Decentralized agent-based modeling

Anonymous Author(s)
Affiliation
Address
email

Abstract

The utility of agent-based models for practical decision making depends upon their ability to recreate populations with great detail and integrate real-world data streams. However, incorporating this data can be challenging due to privacy concerns. We alleviate this issue by introducing a paradigm for secure agent-based modeling. In particular, we leverage secure multi-party computation to enable decentralized agent-based simulation, calibration, and analysis. We believe this is a critical step towards making agent-based models scalable to the real-world application.

1 Introduction

Agent-based modeling (ABMing) is a bottom-up simulation technique wherein a system is modeled through the interaction of autonomous decision-making entities referred to as agents. Due to their granular approach, ABMs are a promising tool for real-world decision-making and policy design and constitute an active field of research across economics [5, 12, 33], biology [44, 27], and epidemiology [6, 52, 35, 32]. Wider adoption of ABMs, however, is hindered by (1) the need for microdata to generate the underlying agent population, and (2) the often large computational resources required to run, calibrate, and analyze an ABM. Recently, there has been significant progress towards developing new design patterns for ABMs, which exploit tensorization [18, 17] and differentiability [19, 3] of simulators. This has alleviated the computational burdens associated with ABM simulation [18], calibration [19, 49], and analysis [48] by granting access to modern computational techniques such as GPU computing and differentiable programming, allowing ABMs to scale to populations comprised of millions of agents [51, 10].

Yet, the increase in computational efficiency for ABMs can be inconsequential if the quality of the underlying population microdata is poor. Currently, prevalent approaches involve the construction of synthetic populations designed to align with a predefined set of summary statistics derived from real-world observations. For instance, in epidemiological ABMs, the population is crafted to replicate summary statistics obtained from census data [45, 13, 6, 15, 46]. However, it is essential to recognize that the limited granularity of census data arises primarily from privacy considerations rather than actual scarcity of available data. As ABMs continue to scale towards one-to-one representations of real-world systems, there remains a fundamental limitation in their modeling potential as long as privacy are not in place. Previous attempts to augment ABM data with additional information, such as mobility or health data, have resulted in data leaks that exposed agents' personal information [1, 34, 21]. These incidents underscore the need for a decentralized approach to ABMing, where each agent's sensitive information is kept confidential throughout the modeling process.

Motivated by this, we introduce a new paradigm for agent-based simulation that ensures the confidentiality of each agent's sensitive information. Leveraging techniques drawn from secure multi-party computation [38], we develop privacy-preserving protocols for the simulation, calibration, and analysis of ABMs. These protocols offer robust security guarantees to agents while preserving the ability of ABMs to effectively model complex systems. Moreover, our methodology enables secure ABMs to

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

take advantage of differentiable programming, allowing them to be integrated into machine learning pipelines, further boosting their modeling capabilities.

In summary, this work constitutes to our knowledge the first protocol for privacy-preserving ABMs that enables their simulation and calibration. We hope that this development will pave the way for the secure and practical utilization of ABMs as valuable tools for policy-making in real-world settings.

2 Agent-based models

Consider an ABM with N agents $A = \{1, 2, \ldots, N\}$. We denote by $\mathbf{z}_i(t)$ the state of agent i at time t which encapsulates both fixed and time-evolving properties of the simulation agents. For instance, $\mathbf{z}$ can represent age and disease status of human agents in epidemiological models; and account balance of firms in a financial auction model. As the simulation proceeds, an agent i updates their state $\mathbf{z}_i(t)$ by interacting with their neighbors $\mathcal{N}_i(t)$ and the environment $\mathcal{E}(t)$. We assume that the interaction of agents with their neighbors can be conceived as message passing on a graph $\mathcal{G} = (V, E)$, where the vertices V of the graph correspond to the agents, the edges $e_{ij} \in E$ connect neighboring agents, and interactions are represented as messages $M_{ij}(t) = M(\mathbf{z}_i(t), \mathbf{z}_j(t), e_{ij}(t), \boldsymbol{\theta}, t)$, where $\boldsymbol{\theta}$ are the ABM structural parameters. This is is indeed the case for contagion models [22], for example, $M_{ij}(t)$ may represent the transmission of infection from agent j to agent i, which may depend on the susceptibility of agent i ($\mathbf{z}_i$), the infectivity of j ($\mathbf{z}_j$), the properties of the virus ($\boldsymbol{\theta}$), and the nature of the interaction (e_{ij}). Thus, at each step t, each agent updates its state following

$$\mathbf{z}_i(t+1) = f\left(\mathbf{z}_i(t), \bigoplus_{j \in \mathcal{N}_i(t)} M_{ij}(t),\ \boldsymbol{\theta}\right), \tag{1}$$

where $\bigoplus$ denotes an aggregation function over all received messages. The specific form of f can be tailored to capture the unique dynamics of the system under investigation, for instance, the diversity of contagion models can be encapsulated by different functional forms of f [22].

During the simulation of an ABM, a central agent (the modeler) collects a time-series of aggregate statistics over agent states, $\mathbf{x}_t = h(\{\mathbf{z}_i(t) \mid i \in A\})$, which can be used to compare the output of the model to ground-truth data. For instance, in epidemiological ABMs, h may correspond to counting the number of infected agents, so that $\{\mathbf{x}_t\}_t$ is a time-series of daily infections.

As we can see, both Equation (1) and the collection of the summary statistics require agents to communicate their state to other agents. In following sections we introduce a methodology to perform these operations in a privacy-preserving manner.

