Rising like a Phoenix: Emerging from the Pandemic and Reshaping Human Endeavors with Digital Technologies ICIS 2023

Advances in Methods, Theories, and Philosophy

Dec 11th, 12:00 AM

Employing Machine Learning to Advance Agent-based Modeling in Information Systems Research

Amirsiavosh Bashardoust
University of Lausanne, amirsiavosh.bashardoust@unil.ch

Negin Safaei
University of Geneva, dorsa.safaei@gmail.com

Kazem Haki
Geneva School of Business Administration, kazem.haki@hesge.ch

Yash Raj Shrestha
University of Lausanne, yashraj.shrestha@unil.ch

Employing Machine Learning to Advance Agent-based Modeling in Information Systems Research

Completed Research Paper

Amirsiavosh Bashardoust
University of Lausanne - Faculty of Business and Economics (HEC Lausanne)
Quartier de Chamberonne, CH-1015, Lausanne, Switzerland
Amirsiavosh.bashardoust@unil.ch

Dorsa Safaei
Geneva School of Social Science, University of Geneva
Route de la Drize 7
1227 Carouge, Switzerland
Negin.safaei@hesge.ch

Kazem Haki
Geneva School of Business Administration (HES-SO, HEG Genève)
Rue de la Tambourine 17
1227 Carouge, Switzerland
Kazem.haki@hesge.ch

Yash Raj Shrestha
University of Lausanne - Faculty of Business and Economics (HEC Lausanne)
Quartier de Chamberonne, CH-1015, Lausanne, Switzerland
Yashraj.shrestha@unil.ch

Abstract

In recent years, computationally intensive theory construction, leveraging big data and machine learning (ML), has gained significant interest in the information systems (IS) community. The integration of computational methods can generate novel methodological paradigms or enhance existing methods. Agent-based modeling (ABM) is one of the computational methods that has recently proliferated in IS research to generate computationally intensive theories. However, ABM is still in nascent state of adoption in IS research and entails some pathological challenges that limit its applicability and robustness. With the goal of advancing ABM in IS research, this article proposes a methodological framework that integrates ML within relevant steps of ABM. The framework is demonstrated in an exemplary IS study, showing its potential for addressing the pathological challenges of ABM. We finally discuss the implications of applying the proposed methodological framework in IS research.

Keywords: Agent-based modeling (ABM), machine learning (ML), simulation, computational methods

Introduction

Information systems (IS) as a scientific discipline seeks to investigate and explain various phenomena at the intersection of social and technology-oriented sciences (Gregor, 2006; Lyytinen & Newman, 2008). As a result, the IS discipline occupies a unique position for leveraging advanced computational techniques to contribute to novel ways of discovering patterns for theory building and testing (Lindberg, 2020; Pentland et al., 2020). Echoing that, in the past few years, *computationally intensive theory construction* that relies on extensive computation, big data, and machine learning (ML) has gained much attention in the IS community (Berente et al., 2019; Shin et al., 2020). Therefore, various editorials in flagship IS journals have illustrated the applicability and value of including computational methods into IS scholar's method repertoire (Berente et al., 2019; Miranda et al., 2022).

On the one hand, (i) computational methods can be used to generate an entirely novel methodological paradigm in IS research. For instance, ML models support generating new theories, developing new measures, comparing competing theories, improving existing theories, assessing the relevance of theories, and accessing the predictability of empirical phenomena in IS research (Shmueli & Koppius, 2011). Further, ML can be effective in extracting features or variables (such as sentiments) from unstructured data (e.g., text, images, videos, and audio) (Shin et al., 2020). On the other hand, (ii) computational methods can also be applied to enhance, innovate, and supplement already existing methods or a subset of steps within those methods. For instance, computational steps can supplement traditional content analysis frameworks (Krippendorff, 2018). Specifically, ML can be used over traditional qualitative methods when analyzing interview data (Choudhury et al., 2020). With this article, we aim to contribute to the latter stream (stream ii) by introducing ML-supported agent-based modeling (ABM)—a methodological framework that introduces ML within the relevant steps of ABM to address challenges that IS scholars encounter when using ABM.

Simulation-based research in general, and ABM in particular, is a computational method on its own, which has been widely adopted in other disciplines and recently started gaining momentum in IS research (Beese et al., 2019; Dong, 2022). As a method, ABM comes with various benefits. It allows for simulating not only the actions and interactions of autonomous agents individually, but it can also capture simultaneous decisions and interactions of multiple agents and emergent complexity of the studied system (Davis et al., 2007). Given its potential in investigating sociotechnical systems, ABM has been employed to study various IS phenomena (Beese et al., 2019; Dong, 2022). ABM has gained even more attention among IS scholars in recent years due to its unique capabilities in capturing nonlinear interactions among human and IT artifacts in an organizational context (e.g., Haki et al., 2020; Sturm et al., 2021) as well as deriving macro-level patterns from nonlinear micro-level interactions (e.g., Nan, 2011; Nan & Tanriverdi, 2017).

Despite its merits, ABM entails some pathological challenges that limit its applicability and robustness (An et al., 2021; Dahlke et al., 2020; Macal, 2016; van der Hoog, 2017; Wilensky & Rand, 2015). For instance, it is challenging to run complex models in an ABM setup due to computational requirements. Further, interpreting and making sense of the generated data from agent-based models is often difficult. In this article, we take stock of such challenges that limit the full-fledged adoption of ABM in IS community and propose four distinct themes of combining ML with ABM. ML is uniquely capable of learning from the data to discover patterns or make inferences on not-seen data (Dahlke et al., 2020). As these capabilities are a huge advantage in compensating ABM's challenges, we offer IS researchers a supplementary ABM framework in which ML is introduced in ABM's constituent steps to address the mounting pathological challenges of ABM. For each step, we discuss how ML could be useful in overcoming the challenges of the given step alongside the hurdles inherent in using ML. To do so, we draw on existing discourses in other disciplines discussing how to couple ML with ABM, and on prior studies that already applied ML in their ABM. We present the supplementary ABM framework along with its demonstration in an exemplary IS study.

Foundations

Agent-based Modeling in IS Research

Simulation is a method that employs computer algorithms to model real-world phenomena into stochastic processes (law, 2007), providing a distinct methodological approach to research. Specifically, simulation aims at representing a complex phenomenon and the environment in which it exists in a more simplified theory logic in the form of a model (Davis et al., 2007). Researchers then use the developed model to run a plethora of virtual experiments under the desired and controlled conditions to investigate the relationships between a set of model's input and output variables (Carley, 2001), which is considered as a powerful tool for theory development (Davis et al., 2007). Simulation as a research method has been widely employed in various disciplines, and in recent years has also proliferated into the IS community (Beese et al., 2019, 2015; Dong, 2022; Za et al., 2018).

Scholars can employ a diversity of techniques for simulation, namely analytical simulations, stochastic processes, system dynamics, genetic algorithms, artificial neural networks, and ABM (Beese et al., 2019; Davis et al., 2007). In this study, we specifically focus on ABM due to its popular adoption and outstanding benefits in IS research (Beese et al., 2019; Dong, 2022).

ABM provides a computational model of the real world, describing the behavior of a collection of autonomous decision making entities called agents (Bonabeau, 2002). ABM enables us to model attributes and behavioral rules of an individual agent to see the output of agents' interactions (Wilensky & Rand, 2015). ABM is employed when agents are autonomous enough to make decisions based on their sensing of the environment and the other agents (law, 2007), and when agents are capable of evolving, learning, and showing nonlinear behavior (Bonabeau, 2002). In this context, multiple agents interact with each other, with respect to their own attributes, behavioral rules, and adaptation capabilities to replicate a complex phenomenon.

