Machine Learning for Design, Week 10

Final Bake-Off

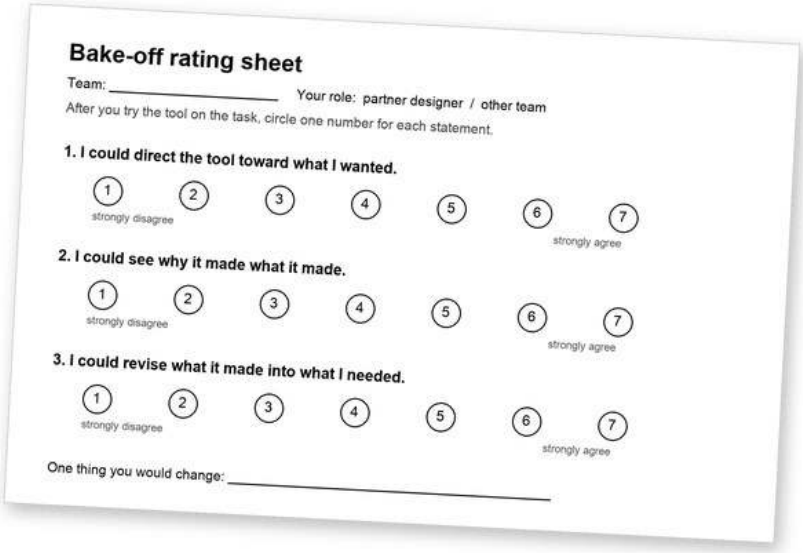
Can your partner's designers use your tool on their real task?

Today

1. The bake-off
2. Hand-ins
3. Looking back
4. The position essay

1. The bake-off

Your partner's designers and another team try each tool on the real task, and rate it



Bake-off rating sheet

Team: _____ Your role: partner designer / other team

After you try the tool on the task, circle one number for each statement.

1. I could direct the tool toward what I wanted.

1 (strongly disagree) 2 3 4 5 6 7 (strongly agree)

2. I could see why it made what it made.

1 (strongly disagree) 2 3 4 5 6 7 (strongly agree)

3. I could revise what it made into what I needed.

1 (strongly disagree) 2 3 4 5 6 7 (strongly agree)

One thing you would change: _____

Schedule

Set up	Presentations	Round 1	Round 2	Critics' notes
20 min	5 min each, then 3 min of questions	40 min	40 min	15 min

Visitors and outside critics: at the presentations, and at the stations in both rounds

Two rounds

Round 1

Team 1's tool	its partner designers, then Team 2
Team 3's tool	its partner designers, then Team 4
Team 5's tool	its partner designers, then Team 6
Team 7's tool	its partner designers, then Team 8

Round 2

Team 2's tool	its partner designers, then Team 3
Team 4's tool	its partner designers, then Team 5
Team 6's tool	its partner designers, then Team 7
Team 8's tool	its partner designers, then Team 1

Each try: **15 minutes** on the real task, then **5 minutes** for the rating sheet. Your team watches and takes notes.

An example with eight teams

The rating sheet

Bake-off rating sheet

Team: _____ Your role: partner designer / other team

After you try the tool on the task, circle one number for each statement.

1. I could direct the tool toward what I wanted.

1 2 3 4 5 6 7

strongly disagree strongly agree

2. I could see why it made what it made.

1 2 3 4 5 6 7

strongly disagree strongly agree

3. I could revise what it made into what I needed.

1 2 3 4 5 6 7

strongly disagree strongly agree

One thing you would change: _____

One sheet from each person who tries your tool

In your paper, next to your own measure from week 3

Statement 1: direct

“I could direct the tool toward what I wanted.”



−1.5



−0.75



0



+0.75



+1.5

Week 5: the simple to ornate slider on the model trained on the 56 chairs, same seed

Statement 2: inspect



“I could see why it made what it made.”