3 Characterizing Privacy

3.1 Threat Model

We assume an honest-but-curious (a.k.a semi-honest) attacker [31] which aims to learn private information about participating agents. This private information is included in an agent's state $\mathbf{z}_j(t)$, interaction trace $\{\mathcal{N}_i(t) \mid \forall t\}$, and neighborhood messages $\{M_{ij}(t) \mid i \in A, j \in \mathcal{N}(i)\}$. For instance, in epidemiological models, this can correspond to the health and demographic traits, and mobility patterns of individual agents. Such attacker can manifest as the coordinating server which wants to surveil agents using the mobility trace or a (sub-group) of adversarial agents which may be incentivized to steal personal health information of agent cohorts. In the context of agent-based modeling, this information can be leaked during message passing over per-step neighborhoods (Equation (1)) and during the collection of summary statistics over the population. The goal of this work is to alleviate such challenges and design a privacy-preserving mechanism which can compute functions over agents' states without revealing private information.

3.2 Secure Multi-party Computation

Secure multi-party computation enables a set of agents to interact and compute a joint function of their private inputs while revealing nothing but the output [38]. MPC protocols are coordinated with a server (MPC server) and are designed to protect against malicious behavior of adversarial participants.

These malicious participants, either an agent or the server, aim to learn private information (of other entities) or cause the result of computation to be incorrect. The idea was first introduced by Yao for the two-party case [57] and generalized to multiparty settings by Goldreich, Micali and Wigderson (GMW) [26]. Among other properties, GMW protocols guarantee 1) *privacy*: so that no entity can learn anything more than its prescribed output and, 2) *correctness*: so that each agent receives the correct output. For instance, in an epidemiological ABM, this would ensure both that the personal disease status of agents is not leaked and also no agent can misrepresent their disease status. We formalize the GMW protocol and provide an intuitive example in the Appendix.

4 Private Simulation of Agent-based Models

First, we present the SECURESUM protocol, which enables the computation of the sum of agents inputs in a private way, based on the GMW protocol (see Subsection 6.1 for a detailed description and an example) in Algorithm 1.

Algorithm 1: SECURESUM

Data: Agents $\{1, \ldots, N\}$ with secret inputs $s_1, \ldots, s_n$, integer $n > \max\{s_1, \ldots, s_n\}$.
Result: The sum of all shares $S = s_1 + \cdots + s_n$.

Splitting secret into shares and distributing:
Each party i generates N shares $s_{i1}, \ldots, s_{iN} \in \mathbb{Z}_n$ which sum up to s_i.
Each party i distributes all their shares $s_{i1}, \ldots, s_{iN} \in \mathbb{Z}_n$ to $1, \ldots, N$, including themselves.
Secure Computation (Addition):
To add the inputs securely, parties simply add their respective shares $\sigma_i = s_{1i} + \cdots s_{Ni} \mod n$.
Reconstruction:
To reveal the final result of the computation, parties collaborate by summing their shares:
$S = (\sigma_1 + \sigma_2 + \cdots + \sigma_n) \mod n$.

This protocol enables the simulation of ABMs for the case where $\bigoplus$ corresponds to addition in Equation (1), which is indeed the case in all contagion models. Furthermore, as long as the agent's update function f is differentiable respect to the structural parameters $\boldsymbol{\theta}$, which is indeed the case for many ABMs [19], each agent can store $\nabla_{\boldsymbol{\theta}} f$ for use during the calibration step. With all this in mind, we present in Algorithm 2, a privacy-preserving protocol for updating the agent's states.

Algorithm 2: SECUREAGENTUPDATE

Data: Agent i with state $\mathbf{z}_i(t)$, Neighboring agent's messages $\{M_{ij}(t) \mid j \in \mathcal{N}(i)\}$, Integer n, State update rule f, ABM parameters $\boldsymbol{\theta}$
Result: New state $\mathbf{z}_i(t+1)$

Agent i calls the SECURESUM protocol with neighbors $\{j \mid j \in \mathcal{N}(i)\}$ and integer n to get the sum $M_i(t) = \sum_{j \in \mathcal{N}(i)} M_{ij}(t)$.
Agent i updates its state $\mathbf{z}_i(t+1) = f\left(\mathbf{z}_i(t), M_i(t), \boldsymbol{\theta}\right)$ and stores the gradient $\nabla_{\boldsymbol{\theta}} f$.

It is worth noting that, in contrast to general applications of the GMW protocol, only the agent who starts the protocol receives the result of the computation, since there is no need for the neighboring agents to have access to that information.

Next, we introduce the SECURESIMULATION protocol in Algorithm 3, where, in addition to performing agent updates, we collect a time-series of aggregate statistics over the agent's population and its gradient respect to the ABM structural parameters $\boldsymbol{\theta}$.

4.1 Private calibration of ABMs

Calibration refers to the process of tuning the set of structural parameters $\boldsymbol{\theta}$ so that ABM outputs $\mathbf{x}$ are compatible with given observational data $\mathbf{y}$. In epidemiological ABMs, for instance, this entails determining values for parameters like the reproduction number R_0 and mortality rates to align with the observed daily infection or mortality data. During the calibration of an ABM, the modeler (central MPC server) requires the ability to evaluate the ABM at different values of $\boldsymbol{\theta}$, and, in the case of

Algorithm 3: SECURESIMULATION

Data: MPC server C, Agents $\{1, \dots, N\}$ with states $\{\mathbf{z}_1, \dots, \mathbf{z}_N\}$, ABM parameters $\boldsymbol{\theta}$, State update rule f, Number of time-steps T

Result: Aggregate statistics $\mathbf{x} = x_1, \dots, x_T$ and gradients $\nabla_{\boldsymbol{\theta}}\mathbf{x}$.

C generates a large enough prime number P and the requested statistics collecting function h; and sends them to all agents along ABM parameters $\boldsymbol{\theta}$.

for t = *1, . . . , T* **do**

    for i = *1, . . . , N* **do**

        Agent i calls the SECUREAGENTUPDATE protocol (Algorithm 2) to compute $\mathbf{z}_i(t+1)$.

