Tutorial & Assignment Week 3 — Managing Sources + Context

Goals: Practice intentionally curating the information AI has access to, explore how source selection changes AI outputs, experiment with persistent instructions and memory, and think about what context is useful to give an AI versus what is better left out.

Due: Submit the indicated portions on Gradescope by Sunday Oct 11th at 11:59pm, and discuss with your TA during tutorial in Week 4.

Hello everyone! This week we're moving beyond the prompt itself and looking at everything else that can shape an AI response: your chat history, saved preferences and instructions, memories, uploaded files, retrieved sources, and purpose-built collections of documents.

As we discussed in lecture, all of these can affect what the model generates. More context is not automatically better context. Part of using AI well is learning to decide what belongs in the system's working context, what does not, and how to tell when your source collection is helping rather than hurting.

This week we'll primarily use Gemini Notebook (previously NotebookLM), plus the AI chat tool of your choice. Gemini Notebook is included in the lecture's tool lineup alongside ChatGPT, Claude, and Gemini.

Assignment outline:

This document is intended to be started in tutorial and completed for homework if necessary. Work through it in the order given.

- First, we'll do a short tutorial exercise about context and source selection.
- Then, you'll build and stress-test a small source-grounded AI system.
- Then, you'll experiment with persistent instructions.
- Finally, you'll complete a short reflection.

Some parts involve submitting material to Gradescope. Others are primarily intended for discussion or "show and tell" with your TA in next week's tutorial.

Part 1: Tutorial group work

Activity 1 — What does the model know?

In lecture, we expanded our picture of the context window. In addition to your current prompt and the AI's response, a chat may also include the system prompt, earlier messages in the conversation, settings, memories about you, uploaded documents, and material retrieved from the web.

With your tutorial group, take a few minutes to make a list of the different kinds of information that might influence an AI response.

For each one, discuss:

- Who chose for that information to be there: you, the AI company, or the AI itself?
- Could you inspect or change it?
- When would including it improve an answer?
- When might it make an answer worse?
- Is there anything you would intentionally **withhold**?

Then discuss as a group: **Is "give the AI as much context as possible" actually good advice? Why or why not?**

Activity 2 — Build a glossary with NotebookLM

We have now accumulated several lectures' worth of AI vocabulary. Instead of giving you another glossary, we're going to have you make one.

Take the slides from Lectures 1–3 and give them to an AI system that can work across multiple files, such as Gemini Notebook (previously called NotebookLM; if you log in with your Stanford email, you should get the Stanford license version that includes things like some extra privacy protections). Ask Notebook to identify the major course terms and concepts and create a glossary for you. You may want to tell it to pay particular attention to vocabulary that is visually emphasized in the slides.

As a group, predict where this could go wrong.

Afterward, inspect the glossary.

Discuss:

- Did it identify the concepts you expected?
- What did it miss?
- Did it include things that do not really deserve to be glossary entries?
- Can you trace each definition back to something actually present in the slides?
- How much do you trust this glossary as a study tool?

You do **not** need to submit the glossary. We'll use this exercise to set up the main assignment.

Part 2: AI tool work

This week, you're going to create a small source-grounded AI system. Choose a topic you genuinely care about or want to understand better. It can be academic, professional, practical, or personal. Ideally, choose something where you can find several sources that contribute **different kinds of useful information**.

Examples might include:

- creating a knowledge base for one of your classes;
- understanding a research topic;
- planning a trip;
- comparing approaches to a hobby or project;
- learning the rules and strategy of a game;
- understanding an organization or industry;
- preparing for an interview;
- collecting material for something you are writing or building.

Avoid choosing a corpus consisting entirely of near-duplicate sources saying the same thing.

Build your corpus

Create a new Gemini Notebook and add a small set of sources about your topic.

There is deliberately **no magic required number of sources** here. Part of the exercise is deciding how much context your system actually needs.

