<u>Large Population Models</u>

differentiable and private learning over million agents

Ayush Chopra

PhD Candidate, MIT

[5-min talk]; [30-min talk]; [resume]

[github.com/AgentTorch/AgentTorch]

Many of the grand societal challenges, such as pandemics, housing crises, or immunosenescence, may be characterized as emergent phenomena resulting from complex interactions within a large population of autonomous agents (eg: humans, cells). Agent-based models (ABMs) aim to understand such complex systems by explicitly modeling the behavior and interaction of individual agents. ABMs use mechanistic rules to simulate stochastic agent behavior and have helped understand population-level effects in several domains. The emergent effects are sensitive to the scale of the input population, simulation parameters, and modeling assumptions. Hence, using ABMs for practical decision-making requires recreating populations with great detail, calibrating to heterogeneous data, and assimilating granular real-world feedback. This utility is constrained by computational and data access bottlenecks. My research, to unlock this potential, introduces protocols for differentiable and private computation that help scale ABMs to million-size populations, compose with DNNs to ingest multi-modal data, and operate in a secure closed loop with the real world. This leads to two themes: a) building realistic simulations (**Generative Populations**), b) minimizing the sim-2-real gap (**Decentralized Intelligence**).

Generative Populations: Conventional ABMs are difficult to scale, tough to calibrate, and prone to misspecification due to hand-crafted rules. My research introduces GradABM - a paradigm utilizing tensorized and differentiable programming to make ABMs compatible with autograd and alleviate these computational bottlenecks. GradABMs can scale to country-size populations in a few seconds on commodity hardware and differentiate through simulations with stochastic dynamics and conditional interventions [8]. This helps utilize gradient-based learning to compose GradABMs with DNNs to calibrate simulation parameters to multi-modal data [4], design expressive interaction rules with neural specification [1] and unlock rapid simulation-free sensitivity analysis [5]. GradABMs enabled rapid policy analysis during COVID-19 to inform prospective vaccination strategies [6,7] and retrospectively quantify the impact of lockdown policies [5] which impacted millions globally [12]. Recently, we developed AgentTorch - a framework to generalize GradABMs across domains and bridge them with autonomous agents (eg:LLMs). AgentTorch formalizes our technical advances in scaling autonomous agent simulations and processing discontinuous computation graphs in a simple Python-API and has unlocked applications in multiple disciplines [3,11]. AgentTorch models are being deployed to tackle population health challenges across the world - helping Olmsted county in Minnesota understand immunologic response of COVID-19 vaccines; and the New Zealand crown fight antimicrobial resistance in their 5 million citizens.

Decentralized Intelligence: The above computational advances are of limited utility if the quality of underlying data is poor. Conventional ABMs rely on synthetic populations generated using sparse summary statistics derived from real-world observations. Privacy, not data sparsity, is the cause of limited granularity as data is siloed with individuals. Ethically crowdsourcing this data can help guide urgent decisions, as demonstrated by contact tracing applications during the pandemic. For example, our MIT-SafePaths protocol provided digital contact tracing to over a million people across 5 US states and territories [10]. As we move forward, there is a need to go from simply accounting for contacts to deriving actionable intelligence from contact data. By crowdsourcing such "contact intelligence" responsibly, we can enable predictive capabilities to inform critical policy decisions. To achieve this, my research introduces DecABM - a paradigm that utilizes secret sharing protocols to execute decentralized simulations, conduct sensitivity analysis, and calibrate macro parameters ***without*** collecting any agents' information [2]. DecABMs provide robust privacy guarantees to each agent's sensitive information without compromising ABM accuracy. Instead of building synthetic populations, DecABM allows the secure integration of real-agents in the modeling process. This citizen intelligence can have a huge impact on policy making, especially in the global south and we are building relevant infrastructure with the WHO [9]. Interfacing high-resolution autonomous agents with decentralized real agents can help aggregate individual insights and make reliable decisions for the collective. I am excited about the potential of LPMs to accelerate scientific discovery and facilitate critical decision-making.

About me: I am a PhD student at MIT supervised by Prof Ramesh Raskar. My research is focused on advancing learning in million-agent systems. Specifically, my thesis is establishing the technical disciplines of differentiable and private agent-based modeling. My research has resulted in publications at top-tier AI conferences (AAMAS, CVPR, ECCV, KDD, IJCAI, etc), journals (BMJ, Vaccine), and best paper awards at CVPR and ICML workshops. My projects have resulted in 25 patents and have been featured by Reuters, Weather Channel, Venture Beat etc. Prior to MIT, I was a researcher at Adobe where I focused on advancing in-browser computer vision and machine learning and also received the Adobe Outstanding Young Engineer Award 2020. I have co-organized workshops and tutorials on multi-agent and collaborative learning at ICLR 2021, ICLR 2023, and AAMAS 2024. I received my MS from MIT in 2022 and BE from Delhi College of Engineering in 2018.

References

[1]: **Ayush Chopra***, Jayakumar Subramanian*, Balaji Krishnamurthy Ramesh Raskar. a framework for learning in agent-based models. Accepted at **AAMAS 2024.** (**Oral**, pdf, github.com/AgentTorch/AgentTorch)

[2]: **Ayush Chopra**, Arnau Quera-borafull, Nurullah Kuru, Mike Wooldridge, Ramesh Raskar. Private Agent-based Modeling. Accepted at **AAMAS 2024 (Oral,** pdf**)**.

[3]: Gauri Gupta, Ritwik Kapila, **Ayush Chopra**, Ramesh Raskar. First 100 days of pandemic: an interplay of pharmaceutical, behavioral and digital interventions. **AAMAS 2024 (Oral,** pdf**)**

[4]: **Ayush Chopra**, Alex Rodriguez, Jayakumar Subramanian, Aditya Prakash, Ramesh Raskar. Differentiable Agent-based Epidemiology. **AAMAS 2023** and **ICML Workshop 2022** (**Best Paper Award**) (**Oral**, pdf)

[5]: Arnau Quera-bofarull, **Ayush Chopra**, Joe Bullock, Ramesh Raskar, Mike Wooldridge. Don't Simulate Twice: One-shot sensitivity analyses via automatic differentiation. **AAMAS 2023 (Oral,** pdf**)**

[6]: **Ayush Chopra**, Alex Rodriguez, Aditya Prakash, Ramesh Raskar, Tom Kingsley. Using neural networks to calibrate agent-based models improves regional evidence for vaccine strategy and policy. **Vaccine 2023 (**pdf**)**

[7]: Santiago Romero-brufau, **Ayush Chopra**, Ramesh Raskar, Tom Kingsley et al. Public Health Impact of delaying 2nd dose of BNT162b2 vaccine. **British Medical Journal 2021** (pdf)

[8]: **Ayush Chopra**, Ramesh Raskar et al. DeepABM: scalable, fast and differentiable agent-based modeling. **WSC 2021** (**Oral**, pdf)

[9]: Ramesh Raskar, **Ayush Chopra**, Shada Al-salamah, Sameer Pujari, Alain Labrique, Batool Baez. Driving Universal Health Equity via the Global Initiative on AI for Health. In Review.

[10]: Safe Paths: A privacy-first approach to contact tracing.(link)

[11]: Joel Dyer, Arnau Quera-bofarull, **Ayush Chopra**, Ani Calinescu, Mike Wooldrige. Gradient-based Calibration of Financial Agent-based Models. **ICAIF 2023**. (pdf)

[12]: Quantifying the effect of delaying the second COVID-19 vaccine dose in England.(link)