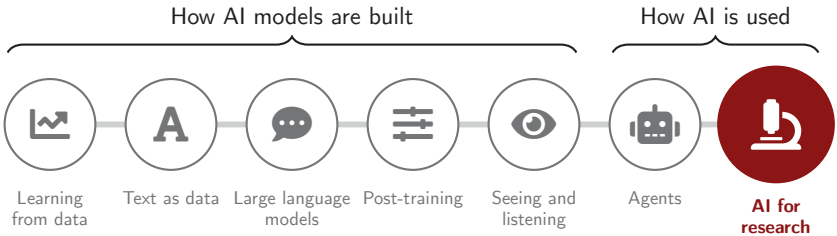
The tech track in review

AI for social science research

POLI 158: AI & Politics · Session 17

Yiqing Xu

The seven Wednesday sessions



Today in five parts



The tech track in review

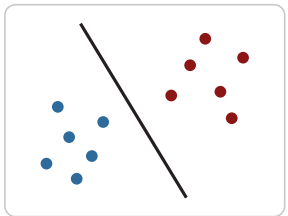
Poll: learning from data

Poll show of hands

A model fits its training data almost perfectly. What do we check next?

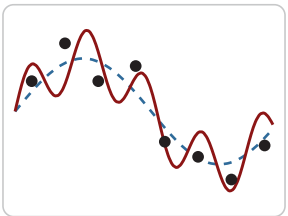
- **A** Whether it has enough layers
- **B** How it does on new data
- **C** How long the training took
- **D** Nothing: a perfect fit is the goal

Session 3: learning from data



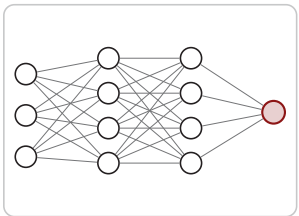
Learning from examples

the rule comes from the data



Overfitting

fits old data, fails on new data



Neural networks

layers of simple units

Poll: text as data

Poll show of hands

On a map of word embeddings, which words sit close together?

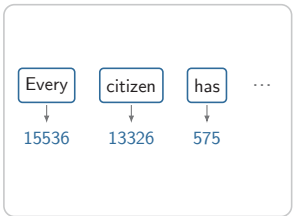
A Words with similar spelling

B Words of the same length

C Words used in similar contexts

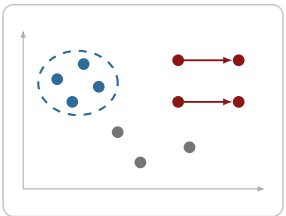
D Words next to each other in the dictionary

Session 5: text as data



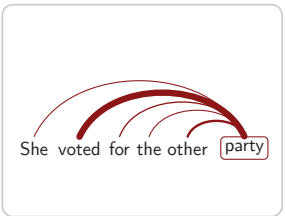
Tokens

text cut into numbered pieces



Embeddings

words as points on a map



Attention

each word looks at the others

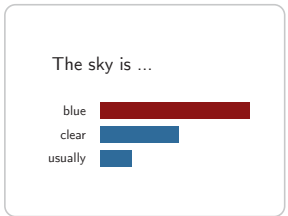
Poll: large language models

Poll show of hands

With ten times more computing, how much does a model's error fall?

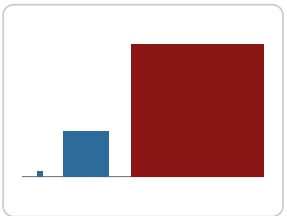
- **A** About a tenth, every time
- **B** About half, every time
- **C** About nine tenths, every time
- **D** A lot at first, then nothing

Session 7: large language models



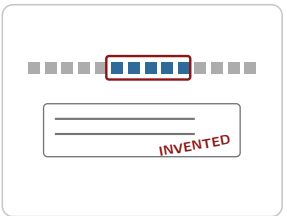
Next-token prediction

one token at a time



Scale

text, size, computing



Context and hallucination

what it sees, what it makes up

Poll: post-training

Poll show of hands

What turns a text predictor into an assistant?

