May 2025: Senior Spring

Simulation fever & applied progress studies

Hype around LLM social simulation, Digital Twins, and Applied Progress Studies.

Right now [May 2025], LLM social simulation is **having its moment**. A16Z Speedrun just put out a request for social simulation startups. Aaru, a flagship company in this space, has partnered with Accenture. YC backed Artificial Societies in W25 to simulate at scale. Spurtest just raised a $4.5 million round for QA simulation. In addition, everyone I've met in the last few weeks seems to be building in this space, in finance, in A/B testing, in consumer growth. LLM-driven social simulation is shedding its obscurity and emerging as one of the few **genuinely new paradigms** born directly out of the ChatGPT era—not as an optimization tool, but as a fundamentally **novel** way to model and understand human dynamics.

But in my eyes, the current hype is missing the point. Social simulation is just a step toward a greater prize: general simulation. To understand why that distinction matters, we need to trace the intellectual roots of simulation, lay out the rationale behind general simulation, and outline some key steps required to get there.

P.S. We will also touch on Progress Studies implications and an urban planning example.

What is LLM social simulation?

LLM social simulation is an extension of agent-based models (ABMs) and complexity science, where interactions between individual "agents" (generalized to people, organizations, etc) at a low level give rise to emergent macro-behaviors. This perspective on "generative social science" is an ongoing field of research that stretches back many decades.

A seminal example is Epstein & Axtell's "Sugarscape" (1996), which showed how heterogeneous agents competing for resources in a grid world can generate complex societal phenomena (e.g,. wealth distributions, cultural diffusion, trade networks) without any central control. This is a *generative* approach to social science, arguing that **if you can grow the phenomenon in a simulation, you have explained it.**

Other examples include Thomas Schelling's segregation model, which illustrated how individual neighborhood preferences lead to global segregation, and von Neumann's cellular automata as evidence of complexity from simplicity.Each of these cases was intertwined with the Santa Fe Institute's continued belief that most social patterns are **emergent properties** of simple agent interaction rather than constructed by some mathematical model of the world (Epstein 1999).

LLMs fundamentally improve this paradigm.

Previously, agents were defined via strict deterministic rules. With LLM simulation, one can model every human (or group/organization) as an LLM agent — a non-deterministic entity — which allows for a better representation of an agent's decision-making process, communication, and adaptation. In this use case, LLMs are a vast improvement in the decision-making process that is paramount to constructing a "good agent".

P.S. If you are interested in learning more, I recommend these papers:

- Generative Agents: Interactive Simulacra of Human Behavior
- Large language models empowered agent-based modeling and simulation: a survey and perspectives

Why should we care?

My personal framing around social simulation is quite different than the corporate-incentive driven, b2b sass edge lords who seek to turn simulation into a tool for creating "value". In my eyes, simulation will be one of the most important technologies to get right in the next few decades. The LLM unlock in the last few years is pivotal in creating **general simulation** → combining humanlike interaction with the systems that govern our worlds. Not just modelling, traffic, pollution, and urban sprawl, but including the micro decisions of a population that lead to large-scale changes in one system

The MIT population lab has a clear vision of why this can be so impactful. They reference Asimov's psychohistory, or the science of predicting the future given changes in large populations. It is clear that "understanding collective human behavior is the key to shaping our shared destiny." While psychohistory remains fiction (and, as sociologist Phil Schrodt quipped, *"there is no psychohistory, and there never will be"*), it inspired real researchers to pursue predictive social science. For instance, the ambitious FuturICT initiative, led by Dirk Helbing (ETH Zurich), explicitly asked if Asimov's vision could be realized. Helbing's proposed "Living Earth Simulator" aimed to harness big data and complexity models to explore *"what-if" scenarios* for global society, not to exactly predict, but to better understand possible futures. This reflects a shift from pure prediction to decision support, as the biggest unlock of working simulation is **to test nearly anything in silico.**

The justification for unified simulation lies in another critical concept, Progress Studies. Simulation is the closest we can get to some kind of **"Applied Progress Studies,"** a loose way of describing tools/systems for applying progress studies research to make a real-world impact. "Applied Progress Studies" has only been mentioned informally until now, so I want to canonize the phrase. The earliest example I could find was a 2021 EA Forum Post that contrasted EA with "a real Applied Progress Studies agenda," noting such an agenda does not yet exist but ought to be articulated. Since then, Alex Guzey has made multiple mentions of the phrase in a BioDojo YouTube interview titled "Implementing Applied Progress Studies" and in his blog

Seeds of Science, treating the label as shorthand for turning progress-studies insights into funding mechanisms, institutional reforms, and immigration policy.