ABM is widely adopted in research mainly owing to its capability of *capturing bottom-up emergence* of real-world phenomena (Bonabeau, 2002). A phenomenon is emergent when its overall behavior is the result of nonlinear interactions among its individual entities (i.e., agents) in response to a turbulent environment (Haki et al., 2020). The bottom-up approach, i.e., simulating the behavior of individual agents and their micro-level interactions with neighboring agents, enables ABM to trace how micro-level interactions lead to macro-level patterns (Nan & Tanriverdi, 2017). Next to ABM's capability of capturing bottom-up emergence, its *flexibility* makes this simulation technique more appealing for research over other simulation techniques (Bonabeau, 2002). In effect, ABM gives scholars the possibility to capture heterogenous agents with different but adaptive attributes and behavioral rule, and run iterative experiments by conveniently changing model parameters to observe emergent outcomes (Bonabeau, 2002).

ABM is employed to serve various research purposes including understanding a phenomenon, prediction, or optimization (Turgut & Bozdag, 2023). It has proven its strength in a wide variety of research fields (Miller, 2015; Sivakumar et al., 2022). Similarly, ABM is gaining traction in IS research to study various topics, such as technological choices with different organizational aspirations (Dong, 2022), the evolution of IS architecture (Haki et al., 2020), and organizational learning (Sturm et al., 2021).

The growing attention to ABM in IS research can be explained when considering specificities of IS phenomena. First, one of the major premises of IS research is to examine the use of technologies in organizational and social contexts from a sociotechnical perspective (Lyytinen & Newman, 2008). Through capturing both human and IT artifacts as agents and their nonlinear interactions in an organizational context, ABM allows us to analyze complex sociotechnical systems in which humans and machines interact with each other to achieve a common goal (Haki et al., 2020; Sturm et al., 2021). Second, after a long tradition of empirically offering macro-level theories, IS scholars started considering individuals as a level of analysis in multi-level research (Lyytinen & Newman, 2008). In this context, ABM's capability of capturing bottom-up emergence allows IS scholars to model nonlinear micro-level interactions that lead to macro-level patterns (Nan, 2011; Nan & Tanriverdi, 2017).

There are two prevailing approaches to ABM, namely theoretical- and empirical-based ABM (Dam et al., 2012). In empirical-based ABM, the model is built upon primary or secondary empirical data to inform agents' attributes, behavioral rules, and interactions (Kavak et al., 2018). The underlying empirical data can be gathered using primary data collection methods, such as survey and observation, or can be retrieved from existing secondary data sources, such as online platforms. Here, agents' attributes, behavioral rules, and interactions are derived using statistical functions or ML methods (Taghikhah et al., 2021). In theoretical-based ABM, the model is built on established theoretical concepts and assumptions, or on existing models, to inform agents' attributes, behavioral rules, and interactions (e.g., Haki et al., 2020; Nan, 2011; Schmid et al., 2021b). The underlying theoretical concepts and assumptions can be extracted from various disciplines (e.g., management science, biology, economics) and translated into computational mathematical parameters of the model. These mathematical if-else rules describe the agents' underlying relations and behavior (Sun et al., 2016; Taghikhah et al., 2021).

The choice between theoretical- or empirical-based ABM approaches is contingent on the research questions as well as the available data, resources, and theoretical frameworks (Sun et al., 2016). On the one hand, empirical-based ABM is often used when the underlying theoretical mechanisms of a system are not well-understood, or when the goal is to predict the system patterns based on observed data. On the other hand, theoretical-based ABM is preferred when there is a strong prior theoretical basis for a system. In comparison to empirical-based models that are data-intensive, theoretical-based models can be developed more quickly. In some cases, a combination of two approaches may be used in order to achieve more comprehensive and accurate results (Sun et al., 2016).

Machine Learning in IS Research

ML provides computers with the ability to learn without being explicitly programmed (Samuel, 1959). ML algorithms can be broadly divided into four classes, namely supervised learning, unsupervised learning, semi-supervised, and reinforcement learning (Dahlke et al., 2020). While in supervised learning, the goal is to make good predictions about future outcomes based on known inputs and known past correct outcomes, in unsupervised learning the dataset does not have explicit examples of correct outputs for given inputs. Instead, the goal is to find patterns or structure in the data without knowing what the correct outputs should be. For instance, regression-based predictions fall under supervised learning and clustering is an example of unsupervised learning. The semi-supervised learning bridges supervised and unsupervised learning by initially training the model on small set of labeled data and iteratively feed unlabeled data to the model and increase its performance. Distinct from (un/semi-)supervised learning, reinforcement learning aims to build entities that learn to make decisions in an environment by interacting with it and receiving feedback on those interactions (i.e., rewards or punishments).

In the last decade, IS discipline has shown great proactiveness in not only studying ML from a sociotechnical vantage point (Sturm et al., 2021), but also using ML as a research tool in testing or building novel IS theories (Shin et al., 2020). As ML models are able to learn from historic data and make predictions, IS scholars have begun to use them to predict outcomes about a phenomenon of interest (Shmueli & Koppius, 2011). Furthermore, ML is effective in extracting features or variables from unstructured data (H. Chen et al., 2012). For example, sentiment analysis models have been used to quantify affective aspects such as valence and arousal in text (Chau et al., 2020), while image classification models facilitate extraction of emotions, age, and other facial features from images (Shin et al., 2020). Once extracted, such features can subsequently be either used as explanatory variables in statistical models (e.g., regression or survival analysis), or for uncovering interesting insights in exploratory data analysis (Aggarwal et al., 2012). As such, ML represents a relatively novel set of tools, which is increasingly being used as a methodological tool or as a supplement to other methods in IS research.

Contribution of Machine Learning to Agent-based Modeling

In various research disciplines, there is a burgeoning discourse on how ABM and ML can be coupled such that the ensemble caters to a more robust tool for research. This discourse advocates three distinct ways through which ABM and ML can be coupled (Dahlke et al., 2020; Sivakumar et al., 2022): employing ML to overcome inherent challenges of ABM; employing ABM to create synthetic data for ML algorithms to overcome inherent challenges of ML; and employing ABM and ML separately in a pipeline to solve subproblems of a complex problem. Given our inquiry's objective, we solely focus on the first way of coupling ABM and ML, to examine how ML could enhance ABM by addressing challenges that IS scholars encounter in using ABM. Therefore, in the remainder of this section, we first discuss challenges of ABM. Subsequently, we explain the themes of using ML in ABM with respect to the discussed ABM's challenges.

Challenges in Agent-based Modeling

Given ABM's specificities as a methodological approach, the challenges of employing ABM in research have been discussed in different disciplines (e.g., An et al., 2021; Dahlke et al., 2020; Macal, 2016). Building on these existing investigations, the major challenges that IS researchers could face when employing ABM are as follows:

Defining agents' attributes and behavioral rules (Challenge I): Defining agents' attributes and behavioral rules presents challenges, especially when agents are expected to replicate human behavior (An et al., 2021; Dahlke et al., 2020). Capturing agents as static and homogeneous entities is an unrealistic assumption when modeling human behavior, and may thus result in flawed insights. As a mitigation strategy, researchers tend to model more sophisticated behaviors for agents to increase simulation accuracy. To achieve this, researchers build their models on insights from behavioral economics and cognitive sciences that consider social factors to make agents' attributes and behavioral rules more realistic (Macal, 2016). While one of the premises of ABM is to accommodate agents with dynamic and heterogeneous behavior, its implementation is challenging due to the lack of relevant data or solid theoretical assumptions (Dahlke et al., 2020).

