THE TRANSPARENCY IMPERATIVE: THE NEED FOR MODEL DOCUMENTATION FOR ENGAGING WITH PUBLIC POLICY FOLLOWING THE EU AI ACT

Michael Belfrage[a b], Fabian Lorig[a b], Christopher Frantz[c], Jason Tucker[d], and Paul Davidsson[a b]

[a]Department of Computer Science and Media Technology, Malmö University, Sweden
[b]Sustainable Digitalisation Research Centre, Malmö University, Sweden
[c]Department of Computer Science, Norwegian University of Science and Technology, Norway
[d]Institute for Futures Studies, Sweden

ABSTRACT

The application of Agent-Based Modeling and Simulation (ABMS) has few established guidelines and often suffers from insufficient model documentation. We assess the prevalence of best practices associated with different types of model documentation in light of the European Union's AI Act (AI Act). Our analysis reveals that best practices are often implemented together but ultimately reinforce the pre-existing view that ABMS frequently lacks adequate model documentation. This deficiency hinders evaluability, making it difficult to conduct quality assurance prior to application and meaningful evaluation post application. We propose a framework that highlights the importance of different types of model documentation and the attributes they enable, which are valuable to both modelers and policy actors, albeit for different reasons. The AI Act provides a valuable opportunity to improve model documentation. By proactively developing and establishing guidelines, we can stay ahead of emerging legal requirements.

Keywords: Documentation, Policy-modeling, Transparency, Responsible ABMS, EU AI Act.

1 INTRODUCTION

The complexity of the COVID-19 pandemic, coupled with the lack of experimental data, highlighted the importance and value of applying Agent-Based Modeling & Simulation (ABMS) in policy-making [1] For good reason, modeling and simulation could be considered the 'third pillar' of scientific inquiry, complementing theoretical and experimental research by integrating aspects of both in a unified methodological approach [2]. Previous research has highlighted both the pitfalls and opportunities of using ABMS in policy-making, while also identifying ways to increase its relevance [3, 4, 5, 6, 7]. However, the COVID-19 pandemic also exposed that several challenges remain, as scientists from the ABMS community voiced concerns regarding the rigor, validation, and access to model documentation for models applied in policy-making [1, 8]. These aspects, as revealed by later work, extend to other fields of application, providing additional insights into the lack of quality assurance and accreditation procedures prior to the use of policy models [9]. This deficiency represents a significant weakness in the application chain [10], increasing the model user's risk, i.e., the risk of erroneous applications [11, 12] – which has, in some cases, led to harmful system-level outcomes and accountability drift [5, 13].

Sufficient model documentation is crucial for the responsible application of ABMS, as it enables effective review and evaluation. ABMS has been used to simulate public policy – a practice known as policy-modeling [3] – in collaboration with policy actors since at least the early 2000s [9]. However, the application of policy models remains in an exploratory phase, with limited established guidelines and standardized practices

Proc. of the 2025 Annual Modeling and Simulation Conference (ANNSIM'25)
May 26-29, 2025, Universidad Complutense de Madrid, Madrid, Spain
J.L. Risco-Martín, G. Rabadi, D. Cetinkaya, R. Cárdenas, S. Ferrero-Losada, and A. Bany Abdelnaby, eds.

[10]. In fact, the lack of guidelines related to model documentation could be one of the potential explanations attributable to the often-noted insufficiency of model documentation. However, without adequate documentation, both quality assurance prior to application and evaluation post-application become difficult, if not impossible. Consequently, inadequate documentation not only complicates the application of policy models, but also hinders understanding and the effective dissemination of information within and between organizations [14]. Additionally, it restricts the modeling community's ability to review and learn from past experiences, thereby limiting opportunities for improvement in future work [12].

The introduction of the European Union's AI Act (AI Act) in 2024 will present additional demands on ABMS in public policy with regard to model documentation. The AI act is novel in that it is the world's first comprehensive law on AI. With case law establishing precedent, it will govern some, if not all, applications of ABMS within European Union (EU) jurisdiction in the future. In addition the requirement to comply with the EU AI Act may be a prerequisite for funding and collaboration in the public and private sectors. Thus, the introduction of the AI Act presents both an opportunity and, depending on its future applicability to ABMS in policy-making, a necessity for the modeling community to revisit best practices and guidelines for policy models. This preparatory work is based on the demand for transparency and explainability in AI applications, particularly in high-risk contexts – a demand that we argue ABMS is well positioned to meet when supported by appropriate model documentation. The ability to evaluate the risks of specific applications and provide sufficient "instructions for use" could also be instrumental for guiding model users while protecting developers from legal repercussions.

