Lecture 14.

Simulating Ourselves and Our Societies With AI

CS 222: AI Agents and Simulations
Stanford University

Joon Sung Park

Announcement

- **Next week, we have two amazing guest lecturers!!**
  - **Monday: Meredith Ringel Morris**
    - Director and Principal Scientist for Human-AI Interaction, Google DeepMind
  - **Wednesday: Serina Chang**
    - Assistant Professor at UC Berkeley, EECS

10 min activity: agent voting!

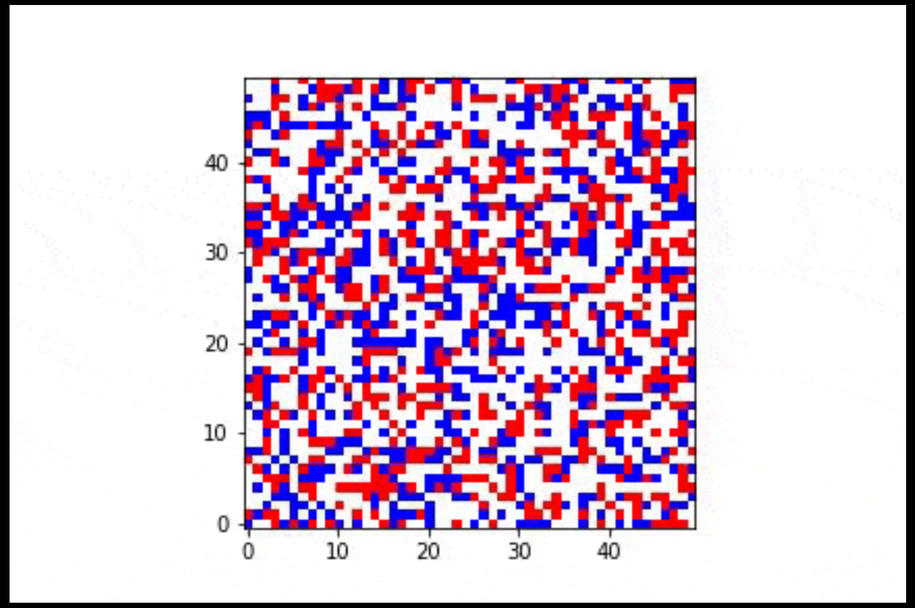
Ver. Fall 2024

Summarizing the quarter

1. Simulations are programs that define an environment and the behaviors of individuals, then output the resulting world

In games (e.g, The Sims)

In movies (e.g, The Matrix)



Agent-based models

T. C. Schelling, Dynamic models of segregation. Journal of Mathematical Sociology 1, 143-186 (1971).

2. Generative AI presents a new opportunity to create more open-ended simulations of human behaviors

J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, M. S. Bernstein, Generative agents: Interactive simulacra of human behavior, in Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (ACM, 2023).

3. The promise of human behavioral simulation is to enable us to address *wicked problems*

Policy Sciences **4** (1973), 155–169

Dilemmas in a General Theory of Planning*

HORST W. J. RITTEL
Professor of the Science of Design, University of California, Berkeley

MELVIN M. WEBBER
Professor of City Planning, University of California, Berkeley

ABSTRACT

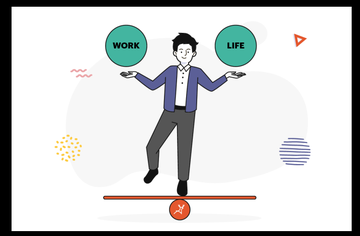
The search for scientific bases for confronting problems of social policy is bound to fail, because of the nature of these problems. They are "wicked" problems, whereas science has developed to deal with "tame" problems. Policy problems cannot be definitively described. Moreover, in a pluralistic society there is nothing like the undisputable public good; there is no objective definition of equity; policies that respond to social problems cannot be meaningfully correct or false; and it makes no sense to talk about "optimal solutions" to social problems unless severe qualifications are imposed first. Even worse, there are no "solutions" in the sense of definitive and objective answers.

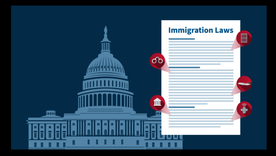
George Bernard Shaw diagnosed the case several years ago; in more recent times popular protest may have already become a social movement. Shaw averred that "every profession is a conspiracy against the laity." The contemporary publics are responding as though they have made the same discovery.