        Agent i gathers its information of interest $h(\mathbf{z}_i(t+1))$ and gradient $\nabla_{\boldsymbol{\theta}} h(\mathbf{z}_i(t+1))$.

    C calls the SECURESUM protocol with all agents to collect the aggregate statistics x_t and their gradients $\nabla_{\boldsymbol{\theta}} x_t$.

C returns the accumulated $\mathbf{x}$ and $\nabla_{\boldsymbol{\theta}}\mathbf{x}$.

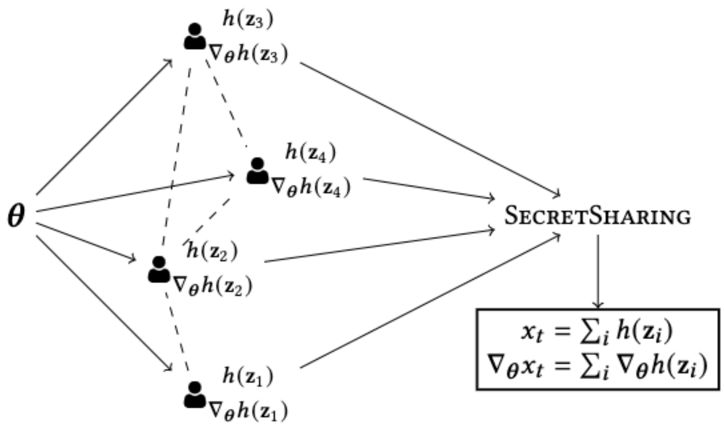

Figure 1: Diagram illustrating the SECURESIMULATION protocol for ABM parameters θ

gradient-assisted calibration, the gradient of the outputs with respect to $\boldsymbol{\theta}$. These values are stored at the MPC server where they are used as part of a ML pipeline to perform calibration. The retrieval of these quantities is seamlessly enabled by the SECURESIMULATION algorithm and so all the standard calibration techniques for ABM (see, e.g., [23, 47] can be easily adapted to this private framework to ensure a decentralized calibration process that respects agent's privacy. Moreover, the ability to retrieve the gradient in a private way enables the use of more advanced gradient-assisted techniques such as generalized variational inference [49].

5 Conclusion

In this paper, we have introduced a new paradigm of decentralized agent-based modeling which enables simulation and calibration on real world data, all without compromising the privacy of the agents involved. Our approach leverages MPC techniques to develop robust privacy-preserving protocols, without compromising the correctness of the ABM output. Our paradigm may be readily integrated into established platforms such as contact-tracing mobile applications, as a means to greatly improving analysis and forecasting of complex systems across diverse domains. Further, we validate by scalability of our simulation and calibration protocols via a decentralized epidemiological ABM, in the appendix.

References

[1] 2021. Indonesia Probes Suspected Data Breach on COVID-19 App. *Reuters* (Aug. 2021).

[2] Philipp Andelfinger. 2021. Differentiable Agent-Based Simulation for Gradient-Guided Simulation-Based Optimization. *arXiv:2103.12476 [cs, eess]* (March 2021). arXiv:2103.12476 [cs, eess]

[3] Gaurav Arya, Moritz Schauer, Frank Schäfer, and Chris Rackauckas. 2022. Automatic Differentiation of Programs with Discrete Randomness. arXiv:2210.08572 [cs, math]

[4] Samuel A. Assefa, Danial Dervovic, Mahmoud Mahfouz, Robert E. Tillman, Prashant Reddy, and Manuela Veloso. 2021. Generating Synthetic Data in Finance: Opportunities, Challenges and Pitfalls. In *Proceedings of the First ACM International Conference on AI in Finance (ICAIF '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. `https://doi.org/10.1145/3383455.3422554`

[5] Robert L. Axtell and J. Doyne Farmer. [n.d.]. Agent-Based Modeling in Economics and Finance: Past, Present, and Future. *Journal of Economic Literature* ([n. d.]). `https://doi.org/10.1257/jel.20221319`

[6] Joseph Aylett-Bullock, Carolina Cuesta-Lazaro, Arnau Quera-Bofarull, Miguel Icaza-Lizaola, Aidan Sedgewick, Henry Truong, Aoife Curran, Edward Elliott, Tristan Caulfield, Kevin Fong, Ian Vernon, Julian Williams, Richard Bower, and Frank Krauss. 2021. June: Open-Source Individual-Based Epidemiology Simulation. *Royal Society Open Science* 8, 7 (July 2021), 210506. `https://doi.org/10.1098/rsos.210506`

[7] Atilim Gunes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. 2018. Automatic Differentiation in Machine Learning: A Survey. *Journal of Machine Learning Research* 18, 153 (2018), 1–43.

[8] Donald Beaver. 1992. Efficient Multiparty Protocols Using Circuit Randomization. In *Advances in Cryptology—CRYPTO'91: Proceedings 11*. Springer, 420–432.

[9] Claudio Bettini, Sergio Mascetti, X Sean Wang, Dario Freni, and Sushil Jajodia. 2009. Anonymity and Historical-Anonymity in Location-Based Services. *Privacy in location-based applications: research issues and emerging trends* (2009), 1–30.

[10] Parantapa Bhattacharya, Jiangzhuo Chen, Stefan Hoops, Dustin Machi, Bryan Lewis, Srinivasan Venkatramanan, Mandy L Wilson, Brian Klahn, Aniruddha Adiga, Benjamin Hurt, et al. 2023. Data-Driven Scalable Pipeline Using National Agent-Based Models for Real-Time Pandemic Response and Decision Support. *The International Journal of High Performance Computing Applications* 37, 1 (2023), 4–27.