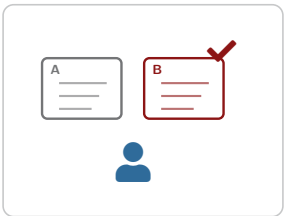
- **A** A larger context window
- **B** More text from the web
- **C** Faster chips
- **D** Example conversations and people's rankings

Session 9: post-training



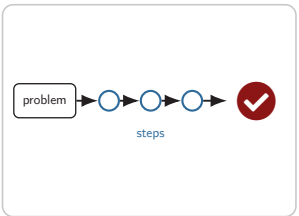
Example conversations

learn from good answers



Learning from comparisons

people pick the better reply



Reasoning models

work in steps, then answer

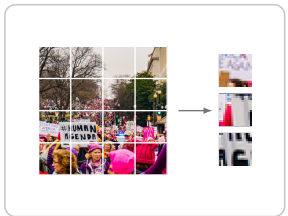
Poll: seeing and listening

Poll show of hands

How does a model take in a photo?

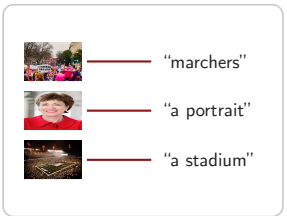
- **A** It reads the file name
- **B** It cuts the photo into patches that become tokens
- **C** It compares the photo with every photo it has seen
- **D** It needs a person to describe the photo first

Session 13: seeing and listening



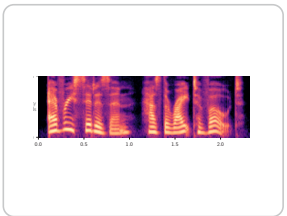
Patches

a photo cut into pieces



A shared map

photos and captions, one map



Sound and video

sound as a picture

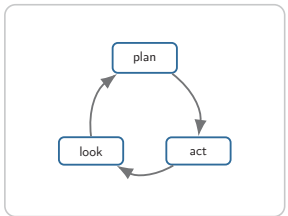
Poll: agents

Poll show of hands

What makes an agent different from a chatbot?

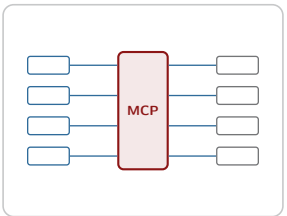
- **A** It uses a larger model
- **B** It was trained on more text
- **C** It uses tools and works in a loop
- **D** It never needs a person to check its work