Computational costs (Challenge II): We are witnessing the wide adoption of empirical-based ABM due to the availability of micro-level data, specifically human behavior data on online platforms (Ye & He, 2016). As such, agent-based models are getting more complex and computationally intensive. Running and evaluating these complex models requires expensive computational infrastructure and immense time (Dahlke et al., 2020), which might not be at the disposal of many researchers. For instance, a model could contain a large number of agents, as is the case in macroeconomics' models (van der Hoog, 2017). Although empirical-based ABM is generally more computationally intensive than theory-based ABM, the latter still could face a similar challenge due to the iterative nature of developing and deploying ABM (Haki et al., 2020). Therefore, building the model, and its validation and optimization, requires many numbers of simulation runs that can exponentially increase the overall computational costs of the model (van der Hoog, 2017).

Calibration and sensitivity analysis of the model (Challenge III): Calibration is an iterative process of finding the optimal values for the model's input parameters that make the model's output closest to the real-world observations (Wilensky & Rand, 2015). Sensitivity analysis, in turn, is the process of analyzing the influence of changes in the model's input parameters on its output (Wilensky & Rand, 2015). With the model's increasing number of input parameters, the traditional calibration and sensitivity analysis techniques (e.g., parameter sweeping) become considerably time-consuming, and to some extent infeasible, due to the resulting intense computational costs (Sivakumar et al., 2022). Therefore, researchers encounter the challenge of ensuring the model's reliability and rigor, and simultaneously performing these calibration and sensitivity analysis tasks in a computationally efficient manner. In addition to being dependent on the size of the model for calibration and sensitivity analysis, nonlinearity complicates the analysis because ABM usually contains nonlinear relationships between its constituent constructs (Wilensky & Rand, 2015).

Verification and validation of the model (Challenge IV): ABM is a type of computational simulation method. Therefore, similar to any other simulation technique, researchers are required to conduct a verification and validation procedure on their agent-based models to ensure the validity of their outputs (Davis et al., 2007). While verification answers the question of whether the model has been implemented correctly, validation addresses the question of whether the model accurately simulates real-world phenomena (Davis et al., 2007). However, verifying a complex ABM is a challenging task as the correctness of the codes should be verified numerically. As for validation, evaluating the model's output that corresponds to real-world observations of the given phenomenon is a difficult task since finding comparable empirical data is not always feasible (An et al., 2021; Wilensky & Rand, 2015).

Interpretation and sensemaking of simulation results (Challenge V): ABM usually generates a large amount of data depending on the number of the model's parameters and the conducted experiments (Dahlke et al., 2020). Therefore, analyzing the generated data, identifying patterns, and making sense of the patterns based on how every single component of the model works, present a considerable challenge in ABM (Dahlke et al., 2020). In particular, as IS scholars mainly aim at producing interpretable and generalizable theories, the sensemaking process of the generated data and turning them into plausible theories becomes a challenging endeavor.

Themes of Coupling Machine Learning with Agent-based Modeling

In this study, we aim to discuss how ML can be employed to advance ABM in IS research. This is an ongoing discussion in other disciplines, and scholars in our neighboring disciplines have already started employing ML in ABM. Drawing on two recent literature reviews on this topic (Dahlke et al., 2020; Sivakumar et al., 2022) as well as exemplary studies that applied ML in ABM in other disciplines (e.g., S. H. Chen et al., 2021; R. Vahdati et al., 2019; Song et al., 2021; van der Hoog, 2017), we derive four themes of coupling ML with ABM to address challenges of ABM.

Employing ML to derive agents' attributes and behavioral rules (ML1): One of the main contributions of ML to ABM is to extract agents' attributes and behavioral rules from empirical data (Zhang et al., 2021). This theme naturally addresses *Challenge I*, in which ML's different algorithm types can come in handy. Specifically, unsupervised learning can be applied to empirical data to cluster agents' attributes to eventually inform the model's parameters (Turgut & Bozdag, 2023). ML can also be employed to identify agents and their behavioral rules from the empirical data, which in this case supervised learning is the main applied ML algorithm type (e.g., Day et al., 2013). Further, reinforcement learning can be useful in the case of defining adaptive agents that learn and adapt during the simulation (Hung & Yang, 2021).

Developing ML surrogate models (ML2): To address *Challenge II* of ABM, an ML surrogate model (a replica) from the agent-based model can be developed. A surrogate model is a black-box representation of a complex simulation, which is trained based on the input and output of several runs of an original simulation model and can serve as its replica (Booker et al., 1999). Therefore, to avoid running a huge ABM simulation, it is possible to build a surrogate model from the simulation and work with the surrogate model to predict the original simulation's output (Sivakumar et al., 2022). Further, surrogate models can be employed to address the ABM's *Challenge III*. In this context, calibration and sensitivity analysis tasks can be performed on the surrogate model and eventually generalized to the original, much larger agent-based model (Lamperti et al., 2018).

Employing ML for exploring the relations between model parameters (ML3): ABM is an iterative approach. That is, in each iteration researchers should analyze the model's output, figure out the relations between all the model's parameters, and modify the model if needed. To achieve this, model calibration and sensitivity analysis tasks should be performed iteratively to gain detailed information on the model's parameters and their relations, to eventually come up with a modified version of the model for the next iteration (Davis et al., 2007). Although calibration and sensitivity analysis play an important role in model modification, they cause *Challenge III* in ABM. To address this challenge, a wide range of ML techniques can be employed, including regression and classifiers to find important variables in the model as well as sophisticated calibration techniques such as Bayesian methods to simplify the model and improve its performance (S. H. Chen et al., 2021; Dyer et al., 2022).

Employing ML to analyze ABM outputs (ML4): *Challenge V* of ABM arises from building complex agent-based models containing a large number of input and output parameters. As such, the "curse of dimensionality" affects the output analysis of ABM (Pereda et al., 2017). ML can be employed to address this challenge because dimension reduction is one of the strengths of ML. Since merely using conventional statistical or graphical techniques may not be useful in high-dimensional analysis, both supervised and unsupervised ML algorithms can help scholars analyze ABM output (Khater et al., 2014; Perry & O'Sullivan, 2018). In this context, ML can be used to find patterns in the output data and reduce the dimensions of the data through, for instance, clustering and classification techniques (Khater et al., 2014; R. Vahdati et al., 2019).

Applying Machine Learning in Different Stages of Agent-based Modeling

Relying on seminal simulation-based, specifically ABM, guidelines (Beese et al., 2019; Davis et al., 2007; Miller, 2015), this section presents all the necessary steps involved in ABM-based research and discusses the potential applications of ML in corresponding steps. An illustration of all the iterative steps along with corresponding challenges and ML themes can be found in Figure 1. Building on the discussed challenges of ABM and the themes of using ML in ABM, it is natural that ML does not apply to all the ABM steps.