Few of the modern professionals seem to be immune from the popular attack—whether they be social workers, educators, housers, public health officials, policemen, city planners, highway engineers or physicians. Our restive clients have been telling us that they don't like the educational programs that schoolmen have been offering, the redevelopment projects urban renewal agencies have been proposing, the law-enforcement styles of the police, the administrative behavior of the welfare agencies, the locations of the highways, and so on. In the courts, the streets, and the political campaigns, we've been hearing ever-louder public protests against the professions' diagnoses of the clients' problems, against professionally designed governmental programs, against professionally certified standards for the public services.

It does seem odd that this attack should be coming just when professionals in

* This is a modification of a paper presented to the Panel on Policy Sciences, American Association for the Advancement of Science, Boston, December 1969.





Wicked problems are complex, ill-defined social or policy challenges that defy straightforward solutions.

H. W. J. Rittel, M. M. Webber, Dilemmas in a general theory of planning. Policy Sciences 4, 155-169 (1973).

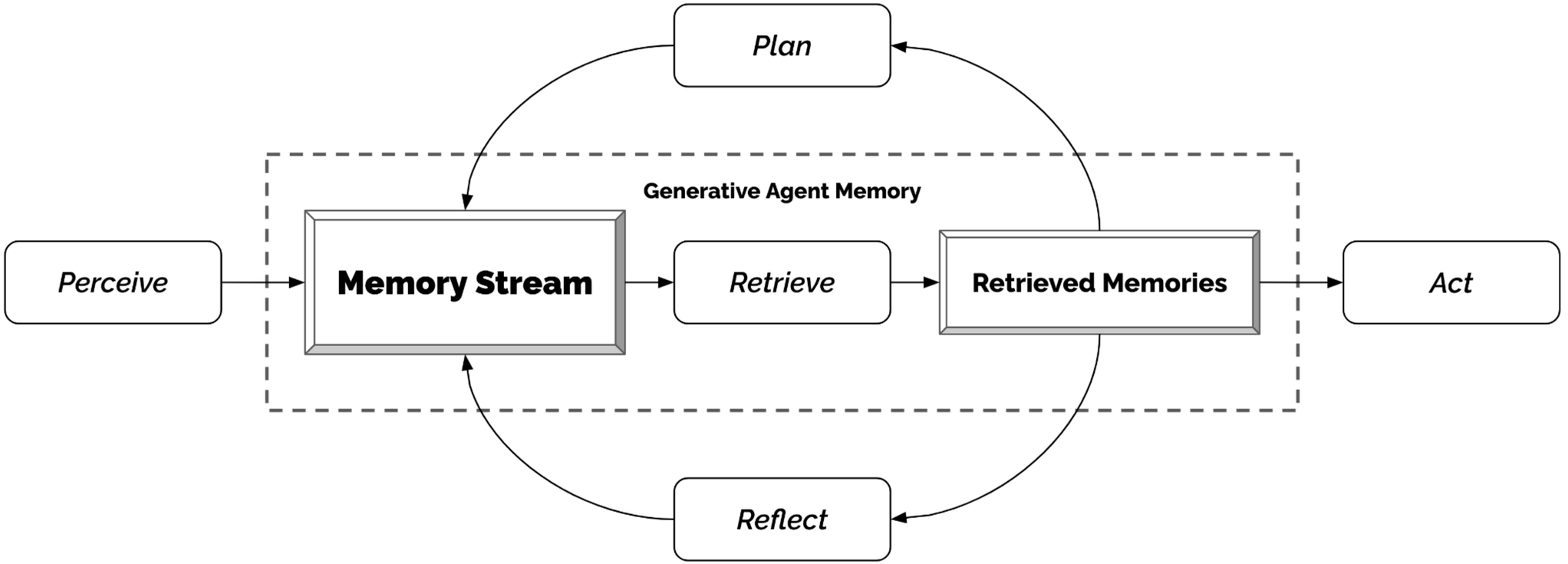
4. To build simulations, you start by understanding the level of analysis you want to conduct

Individuals

Groups

Populations

5. You then build the architecture of individual agents and their behaviors



J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, M. S. Bernstein, Generative agents: Interactive simulacra of human behavior, in Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (ACM, 2023).