In the same way that medicine turns biology into cures, *Applied* Progress Studies might turn growth diagnostics into concrete programs, products, and policies. As such, the biggest piece of leverage in turning progress studies research is experimentation and measurement. Simulation unlocks Applied Progress — here's an in-depth look:

1. Applied Progress Studies *requires* interventions that work
2. Effective interventions emerge from fast feedback cycles
3. Physical trials of any complex system are bottlenecked by time, money, and ethics.
4. Simulation collapses the cycle — iteration speed turns from years into days.

If Applied Progress Studies is medicine for civilization, simulation is its MRI scanner, petri dish, and flight-sim all rolled into one. It supplies the necessary counter-factuals, speed, and transparency that turn grand theories into testable, improvable engineering problems that move our world forward.

What does it take to get here?

In order to build unified prediction models (for Applied Progress Studies), the biggest barrier is the combination of the social simulation and physical systems that govern our world. I have already outlined some of the underlying structure for social simulation, but physical systems would require completely different infrastructure — the **digital twin**. Coined by Michael Grieves in the early 2000s, the digital twin is an exact virtual replica of physical systems that stay in sync with the real world via streaming sensor data. This paradigm quickly took hold in manufacturing, aerospace engineering, and smart city initiatives.

Interestingly, the concept was foreshadowed by computer scientist **David Gelernter** in his 1991 book *"Mirror Worlds."* Gelernter envisioned that "every aspect of life could soon be modeled in a parallel digital simulation", where everything in the real world is continuously fed by data into a **high-fidelity, real-time virtual mirror**. Crucially, Gelernter predicted these "high-tech voodoo dolls" would not just **monitor** the world but also allow us to **predict** possible futures by running "endless simulations of possible future events" inside the model. Gelernter was very much ahead of his time, as the "high-tech voodoo doll" is nearly equivalent to the use case of modern digital twin.

The **current frontier** is expanding this concept from physical systems to **social systems**, creating terms like "social digital twin" or "digital twins of society." A Social Digital Twin, as defined by researchers at Fujitsu and Carnegie Mellon University, "digitally reproduces the relationships and connections between people, goods, the economy and society to offer a **simulation, prediction and decision-making environment**" for solving complex social issues.

In short, the theoretical underpinnings for simulating society come from TWO **converging streams**: ABMs and complexity science provide bottom-up *mechanistic* models of emergent social phenomena, and the digital twin concept provides the *integration of real-time data*, physical systems, and the aspiration to mirror reality in silico.

We have these two different paradigms of simulation:

1. **LLM-Powered Emergent Agent-Based Models (Social Simulation)**
    - Each person, organization, or group is represented by an LLM-driven agent capable of memory, reasoning, and adaptation.
    - Millions of such agents interact under simple rules; their micro-decisions generate emergent macro-patterns (opinion shifts, market booms, mobility trends).
2. **Continuous, Sensor-Driven Digital Twins of Physical Systems**
    - Real-time data streams from IoT devices feed physics-based or differential-equation models of infrastructure and environment (traffic networks, power grids, climate, supply chains).
    - The twin mirrors current state and forecasts near-term dynamics of physical assets with high fidelity, independent of human decision-making logic.

Let's think through an example of a particularly impactful unified system — a model for urban planning. *Note, this is still completely theoretical.*

Consider a scenario where a city wants to implement a new policy aimed at addressing traffic congestion and promoting sustainable transportation in rapidly expanding cities. In order to understand if this policy will be beneficial, we construct two parallel models: **social simulation for human interaction** and **systemic changes via digital twins.**

Each city resident and commuter is modeled as an LLM-driven agent with memory, reasoning, and adaptive behaviors. These agents make decisions about transportation based on personal factors such as cost, convenience, social influence, and environmental considerations. Collectively, these individual choices lead to **emergent patterns**, like shifts toward greater public transit use or increased cycling. For example, ETH Zurich's City Science Lab has successfully developed an AI-powered digital twin of Sarajevo, allowing city planners to conduct bottom-up simulations that significantly influence urban development decisions.