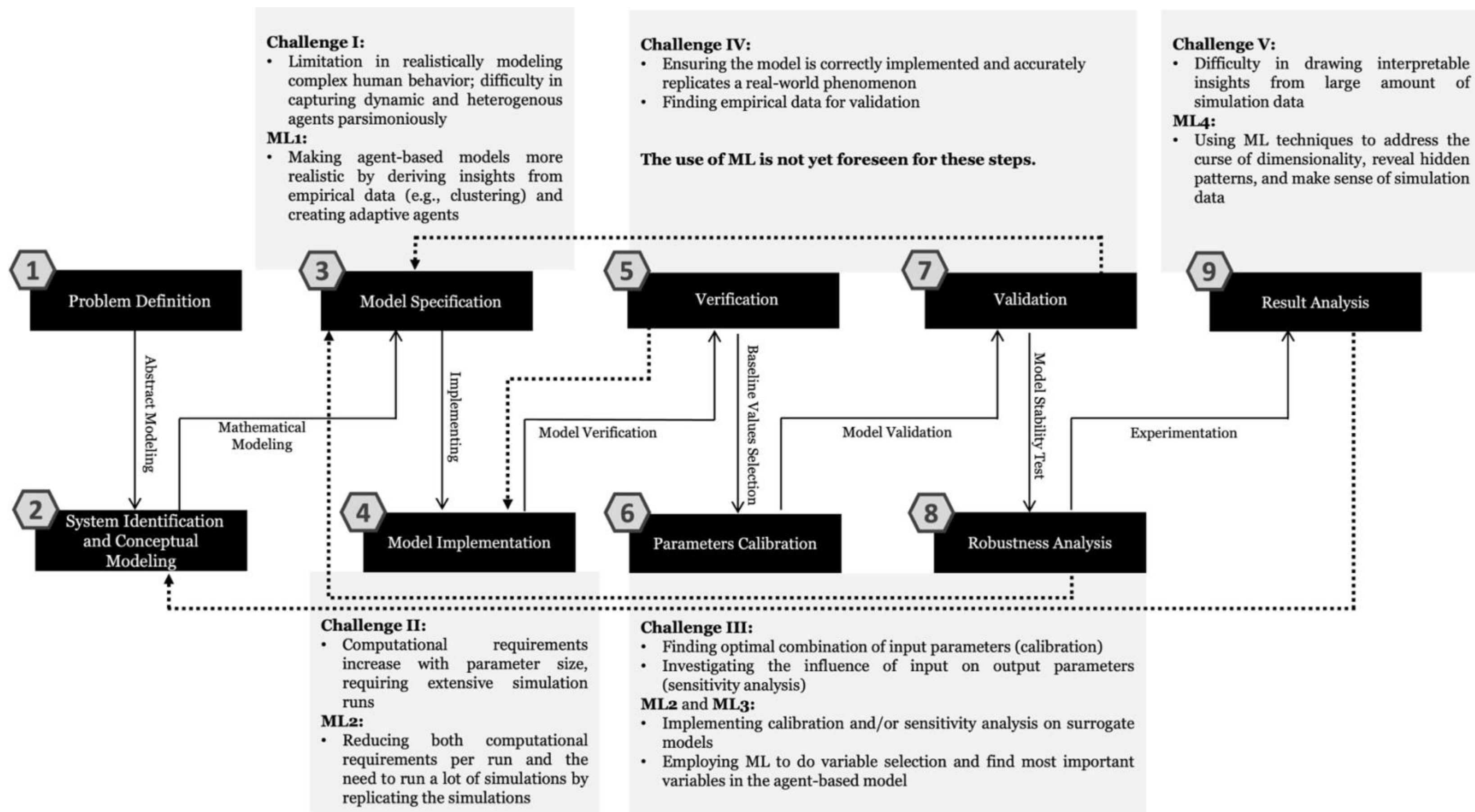

Figure 1. Steps in ABM, Main Challenges in ABM, and the Use of ML to Address Challenges in ABM

However, we intend to provide a comprehensive ABM framework to serve scholars who intend to use ABM in their studies to see how and in which steps ML can advance their ABM approach.

Problem definition and research question (step 1): In order to conduct effective simulation-based research, it is crucial to begin with an intriguing problem from the real world and define a thought-provoking research question (Davis et al., 2007). Since simulation requires the translation of real world to a simulation model, and the translation of simulation results back to the real world (Beese et al., 2019), scholars need to justify the use of simulation, specifically ABM, as a methodological approach. For instance, ABM makes it feasible to experimentally manipulate the independent variables of the study to conduct what-if analyses to an extent that would not be feasible in conventional empirical settings (Nan & Tanriverdi, 2017).

System identification and conceptual modeling (step 2): In this step, the high-level system (i.e., the phenomenon of interest) is identified and its boundaries are set (Beese et al., 2019; Davis et al., 2007; Miller, 2015). The conceptual model is a conceptual representation of a real-world system including information about the background context of the computational model, as well as assumptions, constraints, relationships, and interactions between the system's various components (Robinson, 2008). It is beneficial to first identify the overall goals of the system before focusing on specific modeling objectives. The simulation model may only be able to address a subset of the system's goals. This subset is then translated into the modeling objectives (Robinson, 2008).

Model specification (step 3): In this step, agents' attributes and behavioral rules, interactions among agents, and the environment in which interactions occur are specified (Heppenstall et al., 2011). This specification can take two forms: the different ways an agent can affect its environment and other agents, and the different ways that the environment and other agents affect an agent (Heppenstall et al., 2011). Having specified the agents, interactions, and environment, the conceptual model from the previous step is then transformed into a mathematical representation to be numerically simulated. It requires researchers to purposefully make simplifications to define constructs and relations with mathematical precision (Beese et al., 2019). As such, this simplified model only includes some aspects of the real-world, delineating the boundary conditions of the simulation model to capture the real-world system of interest (Heppenstall et al., 2011).

Modeling human behavior, adaptive agents, and extracting behavioral rules from empirical data are considered the main challenge (*Challenge I*) in ABM (Sivakumar et al., 2022). Specifically, empirical-based ABM could benefit from ML in this step as the agents' behavioral rules and the environment are extracted from empirical data (which is not the case in theory-based ABM). As discussed in the first theme of ML usage in ABM (*ML1*), different ML techniques can contribute to this step. For instance, Kavak et al. (2018) propose an approach to build agents from data and feed them to the simulation engine. To this, agents and their attributes and behavioral patterns can be extracted from raw empirical data. Completely automated extraction of agents is not the only way ML could contribute to this step. Turgut and Bozdag (2023) propose a framework in which different scenarios of incorporating ML in model specification are investigated. Further, reinforcement learning can be applied to create adaptive agents that learn from the environment and from their interactions with other agents (Dahlke et al., 2020). This approach has been applied by Hung and Yang (2021) in investigating demand and supply for water resource management, where the agents can learn and adjust based on their interactions in the system.

Since different ML techniques could be utilized in this step, specifying the goals and possible outcomes of using ML could help researchers narrow down the list of possible options. Researchers may use ML in this step when they want to explore an unknown behavioral rule from the data, create adaptive agents who will learn and adapt in the process of the simulation, or generate a large number of heterogeneous agents. However, researchers should be mindful that the feasibility of using specific ML techniques depends on their data. They hence should account for different features in their data before implementing ML techniques. As suggested by Turgut and Bozdag (2023), researchers should conduct literature research to know which parameters may affect the agents' behavior and include them in the data. Another point worth considering is that, in the case of using supervised learning for extracting agents' behavioral rules, the ML output is very contingent on the provided data, and thus may not be generalizable (Turgut & Bozdag, 2023).

Model implementation (step 4): After model specification, software implementation enables the model to be computationally simulated (Beese et al., 2019). The implementation can be done by general programming languages or by using platforms that support ABM (Macal & North, 2007). The simulation implementation process includes two main parts. In the first part (i.e., setup procedure), the initial state of the model is specified, and the model is executed once. In the second part (i.e., dynamics procedure), the model is repeatedly executed to get the results of agents' interactions with other agents and with the environment (Heppenstall et al., 2011).

However, the computational intensity of agent-based models (especially empirical-based ABM) could be high (*Challenge II*). To address this challenge, a large agent-based model could be replaced by an ML surrogate model (*ML2*). A verified surrogate model could predict the output of the simulation. Troost et al. (2022) compared different surrogate models and provided a solid basis for using surrogate models in ABM. As an example, Chen et al. (2021) used an ML surrogate model in their ABM to study Pantoea bacteria. In their study, bacterial growth was supposed to give an insight into how agents would compete for resources. However, running the simulation for any length of the problem or any timestep was not computationally feasible. To overcome this limitation of ABM, based on the results of over 3000 simulations, they trained a dense neural network on 70% of the data to predict model outputs from the given input parameters. They kept the remaining data as a test dataset for validation purposes. Their validated surrogate model was capable of predicting simulation outputs, even for input parameters that were computationally infeasible using ABM.

