AI and Research Practice

Lab 2: Context and verification when a right answer exists

Derek Harmon

November 17, 2026

BROAD COLLEGE *of* BUSINESS

Today

1. **Motivation**
   Where we are in the series.
2. **Where we left off**
   Agents, agentic architecture, and the terminal.
3. **Types of questions**
   Type 1 vs. Type 2, and the range of Type 1 tasks.
4. **Demos – Type 1 questions**
   Literature and a paper check.
5. **Conclusion**

QUESTIONS ANYTIME
Put them in the Zoom chat.

1 Motivation

Reminder of our goals

1) Most conversations are about permissibility or the future of AI.
    - **We are focused on how people can use these tools to get high-quality output**.

2) Create a community where people can:
    - **develop their intuitions**
    - **learn practical skills**
    - **talk with others that are interested**

An ongoing caveat

I use these tools heavily and have research on the effective uses of these tools. But I am always still learning and cannot claim to always be right.

Three labs

LAB 1 · NOV 10

Using agents and the terminal

Developing agent architecture and working at scale in the terminal.

LAB 2 · TODAY · TYPE 1

Context and verification when a right answer exists

Attaching sources and building checks that improve output accuracy.

LAB 3 · DEC 1 · TYPE 2

Context and verification when no right answer exists

Curating context and running robustness checks against what you intended.

2 Where we left off

Running agents in the terminal is where AI usefulness starts to scale.

What an agent is

A model that works toward a goal by taking actions, seeing the results, and deciding what to do next.

THREE BENEFITS

Independence

Agents that can't see each other don't anchor on each other. A disagreement thus tells you something useful, and a separate agent can reliably do a check.

Focus

Each agent sees only its piece and the material for it. Limits distractions and errors that arise when you provide context that isn't relevant to the task.

Scale

The work can be bigger than one ChatGPT conversation can hold.

SCRIPT

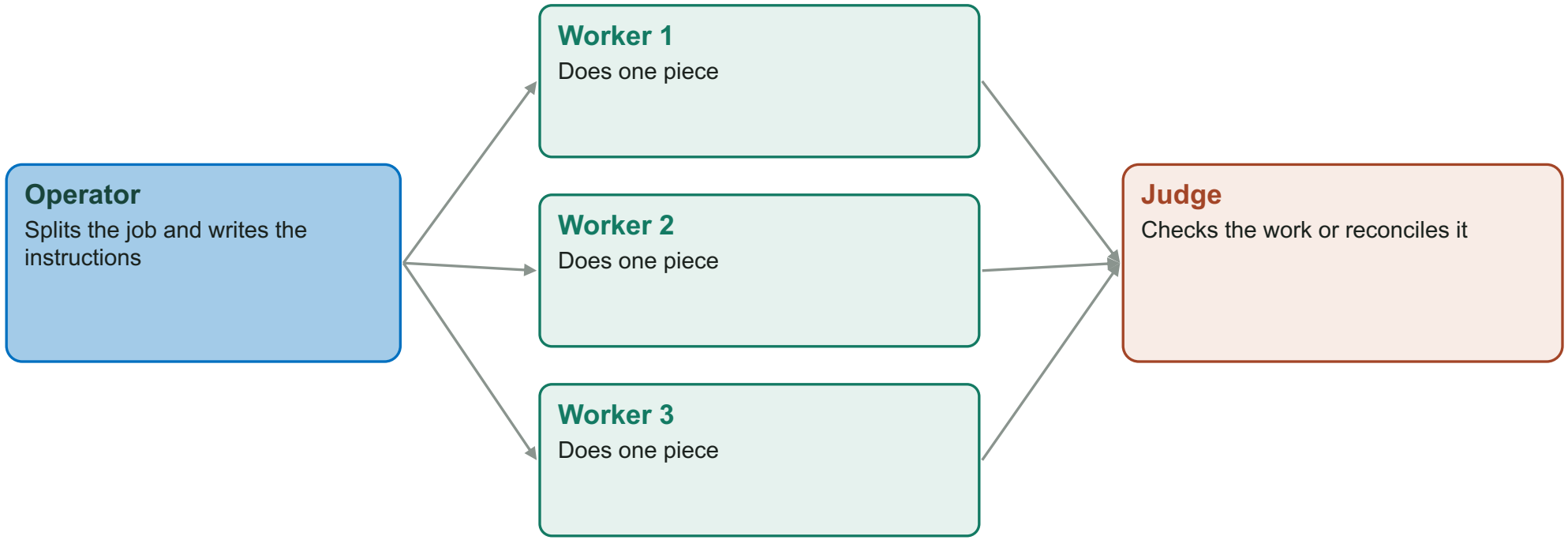
Follows steps written in advance.

