Emerging Directions in Leveraging Machine Intelligence for Explainable and Equity-Focused Simulation Models of Mental Health

Philippe J. Giabbanelli
Virginia Modeling, Analysis & Simulation Center (VMASC)
Old Dominion University
Norfolk, VA, USA
pgiabban@odu.edu

Abstract

Simulation models support policymakers, clinicians, and community members in identifying and evaluating interventions to improve population health. While these models are particularly valuable to measure the fairness of interventions, such measurements may require simulating massive populations in order to isolate effects for specific groups (e.g., by race and ethnicity, gender, age). This can create a computational bottleneck, forcing tradeoffs such as simplifying a model (thus potentially losing accuracy) or running fewer simulations (thus accepting wider confidence intervals) in exchange for sufficiently large populations. In addition, policymakers, clinicians, and community members can be involved at the design stage of a simulation model but its complex set of rules often tends to preclude participation at later stages. This discussion considers the use of Machine Intelligence to tackle both challenges, by automatically scaling up simulations and explaining them to stakeholders. This potential is illustrated through the public health challenge of mental health, focusing on agent-based models for suicide prevention.

Introduction

Modeling & Simulation (M&S) has a long history in mental health and can be broadly divided into two applications. Simulation-based education includes the use of simulated patients to train students and clinicians (Williams et al. 2017; Herrera-Aliaga and Estrada 2022). This discussion focuses on *simulations as decision-support tools for health interventions* (Long and Meadows 2018). In this context, a simulation model proceeds through several stages, from the abstraction of a system to its implementation in a computational format that allows to ask 'what-if questions'. When using models as 'policy sandboxes' (Silverman et al. 2021), what-if questions may pertain to policy *design* by "assessing the relative merits of alternative policy prescriptions in meeting the policy objectives" (Gilbert et al. 2018), or they may serve the needs of policy *evaluation* by comparing changes after policy implementation with respect to a status-quo expectation. What-if analyses have supported a variety of inquiries in the field of mental health, such as how to maintain quality of service in the delivery of mental health services given rising needs (Pierotti et al. 2024), or how to find effective mental health interventions at lower costs (Silverman et al. 2015). Since mental disorders cover a large variety of disturbances in behavior and cognition (e.g., eating disorders, schizophrenia, depression), we focus on suicide prevention as a guiding example of a major public health problem in which simulations are being used and where mental health plays a significant role.

Our recent review documented the increasing use of M&S in suicide research (Schuerkamp et al. 2023), with models serving to study interventions such as school-based mental health literacy programs (Page et al. 2017), changes in the number of psychiatric beds, suicide helpline services, or the duration of antidepressant treatments (Zhang et al. 2023). The heterogeneity of suicide has been documented, showing that suicide ideation can result from different experiences and pathways (Coppersmith et al. 2024). Given this heterogeneity, modelers have often employed Agent-Based Models (ABMs) to simulate individuals. These heterogeneous agents are equipped with their own characteristics, are embedded in different communities (Figure 1), and can express different behaviors (Michail and Witt 2023). ABMs provide powerful virtual laboratories to study the detailed effects of interventions. In particular, they can be used to ensure that a potential intervention does not unintentionally penalize certain population subgroups. For instance, users can compare the effects of interventions over time and across sub-populations based on features such as gender, race and ethnicity, or socio-economic category. Intersectionality should not be neglected in computational assessments of health disparities (Mhasawade, Zhao, and Chunara 2021), so modelers may need to *simulate populations that have sufficient sample sizes at the intersection of specific categories* (e.g., enough agents who are non-Hispanic white males using healthcare services). This is a challenge, as it becomes computationally intensive to handle simulated populations that are large enough to afford detailed analyses at the sub-group level (Huddleston et al. 2022).