When developing a surrogate model, it is important to note that, similar to the ABM process, creating a surrogate is a highly iterative process. Researchers need to consider many factors, such as which ML model to choose and what parameters need to be tuned to increase the performance of the surrogate model. Further, researchers need to conduct the verification and validation tasks for the surrogate model as well, to ensure that the surrogate is a proper replica of the original simulation model. Therefore, it is important that researchers have an estimate of the efforts and the complexity they will add to their research process by utilizing a surrogate model.

Verification (step 5): Any computer simulation should go through a verification and validation procedure to ensure the reliability of the results. From an epistemic standpoint, when researchers implement a model on a computer, they need to provide evidence that the simulation model does not have any errors and is implemented accurately (Beese et al., 2019). Due to the complexity of agent-based models, formal algorithmic code verification may not always be practical. Therefore, researchers often follow recommended processes and software engineering principles to demonstrate the rigor of the implementation process and the simulation model's accuracy (Beese et al., 2019; D'Silva et al., 2008).

Parameters calibration (step 6): Once the model is verified, it is time to fine-tune the model using the calibration of parameters (Beese et al., 2019). Calibration techniques involve testing a wide range of parameter values for sets of initial conditions in simulations. The objective of this iterative step is to find and disregard parameter sets that do not produce outcomes that align with real-world empirical observations. The remaining sets can also be further refined using expert judgment. Another approach involves using multiple secondary sources to design decision-making functions and initial conditions that are likely to generate outcomes that align with real world (Windrum et al., 2007).

Calibration in ABM is a serious challenge (*Challenge III*) especially as the complexity grows with increasing the required number of petameters subject to calibration (Perumal & Zyl, 2022). ML could be beneficial in this step in various ways. The first is to build a surrogate model of the simulation for calibration purposes (*ML2*). For instance, Lamperti et al. (2018) showed that calibration based on a surrogate model has considerably lower computational costs (simultaneously addressing *Challenge II*) while maintaining accuracy. However, using a surrogate is not the only technique to enhance calibration using ML. For instance, Song et al. (2021) used reinforcement learning (*ML3*) to perform calibration in ABM. They selected a proportion of parameters per agent and used a reinforcement learning algorithm to learn agents' behavioral relations. Subsequently, they used the latter as a basis to calibrate their agent-based model with high accuracy.

Validation (step 7): After the calibration process, a validation procedure should be performed. Validation in ABM could be done at micro- or macro-levels (Macal & North, 2005). Macro-level validation ensures the accuracy of the results by analyzing the aggregated output of the model. Micro-level validation ensures that the model does not produce inaccurate and misleading results by testing the behavior and actions of individual agents. Validation is difficult, considering that the macro-level output of the system is the result of a large number of micro-level non-linear agents' interactions. In order to get a meaningful output of the system, it is rudimentary to make confident that agents (at the micro-level) behave as intended (Chaturvedi et al., 2011).

Robustness analysis (step 8): After validating the model, its robustness should be investigated to see whether the simulation is stable and reliable (Davis et al., 2007). Sensitivity analysis, which checks the sensitivity of the model's output to small changes in its input parameters, is one of the most important practices under robustness analysis (Sivakumar et al., 2022). Another practice under robustness analysis is model simplification to reduce the dimensionality of the model while maintaining accuracy (Sivakumar et al., 2022).

As any study has its unique situation, there is no one formula for sensitivity analysis, which advocates one of the main challenges of ABM (*Challenge III*). To address this challenge, Ligmann-Zielinska et al. (2020) investigated the different approaches for sensitivity analysis in ABM. They discuss different approaches, among which using surrogate models (*ML2*) is quite promising. For instance, Walzberg et al. (2022) employed a surrogate model to predict outputs of an agent-based model for input combinations that were computationally infeasible to simulate directly. They then conducted a sensitivity analysis on the agent-based model using the surrogate model. Since a surrogate model is a replica of the agent-based model, researchers conduct sensitivity analysis on the surrogate model and then generalize its results to the original agent-based model. As another way to use ML (*ML3*) in this step, Chen et al. (2021) employed a supervised ML algorithm by training a random forest to perform sensitivity analysis on their agent-based model. This approach allowed them to identify the input parameters that most strongly influenced their model's output.

Result analysis (step 9): After running the model and performing robustness analysis, the results are ensured to be robust and ready for analysis (Beese et al., 2019). Due to the large amount of data generated via simulation (*Challenge V*), researchers use different techniques to analyze the simulation data, and to build a high-level theory or test hypotheses. The fourth theme of coupling ML with ABM (*ML4*) is specifically dedicated to this step. Some studies used ML's clustering techniques (e.g., Khater et al., 2014) for the sensemaking of their results, while others used classification (e.g., R. Vahdati et al., 2019). Classification and clustering are very functional techniques for identifying hidden patterns in data that could not be revealed with normal statistics. For instance, Perry and O'Sullivan (2018) used clustering techniques to identify similar scenarios in the output data. Then in the second layer of analysis, they performed classification using a random forest algorithm for each group of scenarios.

Demonstration

In this section, we showcase how the proposed framework could be employed to use ML in one step of ABM in exemplary IS research (Schmid et al., 2021a; Schmid et al., 2021b). In this exemplary research, the developed agent-based model captures the ecosystem of an incumbent firm to investigate the roles of innovation platform mechanisms in a B2B context. The model contains a plethora of input and output parameters. Among the input parameters, the authors experimentally manipulated a few parameters (we called "experimental" in Figure 2). There are also a dozen other parameters that capture the constituents of the modeled ecosystem (we called "secondary" in Figure 2). For the latter parameters, the authors selected a baseline value (a *single* constant or randomized value) per parameter across all ranges of experimentally manipulated parameters. Among the output parameters, the authors were interested in measuring several dependent parameters (we called "dependent" in Figure 2), and the effects on these parameters can be mediated through a few other parameters (we called "intermediate" in Figure 2). This exemplary study aimed to investigate the effects of "experimental" parameters on "dependent" and "intermediate" parameters.

As the framework suggests, we should identify the steps in which ML could offer benefits with respect to the nature of the agent-based model and the encountered challenges. This exemplary agent-based model is theory-based, such that the conceptual model and the model's specification are developed based on theoretical grounds. Therefore, ML can be barely applied to *step 3*. One could use ML to build a surrogate model to reduce the computational intensity of the agent-based model in *step 4*. However, as this model is not very computationally intensive, a tradeoff between inserting a black box (i.e., the ML surrogate model) into the model, which needs its own verification and validation steps (see *step 4* of the framework), with the gained benefits needs to be considered. Given this situation, there is no compelling reason to use ML in *step 4* as well. Since the given agent-based model has 21 inputs and 17 outputs, using ML in *steps 6*, *8*, and *9* could be considerably beneficial. Due to space limits in this article, we focus our demonstration solely on step 8.

Given the number of parameters, one of the major challenges is to conduct sensitivity analysis to test whether the model's "experimental" parameters had an effect on the "dependent" and "intermediate" parameters that was sensitive to the authors' choice of baseline values for "secondary" parameters. This relates to the ABM's *Challenge III* in *step 8*.

To address the mentioned challenge and have the privilege of access to the generated data, we opted for the second theme of coupling ML with ABM (*ML2*) and developed a surrogate model. We trained the surrogate model on 35 simulation runs using a multilayer perceptron neural network. To do so, we divided our dataset into 80% training data and 20% test data. The model achieved an accuracy of 72% in predicting the output parameters. Subsequently, we performed a gradient-based sensitivity analysis on our surrogate model (Pizarroso et al., 2022). Our surrogate model and sensitivity analysis are implemented using sci-kit learn library in Python. We used MLPregressor for multilayer perceptron neural networks and neuralSens for sensitivity analysis.