AGENT

Gets a goal and works out the steps.

Several agents, wired together

AGENTIC ARCHITECTURE



Each agent has its own goal and its own context. Deciding what each one sees is the design work.

The terminal and its benefits

1 **It runs agents wired together**

Separate contexts, run in parallel, and the checker waits for every worker. No copy and paste.

2 **Your project is the context**

Point it at a folder of papers, data, or drafts. It reads what it needs.

3 **It acts, not just answers**

It runs code on your data, writes and edits files, and checks its own results.

4 **It’s repeatable**

The instructions live in a file you can open and edit. Next month, rerun the same folder.

Let’s pause to discuss the terminal a little more…

Getting used to the terminal

It can feel intimidating.

But I'll show you why it's more familiar than you think.

Getting used to the terminal

1 `Terminal`

The raw terminal

Moves around your folders
`cd ls`

Acts on files: copy, rename, delete, search
`cp mv rm grep`

Runs programs: Python, R, git, Claude Code

You won't need to type these.

2

Who uses it like this?

Developers, data scientists, and researchers running code on a server like HPCC

Everything is plain text, typed in and printed back

So it's fast, repeatable, and works on remote machines

That's also why an AI model can operate it.

3 Chat

ChatGPT/Claude

The chat can reason, write, and code

But it can't open your files, run your programs, or save anything to your computer

What if it could?

4 `Terminal`

Terminal + Codex or Claude

The agent works in the terminal for you

You give instructions in plain English. It types the commands.

By default, it asks before it changes anything and works only in the folder you choose

This is what you saw in Lab 1.

How-tos on the website

Step by step, for when you set this up yourself.

Setting up the terminal derekharmon.com/ai/setup	Open the terminal on a Mac or a Windows PC, install Claude Code or Codex, and run it in a folder.
Setting up Warp derekharmon.com/ai/setup/warp	A free terminal app that’s easier to read and type in, set up for Claude Code or Codex.
Claude Code and Codex, side by side derekharmon.com/ai/setup/agents	Plans, starting, instruction files, permissions, everyday commands, and MSU’s data rules.
Choosing a tool derekharmon.com/ai/tools	Which research moves you can make in a chat window, an agent, an API script, or a model on your laptop.

3 Two types of questions

Two types of questions

LAB 2 · TODAY

TYPE 1 · A RIGHT ANSWER EXISTS

You can check the output against a source: the paper, the data, the filing.

CONTEXT

Bring the external source into your context.

VERIFICATION

Still needed, but now quick.

LAB 3 · DEC 1

TYPE 2 · NO RIGHT ANSWER EXISTS

No source settles it, so you judge the output against what you intended.

CONTEXT

Decide your intention and bring it into your context.

VERIFICATION

Ask for many options and judge them yourself.

The range of Type 1 tasks

Lab 1, Nov 10 Today

Each task has a right answer that lives in an external source you can check.

RESEARCH TASK	#	TASK	EXTERNAL SOURCE
Literature review	**1**	Does this paper exist, and are the authors, year, and journal right?	The DOI record: Crossref, OpenAlex
	2	What did this paper actually find or claim?	The paper
	3	How do these papers' tested relationships fit together in a causal map (a DAG)?	Each paper's hypotheses and results tables
	4	How did the original paper define or measure X: its items, variable construction, or procedure?	The original paper and its appendix
Data collection	**5**	Pull data for a sample from a source like SEC EDGAR or WRDS (Compustat, CRSP, ExecuComp).	The database
	6	Extract fields from documents and records: facts about firms (founding year, headquarters, executives), board members from proxies, deal terms from press releases, policies from journal websites.	The documents, filings, registries
	7	Match firms across datasets by name, CIK, GVKEY, or PERMNO.	The identifier crosswalks

The range of Type 1 tasks

Lab 1, Nov 10 Today

Each task has a right answer that lives in an external source you can check.

