A Practical Guide to Organizing Research with AI Agents

A shared approach for planning, carrying out, checking, and documenting complex social science research projects.

PART I | THE OVERVIEW

The purpose of this approach is to make it possible for AI tools to carry out substantial parts of a research project without losing track of what the researchers want, what evidence is available, or which decisions require scholarly judgment. Researchers remain responsible for research design and substantive choices. AI agents help organize the work, write and run code, produce outputs, and check results.

> The central principle: Decide what the finished research project should contain; take stock of the materials already available; then write a step-by-step plan for getting from the starting materials to the finished outputs. Only after these three pieces are agreed upon should large-scale implementation begin.

The process at a glance

1. Describe the research project and the complete set of desired outputs

↓

2. Review and refine the description until it is precise (SPEC.md)

↓

3. List the data, code, software, and other starting materials (RESOURCES.md)

↓

4. Match each desired output to its required materials; resolve gaps

↓

5. Write an ordered set of small tasks and checks (PLAN.md)

↓

6. Carry out tasks in groups, with researcher decisions at key points

↓

7. Check that every promised output exists and can be reproduced

The work is recorded in three main documents. These documents are not the paper itself: they are shared instructions that allow researchers and AI agents to work consistently, even across many separate sessions.

Document	Question it answers	What it contains
Project specification (SPEC.md)	What exactly must the project produce?	All datasets, analyses, tables, figures, code, and paper-ready materials, with criteria for judging whether each is finished.
Resource inventory (RESOURCES.md)	What do we already have?	Locations and descriptions of existing data, scripts, folders, software, documentation, and their uses.
Work plan (PLAN.md)	How will we get there?	Small tasks, their order, required inputs, expected results, checks, and decisions that require a researcher.

Who makes which decisions?

The researcher decides the scientific questions, sample definitions, empirical strategies, interpretation, and any changes to agreed research outputs. An AI agent may identify options and recommend a course of action, but it must ask before changing substantive decisions. The agents can handle clearly defined technical work and routine checks within those boundaries.

How the work moves forward

The project is completed in groups of related tasks, called waves. One coordinating agent keeps track of the overall plan and communicates with the researcher. Other agents take on smaller assignments, such as preparing a dataset, running a model, or making a figure. When they find a problem or need a decision, they report it to the coordinating agent, which brings the question to the researcher.

> Example: An agent preparing a panel dataset discovers that some geographic units change boundaries over time. It should not quietly choose a method for handling the changes. It should explain the options and their consequences, then ask the researcher to decide before continuing with affected tasks.

PART II | STEP-BY-STEP GUIDE

1. Write down what the finished project must contain

Start with an ordinary-language description of the research question, main argument, population or setting, likely empirical strategy, and expected outputs. The initial description can be incomplete. Its purpose is to give the agent enough context to ask useful questions.

Step 1A. First description from the researcher

Explain what you hope to learn, what evidence you expect to use, and what the final paper should show. Mention planned descriptive statistics, models, comparisons, sensitivity analyses, figures, tables, and supporting materials, even if some details are uncertain.

Step 1B. Questions that fill in the missing details

Ask an agent to read the description and interview you about everything that remains unclear. For example: What is the unit of observation? Which years are included? How is treatment defined? Which outcome is primary? What is the main comparison? Which estimates belong in the paper versus the appendix? What evidence would make you revise the design? The agent should keep track of questions that have not yet been answered.

Step 1C. A critical review

Have the agent take the perspective of a skeptical co-author or reviewer. Its job is to find contradictions, missing pieces, and possible problems. If the main specification uses one sample but the proposed figure appears to use another, that discrepancy should be flagged. If a planned analysis requires data that may not exist, the agent should say so and propose alternatives. Researchers choose how to resolve these issues.

Step 1D. Produce the first specification

Bring the answers and revisions into one organized document. Give each output a short identifier, such as TABLE-01 or FIGURE-03. Describe what it shows, what data and methods it needs, what files should be produced, and how a reviewer can tell that it is complete. Include the underlying estimates and scripts when they are needed to reproduce a table or figure.

> Example: FIGURE-03 — Event-study plot
> The final materials should include a publication-ready figure, the numerical estimates used to draw it, and the code that produces it. The specification states the event-time periods, comparison group, omitted reference period, confidence intervals, sample, and model. A check confirms that the plotted points match the stored estimates and that the figure can be recreated.

2. Take stock of the materials available at the start

Before planning the work, build a clear inventory of what already exists. This is especially important for projects with several data sources, old analysis scripts, shared folders, or work carried out by different collaborators.

Step 2A. Explore the project materials

An agent can inspect the folder structure, read documentation, identify scripts and software requirements, and summarize datasets and their variables. It should ask the researcher about files whose purpose is unclear and about resources outside the main project folder. The inventory should record locations, contents, access limitations, and any known problems. This is a discovery exercise: the agent should not overwrite or reorganize files without permission.

Step 2B. Connect each desired output to its inputs

Now compare the inventory with the specification. For each figure, table, dataset, or report, identify the data sources, existing code, software, and intermediate results it will require. The goal is to make it easy to trace each finished result back to its inputs.

Step 2C. Look for gaps in both directions

First, find outputs for which a required input is missing. A proposed analysis may need a crosswalk, a survey weight, or a year of data that is not available. Second, identify resources that have not been assigned any role. They may be useful, redundant, obsolete, or intentionally outside the scope. The researcher decides how to handle important gaps.

Step 2D. Update the specification

Add references to the relevant resources for every output. Record unresolved gaps rather than pretending they are solved. At the end of this stage, the project description should be tied to actual available materials.