[11] Keith R. Bisset, Jiangzhuo Chen, Xizhou Feng, V.S. Anil Kumar, and Madhav V. Marathe. 2009. EpiFast: A Fast Algorithm for Large Scale Realistic Epidemic Simulations on Distributed Memory Systems. In *Proceedings of the 23rd International Conference on Supercomputing (ICS '09)*. Association for Computing Machinery, New York, NY, USA, 430–439. `https://doi.org/10.1145/1542275.1542336`

[12] Eric Bonabeau. 2002. Agent-Based Modeling: Methods and Techniques for Simulating Human Systems. *Proceedings of the National Academy of Sciences of the United States of America* 99, 10 (2002), 7280–7287. arXiv:3057854

[13] Stanislav S. Borysov, Jeppe Rich, and Francisco C. Pereira. 2019. How to Generate Micro-Agents? A Deep Generative Modeling Approach to Population Synthesis. *Transportation Research Part C: Emerging Technologies* 106 (Sept. 2019), 73–97. `https://doi.org/10.1016/j.trc.2019.07.006`

[14] US Census Bureau. [n.d.]. Why the Census Bureau Chose Differential Privacy. https://www.census.gov/library/publications/2023/decennial/c2020br-03.html.

[15] Kevin Chapuis, Patrick Taillandier, and Alexis Drogoul. 2022. Generation of Synthetic Populations in Social Simulations: A Review of Methods and Practices. *Journal of Artificial Societies and Social Simulation* 25, 2 (2022), 6.

[16] Badr-Eddine Cherief-Abdellatif and Pierre Alquier. 2020. MMD-Bayes: Robust Bayesian Estimation via Maximum Mean Discrepancy. In *Proceedings of The 2nd Symposium on Advances in Approximate Bayesian Inference*. PMLR, 1–21.

[17] Ayush Chopra. 2022. *Decision Making for Populations*. Ph.D. Dissertation. Massachusetts Institute of Technology.

[18] Ayush Chopra, Ramesh Raskar, Jayakumar Subramanian, Balaji Krishnamurthy, Esma S Gel, Santiago Romero-Brufau, Kalyan S Pasupathy, and Thomas C Kingsley. 2021. DeepABM: Scalable and Efficient Agent-Based Simulations via Geometric Learning Frameworks-a Case Study for COVID-19 Spread and Interventions. In *2021 Winter Simulation Conference (WSC)*. IEEE, 1–12.

[19] Ayush Chopra, Alexander Rodríguez, Jayakumar Subramanian, Arnau Quera-Bofarull, Balaji Krishnamurthy, B. Aditya Prakash, and Ramesh Raskar. 2023. Differentiable Agent-based Epidemiology. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems (AAMAS '23)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1848–1857.

[20] Abdoul-Ahad Choupani and Amir Reza Mamdoohi. 2016. Population Synthesis Using Iterative Proportional Fitting (IPF): A Review and Future Research. *Transportation Research Procedia* 17 (Jan. 2016), 223–233. `https://doi.org/10.1016/j.trpro.2016.11.078`

[21] Joseph Cox. 2019. T-Mobile 'Put My Life in Danger' Says Woman Stalked With Black Market Location Data.

[22] Peter Sheridan Dodds and Duncan J. Watts. 2004. Universal Behavior in a Generalized Model of Contagion. *Physical Review Letters* 92, 21 (May 2004), 218701. `https://doi.org/10.1103/PhysRevLett.92.218701`

[23] Joel Dyer, Patrick Cannon, J. Doyne Farmer, and Sebastian Schmon. 2022. Black-Box Bayesian Inference for Economic Agent-Based Models. `https://doi.org/10.48550/arXiv.2202.00625` arXiv:2202.00625 [cs, econ, stat]

[24] Wouter Edeling, Hamid Arabnejad, Robbie Sinclair, Diana Suleimenova, Krishnakumar Gopalakrishnan, Bartosz Bosak, Derek Groen, Imran Mahmood, Daan Crommelin, and Peter V. Coveney. 2021. The Impact of Uncertainty on Predictions of the CovidSim Epidemiological Code. *Nature Computational Science* 1, 2 (Feb. 2021), 128–135. `https://doi.org/10.1038/s43588-021-00028-9`

[25] Enrique Frias-Martinez, Graham Williamson, and Vanessa Frias-Martinez. 2011. An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. IEEE, 57–64.

[26] Oded Goldreich, Silvio Micali, and Avi Wigderson. 2019. How to Play Any Mental Game, or a Completeness Theorem for Protocols with Honest Majority. In *Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali*. 307–328.

[27] Chang Gong, Oleg Milberg, Bing Wang, Paolo Vicini, Rajesh Narwal, Lorin Roskos, and Aleksander S. Popel. 2017. A Computational Multiscale Agent-Based Model for Simulating Spatio-Temporal Tumour Immune Response to PD1 and PDL1 Inhibition. *Journal of The Royal Society Interface* 14, 134 (Sept. 2017), 20170320. `https://doi.org/10.1098/rsif.2017.0320`

[28] Gov.uk. [n.d.]. List of Ethnic Groups. https://www.ethnicity-facts-figures.service.gov.uk/style-guide/ethnic-groups.

[29] Jakob Grazzini, Matteo G. Richiardi, and Mike Tsionas. 2017. Bayesian Estimation of Agent-Based Models. *Journal of Economic Dynamics and Control* 77 (April 2017), 26–47. `https://doi.org/10.1016/j.jedc.2017.01.014`

[30] Daniel Günther, Marco Holz, Benjamin Judkewitz, Helen Möllering, Benny Pinkas, Thomas Schneider, and Ajith Suresh. 2022. Privacy-Preserving Epidemiological Modeling on Mobile Graphs. arXiv:2206.00539 [cs]

[31] Carmit Hazay and Yehuda Lindell. 2010. A Note on the Relation between the Definitions of Security for Semi-Honest and Malicious Adversaries. *Cryptology ePrint Archive* (2010).