RESEARCH TASK	#	TASK	EXTERNAL SOURCE
Coding and measurement	**8**	Code a factual variable from a document: was this a real CEO change, was the new CEO an insider?	The document
	9	Apply a published codebook, wherever the codebook decides the case.	The codebook
	10	Have a second coder re-code a sample.	The same source
Analysis	**11**	Does the code build the variables and sample the way the methods section says?	The code and data
	12	Do the numbers in the text match the tables?	The tables
	13	Do the reported statistics agree with each other: the test statistic, degrees of freedom, p-value, and confidence interval?	The math
	14	Does your replication package run and reproduce every number in the paper?	The data, code, and paper

The range of Type 1 tasks

Lab 1, Nov 10 Today

Each task has a right answer that lives in an external source you can check.

RESEARCH TASK	#	TASK	EXTERNAL SOURCE
Writing and submission	**15**	Does each citation say what the cited paper says?	The cited paper
	16	Do the in-text citations and the reference list match, in the journal's style?	The manuscript, the style guide
	17	Does the manuscript follow the journal's rules: word limits, anonymization, AI disclosure?	The author guidelines
	18	Does the response letter match what the manuscript actually changed?	The manuscript

What we're covering today

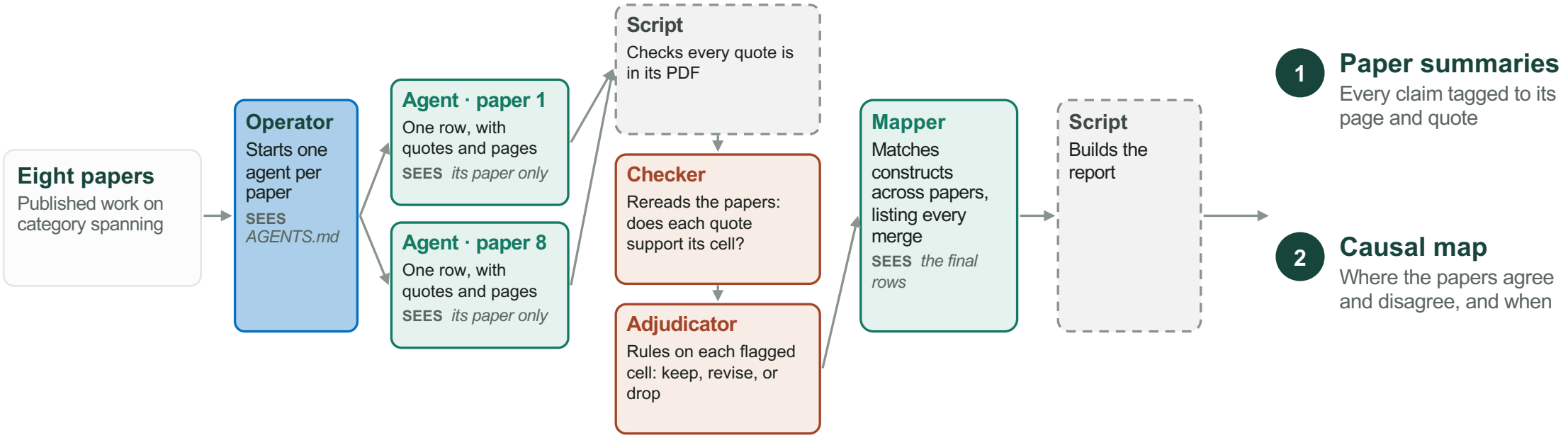
The Type 1 tasks in today's two demos: literature and a paper check.

DEMO	RESEARCH TASK	#	TASK	EXTERNAL SOURCE
DEMO 1	Literature review	2	What did this paper actually find or claim?	The paper
		3	How do these papers' tested relationships fit together in a causal map (a DAG)?	Each paper's hypotheses and results tables
DEMO 2	Analysis	12	Do the numbers in the text match the tables?	The tables
		13	Do the reported statistics agree with each other: the test statistic, degrees of freedom, p-value, and confidence interval?	The math
	Writing and submission	15	Does each citation say what the cited paper says?	The cited paper
		16	Do the in-text citations and the reference list match?	The manuscript
		18	Does the response letter match what the manuscript actually changed?	The manuscript

4 Demos – Type 1 questions

DEMO 1 LITERATURE REVIEW

From articles to a literature review and causal map



Each agent sees only what it needs. Scripts do the exact checks, and agents make the judgment calls.