Desired output	Starting materials	Issue to resolve
Table of descriptive statistics	Survey microdata; variable dictionary; existing cleaning script	Confirm whether the same sample restrictions apply to every column.
Main regression table	Clean analysis dataset; estimation script; model decisions	Clarify clustering and missing-data treatment.
Map of study locations	Geographic identifiers; boundary files; plotting code	Check that boundary definitions match the study years.

3. Turn the project into a clear sequence of tasks

The work plan describes how to produce each output, beginning with the available resources. It is more useful than a simple to-do list because it shows which tasks depend on others and how each completed task will be checked.

Step 3A. Identify the main parts of the work

Begin with broad activities: obtain and inspect data; clean and combine sources; construct measures; produce descriptive results; estimate models; perform additional checks; generate tables and figures; assemble paper materials. Identify which activities can happen at the same time and which must wait.

Step 3B. Divide the broad activities into manageable assignments

Break each activity into tasks small enough for one agent to complete with clear instructions. Each task needs an objective, input files, expected output files, prerequisites, and a note about decisions that require a researcher. For example, "prepare the data" is too broad; "merge county-year unemployment rates into the county-year analysis file and report unmatched records" is a workable task.

Step 3C. Decide how each task will be checked

Before the task is carried out, specify concrete checks. A data merge might require unique county-year identifiers, a report on unmatched records, and a comparison of row counts before and after merging. A regression task might require the expected sample, saved coefficient estimates, and confirmation that the model matches the approved specification. Some checks can be run automatically; others require researcher judgment.

Example task: DATA-07 — Merge two panel datasets
Inputs: Clean county-year panel and annual unemployment data.
Action: Merge by county and year, without altering the source files.
Output: A new analysis dataset and a merge summary.
Checks: No duplicate county-year keys; unmatched records listed; observation counts explained; unemployment values fall within plausible bounds.
Researcher decision: Ask how to handle counties with missing years if the answer changes the study sample.

4. Carry out the work in groups of related tasks

Once the specification, inventory, and work plan have been reviewed, implementation can begin. Rather than asking one agent to carry out the entire project in a single conversation, work is divided across multiple sessions and smaller agents. This reduces confusion and makes it easier to check progress.

The coordinating agent

One agent acts as the main point of contact. It sees the overall plan, chooses tasks that are ready to start, assigns work, reviews reported results, and asks the researcher for decisions. It should know the current status of the project without needing to reread every file and every previous conversation.

The task agents

Other agents receive limited, specific assignments. They should be given the relevant section of the specification, the necessary input files, and the checks they must pass. They report what they changed, where outputs were saved, which checks passed, and what remains uncertain. Where agents work at the same time, their assignments should avoid editing the same files.

Choosing the next group of tasks

A group of tasks should be chosen based on what is ready, what can be done together efficiently, and where a researcher decision is likely to be needed. For example, data cleaning and basic summaries might form one group; model estimation another; and final tables and figures a later group. If a question arises midway through a group, affected work pauses while unrelated work may continue.

Illustrative flow of communication

Researcher — sets priorities and makes scientific decisions
↓
Coordinating agent — tracks the overall plan, assigns work, and reviews results
↓
Task agents — complete individual assignments and run checks
↑
Results, failed checks, and questions — return to the coordinating agent, which asks the researcher when necessary.

5. Keep changes under researcher control

The specification is the agreed description of what the project will produce. Once work begins, agents should treat it as fixed unless the researcher approves a change. This does not mean that the project can never evolve. Research often changes when new evidence appears. It means that changes should be deliberate, documented, and understood by everyone working on the project.

When an agent recommends a change, it should explain what it wants to change, why, which outputs and tasks would be affected, and what alternatives exist. The researcher approves or rejects the proposal. If approved, the specification and work plan are updated together so they do not contradict one another.

> Example: A planned subgroup analysis turns out to have too few observations for meaningful estimation. The agent documents the problem and proposes alternatives, such as changing the subgroup definition, presenting descriptive results only, or dropping the analysis. The researcher chooses; the agent does not silently substitute a different analysis.

6. Keep the shared instructions and history organized

Store the three documents with the project files in a Git repository. Git is a system that records changes to files over time. This makes it possible to see what changed, when it changed, and which version of the plan was being followed. It also helps collaborators avoid losing track of approved decisions.

A short file called CLAUDE.md should explain the working rules to any new AI session. It should tell the agent where the three main documents live, which one is authoritative, what stage the project is in, what it should read before starting, what it is allowed to change, how to record its work, and when it must ask the researcher for approval. This file is a set of operating instructions, not a replacement for the detailed documents.

Illustrative project folder

```
project/
CLAUDE.md — instructions for AI agents
governance/SPEC.md — agreed final outputs
governance/RESOURCES.md — starting materials and their locations
governance/PLAN.md — tasks, order, and checks
governance/CHANGELOG.md — important approved changes
data/ — datasets
src/ — scripts and analysis code
outputs/ — tables, figures, and results
tests/ — procedures that check the work
docs/ — notes and paper materials
```

7. Decide when the project is complete

The project is not complete simply because every agent reports that its assignment is finished. The final review should confirm that every promised output exists, the checks have passed, important questions have been resolved, and the results can be reproduced from the documented data and code.

For each paper table or figure, it should be possible to identify the script that created it, the intermediate results it used, the original inputs, and the approved methodological choices. Any limitations that remain should be visible to the researchers rather than hidden inside an agent conversation.

> Completion checklist: All specified outputs exist; each is connected to its underlying materials; required checks have passed; researcher approvals are recorded; scripts and documentation support reproduction; and the final set of outputs is ready to be used in writing the paper.

The overall aim is a research process in which AI tools can do substantial, well-defined work while researchers retain control over scientific choices and can understand, verify, and reproduce the results.