As emphasized by Gilbert and colleagues, "communication is necessary to clearly explain results, and their limitations, ensure that the outputs are used appropriately, and build confidence in the modeling process and outputs" (Gilbert et al. 2018). In a similar way, Grimm *et al.* consider that decision-makers need information to evaluate

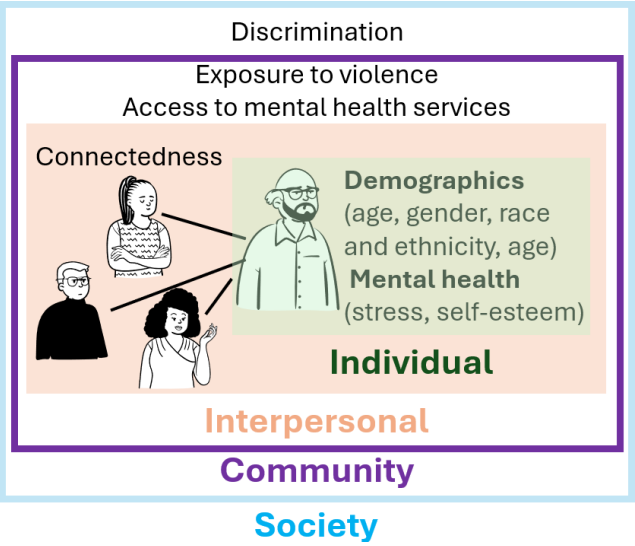

Figure 1: An Agent-Based Model (ABM) consists of individual entities (agents) who interact among themselves and with the environment. An ABM can represent population heterogeneity as agents have different demographic characteristics and other relevant traits. This example shows sample features of an ABM for mental health at several levels, using the social-ecological categorizations.

a model before using its outputs: purpose of the model, its organization, and evidence that it works (Grimm et al. 2020). We argue that communication is essential when using simulations for interdisciplinary and cross-sectoral problems such as mental health: subject-matter experts from different domains (e.g., epidemiology, psychology) may be involved in building and validating the model, and cross-sector collaboration is necessary to implement interventions in multilevel healthcare systems. In addition, 'communicating a model' is not a one-way process: feedback from subject-matter experts and key actors can provide clarifications and sustain engagement throughout the modeling process. A recent review in a different field reported that participants were not engaged in the modeling process beyond requirements elicitation and design (Manellanga and David 2024). While there is no current review dedicated to communication or engagement at different stages of simulation modeling for mental health, we posit that there is a similar *paucity of methods to communicate results along with the modeling process that produced them*.

In this paper, we identify potential means to address the challenges of equity and explainability in simulation models for mental health through new uses of machine intelligence. This discussion builds on emerging technology and pilot studies conducted since our prior vision paper dedicated to cross-pollination between machine learning and M&S (Giabbanelli 2022).

Scaling-Up Computationally Intensive Simulations for Health Equity

In the context of Agent-Based Modeling, the notion of 'fairness' or 'equity' has been operationalized in different ways. For example, it can mean that all agents would receive a benefit (although some may benefit noticeably *more* than others) and that it is economically viable for a service provider (Thorve et al. 2024). Williams and colleagues suggest that assessing equity is a shift from the dominant approach in the literature of averaging results across agents (Williams et al. 2022). Their review of 141 ABMs focused on equity and related notions found that most studies ($n = 60$) approach equity using a distributional approach, whereby outcomes are stratified by group identity. In our context, it means that a mental health intervention is considered more equitable when the groups receive similar benefits. Fewer studies ($n = 40$) performed complementary analyses to examine the conditions that gave rise to inequalities among groups. This observation echoes the findings of a review by Boyd *et al.*, who found that ABMs of inequalities in health tended to focus on differences in health behaviors and did not systematically delve into the causes of these inequalities (Boyd et al. 2022). A simulation may not be able to show that an intervention yields equitable outcomes: for instance, an intervention such as after-school programs can produce different impacts due to structural racism and disinvestment in some communities. In this case, analyzing the causes of inequalities provides a broader context to situate the simulated outcomes.