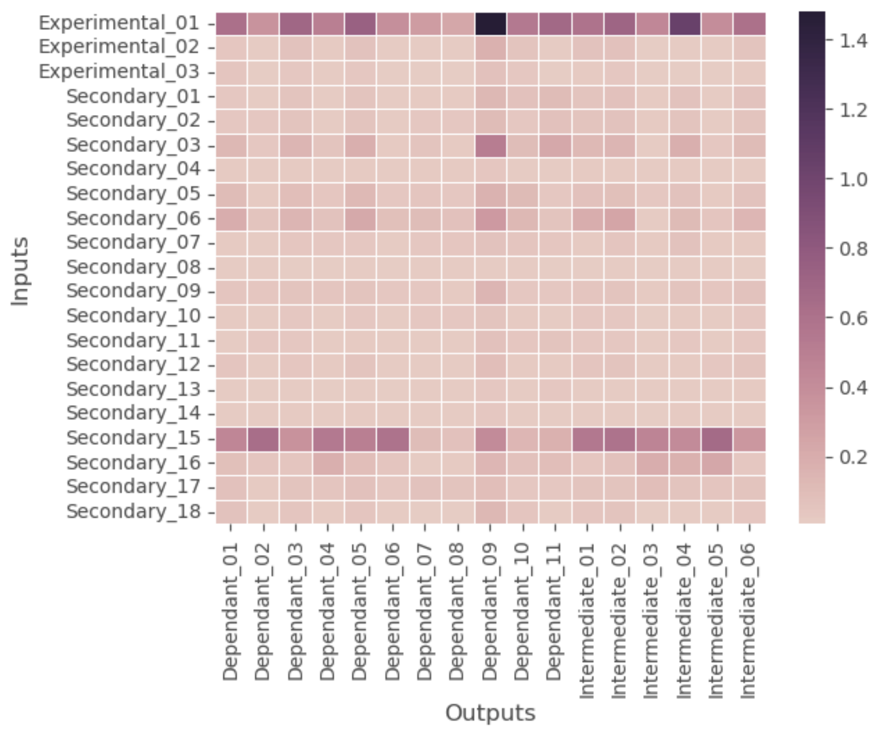

Figure 2. Sensitivity Heatmap Generated by a Surrogate Model for an Exemplary IS Study

Figure 2 illustrates the heatmap of the mean squared sensitivity of each output parameter with respect to each input parameter. As previously noted, sensitivity analysis in this exemplary study aims to test whether "dependent" and "intermediate" parameters are sensitive to the choice of baseline values for "secondary" parameters. As shown in Figure 2, the choice of values for "secondary" parameters 15, 06, and 03 has a considerable effect on "dependent" and "intermediate" parameters. The latter calls for prudent theoretically- or empirically-sound reasonings to justify why the authors opted for a specific value as a baseline for these "secondary" parameters. Therefore, our developed surrogate model can supplement the undertaken robustness analysis step (*step 8*) in this exemplary study.

Discussion and Conclusion

Computationally intensive theory construction has gained much attention in the IS community in recent years. Editorials in flagship IS journals have encouraged the adoption of computational methods for investigating next-generation IS phenomena (Berente et al., 2019; Miranda et al., 2022; Shrestha et al., 2021). In this article, we respond to this call by introducing ML within the ABM steps aimed at addressing pathological challenges that IS scholars encounter in using ABM.

Our article and the accompanying methodological novelty contribute to the prevailing discourse on computationally intensive theory construction in several ways. First, our proposed framework reinforces methodological pluralism and embraces the advice of combining disparate computational techniques from different methodological families aimed at gaining deeper and more robust insights (Miranda et al. 2022). As a result, our framework enhances researchers' reflexibility and curbs the danger associated with using single data-driven theory construction methods, such as researchers simply "rationalizing" what data analysis is able to show (Berente et al. 2019, p. 56). Second, by providing a stepwise guideline to follow, our intention is to make computationally intensive theory construction more accessible as well as standard in its adoption and use. Such standards should not only enhance the robustness of findings but also ease replicability of research findings, akin to research in natural sciences. Further, our step-wise guidelines should be useful for reviewers in carefully evaluating the rigor of papers applying computational methods. Finally, we contribute to a crucial step in computational theory construction – i.e., conversion of patterns into theory by introducing machine learning. A key task of researchers in computationally intensive theory development is creatively curating, assembling, and aligning patterns in the data, identifying regularities, and converting them into interpretable theories. Patterns that surfaced from ABM are not themselves theories, and given the complexities associated with a vast universe of potential parameters, theorizing is often difficult. The application of ML in augmenting pattern detection and theorizing contributes to the discourse on human-AI collaboration for scientific discovery (Choudhary et al., 2023; Shrestha et al., 2021). Moreover, we contribute to current guidelines of simulation-based studies in IS literature (Beese et al., 2019; Dong, 2022) by integrating ML into relevant steps of ABM-based studies.

Although ABM has gained rapid proliferation in IS research in recent years (e.g., Haki et al., 2020; Nan, 2011; Sturm et al., 2021), its methodological adoption remains in a nascent state as compared to other disciplines, such as economics and biology. Drawing on contemporary discourse in other disciplines, we take stock of five core challenges that ABM faces, which include defining agents' attributes and behavioral rules, computational costs, calibration and sensitivity analysis of the model as well as verification and validation of the model, and interpretation and sensemaking of simulation results. As a core contribution of our inquiry, we propose a framework for employing ML in relevant steps of ABM with the goal of mitigating identified challenges of ABM, specified into four distinct themes. This framework can be useful for IS researchers in not only highlighting potential challenges of ABM, but also mitigating these challenges by employing recent advances in ML. We illustrate the coupling of ML with ABM stepwise to provide more practical illustrations followed by an empirical demonstration.

The proposed framework offers a comprehensive roadmap to be applied in various research contexts. Nevertheless, the context-specific applicability of the framework is contingent on various factors, such as the type of ABM (i.e., empirical- or theoretical-based), the complexity of the agent-based model (e.g., the number of parameters), or the main challenge scholars encounter in the context of their study (e.g., sensitivity analysis, interpretation of results). Specifically, it is not indispensable to apply ML in all the steps that we specified. That is, depending on the nature of their model and contingencies, IS scholars can be selective in which step of ABM they employ ML.

The combination of ML and ABM generates flexibility in producing novel avenues for both empirical- and theoretical-based ABM. ML can inform empirical-based ABM using vast amounts of social media and other digital trace data, that can effectively improve the calibration of models and bring them closer to "real world". Similarly, sensemaking of large amount of simulated data using pattern recognition can produce "novel" and interesting insights that could have often escaped human researchers' eyes. With the proposed themes and demonstration, we intend to provide authors and editorial boards with an accessible starting point in identifying synergies between ML and ABM. While the opportunities are limitless, one should also be cautious in maintaining methodological rigor and epistemological assumptions when adopting novel methodological techniques. Authors and editorial boards are especially required to be cautious of the pitfalls in combining methods and carefully consider when it jeopardizes methodological assumptions of the phenomenon under investigation. Further, similar to other mixed-methods inquiries, authors should aim at displaying transparency of steps followed for both techniques (ABM and ML). Given the complex nature of both ABM and ML, this is likely to bring more challenges to authors and reviewers for ensuring and demonstrating the reliability of both models.