6. And the environment in which the agents can interact with one another

DISCOVERYWORLD: A Virtual Environment for Developing and Evaluating Automated Scientific Discovery Agents

Generative Agents: Interactive Simulacra of Human Behavior

Joon Sung Park
Stanford University
Stanford, USA
joonspk@stanford.edu

Joseph C. O'Brien
Stanford University
Stanford, USA
jobrien3@stanford.edu

Carrie J. Cai
Google Research
Mountain View, CA, USA
cjcai@google.com

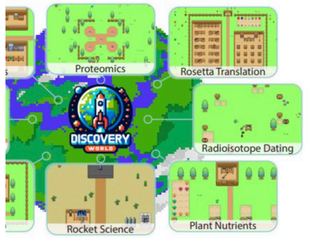
Meredith Ringel Morris
Google Research
Seattle, WA, USA
merrie@google.com

Percy Liang
Stanford University
Stanford, USA
pliang@cs.stanford.edu

Michael S. Bernstein
Stanford University
Stanford, USA
msb@cs.stanford.edu

Figure 1: Generative agents create believable simulacra of human behavior for interactive applications. In this work, we demonstrate generative agents by populating a sandbox environment, reminiscent of The Sims, with twenty-five agents. Users can observe and intervene as agents they plan their days, share news, form relationships, and coordinate group activities.

exandre Côté[‡], Tushar Khot[♠] Erin Bransom[♠], Bhavana Dalvi Mishra[♠],
twa Prasad Majumder[♠], Oyvind Tafjord[♠], Peter Clark[♠]
rtificial Intelligence [†]Microsoft Research [‡]University of Arizona
peterj@allenai.org



D is a virtual environment for developing and evaluating discovery agents, with
ad variety of different topics such as those shown above.

Abstract

fic discovery promises to accelerate progress across scientific
r, developing and evaluating an AI agent's capacity for end-
easoning is challenging as running real-world experiments is
expensive or infeasible. In this work we introduce DISCOV-
irst virtual environment for developing and benchmarking an
erform complete cycles of novel scientific discovery. DISCOV-
ains a variety of different challenges, covering topics as diverse
ng, rocket science, and proteomics, to encourage development
y skills rather than task-specific solutions. DISCOVERYWORLD
sive, simulated, text-based environment (with optional 2D visual
s 120 different challenge tasks, spanning eight topics each with
iculty and several parametric variations. Each task requires an
otheses, design and run experiments, analyze results, and act
DISCOVERYWORLD further provides three automatic metrics

mmunicative Agents for Software Development

n Qian[♠] Xin Cong[♠] Wei Liu[♠] Cheng Yang[♣] Weize Chen[♠] Yusheng Su[♠]
Dang[♠] Jiahao Li[♦] Juyuan Xu[▲] Dahai Li[★] Zhiyuan L
[♠]Tsinghua University [♣]Beijing University of Posts and Telec
[♦]Dalian University of Technology [▲]Brown University [★]
qianc62@gmail.com liuzy@tsinghua.edu.cn sms@ts

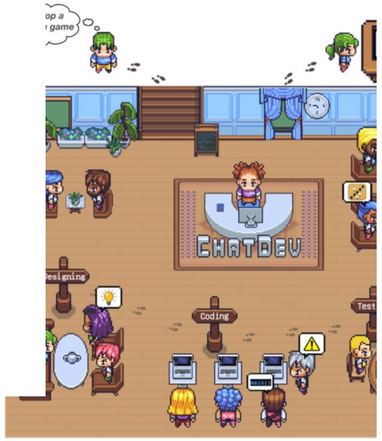
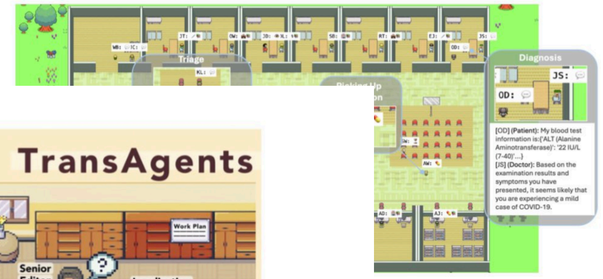