[32] Robert Hinch, William J. M. Probert, Anel Nurtay, Michelle Kendall, Chris Wymant, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, Lele Zhao, Andrea Stewart, Luca Ferretti, Daniel Montero, James Warren, Nicole Mather, Matthew Abueg, Neo Wu, Olivier Legat, Katie Bentley, Thomas Mead, Kelvin Van-Vuuren, Dylan Feldner-Busztin, Tommaso Ristori, Anthony Finkelstein, David G. Bonsall, Lucie Abeler-Dörner, and Christophe Fraser. 2021. OpenABM-Covid19—An Agent-Based Model for Non-Pharmaceutical Interventions against COVID-19 Including Contact Tracing. *PLOS Computational Biology* 17, 7 (July 2021), e1009146. `https://doi.org/10.1371/journal.pcbi.1009146`

[33] John H. Holland and John H. Miller. 1991. Artificial Adaptive Agents in Economic Theory. *The American Economic Review* 81, 2 (1991), 365–370. arXiv:2006886

[34] Bennett Cyphers and Jason Kelley. 2021. Illinois Bought Invasive Phone Location Data From Banned Broker Safegraph. https://www.eff.org/deeplinks/2021/08/illinois-bought-invasive-phone-location-data-banned-broker-safegraph.

[35] Cliff C. Kerr, Robyn M. Stuart, Dina Mistry, Romesh G. Abeysuriya, Katherine Rosenfeld, Gregory R. Hart, Rafael C. Núñez, Jamie A. Cohen, Prashanth Selvaraj, Brittany Hagedorn, Lauren George, Michał Jastrzębski, Amanda S. Izzo, Greer Fowler, Anna Palmer, Dominic Delport, Nick Scott, Sherrie L. Kelly, Caroline S. Bennette, Bradley G. Wagner, Stewart T. Chang, Assaf P. Oron, Edward A. Wenger, Jasmina Panovska-Griffiths, Michael Famulare, and Daniel J. Klein. 2021. Covasim: An Agent-Based Model of COVID-19 Dynamics and Interventions. *PLOS Computational Biology* 17, 7 (July 2021), e1009149. `https://doi.org/10.1371/journal.pcbi.1009149`

[36] Kamlesh Khunti, Awadhesh Kumar Singh, Manish Pareek, and Wasim Hanif. 2020. Is Ethnicity Linked to Incidence or Outcomes of Covid-19? *BMJ* 369 (April 2020), m1548. `https://doi.org/10.1136/bmj.m1548`

[37] Jeremias Knoblauch, Jack Jewson, and Theodoros Damoulas. 2022. An Optimization-Centric View on Bayes' Rule: Reviewing and Generalizing Variational Inference. *The Journal of Machine Learning Research* 23, 1 (2022), 5789–5897.

[38] Yehuda Lindell. 2020. Secure Multiparty Computation (MPC). *Cryptology ePrint Archive* (2020).

[39] Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*.

[40] Yanir Marmor, Alex Abbey, Yuval Shahar, and Osnat Mokryn. 2023. Assessing Individual Risk and the Latent Transmission of COVID-19 in a Population with an Interaction-Driven Temporal Model. *Scientific Reports* 13, 1 (Aug. 2023), 12955. `https://doi.org/10.1038/s41598-023-39817-9`

[41] Christopher A. Martin, David R. Jenkins, Jatinder S. Minhas, Laura J. Gray, Julian Tang, Caroline Williams, Shirley Sze, Daniel Pan, William Jones, Raman Verma, Scott Knapp, Rupert Major, Melanie Davies, Nigel Brunskill, Martin Wiselka, Chris Brightling, Kamlesh Khunti, Pranab Haldar, and Manish Pareek. 2020. Socio-Demographic Heterogeneity in the Prevalence of COVID-19 during Lockdown Is Associated with Ethnicity and Household Size: Results from an Observational Cohort Study. *eClinicalMedicine* 25 (Aug. 2020). `https://doi.org/10.1016/j.eclinm.2020.100466`

[42] Derek Meyer and George G. Vega Yon. 2023. epiworldR: Fast Agent-Based Epi Models. *Journal of Open Source Software* 8, 90 (Oct. 2023), 5781. `https://doi.org/10.21105/joss.05781`

[43] Shakir Mohamed, Mihaela Rosca, Michael Figurnov, and Andriy Mnih. 2020. Monte Carlo Gradient Estimation in Machine Learning. *Journal of Machine Learning Research* 21, 132 (2020), 1–62.

[44] Alexander Mordvintsev, Ettore Randazzo, and Craig Fouts. 2022. Growing Isotropic Neural Cellular Automata. `https://doi.org/10.48550/arXiv.2205.01681` arXiv:2205.01681 [cs, q-bio]

[45] Kirill Müller and Kay W. Axhausen. 2010. Population Synthesis for Microsimulation: State of the Art. *Arbeitsberichte Verkehrs- und Raumplanung* 638 (Aug. 2010). `https://doi.org/10.3929/ethz-a-006127782`

[46] Marco Pangallo, Alberto Aleta, R. Maria del Rio Chanona, Anton Pichler, David Martín-Corral, Matteo Chinazzi, François Lafond, Marco Ajelli, Esteban Moro, Yamir Moreno, Alessandro Vespignani, and J. Doyne Farmer. 2022. The Unequal Effects of the Health-Economy Trade-off during the COVID-19 Pandemic. `https://doi.org/10.48550/arXiv.2212.03567` arXiv:2212.03567 [physics, q-fin]

[47] Donovan Platt. 2020. A Comparison of Economic Agent-Based Model Calibration Methods. *Journal of Economic Dynamics and Control* 113 (April 2020), 103859. `https://doi.org/10.1016/j.jedc.2020.103859`

[48] Arnau Quera-Bofarull, Ayush Chopra, Joseph Aylett-Bullock, Carolina Cuesta-Lazaro, Anisoara Calinescu, Ramesh Raskar, and Michael Wooldridge. 2023. Don't Simulate Twice: One-shot Sensitivity Analyses via Automatic Differentiation. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems (AAMAS '23)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1867–1876.