Simultaneously, a digital twin captures the state of physical infrastructure through real-time data from IoT sensors, GPS systems, and traffic cameras. This virtual replica mirrors roads, transit routes, and bike paths, forecasting issues like congestion or maintenance requirements. However, directly presenting this raw sensor data to simulated human agents is unrealistic. Humans don't process data as exact vehicle counts or precise speeds. Instead, they perceive abstracted information, such as recognizing congestion nearby. There must be an intentional transformation ("human-oriented obfuscation") of the digital twin's precise sensor data into a simplified, subjective form → an "agent-state" that aligns with human perception and decision-making.

This approach, integrating precise digital twin data with abstracted agent perceptions via LLM-driven models, is *one (albeit a personal favorite)* of many potential solutions for merging these two methodologies. Through unification, urban planners can quickly simulate and assess novel policy interventions such as congestion pricing, infrastructure upgrades, or increased transit frequency.

However, improvement is necessary to get to unified simulation models. On the social simulation side, LLM-driven agents often exhibit biased or overly generalized behaviors, failing to capture complex individual differences and group dynamics adequately. Generative agents are prone to stereotype generalizations, especially in the context of LLM agents. In addition, LLM personas will never act as outliers, a significant sample in any simulation process.

Additionally, scaling simulations to encompass realistically large populations remains computationally demanding. Frequent updates required by dynamic city simulations present substantial computational challenges. Scaling LLM multi-agent simulations to millions of agents is a barrier that must be overcome. For digital twins, managing real-time data streams presents challenges in latency and processing capacity, diminishing accuracy and reliability. The high costs and logistical complexities of deploying comprehensive IoT infrastructures worsen these issues in under-resourced urban environments.

Ultimately, if unified simulation can overcome these barriers (and many more not listed here), it will be a pivotal technology of the next few decades. It will empower decision-makers to craft policies and infrastructure improvements at an unprecedented pace. By dramatically reducing the feedback loop between idea, implementation, and evaluation, **simulation can be the ultimate lever of Applied Progress Studies**, turning theory into **practical, measurable outputs.**

Simulation fever & applied progress studies

Right now, LLM social simulation is having its moment. A16Z Speedrun just put out a request for social simulation startups. Aaru, a flagship company in this space, has partnered with Accenture. YC backed Artificial Societies in W25 to simulate at scale. Spurtest just raised a $4.5 million round for QA simulation. In addition, everyone I've met in the last few weeks seems to be building in this space, in finance, in A/B testing, in consumer growth.
LLM-driven social simulation is shedding its obscurity and emerging as one of the few genuinely new paradigms born directly out of the ChatGPT era—not as an optimization tool, but as a fundamentally novel way to model and understand human dynamics.
But in my eyes, the current hype is missing the point. Social simulation is just a step toward a greater prize: general simulation. To understand why that distinction matters, we need to trace the intellectual roots of simulation, lay out the rationale behind general simulation, and outline some key steps required to get there.
P.S. We will also touch on Progress Studies implications and an urban planning example.
What is LLM social simulation?
LLM social simulation is an extension of agent-based models (ABMs) and complexity science, where interactions between individual "agents" (generalized to people, organizations, etc) at a low level give rise to emergent macro-behaviors. This perspective on "generative social science" is an ongoing field of research that stretches back many decades.
A seminal example is Epstein & Axtell's "Sugarscape" (1996), which showed how heterogeneous agents competing for resources in a grid world can generate complex societal phenomena (e.g,. Wealth distributions, cultural diffusion, trade networks) without any central control. This is a generative approach to social science, arguing that if you can grow the phenomenon in a simulation, you have explained it.
Other examples include Thomas Schelling's segregation model, which illustrated how individual neighborhood preferences lead to global segregation, and von Neumann's cellular automata as evidence of complexity from simplicity. Each of these cases was intertwined with the Santa Fe Institute's continued belief that most social patterns are emergent properties of simple agent interaction rather than constructed by some mathematical model of the world (Epstein 1999).
LLMs fundamentally improve this paradigm.
Previously, agents were defined via strict deterministic rules. With LLM simulation, one can model every human (or group/organization) as an LLM agent — a non-deterministic entity — which allows for a better representation of an agent's decision-making process, communication, and adaptation. In this use case, LLMs are a vast improvement in the decision-making process that is paramount to constructing a "good agent".
P.S. If you are interested in learning more, I recommend these papers:

- Generative Agents: Interactive Simulacra of Human Behavior
- • Large language models empowered agent-based modeling and simulation: a survey and perspectives

- Why should we care?
- My personal framing around social simulation is quite different than the corporate-incentive driven, b2b sass edge lords who seek to turn simulation into a tool for creating “value”. In my eyes, simulation will be one of the most important technologies to get right in the next few decades. The LLM unlock in the last few years is pivotal in creating general simulation combining humanlike interaction with the systems that govern our worlds. Not just modelling, traffic, pollution, and urban sprawl, but including the micro decisions of a population that lead to large-scale changes in one system
- The MIT population lab has a clear vision of why this can be so impactful. They reference Asimov’s psychohistory, or the science of predicting the future given changes in large populations. It is clear that “understanding collective human behavior is the key to shaping our shared destiny.” While psychohistory remains fiction (and, as sociologist Phil Schrodt quipped, “there is no psychohistory, and there never will be”), it inspired real researchers to pursue predictive social science. For instance, the ambitious FuturICT initiative, led by Dirk Helbing (ETH Zurich), explicitly asked if Asimov’s vision could be realized. Helbing’s proposed “Living Earth Simulator” aimed to harness big data and complexity models to explore “what-if” scenarios for global society, not to exactly predict, but to better understand possible futures. This reflects a shift from pure prediction to decision support, as the biggest unlock of working simulation is to test nearly anything in silico.
- The justification for unified simulation lies in another critical concept, Progress Studies. Simulation is the closest we can get to some kind of “Applied Progress Studies,” a loose way of describing tools/systems for applying progress studies research to make a real-world impact. “Applied Progress Studies” has only been mentioned informally until now, so I want to canonize the phrase. The earliest example I could find was a 2021 EA Forum Post that contrasted EA with “a real Applied Progress Studies agenda,” noting such an agenda does not yet exist but ought to be articulated. Since then, Alex Guzey has made multiple mentions of the phrase in a BioDojo YouTube interview titled “Implementing Applied Progress Studies” and in his blog Seeds of Science, treating the label as shorthand for turning progressstudies insights into funding mechanisms, institutional reforms, and immigration policy.
- In the same way that medicine turns biology into cures, Applied Progress Studies might turn growth diagnostics into concrete programs, products, and policies. As such, the biggest piece of leverage in turning progress studies research is experimentation and measurement. Simulation unlocks Applied Progress — here’s an in-depth look:
- 1. Applied Progress Studies requires interventions that work
- 2. Effective interventions emerge from fast feedback cycles
- 3. Physical trials of any complex system are bottlenecked by time, money, and ethics.

- 4. Simulation collapses the cycle — iteration speed turns from years into days.
- If Applied Progress Studies is medicine for civilization, simulation is its MRI scanner, petri dish, and flightsim all rolled into one. It supplies the necessary counter-factuals, speed, and transparency that turn grand theories into testable, improvable engineering problems that move our world forward.
- What does it take to get here?
- In order to build unified prediction models (for Applied Progress Studies), the biggest barrier is the combination of the social simulation and physical systems that govern our world. I have already outlined some of the underlying structure for social simulation, but physical systems would require completely different infrastructure — the digital twin. Coined by Michael Grieves in the early 2000s, the digital twin is an exact virtual replica of physical systems that stay in sync with the real world via streaming sensor data. This paradigm quickly took hold in manufacturing, aerospace engineering, and smart city initiatives.
- Interestingly, the concept was foreshadowed by computer scientist David Gelernter in his 1991 book “Mirror Worlds.” Gelernter envisioned that “every aspect of life could soon be modeled in a parallel digital simulation”, where everything in the real world is continuously fed by data into a high-fidelity, real-time virtual mirror. Crucially, Gelernter predicted these “high-tech voodoo dolls” would not just monitor the world but also allow us to predict possible futures by running “endless simulations of possible future events” inside the model. Gelernter was very much ahead of his time, as the “high-tech voodoo doll” is nearly equivalent to the use case of modern digital twin.
- The current frontier is expanding this concept from physical systems to social systems, creating terms like “social digital twin” or “digital twins of society.” A Social Digital Twin, as defined by researchers at Fujitsu and Carnegie Mellon University, “digitally reproduces the relationships and connections between people, goods, the economy and society to offer a simulation, prediction and decision-making environment” for solving complex social issues.
- In short, the theoretical underpinnings for simulating society come from TWO converging streams: ABMs and complexity science provide bottom-up mechanistic models of emergent social phenomena, and the digital twin concept provides the integration of real-time data, physical systems, and the aspiration to mirror reality in silico.
- We have these two different paradigms of simulation:
- 1. LLMPowered Emergent AgentBased Models (Social Simulation)
- Each person, organization, or group is represented by an LLMdriven agent capable of memory, reasoning, and adaptation.
- Millions of such agents interact under simple rules; their microdecisions generate emergent macropatterns (opinion shifts, market booms, mobility trends).
- 2. Continuous, SensorDriven Digital Twins of Physical Systems