Figure 1: ChatDev, our virtual chat-powered company for software development, brings together "software agents" from diverse social identities, including chief officers, professional programmers, test engineers, and art designers. When presented with a preliminary task by a human "client" (*e.g.*, "*develop a gomoku game*"), the software agents at ChatDev engage in effective communication and mutual verification through collaborative chatting. This process enables them to automatically craft comprehensive software solutions that encompass source codes, environment dependencies, and user

Agent Hospital: A Simulacrum of Hospital with Evolvable Medical Agents

JUNKAI LI[†#], SIYU WANG[†], MENG ZHANG[†], WEITAO LI[†#], YUNGHWEI LAI[†], XINHUI KANG[†#], WEIZHI MA[†], and YANG LIU[#†]



ıospital in which patients, nurses, and
els. Agent Hospital simulates the whole
ge, registration, consultation, medical
and post-hospital follow-up visit. An
ing treatment performance over time
rld evaluations.

Figure 2: TRANSAGENTS, a multi-agent virtual company for literary translation.

J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, M. S. Bernstein, Generative agents: Interactive simulacra of human behavior, in Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (ACM, 2023).
C. Qian, W. Liu, H. Liu, N. Chen, Y. Dang, J. Li, C. Yang, W. Chen, Y. Su, X. Cong, J. Xu, D. Li, Z. Liu, M. Sun, ChatDev: Communicative Agents for Software Development, in Proceedings of the 2024 Annual Conference of the Association for Computational Linguistics (ACL 2024).
P. Jansen, M.-A. Côté, T. Khot, E. Bransom, B. Dalvi Mishra, B. P. Majumder, O. Tafjord, P. Clark, DISCOVERYWORLD: A Virtual Environment for Developing and Evaluating Automated Scientific Discovery Agents. Preprint (2024).
J. Li, S. Wang, M. Zhang, W. Li, Y. Lai, X. Kang, W. Ma, Y. Liu, Agent Hospital: A Simulacrum of Hospital with Evolvable Medical Agents. Preprint (2024).

7. So far, we have evaluated the success of simulations by testing their believability and their ability to predict known phenomena

The Role of Emotion
in Believable Agents

Joseph Bates

There is a notion in the Arts of "believable character." It does not mean an honest or reliable character, but one that provides the illusion of life, thus permitting the audience's suspension of disbelief

The idea of believability has long been studied and explored in literature, theater, film, radio drama, and other media. Traditional character animators are among those artists who have sought to create believable characters, and the Disney animators of the 1930s made great strides toward this goal. The first page of the enormous classic reference work on Disney animation [12] begins with these words:

> Disney animation makes audiences really believe in... characters, whose adventures and misfortunes make people laugh — and even cry. There is a special ingredient in our type of animation that produces drawings that appear to think and make decisions and act of their own volition; it is what creates the illusion of life.

Many artificial intelligence researchers have long wished to build robots, and their cousins called "agents," that seem to think, feel, and live. These are creatures with whom you'd want to share some of your life—as with a companion, or a social pet. For instance, in his 1985 American Association of Artificial Intelligence (AI) presidential address [3], Woody Bledsoe told of his continuing dream to build a computer friend. He spoke of the "excitement of seeing a machine act like a human being, at least in many ways," of building a machine that could "understand, act autonomously, think, learn, enjoy, hate" and which "liked to walk and play Ping-Pong, especially with me."

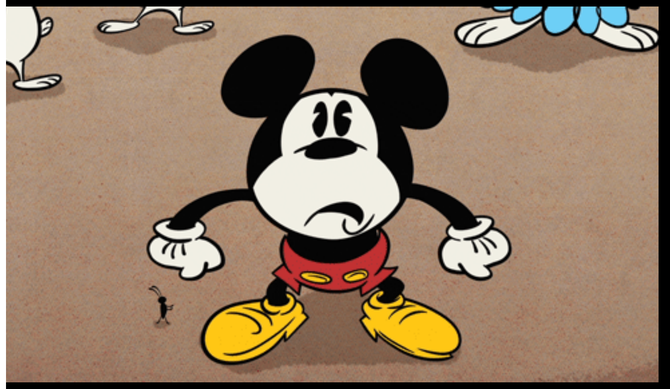
Woody Bledsoe is hardly alone. Further reading on the dreams of animators and AI researchers finds both groups speaking of thinking, feeling, living creatures, of creating at least the illusion of life, of building apparently autonomous entities that people, especially their creators, would genuinely care about. Both groups also speak of achieving these dreams by finding the essence of the creatures to be simulated, and reconstructing that essence in the medium of the artist's or scientist's choice.