Top row:
made by the model

Bottom row:
closest of the 56

Week 5: closest chair by DINOv2 image similarity

Statement 3: revise

“I could revise what it made into what I needed.”



first image



asked in words

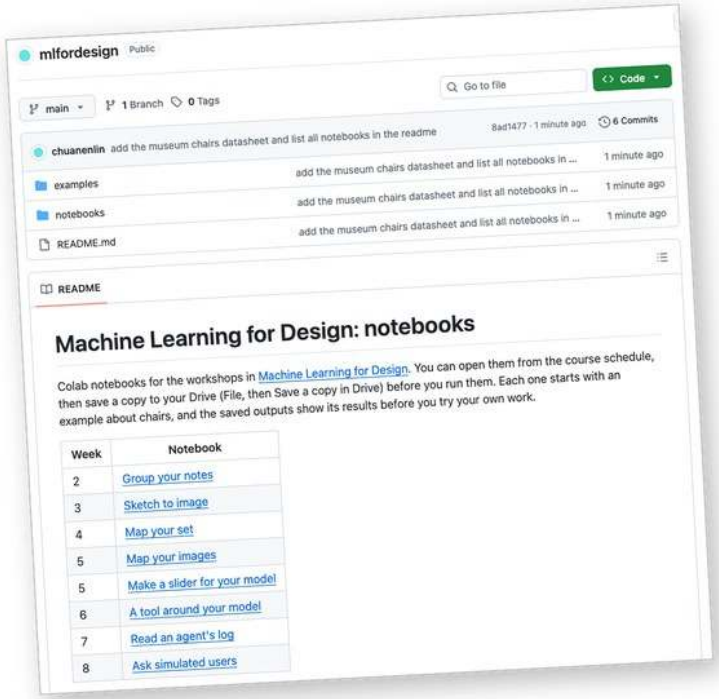


sketch redrawn

Week 3: “make the legs and back posts solid oak instead of black metal” with FLUX.1 Kontext [pro], and the redrawn sketch with FLUX.1 Canny [dev]

2. Hand-ins

Due with the bake-off: two videos, the paper, and a public code repository



Two videos

Project video, 3 to 4 minutes
The tool in use on the real task

Process video, 2 to 3 minutes
How you built it, and where AI helped or did not

Claude Code, with CLAUDE.md

1. Show the app file and git state
2. Check installed Gradio version and Gallery options
   Gradio won't import on this Mac, so I can't check its options here. I'll stick to Gallery options that have been in Gradio since version 4.
3. Edit app.py
4. Check app.py compiles and show diff size
   added the Make strip button. app.py compiles; per your CLAUDE.md i didn't load the model or run the tool.