[49] Arnau Quera-Bofarull, Ayush Chopra, Anisoara Calinescu, Michael Wooldridge, and Joel Dyer. 2023. Bayesian Calibration of Differentiable Agent-Based Models. *arXiv preprint arXiv:2305.15340* (2023). arXiv:2305.15340

[50] Arnau Quera-Bofarull, Joel Dyer, Anisoara Calinescu, J. Doyne Farmer, and Michael Wooldridge. 2023. BlackBIRDS: Black-Box Inference foR Differentiable Simulators. *Journal of Open Source Software* 8, 89 (Sept. 2023), 5776. `https://doi.org/10.21105/joss.05776`

[51] Arnau Quera-Bofarull, Joel Dyer, Anisoara Calinescu, and Michael Wooldridge. 2023. Some Challenges of Calibrating Differentiable Agent-Based Models. `https://doi.org/10.48550/arXiv.2307.01085` arXiv:2307.01085 [cs, q-fin, stat]

[52] Santiago Romero-Brufau, Ayush Chopra, Alex J Ryu, Esma Gel, Ramesh Raskar, Walter Kremers, Karen S Anderson, Jayakumar Subramanian, Balaji Krishnamurthy, Abhishek Singh, et al. 2021. Public Health Impact of Delaying Second Dose of BNT162b2 or mRNA-1273 Covid-19 Vaccine: Simulation Agent Based Modeling Study. *BMJ (Clinical research ed.)* 373 (2021).

[53] Ludger Ruschendorf. 1995. Convergence of the Iterative Proportional Fitting Procedure. *The Annals of Statistics* 23, 4 (1995), 1160–1174. arXiv:2242759

[54] Vincent Stimper, David Liu, Andrew Campbell, Vincent Berenz, Lukas Ryll, Bernhard Schölkopf, and José Miguel Hernández-Lobato. 2023. Normflows: A PyTorch Package for Normalizing Flows. *Journal of Open Source Software* 8, 86 (2023), 5361. `https://doi.org/10.21105/joss.05361`

[55] Guus ten Broeke, George van Voorn, and Arend Ligtenberg. 2016. Which Sensitivity Analysis Method Should I Use for My Agent-Based Model? *Journal of Artificial Societies and Social Simulation* 19, 1 (2016), 5.

[56] Zhibo Wang, Xiaoyi Pang, Yahong Chen, Huajie Shao, Qian Wang, Libing Wu, Honglong Chen, and Hairong Qi. 2018. Privacy-Preserving Crowd-Sourced Statistical Data Publishing with an Untrusted Server. *IEEE Transactions on Mobile Computing* 18, 6 (2018), 1356–1367.

[57] Andrew C Yao. 1982. Protocols for Secure Computations. In *23rd Annual Symposium on Foundations of Computer Science (Sfcs 1982)*. IEEE, 160–164.

[58] Shuli Zhou, Suhong Zhou, Zhong Zheng, and Junwen Lu. 2021. Optimizing Spatial Allocation of COVID-19 Vaccine by Agent-Based Spatiotemporal Simulations. *GeoHealth* 5, 6 (2021), e2021GH000427. `https://doi.org/10.1029/2021GH000427`

6 Appendix

6.1 The GMW protocol

The GMW protocol uses additive secret sharing to communicate (or aggregate) private inputs across the participant entities. The key insight is to divide a secret input into multiple shares in such a way that the secret can be reconstructed only when a sufficient number of shares are combined together. The scheme supports diverse aggregation queries such as secure addition, or secure multiplication [8] of the secrets held by the participating agents. Here we focus on the addition case and we assume that all participating agents are required to compute the secret, usually denoted by $t = N$, but the same methodology can be extended to multiplication and composite queries (see, e.g., [38]).

Consider N agents holding private values s_i. We want to compute the sum $\sum_i s_i$ without any agent j acquiring knowledge about $s_{\{k \neq j\}}$. To setup the protocol, the agents agree an integer $n > \max\{s_1, \ldots, s_N\}$ defining the finite group $\mathbb{Z}_n$ on which all computations will be carried [1]. Each agent i then samples $N-1$ random numbers, $r_{ij} \sim \mathcal{U}\{0, n-1\}$, such that the input is divided into N shares, s_{ij} defined by

$$s_i = \sum_{j=1}^{N} s_{\mathrm{ij}} \pmod n = \sum_{j=1}^{N-1} r_{ij} + \left(s_i - \sum_{j=1}^{N-1} r_{ij} \right) \pmod n. \tag{2}$$

Each agent then sends each share of their secret to each corresponding agent; agent i sends s_{i1} share to agent 1, s_{i2} share to agent 2, etc. Locally, each agent performs the sum

$$\sigma_k = \sum_{i=1}^{N} s_{ik} \pmod n. \tag{3}$$

Finally, all values σ_k are shared so that the reconstructed sum, $S = \sum_k \sigma_k \pmod n$, can be computed which corresponds to the sum of the agent inputs s_i by construction. Typically, this reconstruction may be conducted by a central MPC server or a trusted agent. We summarize the protocol in Algorithm 1 and we provide an illustrating example below.