As AI researchers tried to find these essential qualities of humanity, they gravitated toward reasoning, problem solving, learning via concept formation, and other qualities apparently central to the capacity we call intelligence. Perhaps this happened because these qualities are characteristic of the idealized scientist, and thus are valued by the communities of which the researchers were part.

Artists, in particular the character animators, also tried to understand and express the essence of humanity in their constructions. Character animators had to be especially analytic, because they had to produce human life from nothing more than individual, hand-drawn, flat-shaded line drawings, moved frame by frame, without being able to rely on a human actor to portray the character. The practical requirement of producing hundreds of thousands of these drawings forced animators to use extremely simple, nonrealistic imagery, and to seek and abstract precisely that which was crucial.

It can be argued that while scien-

122 July 1994/Vol.37, No.7 COMMUNICATIONS OF THE ACM



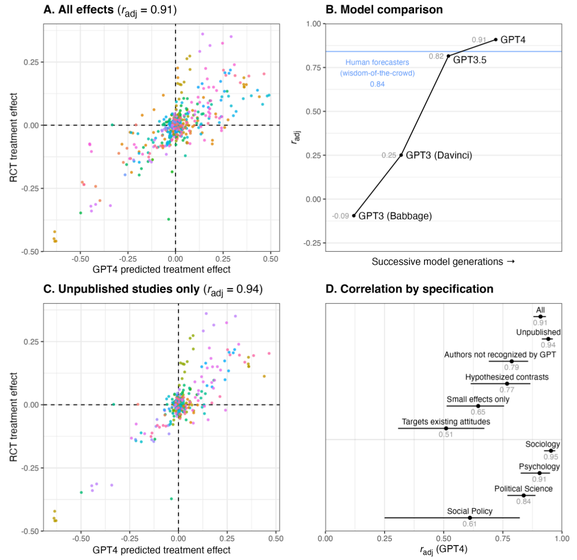


Figure 2: **LLMs accurately predict treatment effects in text-based social science experiments conducted in the US.** (a) In a dataset of 70 text-based experiments with 476 effects, LLM-derived estimates of treatment effects pooled across many prompts were strongly correlated with original treatment effects ($r = 0.85$; $r_{adj} = 0.91$). (b) The accuracy of LLM-derived predictions improved across generations of LLMs, with accuracy surpassing predictions collected from the general population. (c) LLM-derived predictions remained highly accurate for studies that could not have been in the LLM training data given they were not published prior to the LLM training data cutoff date. (d) In robustness check analysis of various subsets of experiments, accuracy of LLM-derived predictions remained high. In panels A and C, different colors depict different studies.

J. Bates, The Role of Emotion in Believable Agents. Commun. ACM 37, 122-125 (1994).
A. Ashokkumar, L. Hewitt, I. Ghezae, R. Willer, "Predicting Results of Social Science Experiments Using Large Language Models" (2024).

8. Going forward, we ought to establish a scientific foundation for simulations that will allow us to trust simulations of unseen worlds. Agent banks might serve this purpose

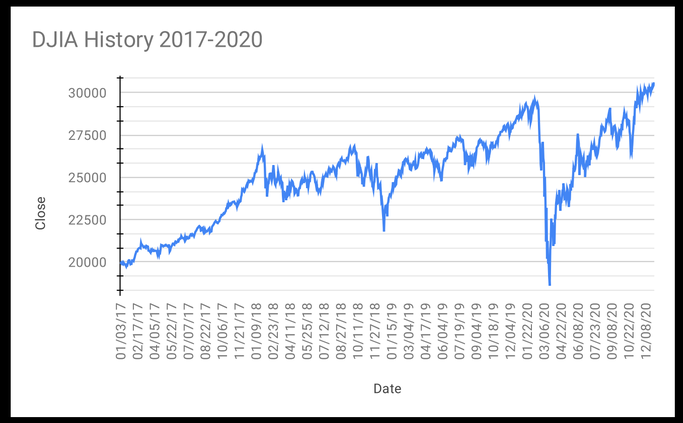
Please note that the responses you share will be shared with your classmates, so you do not have to share any information you ar
to answer, simply write "Prefer not to answer" in the response column.