Large-scale simulations are needed to assess equity using a distributional approach. In other words, if the simulation needs sufficiently large subgroups for various combination of features (e.g., age, gender, race and ethnicity, socio-economic status) then the total number of simulated agents would be large. There are several ways in which machine learning could help with the associated computational challenges. In this paper, we suggest that the choice of approach would depend on whether end-users are potentially interested in *many subgroups during later inquiries* (post-simulation) or whether *few specific subgroups have been identified prior to performing simulations* (pre-simulation).

If analyses can concern many subgroups that are not selected a priori, then the whole population needs to be simulated and machine learning techniques such as *surrogate modeling* can be appropriate (Figure 2a). In this approach, a few (computationally expensive) simulations are performed and their results provide training data, which is fed to a machine learning algorithm (e.g., neural network, support vector machine) to predict the results of other simulations at a fraction of the computational cost (Angione, Silverman, and Yaneske 2022). This is a well-known approach in engineering and in health. For example, we showed that machine learning algorithms could *predict the cost* of a simulation, and if it is too computationally expensive given a specific combination of parameter values, then a surrogate model would be employed to *predict the result* (Fisher et al. 2020). To the best of our knowledge, surrogate models for simulations have not yet been used in models of suicide or

in the broader literature on mental health. A recent review on ABMs for policy research in substance abuse (an important risk factor for suicide) included surrogate models among future directions for the field (Zhong, Li, and Mangoni 2023), but this continues to be an untapped potential.

If analyses are devoted to a few pre-selected groups, then it is not necessary to simulate the *whole* population. However, it does not mean that we can *only* simulate the groups of interest. For example, consider that policymakers are concerned with the increase of firearm suicide rates for non-Hispanic Black and Hispanic people, thus they wish to evaluate interventions specifically in these two subgroups. Removing every other agents from the simulations would lead to an incorrect estimate for the two subgroups of interests, since their experiences (e.g., discrimination) are shaped in part through their *interactions* with other groups. In this case, we recommend using mixed model granularities/resolutions: the subgroups of interest must be highly detailed (i.e., large populations of agents with the desired characteristics) and a low fidelity model can be used for each of the other subgroups (i.e., they get fewer agents). While ABMs at different population granularities (e.g., few agents vs. many agents) have been analyzed (Guizani et al. 2019), less attention has been devoted to ABMs with mixed population granularities. Groups that are not the focal point but need to exist in order to provide relevant interactions can be simplified (Figure 2b) either by aggregating several agents into a representative one (Lippe et al. 2019; Wise et al. 2023) or by 'downscaling' through the use of a sample of agents (Hosszú et al. 2024). So far, neither technique has been employed in simulation models for mental health.

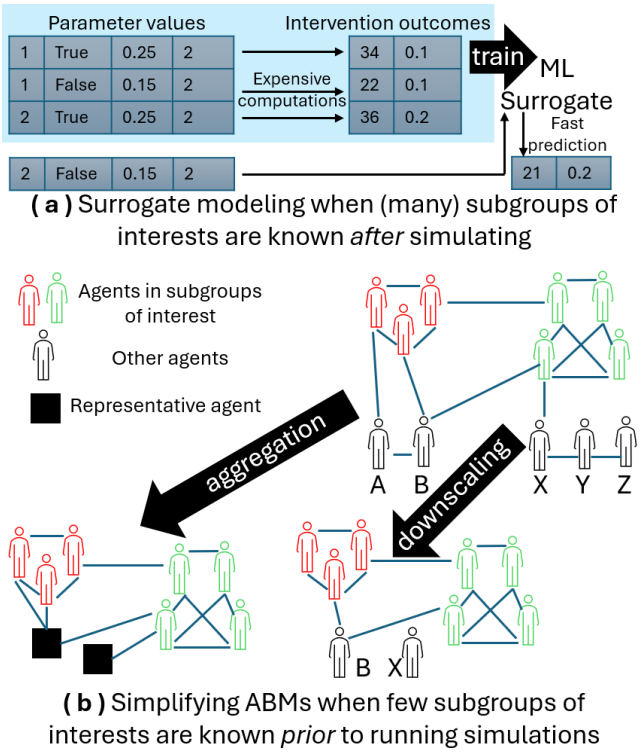

Figure 2: Health equity may require larger subgroups, which increases the total population size. To produce results within a time frame acceptable to stakeholders and given the resources available, we can either use surrogate modeling (a) or different types of simplifications (b).