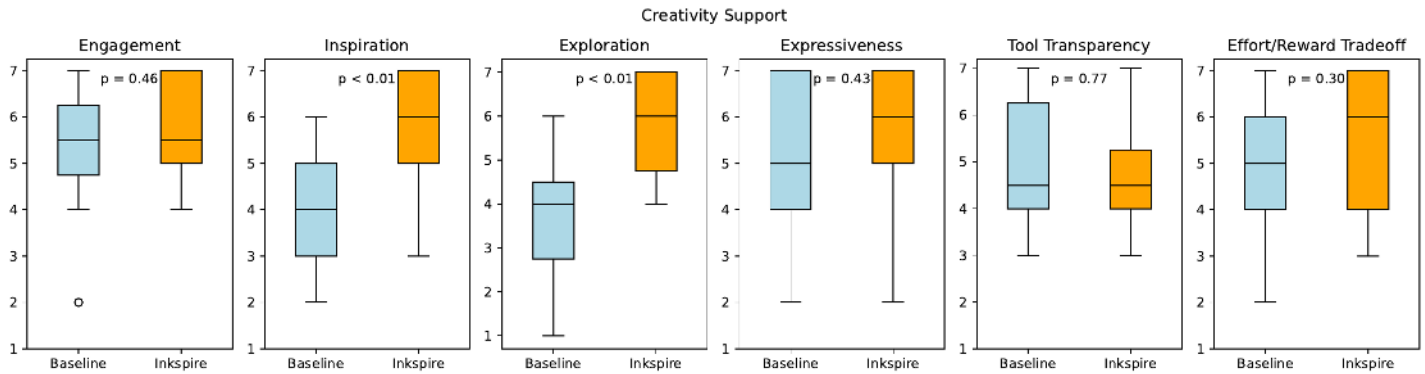
Claude Code, without CLAUDE.md

1. Show repo status, history and files
2. Read the app source and the start of the transcript
3. Look around the course folder and check installed pac…
4. Look at sibling folders and run scripts for a working…
5. Read the existing app test harness and check the cour…
6. Check how the course notebook's version of the tool h…
7. Read the notebook's tool cell
8. Edit app.py
9. Edit app.py
10. Run the interface with a fake model and screenshot th…
11. Rerun the browser test with the system Chrome
12. Read strip-done.png
13. Read strip-midway.png
14. Look at the Gallery component's preview and height op…
15. Find the Gallery frontend source to see when it opens…
16. Search the compiled Gallery code for the single-image…
17. Look at how the compiled Gallery sets the selected im…
18. Screenshot three gallery settings to compare their la…
19. Put the six crops on one comparison sheet
20. Read compare.png
21. Check the Gallery fit_columns option and its CSS mini…
22. Screenshot fit_columns off, in the right column and a…
23. Fix the test's text swap and rerun both layouts
24. Combine the four new screenshots into one sheet
25. Read compare2.png
26. Crop the full-width strip area at full resolution
27. Read v5-done-crop.png
28. Try a fit-to-content gallery height at full width
29. Read compare3.png
30. Click one strip image and screenshot the enlarged view
31. Read v6-clicked.png
32. Check which Gradio releases have the Gallery options …
33. Edit app.py
34. Edit app.py
35. Edit app.py
36. Test the final app page on Gradio 6.32 and Gradio 5.49
37. Put the Gradio 6 and 5 screenshots on one sheet
38. Read final-compare.png
39. Edit app.py
40. Edit app.py
41. Edit app.py
42. Rerun both Gradio versions with the grey-card strip a…
43. Read final-compare.png
44. Show the final change to app.py and check it compiles

Week 6: the same request to Claude Code with an instruction file (top, 4 steps) and without one (44 steps, stopped at 10 minutes)

Paper: 4 to 6 pages, CHI format

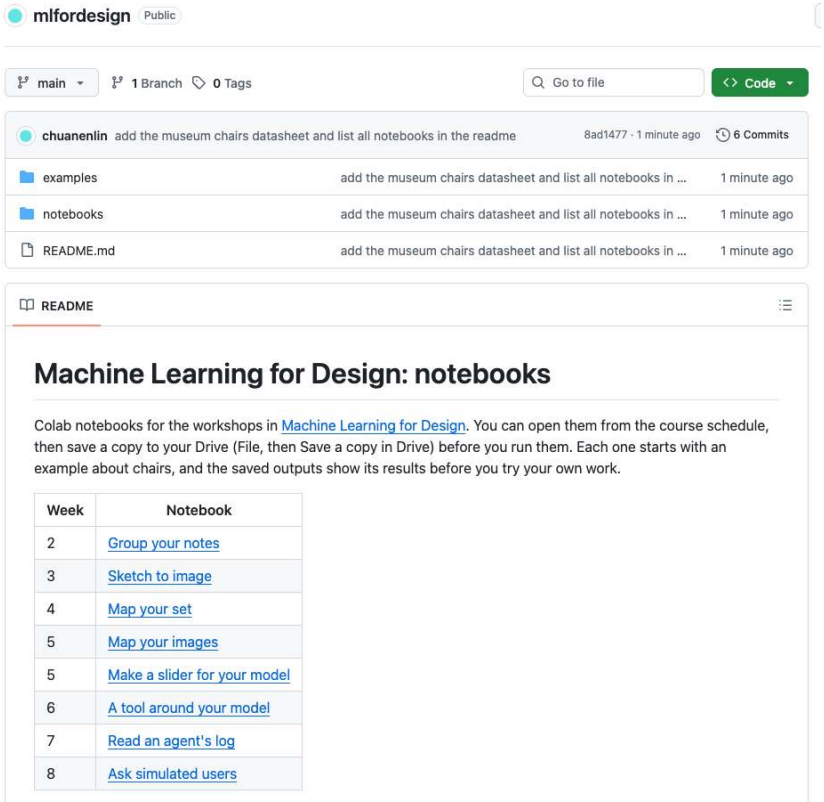
1. Introduction: the design activity, and what better means
2. Related tools and papers
3. The tool
4. Partner tests and the bake-off
5. Results: the ratings and your own measure
6. Discussion and limitations