DEMO 2 PAPER ACCURACY AUDIT

Checking the accuracy of my own paper (and revision)

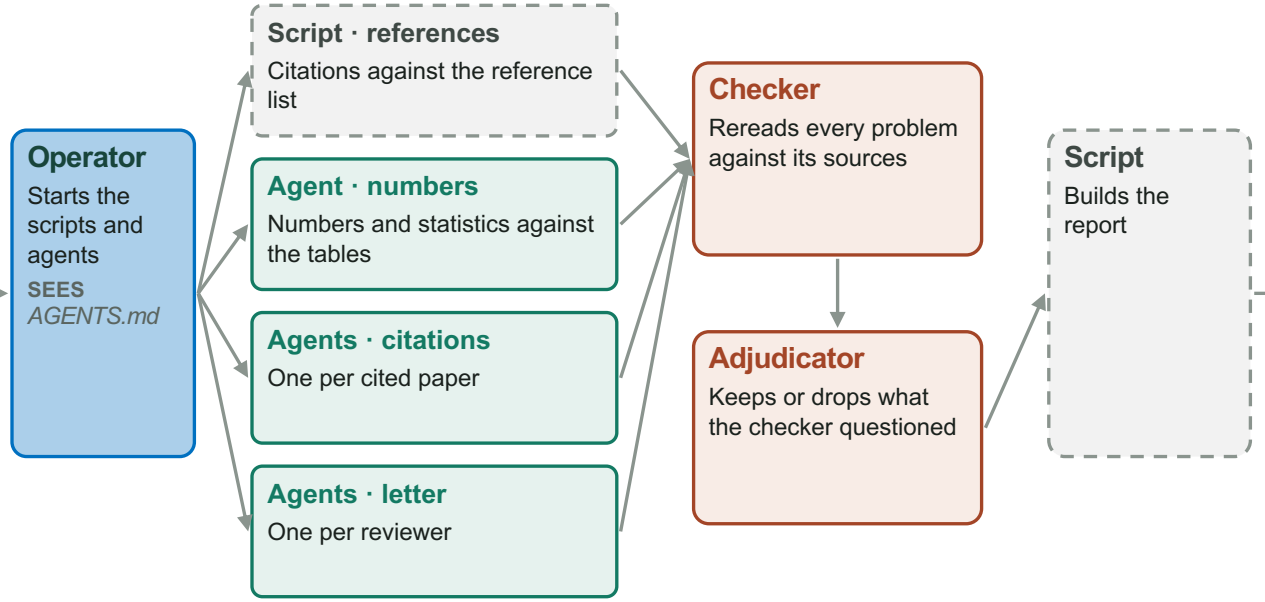
INPUT

My paper, Round 1
Manuscript, tables, and the earlier version

Response letter
What we told the reviewers we changed

Four cited papers
The ones mapped in Demo 1

AGENTIC ARCHITECTURE scripts in grey · agents in color



OUTPUT: ONE REPORT

1. **A count per check**
   How many items were checked, and how many problems hold up
2. **Every problem**
   Tagged to its page and quote, with the checker's and adjudicator's rulings

Each agent sees only what it needs. Scripts do the exact checks, and agents make the judgment calls.

What one run of each demo used

	RUN TIME	5-HOUR LIMIT	WEEKLY LIMIT
Demo 1 Literature: eight papers	7 min	51%	8%
Demo 2 Paper check: five checks	6 min	37%	6%
Both demos Including the live builds	13 min	90%	14%

Measured with Codex's usage meter in the Nov 15 rehearsal: ChatGPT Edu, GPT-6.1-Sol at medium effort. Runs vary. Earlier runs of Demo 1 used 29% to 43% of the 5-hour limit.

Open question: past either limit, Codex keeps running on monthly credits (32,000 a month on my account). MSU doesn't say who pays for those overages.

5 Conclusion

Three labs

LAB 1 · NOV 10

Using agents and the terminal

Developing agent architecture and working at scale in the terminal.

LAB 2 · TODAY · TYPE 1

Context and verification when a right answer exists

Attaching sources and building checks that improve output accuracy.

LAB 3 · DEC 1 · TYPE 2

Context and verification when no right answer exists

Curating context and running robustness checks against what you intended.

At Lab 3 on Dec 1, I'll share what I'm planning after these three labs.

Thank you!



TODAY'S RECORDING AND SLIDES

derekharmon.com/ai/lab/2

Posted after the session.

GUIDE, DEMOS, SLIDES, AI POLICIES, AND EVERYTHING ELSE

derekharmon.com/ai

The setup guide and the tools table are here too.