Copy this spreadsheet and answer the survey questions. Once you are done, download it as a CSV file and submit it on Canvas.

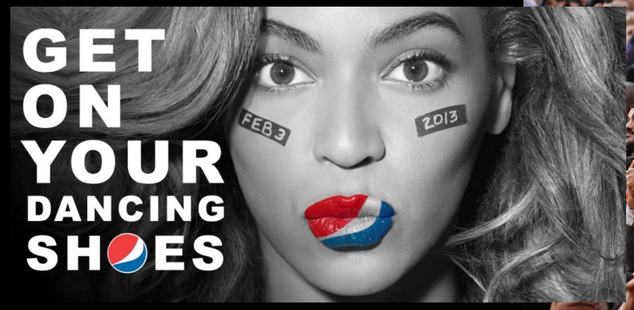
	Question	Options
Q1	What is your age group?	18-24, 25-34, 35-44, 45-54, 55+
Q2	What is your gender?	Male, Female, Non-binary, Prefer not to say, Other
Q3	Where did you grow up?	Urban, Suburban, Rural, Small town, Multiple locations
Q4	Which of the following activities do you spend the most time on?	Work, Family time, Socializing, Hobbies, Relaxing, Exercising
Q5	Which value is most important to you?	Integrity, Compassion, Ambition, Independence, Community
Q6	How would you describe the personality of your closest friend or fam	Extroverted, Introverted, Thoughtful, Outgoing, Analytical
Q7	How do you typically act in unfamiliar social contexts?	Confident, Reserved, Friendly, Neutral, Awkward
Q8	If you had infinite money, how would you spend most of your time?	Traveling, Pursuing hobbies, Helping others, Investing, Working on p
Q9	What is your favorite hobby?	Reading, Sports, Arts & crafts, Traveling, Video games
Q10	How would you describe your political affiliation?	Liberal, Conservative, Moderate, Libertarian, Apolitical
Q11	How many places have you lived in?	1, 2-3, 4-5, More than 5
Q12	What is most important to you in social relationships?	Trust, Fun, Loyalty, Intellectual connection, Shared experiences
Q13	How would you describe your childhood?	Happy, Difficult, Balanced, Adventurous, Strict
Q14	What is your MBTI type?	[Pick one of the 16 categories], I don't know
Q15	What is your primary goal for the next 5 years?	Career growth, Personal development, Family, Financial stability, Tra
Q16	What do you fear the most?	Failure, Rejection, Loneliness, Uncertainty, Loss
Q17	Have you experienced any childhood trauma that affects you today?	Yes, No, Unsure
Q18	How often do you experience intrusive thoughts?	Never, Rarely, Occasionally, Frequently, Constantly
Q19	What has been one of the most meaningful events in your life?	Birth of a child, Graduation, Loss of a loved one, Marriage, Moving t
Q20	Have you experienced tension growing up between different cultural	Yes, No, Somewhat, I'm unsure
Q21	When solving a difficult situation, what is your primary approach?	Logical analysis, Asking for help, Intuition, Trial and error, Avoidance
Q22	How would you describe your religious or spiritual beliefs?	Strongly religious, Spiritual but not religious, Atheist, Agnostic, Unde
Q23	What is your most prized possession?	Family heirloom, Car, Home, Tech gadget, Jewelry
Q24	What is your biggest career aspiration?	Becoming a leader in my field, Achieving work-life balance, Financia
Q25	How do you solve difficult situations?	Analyzing all options, Relying on instinct, Seeking advice, Procrastin
Q26	What trait do you value most in friends?	Loyalty, Humor, Intelligence, Empathy, Honesty
Q27	What would you with $100?	Buy something special, Save it, Nothing