Explaining Simulations to Clinicians, Policymakers, and Community Members

Several studies in mental health have long shown that healthcare providers and patients do not always comply to guidelines (Hepner et al. 2007) even if they initially intended to follow them or viewed them positively (Rebergen et al. 2006). Barriers to compliance can be related to insufficiently understanding the guidelines (knowledge), disagreeing with its content (attitude), or a variety of external causes (Lugtenberg et al. 2016). Given this backdrop, it stands to reason that simulations will face an even more challenging situation: we have to clearly convey how a sophisticated machinery (e.g., agent-based model and machine learning) produced a set of results based on the interdisciplinary expertise that went into the model design. Without clear explanations, misunderstandings and disagreements may prevent a model from making a difference in practice. Our usability study on a model of physical and mental well-being with experienced policymakers demonstrated the challenges of explaining the inner workings of a simulation model through visualizations (e.g., node and link diagrams), as these formats were unusual to the target audience (Giabbanelli and Vesuvala 2023). Thanks to the emergence of Large Language Models (LLMs) such as GPT, it is now possible to generate explanations in a format that transcends the different groups of users: textual reports. In addition, the quality of these reports can be automatically assessed for factuality or readability, thanks to fast-paced progress upon early generations of GPT where hallucinations or fluency were major concerns.

Recent studies have provided several prototypes that contribute to diverse facets of explaining simulation models. As mentioned in the introduction, a simulation model starts as an abstraction of a system, for instance as a graphical representation that lists relevant constructs and details the nature of their interrelationships. The task of *graph-to-text* can turn such representations into reports using LLMs. A potential concern for specialized domains such as mental health is that an exclusively reliance on the knowledge model of a LLM may lead to hallucinations or misunderstandings in the generated text. Phatak and colleagues have shown that providing a handful of examples would already alleviate these issues in the case of generating explanations of a suicide model (Phatak et al. 2024). While the translation of a design into text has now been covered by multiple studies, the explanation of results produced by Agent-Based Models has received less attention. Lynch *et al.* have shown that the events experienced by agents could be turned into short narratives in the form of tweets, which resembled real-world tweets in several aspects (Lynch et al. 2023).

While these studies demonstrate the feasibility of explaining the journey of agents or the intricate design of a model,

numerous questions remain. First, proof-of-concept studies have shown that we can produce text, but readers were notably absent. We thus need extensive user studies with several stakeholder groups (e.g., community members, clinicians, policymakers) to examine whether a generated narrative is clear for individuals, and whether providing it increases confidence in a model. The heterogeneous profile of stakeholder groups also suggests that the generated text should be customized based on their needs. Second, the ability of turning the experiences of an agent into text does not mean that we are ready to transform the complete output of a simulation into text. Indeed, a simulation may consists of millions of agents, and the simulation may be performed several times to account for randomness. Since it is impossible to express the voices of all simulated agents, there should be a simplification. Should we pick a handful of agents, and if so, on which basis should they be selected for narratives? For example, we could produce narratives to exemplify different pathways to suicide ideation and attempt, such as involving substance use and abuse or adverse childhood experiences. Or should we instead produce narratives from a multitude of agents and then apply summarization algorithms to create a narrative that is just as long as a user desires? The future of the field presents several potential tradeoffs that are yet to be explored in the context of mental health.

Acknowledgments

The author is indebted to many students and collaborators with whom discussions have contributed to the views expressed in this paper. In particular, this paper has benefited from joint projects with Dr. Ketra Rice, Dr. Ameeta Agrawal, and several group alumni (Tyler Gandee, Anish Shrestha, Ryan Schuerkamp).

References

Angione, C.; Silverman, E.; and Yaneske, E. 2022. Using machine learning as a surrogate model for agent-based simulations. *Plos one*, 17(2): e0263150.