Think about:

- What does each source contribute?
- Are you including multiple perspectives or types of evidence?
- Are some sources redundant?
- Is anything important missing?
- Are you adding something simply because it is available, rather than because it is useful?

Then ask your Notebook several questions that you think a good corpus on this topic **should** be able to answer.

The point here is not merely to make a big pile of documents. The lecture distinguishes grounding a model from simply filling its context window, and introduces retrieval-augmented generation as another way of supplying relevant information.

Deliverable: Ask your Notebook to write a short description of itself, answering the following questions:

- What is this collection of sources designed to help someone understand or accomplish?
- What kinds of questions is it well equipped to answer, and what are its main limitations?

Submit the summary of your notebook as well as a link to it in this week's Gradescope assignment.

Part 3: Non-AI tool work

Poison the well

Now deliberately add **one bad source**. "Bad" can mean several things here. For example, the source could be:

- factually unreliable;
- irrelevant to your topic;
- very low quality;
- outdated;
- confidently misleading;
- strongly contradictory to the rest of your corpus.

If you cannot think of one, you are also welcome to use this wonderful source I dredged up from the depths of the internet that is as old as I am:
https://stanfordmag.org/contents/have-you-heard-the-one-about-the-squirrels

Before you add it, ask your Notebook 2–3 questions and save its answers. Then add the bad source and ask the **same questions again**, and compare the results.

The purpose is not to "break" the system. We want you to see whether one source that does not belong changes the behavior of the whole collection.

Deliverable: Submit a short response (3-4 sentences) to gradescope answering at least one of the following questions:

- What bad source did you add, and why did it count as "bad" for this corpus?
- What changed, if anything, when you asked the same questions again?
- Did the Notebook ignore the source, appropriately qualify it, accidentally incorporate it, or do something else? Give one concrete example.
- After this experiment, what would you change about the way you curate sources for an AI system in the future?

Persistent Instructions

So far, we have mostly changed an AI's behavior by changing what we put into the current conversation. But major AI tools can also preserve instructions or preferences across conversations. *Questions adapted from @artefactohq on X and @whighsmith on X*

The lecture uses examples such as ChatGPT personalization settings and custom instructions, and Gemini's **Gems**, which allow a set of instructions to function essentially like a standing system prompt for a new chat.

We're going to experiment with what happens when one of those instructions becomes persistent.

Choose **one** of the following instructions, and add it to your favorite model's permanent memory (you can always delete it when done with this exercise):

- Never validate my question before answering. No "great point," "absolutely," or "that's a really interesting question." Start with the answer or with a correction.
- If my assumption is wrong, say so before answering.
- If you are uncertain, tell me how confident you are and explain the source of the uncertainty. Never present a guess as a fact.
- If I push back on your answer, do not change it merely because I seem unhappy. Change it when I provide new evidence or reasoning.
- After an important answer, add one sentence identifying a useful question I did not ask.
- If I ask you to agree with something that is wrong, disagree.
- Create your own instruction inspired by these.

Now take a prompt from an earlier week that produced an answer you cared about. Run it again.

Deliverable: Upload the following to gradescope:

- The instruction you chose
- The prompt you retested
- What changed in the output, and whether or not you would keep this instruction long term (3-4 sentences)

Now, remove the instruction if you don't want it influencing future chats.

Part 4: Reflection — What should stay yours?

One theme from lecture is that AI can help you do more without actually changing the underlying model: changing the context changes the model's next-token predictions, not its learned weights.

But there is another side to managing context: **what do you choose not to outsource?**

Write one short paragraph, approximately **4–6 sentences**, responding to:

After this week's exercises, what is one kind of context you think you should actively curate rather than simply hand over to AI?

You can interpret "context" broadly: documents, personal memories, preferences, instructions, prior work, sources, examples, or even your own judgment.

Then explain one concrete habit you could adopt to maintain control over that context.

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. Notebook summary + link
2. Bad-source experiment
3. Persistent-instruction experiment
4. Reflection