9. Doing so promises to help us address new sets of social scientific questions that are too difficult to tackle today

Phantom traffic jams

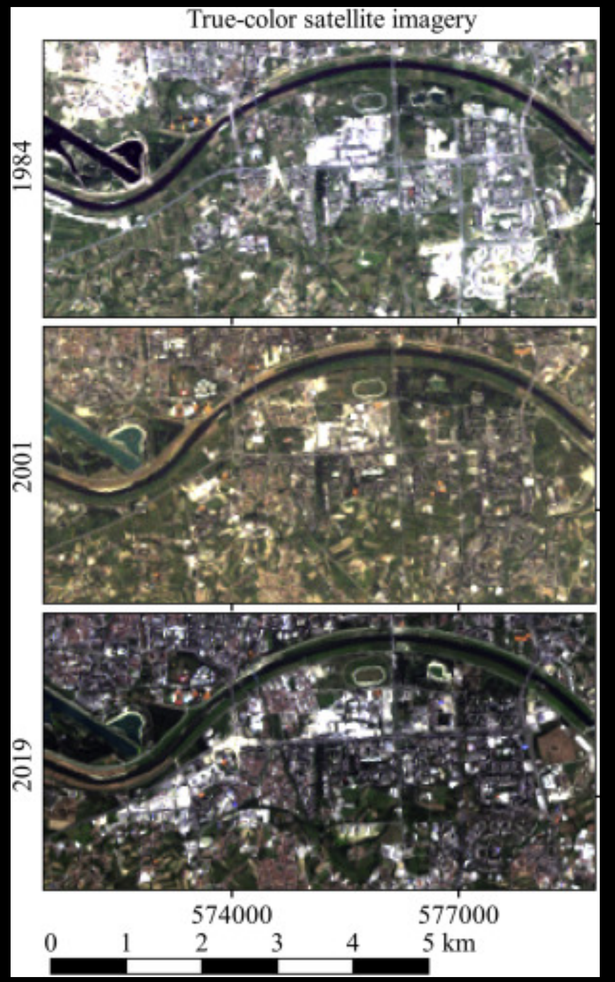


Market crash

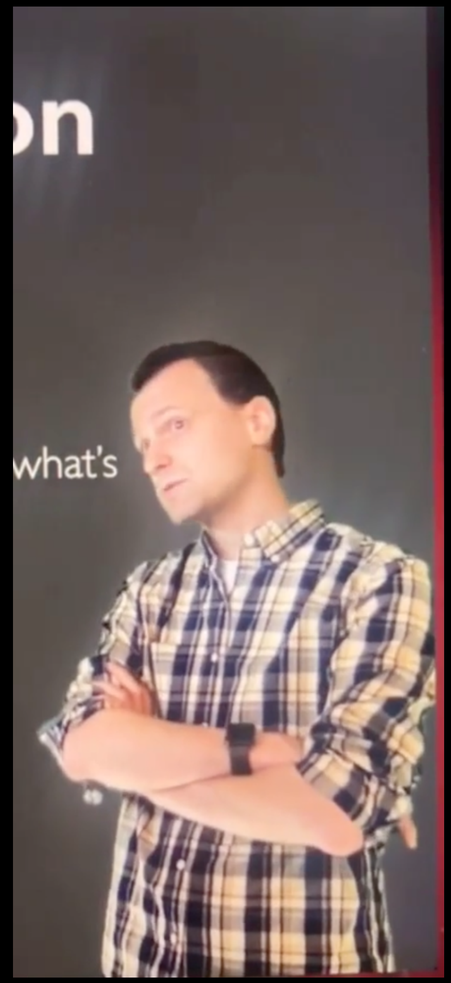


Consumer behavior

Social movement



Urban growth



Viral content

1. Simulations define an environment for individuals, then output their interactions.
2. Generative AI presents an opportunity to create more open-ended simulations.
3. The promise is to enable us to tackle wicked problems.
4. To build simulations, you start by choosing the level of analysis you want to conduct.
5. You then build the individual agents and their environment.
6. So far, we have evaluated the success of simulations by testing their believability and their ability to predict known phenomena.
7. Going forward, we should establish a scientific foundation (e.g., an agent bank) for simulations that will allow us to trust simulations of unseen worlds.
8. Doing so promises to help us tackle wicked problems that are too difficult to address today.