Against this backdrop, where examples from the field of ABMS highlight the need for increased model documentation while the EU is strengthening legal requirements for greater transparency, meeting this demand becomes essential. To address this need, this paper explores the relevance of the AI Act to the application of ABMS-based policy models and how model documentation can play a crucial role in ensuring their practical utility and compliance within the EU policy context. This is achieved through an evaluability assessment of model documentation, identifying key components that enable transparency and systematic evaluation [15, 16]. By identifying different types of model documentation from the modeling and simulation (M&S) literature, the study investigates the prevalence of best practices in applied policy-modeling projects. This work further proposes a framework for conceptualizing the attributes and validity assessments enabled by different sets of model documentation, along with the implications of these attributes, such as transparency, for both modelers and policy actors. We hope this work will inspire further methodological development and the establishment of best practices for the application of policy models, while contributing to research that will shape the future direction of applied ABMS and policy-modeling.

How can the application of ABMS-based policy models be best aligned with the AI Act?

1. How commonly are best practices related to different types of model documentation implemented in applied policy modeling?
2. How can various combinations of model documentation effectively support the needs of both modelers and policy actors?
3. What kind of model documentation is necessary to ensure transparency and evaluability in applied policy models?

2 ON THE RELEVANCE OF THE EU AI ACT TO ABMS

The EU AI Act, which was adopted in 2024, comes into full force in 2026 [17]. Following its adoption, the EU AI Act enters a phase of gradual implementation, which includes standardization efforts, regulatory guidance, and clarification of definitions through secondary legislation and guidelines. The implementation period gives AI developers, providers, and users time to align their systems with the new requirements. The AI Act has extraterritorial reach, meaning its requirements apply not only to actors within EU member

states, but also to any entity outside wishing to operate AI systems in the EU. This exemplifies the 'Brussels Effect', a phenomenon whereby EU norms and legislation impacts beyond it borders, likely making its influence significant. This is due to the EU AI Act being at the vanguard of the emerging AI governance globally [18]. Further, beyond the legal dimension, the AI Act is based on the demand for great transparency in AI systems. Thus, new soft law and norms around the use of AI are being established.

Given the gradual implementation phase, the extent to which ABMS fall under the AI Act remains uncertain. The AI Act defines AI systems as follows:

'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments [19].

There are different components included in the definition, but the most central for determining which systems based on ABMS are covered are:

- *operate with varying levels of autonomy*: When used for policy support, ABMS systems are typically not run autonomously. There is usually an analyst or researcher who supervises the execution of the simulation and analyzes the results after the execution. That is, ABMS systems do not perform any actions in the real world. Although the computational agents used in an ABMS simulation model may be considered to act autonomously, they do this only in a virtual environment, as noted in the definition, with no immediate effects in the real world. This is in contrast to other applications of agent technology, e.g., Multi-agent Systems, which can be used in autonomous applications, e.g., self-driving cars.
- *may exhibit adaptiveness after deployment*: Most ABMS systems are not adaptive in the sense of being self-learning systems. That said, it should be noted that recent developments invite the possibility to integrate machine learning (ML) in different ways within agent-based simulations. However, the use of "may" in the definition suggests optionality or lesser stringency, yet it remains an attribute that ABMS have the potential to fulfill.
- *infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*: This seems to be true for most ABMS systems, e.g. they may be used to generate recommendations and predictions about future states of the modeled system. As model output is later used to inform policy-making, it seems reasonable to assume this meets the definition for the "influence" of physical or virtual environments.

Judging from this definition, ABMS can be modeled and applied in a way that appears to fulfill all the aforementioned criteria in the EU AI Act's definition of AI. However, the interpretation of autonomy is the key attribute, and from this perspective, it will determine whether which, if not all, ABMS fall under the scope of the AI Act. Relatedly, this also hinges on the interpretation of "influence." When applied, ABMS aims to address real-world problems by generating recommendations or predictions based on the behavior of autonomous agents in virtual environments, from which insights are inferred and then applied in reality. However, these insights require a human-in-the-loop approach as they involve complex analysis, interpretation, and evaluation before being transferred to the real world. However, one must remember that this is not simply a legal or technical issue – it is a political one. The L'Aquila trial serves as a stark reminder of this [13]. In sum, the definition remains unclear, seemingly designed with a "catch-all" intent to include various systems, leaving its interpretation to legal institutions. Given this broad scope, case law and precedent will likely determine the AI Act's applicability.