It is also important to note that ML is no "magic bullet" and is accompanied by its own challenges. For instance, ML requires model building and evaluation procedures comprising hyperparameter tuning, cross-validation, and model selection steps, which need to be carefully executed. Further, there exists a plethora of available algorithms, approaches, and techniques, and choosing the right set requires specific expertise. ML methods themselves suffer from pathological issues, such as the explainability of their underlying models and their findings (Rai, 2020). As a result, the application of ML could risk introducing errors and unintended side effects into the ABM findings. Additionally, reporting two equally rich methodological exercises while maintaining transparency and rigor will occupy a substantial real state of the manuscript, potentially cannibalizing space from other sections such as literature review and theory. Therefore, as an IS researcher aiming to combine ML and ABM, it is important to comprehend the pros and cons of both methods and identify a synergetic match based on the given study's objectives and contingencies. The use of both ABM and ML could also produce difficulties in review processes and delays may be faced in identifying experts in either or both methods.

Moreover, as the technology behind tools and the available data continue to evolve, so should the norms and conventions around coupling ML with ABM. This is particularly the case with advances in ML, which has recently shown exponential progress. IS researchers are required to stay in touch with such advances with a curious eye for borrowing them into their methodological toolkit. This suggests that the framework we offer is meant to offer a novel perspective and a scaffold, and should not be viewed as a final state. In other words, our framework should not be adopted dogmatically, but it should allow for the emergence of the creative imperative for this genre.

As for the status quo of using ABM in IS research, while theoretical-based ABM is common (Beese et al., 2019; Dong, 2022), empirical-based ABM is not much adopted. However, the latter has great potential for computationally intensive theory construction, particularly due to the availability of a large amount of data on, for instance, online platforms. Following trends in other disciplines, empirical-based ABM may soon gain IS scholars' attention, where coupling ML with ABM may reveal its full potential in, for example, deriving agents' attributes and behavioral rules from empirical data. Therefore, our suggested integration of ML into ABM steps may also serve IS scholars who are interested in employing empirical-based ABM in their prospective research inquiries.

Notes:

- The first and second authors have the same contribution. Therefore, they are sorted alphabetically.
- Dorsa Safaei is also affiliated to the Geneva School of Business Administration (HES-SO, HEG Genève), Rue de la Tambourine 17, 1227 Carouge, Switzerland.

References

Aggarwal, R., Gopal, R., Gupta, A., & Singh, H. (2012). Putting Money Where the Mouths Are: The Relation Between Venture Financing and Electronic Word-of-Mouth. *Information Systems Research*, *23*(3-part-2), 976–992.

An, L., Grimm, V., Sullivan, A., Turner II, B. L., Malleson, N., Heppenstall, A., Vincenot, C., Robinson, D., Ye, X., Liu, J., Lindkvist, E., & Tang, W. (2021). Challenges, tasks, and opportunities in modeling agent-based complex systems. *Ecological Modelling*, *457*, 109685.

Beese, J., Haki, K., & Aier, S. (2015, December). On the Conceptualization of Information Systems as Socio-Technical Phenomena in Simulation-Based Research. *Proceedings of the International Conference on Information Systems (ICIS) 2015*.

Beese, J., Haki, M. K., Aier, S., & Winter, R. (2019). Simulation-Based Research in Information Systems. *Business & Information Systems Engineering*, *61*(4), 503–521.
Berente, N., Seidel, S., & Safadi, H. (2019). Research Commentary—Data-Driven Computationally Intensive Theory Development. *Information Systems Research*, *30*(1), 50–64.
Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, *99*(suppl_3), 7280–7287.
Booker, A. J., Dennis, J. E., Frank, P. D., Serafini, D. B., Torczon, V., & Trosset, M. W. (1999). A rigorous framework for optimization of expensive functions by surrogates. *Structural Optimization*, *17*(1), 1–13.
Carley, K. M. (2001). Computational Approaches to Sociological Theorizing. In J. H. Turner (Ed.), *Handbook of Sociological Theory* (pp. 69–83). Springer US.
Chaturvedi, A. R., Dolk, D. R., & Drnevich, P. L. (2011). Design Principles for Virtual Worlds. *MIS Quarterly*, *35*(3), 673–684.
Chau, M., Li, T. M., Wong, P. W., Xu, J. J., Yip, P. S., & Chen, H. (2020). Finding People with Emotional Distress in Online Social Media: A Design Combining Machine Learning and Rule-Based Classification. *MIS Quarterly*, *44*(2), 933–956.
Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, *36*(4), 1165–1188.
Chen, S. H., Londoño-Larrea, P., McGough, A. S., Bible, A. N., Gunaratne, C., Araujo-Granda, P. A., Morrell-Falvey, J. L., Bhowmik, D., & Fuentes-Cabrera, M. (2021). Application of Machine Learning Techniques to an Agent-Based Model of Pantoea. *Frontiers in Microbiology*, *12*, 726409.
Choudhary, T., Gujar, S., Goswami, A., Mishra, V., & Badal, T. (2023). Deep learning-based important weights-only transfer learning approach for COVID-19 CT-scan classification. *Applied Intelligence*, *53*(6), 7201–7215. https://doi.org/10.1007/s10489-022-03893-7
Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal*, *41*(8), 1381–1411.
Dahlke, J., Bogner, K., Mueller, M., Berger, T., Pyka, A., & Ebersberger, B. (2020a). *Is the Juice Worth the Squeeze? Machine Learning (ML) In and For Agent-Based Modelling (ABM)*. arXiv.Dam, K. H. van, Nikolic, I., & Lukszo, Z. (2012). *Agent-Based Modelling of Socio-Technical Systems*. Springer Science & Business Media.
Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2007). Developing Theory Through Simulation Methods. *Academy of Management Review*, *32*(2), 480–499.
Day, T. E., Ravi, N., Xian, H., & Brugh, A. (2013). An Agent-Based Modeling Template for a Cohort of Veterans with Diabetic Retinopathy. *PLOS ONE*, *8*(6), e66812.
Dong, J. Q. (2022). Using Simulation in Information Systems Research. *Journal of the Association for Information Systems*, *23*(2), 408–417.
D'Silva, V., Kroening, D., & Weissenbacher, G. (2008). A Survey of Automated Techniques for Formal Software Verification. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *27*(7), 1165–1178.
Dyer, J., Cannon, P., Farmer, J. D., & Schmon, S. M. (2022). *Calibrating Agent-based Models to Microdata with Graph Neural Networks* (arXiv:2206.07570). arXiv.
Gregor, S. (2006). The Nature of Theory in Information Systems. *MIS Quarterly*, *30*(3), 611–642.
Haki, K., Beese, J., Aier, S., & Winter, R. (2020). The Evolution of Information Systems Architecture: An Agent-Based Simulation Model. *MIS Quarterly*, *44*(1), 155–184.
Heppenstall, A. J., Crooks, A. T., See, L. M., & Batty, M. (2011). *Agent-Based Models of Geographical Systems*. Springer Science & Business Media.
Hung, F., & Yang, Y. C. E. (2021). Assessing Adaptive Irrigation Impacts on Water Scarcity in Nonstationary Environments—A Multi-Agent Reinforcement Learning Approach. *Water Resources Research*, *57*(9), e2020WR029262.
Kavak, H., Padilla, J. J., Lynch, C. J., & Diallo, S. Y. (2018). Big data, agents, and machine learning: Towards a data-driven agent-based modeling approach. *Proceedings of the Annual Simulation Symposium*, 1–12.
Khater, M., Murariu, D., & Gras, R. (2014). Contemporary evolution and genetic change of prey as a response to predator removal. *Ecological Informatics*, *22*, 13–22.
Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology*. SAGE Publications.
Lamperti, F., Roventini, A., & Sani, A. (2018). Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control*, *90*, 366–389.

law, A. (2007). Simulation modeling & analysis. *Physical Sciences Books*.

