Tutorial & Assignment Week 5A — ChatGPT Work/Claude Cowork

Goals: Use ChatGPT Work/Claude Cowork with connectors to simplify something in your life!

Due: Submit the indicated portions on Gradescope by Sunday Oct 18th at 11:59pm, and be ready to discuss your work with your TA during tutorial in Week 6.

Hello everyone! This week, most of the tutorial will be yours to work on something useful to you. We will start by discussing what you would let an agent do, look at how agents use connectors and browsers, and then give you time to try a task of your own.

In Week 4, you practiced evaluating an AI's research or the work behind a result. This week, an agent may also act on your files or in other apps. You will decide what access it needs, where you want a checkpoint, and how to tell whether it did what you asked.

Assignment outline:

This document is intended to be started in tutorial and completed for homework if necessary. Work through it in the order given. Choose either Assignment 5A or 5B; you only need to complete one version.

- Discuss what you would let an agent access and do
- Get a better understanding of connectors, browser agents, prompt injection, and action records
- Meet classmates with related project interests, then explore a task of your own
- Submit a short account of your exploration, and a reflection

Part 1: Tutorial group work

Group Activity 1 — What mitigations would fix this?

Below, you'll find each of the specific scenarios we discussed in class. Now, instead of thinking about what could go wrong, we're going to think about how to fix it. For each of the scenarios, think about:

1. What mitigations (such as placement of human-in-the-loop) would fix the risks?
2. Or, are there no mitigation steps possible, and this is just a bad idea"

Situation one	Situation two	Situation three

Notify maintenance of dorm window leak Goal: send an issue report, including photos of rainwater intrusion Context: maintenance email address Tools: Phone photos connector, Outlook connector Permissions: can view all photos, send emails Memory: past maintenance requests Actions: locate photos, identify issues depicted in photos, compose email report, send, monitor for response Stopping conditions: receive email confirmation of work completed, or ask you for help if no response in 24 hours	Job hunt assistant Goal: send polite and informed-sounding replies to recruiter outreach on LinkedIn Context: your career interests Tools: LinkedIn connector, Web search connector Permissions: read, write, and send LinkedIn messages, web search Memory: previous LinkedIn recruiter responses you've written Actions: monitor for new recruiter messages, web search the company's name to glean facts about the company to assess match and inform the reply, compose and send Stopping conditions: run for one month	Disk cleanup Goal: delete files in Downloads that are untouched at least 6 months to free up space Context: current storage usage Tools: access to laptop files Permissions: read and permanently delete any files on laptop Memory: n/a Actions: scan files starting with oldest, delete, check total usage Stopping conditions: file usage below target

Deliverable: You don't need to submit anything for this portion of the assignment.

Part 2: Individual AI tool work

Individual Activity 1 — Three things!

If you've ever taken an improv class, you've heard of the exercise "three things". In that exercise, you basically say the first three things that come to mind. To help you start brainstorming your final project, we're asking you to brainstorm three topics/realms/domains/ideas for your project, and submit them to gradescope.

They don't have to be thought out or even what you end up choosing- just write down a few of your ideas!

Deliverable: Submit your three ideas to gradescope (just type them).

Individual Activity 2 — Figure out your problem… and run!

Use Claude Cowork or ChatGPT Work to try a task involving files or a connected app. You only need to use one tool. Choose the files or connector that fit your task, and check what access you are granting before you begin.

For example, ask the agent to organize copies of project files, compare documents and produce a useful summary, or act as a chief of staff by reviewing selected messages and calendar information and proposing a plan for your week. You could also use it to triage the gazillion emails you get from all of the

Stanford mailing lists. You can use selected or sample materials if you prefer not to connect a personal account.

Choose something you would actually find useful- you get to spend 50 minutes working on a fun AI tool! An unfinished task, a failed attempt, or a result that changes your mind can all be worthwhile if you can explain what you learned.

Deliverable, all copy and pasted into Gradescope:

1. Name the tool and the task you chose
2. Copy and paste your opening prompt.
3. Write a short paragraph explaining what you tried, what access and permissions you allowed, and one thing you noticed about how the agent worked. Include any important checkpoint, failure, or unexpected action.

Part 3: Non-AI tool work

Check what the agent actually did

Step away from the AI and choose one important action or result to check yourself. Compare what the agent said it did with the available action record and the resulting state of the file, app, or website. Asking the agent whether it is correct is not an independent check.

For example, open the source messages behind its list of deadlines, compare an edited file with the original, or inspect a calendar draft to check the date, time, and attendees. If it only proposed an action, check the proposal and make clear that it did not carry it out.

Use ordinary tools to perform the check. If the action record is missing or incomplete, say so and explain what you could verify directly. You do not need to show that the agent was wrong; a result that checks out is useful too.

PS- if you find yourself with more time here, start thinking about/working on your final project!

Deliverable, in a 4-6 sentence paragraph copied and pasted into Gradescope:

1. What action or result you checked, and what the agent claimed.
2. How you checked it, including what the visible action record showed, if available.
3. What you found, whether it matched the agent's account, and what remains uncertain.

Part 4: Reflection — Trust and agency

Write one short paragraph, approximately 4–6 sentences, responding to at least one of the mini questions below about what happened when you began using an agent:

How would your work associates, friends, and family feel if they knew that their interactions with you were being mediated through an LLM schedule manager, or that your email responses were being largely auto-generated? Over time, could you lose touch with your own life goals and priorities, becoming more of a passenger in the car of your life?)

Deliverable: Copy and paste your reflection into Gradescope.

Submission

You should have four items to submit to Gradescope. Be ready to discuss these with your TA next week:

1. Your three project ideas
2. A short account of your exploration, with your tool, task, opening prompt, and what you noticed about how the agent worked.
3. An independent check of one important action or result.
4. A trust and agency reflection.