Boyd, J.; Wilson, R.; Elsenbroich, C.; Heppenstall, A.; and Meier, P. 2022. Agent-based modelling of health inequalities following the complexity turn in public health: a systematic review. *International Journal of Environmental Research and Public Health*, 19(24): 16807.

Coppersmith, D. D.; Kleiman, E. M.; Millner, A. J.; Wang, S. B.; Arizmendi, C.; Bentley, K. H.; DeMarco, D.; Fortgang, R. G.; Zuromski, K. L.; Maimone, J. S.; et al. 2024. Heterogeneity in suicide risk: Evidence from personalized dynamic models. *Behaviour research and therapy*, 180: 104574.

Fisher, A.; Adhikari, B.; Zhai, C.; Morgan, J. E.; Mago, V. K.; and Giabbanelli, P. J. 2020. Predicting the resource needs and outcomes of computationally intensive biological simulations. In *2020 Spring Simulation Conference (SpringSim)*, 1–12. IEEE.

Giabbanelli, P. J. 2022. Hybrid Models That Combine Machine Learning and Simulations. *Computing in Science & Engineering*, 24(5): 72–76.

Giabbanelli, P. J.; and Vesuvala, C. X. 2023. Human factors in leveraging systems science to shape public policy for obesity: A usability study. *Information*, 14(3): 196.

Gilbert, N.; Ahrweiler, P.; Barbrook-Johnson, P.; Narasimhan, K. P.; and Wilkinson, H. 2018. Computational modelling of public policy: Reflections on practice. *Journal of Artificial Societies and Social Simulation*, 21(1).

Grimm, V.; Johnston, A. S.; Thulke, H.-H.; Forbes, V.; and Thorbek, P. 2020. Three questions to ask before using model outputs for decision support. *Nature Communications*, 11(1): 4959.

Guizani, N.; Elghariani, A.; Kobes, J.; and Ghafoor, A. 2019. Effects of social network structure on epidemic disease spread dynamics with application to ad hoc networks. *IEEE Network*, 33(3): 139–145.

Hepner, K. A.; Rowe, M.; Rost, K.; Hickey, S. C.; Sherbourne, C. D.; Ford, D. E.; Meredith, L. S.; and Rubenstein, L. V. 2007. The effect of adherence to practice guidelines on depression outcomes. *Annals of internal medicine*, 147(5): 320–329.

Herrera-Aliaga, E.; and Estrada, L. D. 2022. Trends and innovations of simulation for twenty first century medical education. *Frontiers in public health*, 10: 619769.

Hosszú, Z.; Borsos, A.; Mérő, B.; and Vágó, N. 2024. The More the Merrier?-the Optimal Choice of Scaling in Economic Agent-Based Models. *Available at SSRN 4751602*.

Huddleston, J.; Galgoczy, M. C.; Ghumrawi, K. A.; et al. 2022. Design and Deployment of a Simulation Platform: Case Study of an Agent-Based Model for Youth Suicide Prevention. In *2022 Winter Simulation Conference (WSC)*, 2582–2593. IEEE.

Lippe, M.; Bithell, M.; Gotts, N.; Natalini, D.; Barbrook-Johnson, P.; Giupponi, C.; Hallier, M.; Hofstede, G. J.; Le Page, C.; Matthews, R. B.; et al. 2019. Using agent-based modelling to simulate social-ecological systems across scales. *GeoInformatica*, 23(2): 269–298.

Long, K. M.; and Meadows, G. N. 2018. Simulation modelling in mental health: A systematic review. *Journal of Simulation*, 12(1): 76–85.

Lugtenberg, M.; Van Beurden, K. M.; Brouwers, E. P.; Terluin, B.; van Weeghel, J.; van der Klink, J. J.; and Joosen, M. C. 2016. Occupational physicians' perceived barriers and suggested solutions to improve adherence to a guideline on mental health problems: analysis of a peer group training. *BMC health services research*, 16: 1–11.