Q: In future iterations of this course, are there topics you wish we had covered?

pollev.com/helenav330

Future work. 1 ~ 32 years

So... where is the field headed? Figuring that out is a wicked problem in itself, but let me speculate

(Let this serve as my pre-registration — the slides are posted to Github :))

Year 1. Scientific Foundation and Models of Individuals

- Currently, the field of AI agents and simulations is working to establish a 'scientific foundation' for simulations.
  - What are the right building blocks for simulations?
  - How can we build robust simulations, and how do we determine whether a simulation is flawed?
- Different bets are being placed on what should be considered the right building blocks.

Year 2. Models of Interactions

- In the next couple of years, I suspect that we will begin to more seriously delve into building and evaluating agent interactions.
  - These are necessary building blocks if we want to develop generative agent-based models that involve multiple agents.

Year 4. Merging of Tool-Based Agents and Simulation Agents

- Currently, there is a subtle divergence within the 'agent' community.
  - Tool-based agents aim to automate tasks, while simulation agents aim to simulate and predict interactions.
  - I posit that the core ingredient for advancing tool-based agents (and realizing Mark Weiser’s vision) is simulations.
- In four years’ time, both approaches will have 'matured' enough for a serious convergence to occur.
- This will unlock a wave of new applications and opportunities in the medium term.

Year 8. Societal Simulations

- There is a significant promise that the field of simulation is making: creating large, multi-agent simulations of societies to address wicked problems.
    - Currently, this is far out of reach, with still-weak models and no comprehensive models to represent the world.
- By year eight, I suspect we will likely have the ingredients to enable semi-large (1 million) societal simulations.
- If this field were to win a Nobel Prize, the prize-winning (or catalyzing) work, akin to Schelling's, would likely emerge around this time.

Year 16. Simulation as a New Computing Platform

- By year 16, neither AI nor simulations will be considered 'new' by any means; they will be facts of life.
  - This implies that the underlying technology will also have matured (with perhaps a few very large central models and many smaller, highly performant models).
- I posit that we will find multi-agent simulations using many smaller models to be uniquely powerful, rather than relying on a single large model.
  - This will be especially true for scientific problems requiring diverse perspectives and for answering wicked problems.
- Where a large central model will function like a CPU, simulations will play the role of a GPU.

Year 32. Multiverse

- I hope that simulation will be viewed as the killer application of AI.
- What is initially a 'killer application' of a platform often becomes a platform itself.
- Applications built on simulation will leverage our ability to create countless multiverses in simulations to help us navigate our future.

References

- T. C. Schelling, Dynamic models of segregation. Journal of Mathematical Sociology 1, 143-186 (1971).
- J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, M. S. Bernstein, Generative agents: Interactive simulacra of human behavior, in Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (ACM, 2023).
- H. W. J. Rittel, M. M. Webber, Dilemmas in a general theory of planning. Policy Sciences 4, 155-169 (1973).
- C. Qian, W. Liu, H. Liu, N. Chen, Y. Dang, J. Li, C. Yang, W. Chen, Y. Su, X. Cong, J. Xu, D. Li, Z. Liu, M. Sun, ChatDev: Communicative Agents for Software Development, in Proceedings of the 2024 Annual Conference of the Association for Computational Linguistics (ACL 2024).
- P. Jansen, M.-A. Côté, T. Khot, E. Bransom, B. Dalvi Mishra, B. P. Majumder, O. Tafjord, P. Clark, DISCOVERYWORLD: A Virtual Environment for Developing and Evaluating Automated Scientific Discovery Agents. Preprint (2024).
- J. Li, S. Wang, M. Zhang, W. Li, Y. Lai, X. Kang, W. Ma, Y. Liu, Agent Hospital: A Simulacrum of Hospital with Evolvable Medical Agents. Preprint (2024).

References

- J. Bates, The Role of Emotion in Believable Agents. Commun. ACM 37, 122-125 (1994).
- A. Ashokkumar, L. Hewitt, I. Ghezae, R. Willer, "Predicting Results of Social Science Experiments Using Large Language Models" (2024).

CS 222: AI Agents and Simulations

Stanford University

Joon Sung Park