6.1.1 Additive secret sharing example

Consider $N = 3$ agents—Alice, Bob, and Carol— holding private values $s_A = 2$, $s_B = 3$, and $s_C = 5$. They wish to compute the sum of these values without disclosing their individual inputs. They agree on an integer $n = 11$, defining a finite group $\mathbb{Z}_n$. First, the agents generate 3 shares each, by sampling 2 random numbers from $\mathbb{Z}_n$. For instance, Alice generates random numbers 7 and 5, so that

$$s_A = s_{AA} + s_{AB} + s_{AC} = 7 + 5 + 1 \pmod{11} = 2, \tag{4}$$

and similarly for Bob and Carol with $s_B = 2 + 0 + 1 \pmod{11}$, and $s_C = 3 + 1 + 1 \pmod{11}$. Second, the agents communicate with each other to keep one of the shares and send the other two to the other two agents and perform the sum of the received shares. For example, Alice receives s_{BA} from Bob and s_{CA} from Carol and computes

$$\sigma_A = s_{AA} + s_{BA} + s_{CA} = 7 + 2 + 3 \pmod{11} = 1 \pmod{11}, \tag{5}$$

and similarly for Bob and Carol with $\sigma_B = 5 + 0 + 1 \pmod{11} = 6 \pmod{11}$ and $\sigma_C = 1 + 1 + 1 \pmod{11} = 3 \pmod{11}$. Finally, the secret can be reconstructed by doing $S = \sigma_A + \sigma_B + \sigma_C = 10 \pmod{11}$ as expected.

In the following section, we apply the GMW protocol to generalize the above insight to share information containing agent's private information to other agents or a central MPC server, providing protocols for the computation of agent updates (Equation (1)), and gradients in a secure way, enabling privacy-preserving simulation, calibration, and analysis of ABMs.

[1]The choice to perform finite group arithmetics is so that no information about the secret can be gained by holding $< N$ shares.

7 Case Study: Privacy-preserving Epidemiology

In this section, we aim to illustrate a practical example where this new ABM methodology could be deployed, by showing a simulation and calibration of a decentralized, privacy-preserving, agent-based SIR model.

The model follows a standard parameterization where agents' interactions are specified through a contact graph $\mathcal{G}$, which in this case is only locally defined by each agent having access to their neighbors. Each agent has 3 possible states, 0 (Susceptible), 1 (Infected), and 2 (Recovered). We initialize the simulation by infecting a fraction I_0 of agents, which are sampled uniformly from the population, while the remaining agents are considered to be susceptible. Following the notation introduced in Section 2, at each time-step, agent i updates its state following Equation (1) with

$$M_{ij}(t) = I_j(t) \tag{6}$$

where $I_j(t)$ is the infected status of the neighbor (0 or 1), so that

$$\begin{aligned} z_i(t+1) = \mathbb{1}_{\{z_i=0\}} \cdot \text{Bernoulli}\left(p_{\text{inf}}^{(i)}(t)\right) \; + \\ \mathbb{1}_{\{z_i=1\}} \cdot \left(1 + \text{Bernoulli}\left(p_{\text{rec}}^{(i)}\right)\right) \; + \\ \mathbb{1}_{\{z_i=2\}} \cdot 2 \end{aligned} \tag{7}$$

with

$$p_{\text{inf}}^{(i)}(t) = 1 - \exp\left(-\frac{\beta\, S_i \Delta t}{n_i} \sum_{j \in \mathcal{N}(i)} I_j(t)\right), \tag{8}$$

where $\mathcal{N}(i)$ is the set of neighbors of agent i, S_i is the susceptibility of agent i, $n_i = \#\mathcal{N}(i)$ is the total number of neighbors, Δt is the duration of the time-step, and β is a structural parameter of the ABM called the effective contact rate. Infected agents can recover at each time-step with recovery rate γ, so that

$$p_{\text{rec}}^{(i)} = 1 - \exp\left(-\gamma \Delta t\right). \tag{9}$$

For the case of a complete graph, the model reduces to the standard ODE-based SIR model with $R_0 = \beta/\gamma$ as the basic reproduction number. The model is run for n_t time-steps.

To ground the example on real data, we consider the contact graph of the city of XXX, extracted from the June ABM model [6] to determine the the neighborhood of each agent, $\mathcal{N}(i)$. This contact graph includes the interactions of agents in households, companies, and schools and it is based on English census data. The choice of parameter values for the experiment is given in Table 1.

Parameter	Value
β	0.5 day^{-1}
γ	0.1 day^{-1}
I_0	0.01
Δt	1 day
n_t	60
$\mathcal{G}$	XXX

Table 1: Parameter values for the considered agent-based SIR model.

7.1 Private policy assessment with ABMs

We first consider the application of the SECURESIMULATION protocol (Algorithm 3). Let us pose a situation where a policy maker wants to study the efficacy of mask-wearing at different compliance levels using agent-based simulation. We introduce a slight modification to Equation (8) to incorporate a reduction on the infection probability due to mask-wearing with certain compliance α,

$$p_{\text{inf}}^{(i)}(t) = 1 - \exp\left(-\frac{\beta\, S_i \Delta_t}{n_i} \sum_{j \in \mathcal{N}(i)} I_j(t)(1 - c_j)\right), \tag{10}$$

where $c_j \sim \text{Bernoulli}(\alpha)$, so that $\alpha_i = 1$ corresponds to full compliance where there is no transmission. Note that we are assuming, complete protection against infection when wearing a mask. We proceed to execute 3 simulations for 3 different values of α. At each simulation, α is sent to the agents, where they locally compute their own compliance to the measure. The SecureSimulation protocol is then used to run the simulation and retrieve the aggregate statistic of interest, $\mathbf{x}$, which in this case is the number of infections over time. The results are shown in Figure 2, where we observe that little transmission occurs when compliance is above 75%.