Lynch, C. J.; Jensen, E. J.; Zamponi, V.; O'Brien, K.; Frydenlund, E.; and Gore, R. 2023. A structured narrative prompt for prompting narratives from large language models: Sentiment assessment of chatgpt-generated narratives and real tweets. *Future Internet*, 15(12): 375.

Manellanga, R.; and David, I. 2024. Paving the way for collaborative modeling of sustainable systems: lessons from participatory modeling.

Mhasawade, V.; Zhao, Y.; and Chunara, R. 2021. Machine learning and algorithmic fairness in public and population health. *Nature Machine Intelligence*, 3(8): 659–666.

Michail, M.; and Witt, K. 2023. Unleashing the potential of systems modeling and simulation in supporting policy-making and resource allocation for suicide prevention. *Crisis*, 44(4).

Page, A.; Atkinson, J.-A.; Heffernan, M.; McDonnell, G.; and Hickie, I. B. 2017. A decision-support tool to inform Australian strategies for preventing suicide and suicidal behaviour. *Public Health Research and Practice*.

Phatak, A.; Mago, V. K.; Agrawal, A.; et al. 2024. Narrating Causal Graphs with Large Language Models. In *Proceedings of the 2024 Hawaii International Conference on Systems Science (HICSS)*.

Pierotti, L.; Cooper, J.; James, C.; Cassels, K.; Gara, E.; Denholm, R.; and Wood, R. 2024. Can computer simulation support strategic service planning? Modelling a large integrated mental health system on recovery from COVID-19. *International Journal of Mental Health Systems*, 18(1): 12.

Rebergen, D.; Hoenen, J.; Heinemans, A.; Bruinvels, D.; Bakker, A.; and van Mechelen, W. 2006. Adherence to mental health guidelines by Dutch occupational physicians. *Occupational Medicine*, 56(7): 461–468.

Schuerkamp, R.; Liang, L.; Rice, K. L.; and Giabbanelli, P. J. 2023. Simulation models for suicide prevention: a survey of the state-of-the-art. *Computers*, 12(7): 132.

Silverman, B. G.; Hanrahan, N.; Bharathy, G.; Gordon, K.; and Johnson, D. 2015. A systems approach to healthcare: agent-based modeling, community mental health, and population well-being. *Artificial intelligence in medicine*, 63(2): 61–71.

Silverman, E.; Gostoli, U.; Picascia, S.; Almagor, J.; McCann, M.; Shaw, R.; and Angione, C. 2021. Situating agent-based modelling in population health research. *Emerging Themes in Epidemiology*, 18: 1–15.

Thorve, S.; Mortveit, H.; Vullikanti, A.; Marathe, M.; and Swarup, S. 2024. Assessing Fairness of Residential Dynamic Pricing for Electricity using Active Learning with Agent-based Simulation. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, 1827–1836.

Williams, B.; Reddy, P.; Marshall, S.; Beovich, B.; and McKarney, L. 2017. Simulation and mental health outcomes: a scoping review. *Advances in Simulation*, 2: 1–8.

Williams, T. G.; Brown, D. G.; Guikema, S. D.; Logan, T. M.; Magliocca, N. R.; Müller, B.; and Steger, C. E. 2022. Integrating equity considerations into agent-based modeling: A conceptual framework and practical guidance. *Journal of Artificial Societies and Social Simulation*, 25(3).

Wise, S.; Milusheva, S.; Ayling, S.; and Smith, R. M. 2023. Scale matters: Variations in spatial and temporal patterns of epidemic outbreaks in agent-based models. *Journal of Computational Science*, 69: 101999.

Zhang, C.; Zafari, Z.; Slejko, J. F.; Castillo, W. C.; Reeves, G. M.; and Dosreis, S. 2023. Impact of undertreatment of depression on suicide risk among children and adolescents with major depressive disorder: A microsimulation study. *American journal of epidemiology*, 192(6): 929–938.

Zhong, X.; Li, X.; and Mangoni, S. 2023. A Review of Agent-Based Modeling Applications in Substance Abuse Policy Research. In *2023 Winter Simulation Conference (WSC)*, 150–161. IEEE.