A results figure with one chart for each factor

Lin et al., CHI 2025, Fig. 5 (from week 9)

Code repository



Public, with a README that says:

What the tool does

How to run it

Which models it calls, and what they cost

The course's notebook repository, github.com/chuanenlin/mlfordesign

Rubric: bake-off and final paper

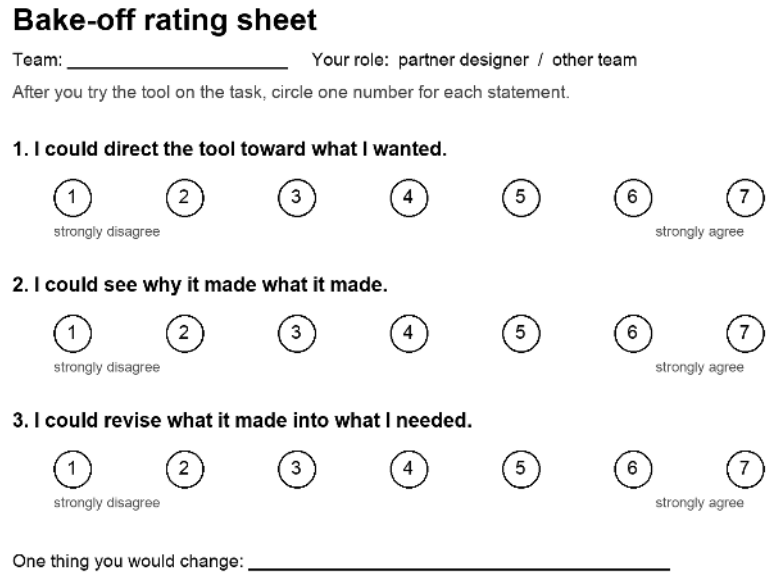
Partner fit: **10**

Ratings: **10**

Paper: **10**

Code: **5**

Videos: **5**



Bake-off rating sheet

Team: _____ Your role: partner designer / other team

After you try the tool on the task, circle one number for each statement.