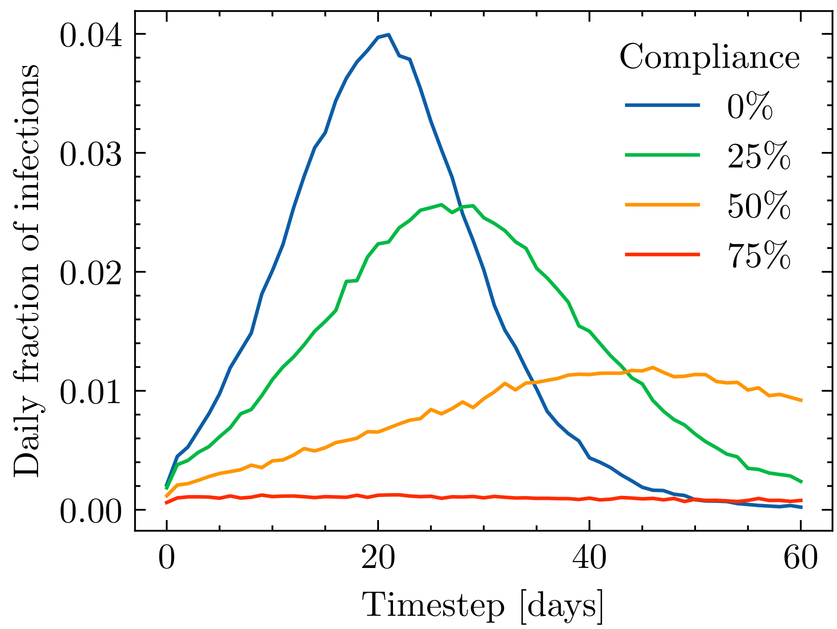

Figure 2: Infection curves for different levels of compliance: 0% (blue), 25% (green), 50% (orange), 75% (red). The number of infections has been normalized to the number of agents N.

Thus we observe that within our privacy-preserving methodology, the policy maker could still have access to the same level of insight than a traditional ABM, all while protecting the individual agent's privacy.

7.2 Private calibration of ABMs

Next, we pose a situation where we want to calibrate our ABM with structural parameters $\boldsymbol{\theta} = (\beta, \gamma)$ to observed ground-truth data. For simplicity, we present the calibration of the β parameter given an observed curve of infections ($\mathbf{y}$), obtained by running the ABM model with the baseline parameters in Table 1.

The first step is to compute the gradient $\nabla_{\boldsymbol{\theta}} \mathbf{x}$, where $\mathbf{x}$ is the number of daily infections and $\theta = \beta$. We note that this gradient can be approximated by the gradient of the average number of new infections with respect to β,

$$\frac{\partial x_t}{\partial \beta} \approx \frac{\partial\, \mathbb{E}[\Delta I(t)]}{\partial \beta} = \sum_{i=1}^{N} \chi_i(t) \exp(-\chi_i(t)/\beta), \tag{11}$$

where

$$\chi_i(t) = \exp\left(-\frac{\beta\, S_i \Delta_t}{n_i} \sum_{j \in \mathcal{N}(i)} I_j(t) \right). \tag{12}$$

The gradient can be safely retrieved by a central agent by performing the SecretSharing protocol across all agents as described in Algorithm 3. We thus conduct GVI by considering $\mathcal{Q}$ to be a masked-autoregressive normalizing flow, and assume the prior is a normal distribution with $\mu = 0.7$ and $\sigma = 0.5$.

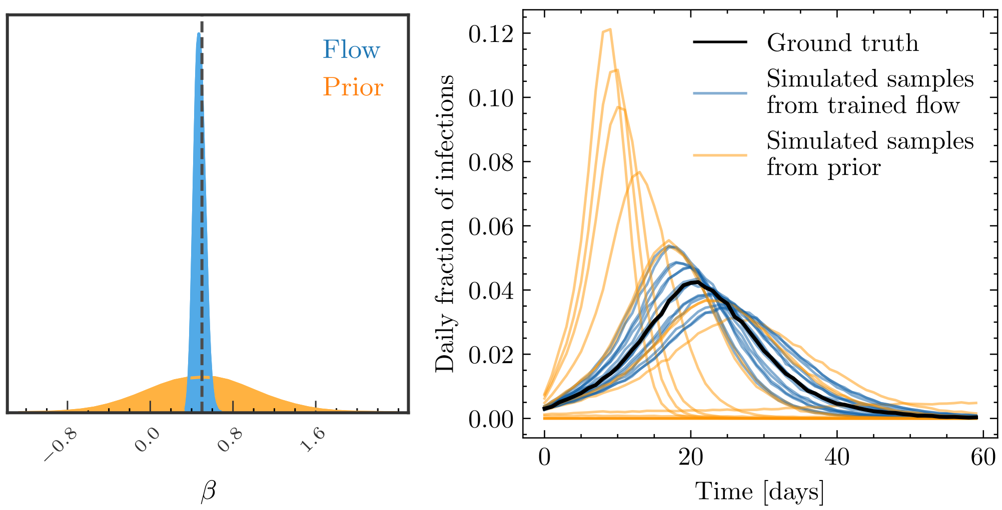

Figure 3: Left: Probability density plot for the trained normalizing flow (blue) against the prior distribution (orange). Ground-truth value is marked as a dashed black line. Right: Results from simulating β samples from the trained flow (blue) and prior (orange) compared to the ground-truth data (black). The number of infections has been normalized to the number of agents N.

Figure 3 (left) shows the trained normalizing flow which correctly assigns high probability mass to the ground-truth value. To further evaluate the goodness of the fit, we plot simulated runs from ABM parameters sampled from the trained flow in Figure 3 (right), where we compare it to runs simulated from prior samples.

This experiment highlights how privacy-preserving ABM can be integrated into probabilistic programming pipelines, like the considered case where we have used the Bayesian gradient-assisted inference algorithms in the BLACKBIRDS software package. This opens the door into integrating ABM insight into more complex ML pipelines leveraging heterogeneous data streams to boost the model's insight capabilities.