Session 15: agents



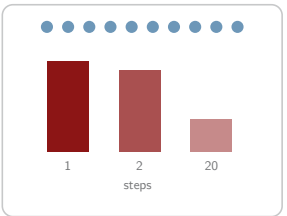
The agent loop

plan, act, look, repeat



Tools and the harness

the program around the model



Long tasks

errors add up over steps

From data to agents



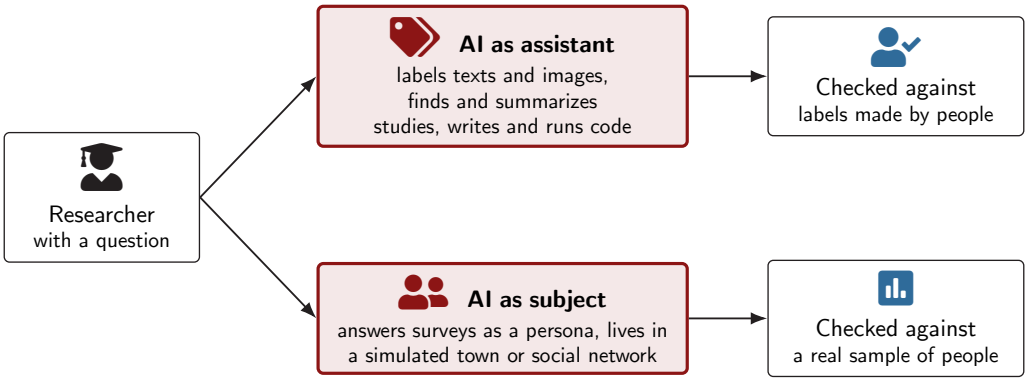
AI for social science research

TODAY'S QUESTION

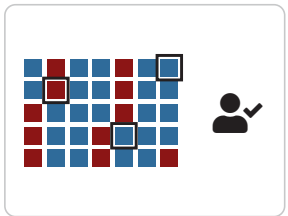
Would you trust a poll of 1,000 AI personas?

For which questions, if any?

Two jobs for AI in research

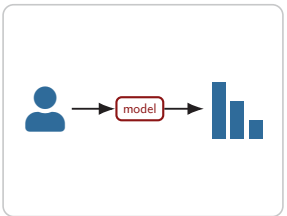


Three concepts in today's video



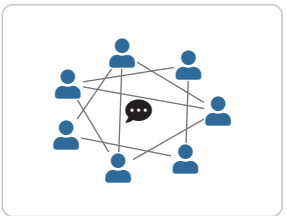
AI as research assistant

checked against people's labels



AI as survey respondent

a "silicon sample"



Simulated societies

many AI agents that interact

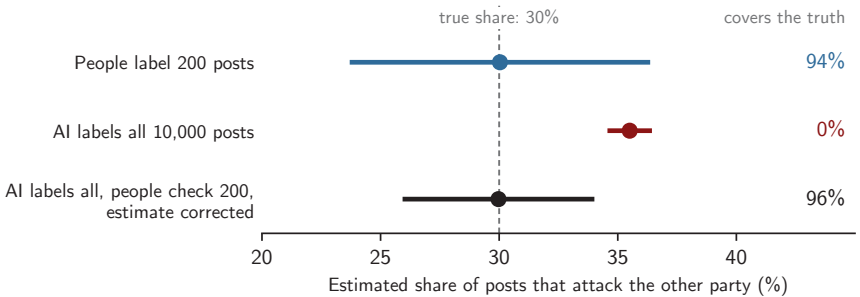


AI in social science research

Video

Correcting an estimate built on AI labels

Example: what share of 10,000 politicians' posts attack the other party? True share: 30%.



Our simulation, 2,000 simulated studies. Bars: 95% intervals. Right: how often the interval contains 30%. Method: Egami et al. (2023).

Research agents and coding agents



Research agent

You ask: "Find studies since 2023 on chatbots and political persuasion. Summarize what they found."

The agent searches, reads, and writes a report with citations.

You check: open two citations. Do they say that?



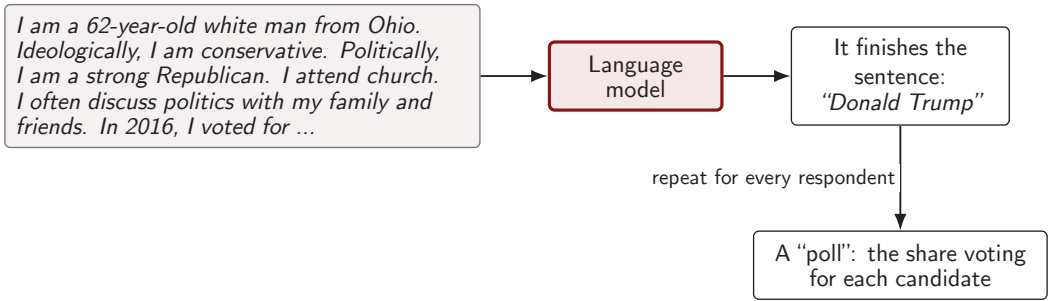
Coding agent

You ask: "Clean this survey file and make a table of turnout by age."

The agent writes code, runs it, and fixes its own errors.

You check: rerun the code. Check five rows by hand.