Ligmann-Zielinska, A., Siebers, P.-O., Magliocca, N., Parker, D. C., Grimm, V., Du, J., Cenek, M., Radchuk, V., Arbab, N. N., Li, S., Berger, U., Paudel, R., Robinson, D. T., Jankowski, P., An, L., & Ye, X. (2020). 'One Size Does Not Fit All': A Roadmap of Purpose-Driven Mixed-Method Pathways for Sensitivity Analysis of Agent-Based Models. *Journal of Artificial Societies and Social Simulation*, *23*(1), 6.

Lindberg, A. (2020). Developing Theory Through Integrating Human and Machine Pattern Recognition. *Journal of the Association for Information Systems*, *21*(1), 90–116.

Lyytinen, K., & Newman, M. (2008). Explaining information systems change: A punctuated socio-technical change model. *European Journal of Information Systems*, *17*(6), 589–613.

Macal, C. M. (2016). Everything you need to know about agent-based modelling and simulation. *Journal of Simulation*, *10*(2), 144–156.

Macal, C. M., & North, M. J. (2005). Tutorial on agent-based modeling and simulation. *Proceedings of the Winter Simulation Conference, 2005.*, 14 pp.

Macal, C. M., & North, M. J. (2007). Agent-based modeling and simulation: Desktop ABMS. *2007 Winter Simulation Conference*, 95–106.

Miller, K. D. (2015). Agent-Based Modeling and Organization Studies: A critical realist perspective. *Organization Studies*, *36*(2), 175–196.

Miranda, S., Berente, N., Seidel, S., Safadi, H., & Burton-Jones, A. (2022). Editor's Comments: Computationally Intensive Theory Construction: A Primer for Authors and Reviewers. *MIS Quarterly*, *46*(2), iii–xviii.

Nan. (2011). Capturing Bottom-Up Information Technology Use Processes: A Complex Adaptive Systems Model. *MIS Quarterly*, *35*(2), 505–532.

Nan, N., & Tanriverdi, H. (2017). Unifying the role of IT in hyperturbulence and competitive advantage via a multilevel perspective of IS strategy. *MIS Quarterly*, *41*(3), 937-A8.

Pentland, B. T., Peng Liu, Kremser, W., & Hærem, T. (2020). The Dynamics of Drift in Digitized Processes. *MIS Quarterly*, *44*(1), 19–47.

Pereda, M., Santos, J. I., & Galán, J. M. (2017). A Brief Introduction to the Use of Machine Learning Techniques in the Analysis of Agent-Based Models. In C. Hernández (Ed.), *Advances in Management Engineering* (pp. 179–186). Springer International Publishing.

Perry, G. L. W., & O'Sullivan, D. (2018). Identifying Narrative Descriptions in Agent-Based Models Representing Past Human-Environment Interactions. *Journal of Archaeological Method and Theory*, *25*(3), 795–817.

Perumal, R., & Zyl, T. L. van. (2022). Surrogate-assisted strategies: The parameterisation of an infectious disease agent-based model. *Neural Computing and Applications*.

Pizarroso, J., Portela, J., & Muñoz, A. (2022). NeuralSens: Sensitivity Analysis of Neural Networks. *Journal of Statistical Software*, *102*(7), 1–36.

R. Vahdati, A., Weissmann, J. D., Timmermann, A., Ponce de León, M. S., & Zollikofer, C. P. E. (2019). Drivers of Late Pleistocene human survival and dispersal: An agent-based modeling and machine learning approach. *Quaternary Science Reviews*, *221*, 105867.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141.

Robinson, S. (2008). Conceptual modelling for simulation Part I: Definition and requirements. *Journal of the Operational Research Society*, *59*(3), 278–290.

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, *3*(3), 210–229.

Schmid, M., Haki, K., Tanriverdi, H., & Aier, S. (2021b). *Taming Complexity in Business Ecosystems: Investigating the Role of Platforms*. In 42nd International Conference on Information Systems (ICIS 2021), Austin, TX.

Schmid, M., Haki, K., Tanriverdi, H., Aier, S., & Winter, R. (2021a). *Platform Over Market—When Is Joining a Platform Beneficial?* Proc. Twenty-Ninth European Conference on Information Systems (ECIS 2021), Marrakesh, Morocco. https://www.alexandria.unisg.ch/263237/

Shin, D., He, S., Lee, G., WHINSTON, A., Centintas, S., & Lee, K.-C. (2020). Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach. *Management Information Systems Quarterly*, *44*(4), 1459–1492.

Shmueli, G., & Koppius, O. R. (2011). Predictive Analytics in Information Systems Research. *MIS Quarterly*, *35*(3), 553–572.

Shrestha, Y. R., He, V. F., Puranam, P., & von Krogh, G. (2021). Algorithm Supported Induction for Building Theory: How Can We Use Prediction Models to Theorize? *Organization Science*, *32*(3), 856–880. https://doi.org/10.1287/orsc.2020.1382

Sivakumar, N., Mura, C., & Peirce, S. M. (2022). Innovations in Integrating Machine Learning and Agent-Based Modeling of Biomedical Systems. *Frontiers in Systems Biology*, *2*, 41.

Song, B., Xiong, G., Yu, S., Ye, P., Dong, X., & Lv, Y. (2021). Calibration of Agent-Based Model Using Reinforcement Learning. *2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI)*, 278–281.

Sturm, T., Gerlach, J., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., Nan, N., & Buxmann, P. (2021). Coordinating Human and Machine Learning for Effective Organization Learning. *MIS Quarterly*, *45*(3), 1581–1602.

Sun, Z., Lorscheid, I., Millington, J. D., Lauf, S., Magliocca, N. R., Groeneveld, J., Balbi, S., Nolzen, H., Müller, B., Schulze, J., & Buchmann, C. M. (2016). Simple or complicated agent-based models? A complicated issue. *Environmental Modelling & Software*, *86*, 56–67.

Taghikhah, F., Filatova, T., & Voinov, A. (2021). Where Does Theory Have It Right? A Comparison of Theory-Driven and Empirical Agent Based Models. *Journal of Artificial Societies and Social Simulation*, *24*(2), 4.

Troost, C., Parussis-Krech, J., Mejaíl, M., & Berger, T. (2022). Boosting the Scalability of Farm-Level Models: Efficient Surrogate Modeling of Compositional Simulation Output. *Computational Economics*, 1–39.

Turgut, Y., & Bozdag, C. E. (2023). A framework proposal for machine learning-driven agent-based models through a case study analysis. *Simulation Modelling Practice and Theory*, *123*, 102707.

van der Hoog, S. (2017). Deep learning in (and of) agent-based models: A prospectus. *ArXiv Preprint ArXiv:1706.06302*.

Walzberg, J., Cooperman, A., Watts, L., Eberle, A. L., Carpenter, A., & Heath, G. A. (2022). Regional representation of wind stakeholders' end-of-life behaviors and their impact on wind blade circularity. *IScience*, *25*(8), 104734.

Wilensky, U., & Rand, W. (2015). *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. MIT Press.

Windrum, P., Fagiolo, G., & Moneta, A. (2007). Empirical Validation of Agent-Based Models: Alternatives and Prospects. *Journal of Artificial Societies and Social Simulation*, *10*(2), 8.

Ye, X., & He, C. (2016). The new data landscape for regional and urban analysis. *GeoJournal*, *81*(6), 811–815.

Za, S., Spagnoletti, P., Winter, R., & Mettler, T. (2018). Exploring Foundations for Using Simulations in IS Research. *Communications of the Association for Information Systems*, *42*(1), 268–300.

Zhang, W., Valencia, A., & Chang, N.-B. (2021). Synergistic Integration Between Machine Learning and Agent-Based Modeling: A Multidisciplinary Review. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(5), 2170–2190.