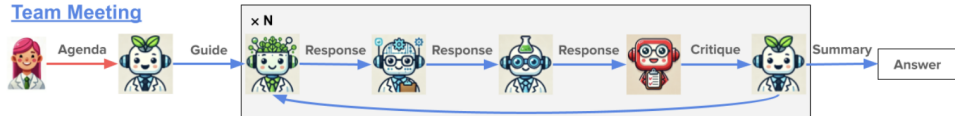
Towards a virtual Lab

1. **Core Definition** : A secure environment with access to specialized agents which agents can work within. Agents can coordinate and collaborate, learn autonomously and solve complex tasks.

Key takeaways

1. **Transformation**: Researchers are becoming managers adding value by providing expertise and guidance to the agents.
2. **Security & Trust** : Done in a sandboxed environment where AI agents can run analysis. Option to protect data from LLM provider.
3. **Empowerment**: Knowing the what you want to do and how to verify lowers the barrier to using new toolset

Discussion Points

1. **Quick Wins**
   - Which tasks to automate first? (merger checking, data downloading, closure researcher)
   - What's easily achievable?
   - How to measure success?
2. **Challenges**
   - Handling hallucinations
   - Maintaining quality
3. **Next Steps**
   - Choose tasks to focus an AI-enhacned workflow
   - Which tools should the team adopt?