1. I could direct the tool toward what I wanted.

1 2 3 4 5 6 7

strongly disagree — strongly agree

2. I could see why it made what it made.

1 2 3 4 5 6 7

strongly disagree — strongly agree

3. I could revise what it made into what I needed.

1 2 3 4 5 6 7

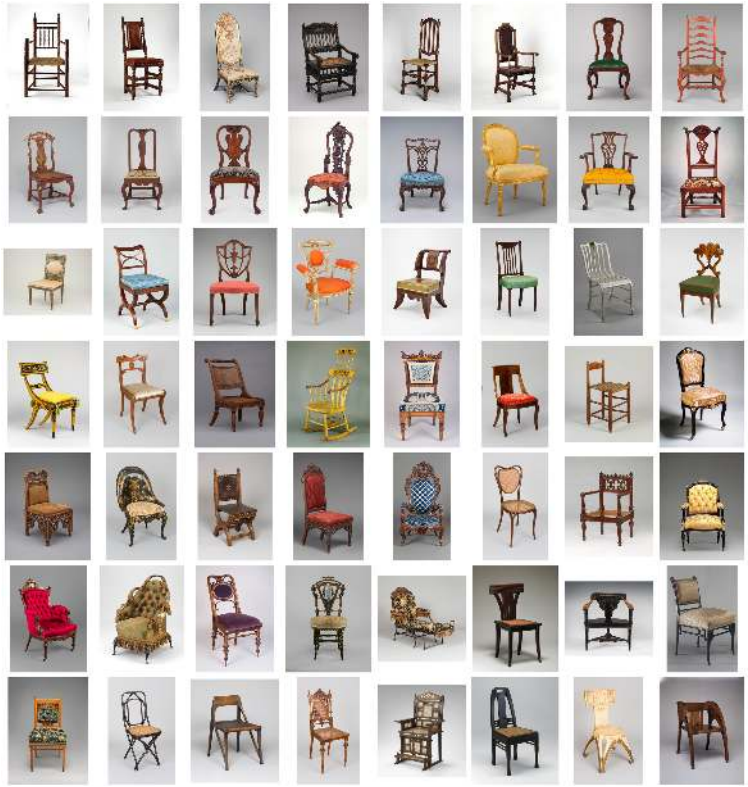
strongly disagree — strongly agree

One thing you would change: _____

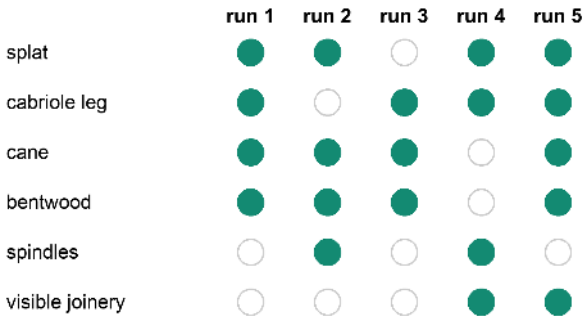
Ratings: all the sheets, including the low ones

3. Looking back

The 56 museum chairs,
from week 1 to week 9



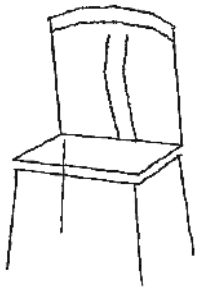
Part 1: Designing with AI



Week 1: the same themes in five runs of one prompt



Week 2: a mood board by hand, and the same brief as a search





Week 3: a rough sketch made into an image

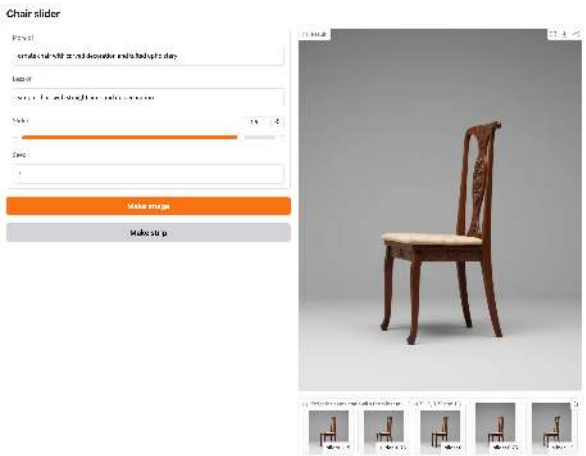
Part 2: Appropriating AI



Week 4: the 56
chairs on a CLIP map



Week 5: chairs from a
model trained on the 56



Week 6: a slider tool,
built with a coding agent

Part 3: Agentic Design



Week 7: a design agent's rounds of done or revise



Week 8: the chairs seven simulated visitors picked



slider prompt

Week 9: measuring our slider against the prompt

4. The position essay

About 1,000 words, due in finals week



Can AI design?

Whether and how AI can design, with evidence from your own work in the course

For example: a run, a rating, or something your partner's designers said

	Checklist by	Images	Failed checks	Ended by
	Run 1: the agent	4	8 of 20	my round limit
	Run 2: the agent	3	3 of 15	agent said done
	Run 3: me	4	3 of 16	agent said done
	Run 4: the agent	8	14 of 40	my round limit

Week 7: four runs of the design agent

Thank you






Position essay: due in finals week