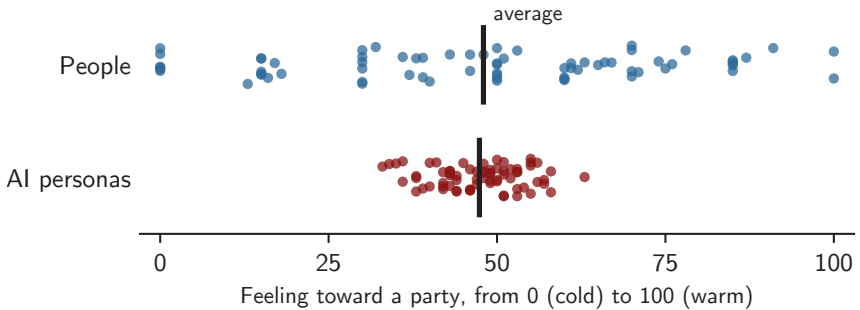
A silicon sample



- Backstory: the traits of one real survey respondent.

Illustration of the method in Argyle et al. (2023), who call the result a "silicon sample."

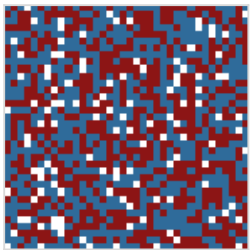
Spread of answers: people and AI personas



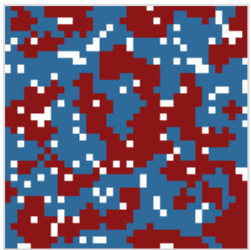
Made-up numbers showing the pattern that Bisbee et al. (2024, *Political Analysis*) report for feeling thermometers in the ANES.

Schelling's segregation model

Start: random
same-group neighbors: 48%



After households move
same-group neighbors: 75%



- Rule: move if too few neighbors are like you.
- Our run: fewer than a third.
- Red, blue: two groups. White: empty homes.

Demo: `https://nifty.stanford.edu/2014/mccown-schelling-model-segregation/`

Our simulation of Schelling's (1971) model: a 36 by 36 grid, 10% of homes empty, eight neighbors per home.

Generative agents

A simulated town (2023)

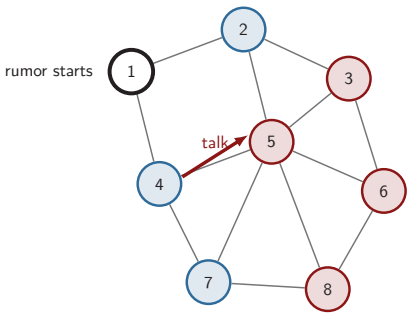
- 25 agents with memory, plans, reflection
- One agent is told to plan a party
- 13 of 25 knew about it
- 5 of the 12 invited came

Agents built from 1,052 people (2024)

- Two-hour interview by an AI interviewer
- One agent built per person
- Benchmark: people retake a survey two weeks later
- Agents reach 83% of it; demographics alone, 74%

Source: Park et al. (2023), a conference paper; Park et al. (2024), arXiv:2411.10109, version 3 (June 2026), a preprint.

A rumor in a simulated dorm



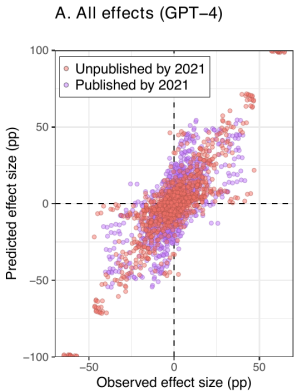
blue: Democratic personas

red: Republican personas

- Rumor: Barack Obama was born outside the U.S.
- Personas: Stanford first-year students.
- Each round: an interviewer scores belief from 0 to 1.
- Round 4: official Hawaii records. Round 8: talk radio.

Built with the llm-society Python library.