- Realtime data streams from IoT devices feed physicsbased or differentialequation models of infrastructure and environment (traffic networks, power grids, climate, supply chains).
- The twin mirrors current state and forecasts nearterm dynamics of physical assets with high fidelity, independent of human decisionmaking logic.
- Let's think through an example of a particularly impactful unified system — a model for urban planning. Note, this is still completely theoretical.
- Consider a scenario where a city wants to implement a new policy aimed at addressing traffic congestion and promoting sustainable transportation in rapidly expanding cities. In order to understand if this policy will be beneficial, we construct two parallel models: social simulation for human interaction and systemic changes via digital twins.
- Each city resident and commuter is modeled as an LLM-driven agent with memory, reasoning, and adaptive behaviors. These agents make decisions about transportation based on personal factors such as cost, convenience, social influence, and environmental considerations. Collectively, these individual choices lead to emergent patterns, like shifts toward greater public transit use or increased cycling. For example, ETH Zurich's City Science Lab has successfully developed an AI-powered digital twin of Sarajevo, allowing city planners to conduct bottom-up simulations that significantly influence urban development decisions.
- Simultaneously, a digital twin captures the state of physical infrastructure through real-time data from IoT sensors, GPS systems, and traffic cameras. This virtual replica mirrors roads, transit routes, and bike paths, forecasting issues like congestion or maintenance requirements. However, directly presenting this raw sensor data to simulated human agents is unrealistic. Humans don't process data as exact vehicle counts or precise speeds. Instead, they perceive abstracted information, such as recognizing congestion nearby. There must be an intentional transformation ("human-oriented obfuscation") of the digital twin's precise sensor data into a simplified, subjective form an "agent-state" that aligns with human perception and decision-making.
- This approach, integrating precise digital twin data with abstracted agent perceptions via LLM-driven models, is one (albeit a personal favorite) of many potential solutions for merging these two methodologies. Through unification, urban planners can quickly simulate and assess novel policy interventions such as congestion pricing, infrastructure upgrades, or increased transit frequency.
- However, improvement is necessary to get to unified simulation models. On the social simulation side, LLM-driven agents often exhibit biased or overly generalized behaviors, failing to capture complex individual differences and group dynamics adequately. Generative agents are prone to stereotype generalizations, especially in the context of LLM agents. In addition, LLM personas will never act as outliers, a significant sample in any simulation process.

- Additionally, scaling simulations to encompass realistically large populations remains computationally demanding. Frequent updates required by dynamic city simulations present substantial computational challenges. Scaling LLM multi-agent simulations to millions of agents is a barrier that must be overcome. For digital twins, managing real-time data streams presents challenges in latency and processing capacity, diminishing accuracy and reliability. The high costs and logistical complexities of deploying comprehensive IoT infrastructures worsen these issues in under-resourced urban environments.
- Ultimately, if unified simulation can overcome these barriers (and many more not listed here), it will be a pivotal technology of the next few decades. It will empower decision-makers to craft policies and infrastructure improvements at an unprecedented pace. By dramatically reducing the feedback loop between idea, implementation, and evaluation, simulation can be the ultimate lever of Applied Progress Studies, turning theory into practical, measurable outputs.

Thanks for reading Philomathia! Subscribe for free to receive new posts and support my work.

Subscribe

Thanks to Harsehaj, Madhu, and Saurish for reviewing this post!