Case: predicting survey experiments



- GPT-4 simulated Americans in 70 survey experiments
- Predicted and real effects: correlation 0.85
- Average forecast of 2,659 lay people: 0.84
- Predicted effects roughly twice as large

Source: Ashokkumar, Hewitt, Ghezae, and Willer (2026), *Nature* 656: 115–122. Figure: their Figure 1A; each dot is one effect.

BACK TO TODAY'S QUESTION

For which questions, if any,
would you trust a poll of 1,000 AI personas?

Where to go next at Stanford



CS 221
Artificial Intelligence: Principles and Techniques



CS 229
Machine Learning



CS 230
Deep Learning



CS 224N
Natural Language Processing with Deep Learning



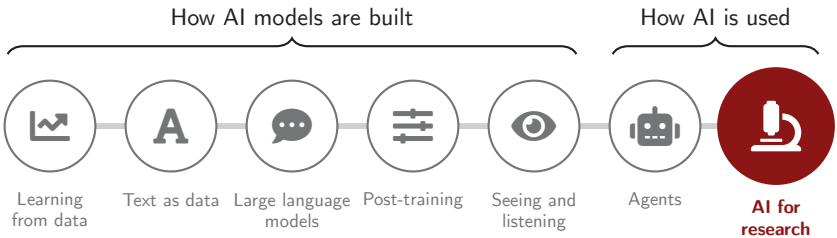
CS 234
Reinforcement Learning (not scheduled in 2026–27)



STATS 315A
Modern Applied Statistics: Learning

Stanford Explore Courses, 2026–27 catalog (checked November 17, 2026). CS: Computer Science; STATS: Statistics.

Final presentations



Monday, November 30 and Wednesday, December 2

Sources

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., and Wingate, D. (2023). "Out of One, Many: Using Language Models to Simulate Human Samples." *Political Analysis* 31(3): 337–351.

Ashokkumar, A., Hewitt, L., Ghezae, I., and Willer, R. (2026). "Large Language Models Can Predict the Results of Social Science Experiments." *Nature* 656: 115–122. Accepted manuscript: PsyArXiv, June 17, 2026.

Bisbee, J., Clinton, J. D., Dorff, C., Kenkel, B., and Larson, J. M. (2024). "Synthetic Replacements for Human Survey Data? The Perils of Large Language Models." *Political Analysis* 32(4): 401–416.

Egami, N., Hinck, M., Stewart, B. M., and Wei, H. (2023). "Using Imperfect Surrogates for Downstream Inference." *Advances in Neural Information Processing Systems (NeurIPS)* 36. arXiv:2306.04746.

Kaplan, J., et al. (2020). "Scaling Laws for Neural Language Models." arXiv:2001.08361.

McCown, F. (2014). "Schelling's Model of Segregation." Nifty Assignments. nifty.stanford.edu/2014/mccown-schelling-model-segregation.

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., and Bernstein, M. S. (2023). "Generative Agents: Interactive Simulacra of Human Behavior." *Symposium on User Interface Software and Technology (UIST '23)*.

Park, J. S., et al. (2024). "Generative Agent Simulations of 1,000 People." arXiv:2411.10109v1. Version 3 (2026, June 28): "LLM Agents Grounded in Self-Reports Enable General-Purpose Simulation of Individuals."

Qin, T. *llm-society*, a Python library for information-diffusion simulations with language-model agents, version 0.3.3. pypi.org/project/llm-society.

Schelling, T. C. (1971). "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1(2): 143–186.

Stanford University. *Explore Courses*, 2026–27 catalog. explorecourses.stanford.edu.

Review pages: drawings from the decks of Sessions 3 to 15. Photos (Wikimedia Commons): Ted Eytan, Women's March, 2017, CC BY-SA 2.0, cropped and resized, with patches adapted from it; U.S. Congress and U.S. Marine Corps, public domain.

Icons: Font Awesome Free 5 (CC BY 4.0). Video: animated with Manim Community Edition; narration by Kokoro-82M, a synthetic voice.