While not directly related to the definition, it is also important to note that the AI Act follows a risk-based regulatory approach. However, there remains a lack of clarity on how risk will be assessed in various contexts. In a "high-risk application," there is a greater demand for transparency (see Annex 3, EU AI

Act) [19]. Regarding the use of AI by public authorities, significant criticism has been raised over the lack of transparency norms in its application and use in decision-making [20, 21]. Public pressure, alongside the new legislative requirements of the AI Act, will likely increase the demand for ABMS to be highly transparent when used in the public sector, even in low-risk contexts. Furthermore, the AI Act regulates the context of the application rather than the technology itself. Thus, clarity on the design and intended purpose of ABMS is vital to prevent their use in unintended contexts and to avoid potential compliance issues with the legislation.

The safest course of action is to assume that the AI Act will apply to all ABMS in policy-making and to be prepared to ensure compliance if relevant case law develops. Regardless of which ABMS ultimately fall within the regulatory scope of the AI Act, pursuing transparency in all ABMS applications remains a worthwhile goal. Transparency, in turn, supports quality assurance and systematic evaluation of applied ABMS, making it essential. Building on this, we now conduct an evaluability assessment to determine what is needed for transparent ABMS applications.

3 EVALUABILITY OF POLICY MODELS

Transparency is also crucial for achieving evaluability, and also necessary to comply with the AI Act, so that quality assurance of models can be performed before application and evaluation afterward. In this context, evaluability refers to the ability to evaluate the policy intervention [15, 22], and the model in a credible and reliable manner. However, if all the requirements necessary for evaluations are met, quality assurance and credibility assessment prior to application can also be performed. Model credibility refers to "a measure of confidence in the model's inferential capability" with a specific application in mind [12]. A credibility assessment involves ensuring that verification has been performed to confirm that the algorithm has been correctly implemented and is (ideally) free of bugs, including logical errors, syntax errors, and issues related to deterministic execution. Additionally, thorough validation and appropriate model testing must be conducted to ensure an adequate goodness of fit between the model and its referent, i.e., the real-world system [11]. The goal of this process is depicted in Figure 1, which shows that an increase in the deviation between the model and the referent reduces alignment, all to increase the likelihood that the element of interest for application falls within the alignment area.

With lacking model credibility, recommendations based on the model's output may be questioned, thereby reducing its impact and utility in shaping effective policies. This skepticism is understandable, as policy actors are responsible for the policy outcome [23]. However, if the model is found to be credible, the organization might seek to accredit the model for use. Usually, accreditation is achieved through the completion of the accreditation plan, which includes acceptability criteria for various model tests. Accordingly, accreditation serves as a certification procedure granted when a specific application is deemed suitable for use [24]. Assuming that the purpose of the model is clearly defined [25], the conducted simulation experiments meet the acceptability criteria and are reproducible [11], the data-generating process producing the results is accessible [1], and the results are reported in a nuanced way [12], an assessment concerning its application could be performed. This step of quality assurance is critical to ensure that models also undergo external review prior to application – which is scientifically customarily conducted after a publication has been submitted for peer review following the model's application – thereby minimizing the model user's risk [11]. Additionally, in the complete absence of model documentation, the modelers may be the only ones capable of justifying the application, thereby substantially increasing the risk of accountability drift.

The evaluability of policy models is also indispensable for evaluation post-application. This requires that the policy is evaluable, meaning it must have clearly defined objectives, allowing the deliverables to be assessed after implementation [15]. In the modeling context, this involves evaluating both the simulated intervention and the physical intervention to identify potential sources of error and gain insights for future improvements. The motivation for this dual evaluation is that any deviations between the model and the referent could stem

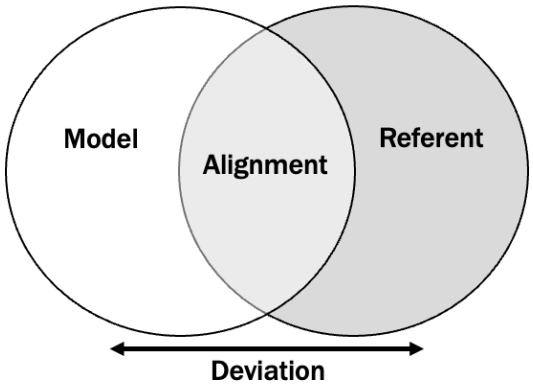

Figure 1: Minimizing the deviation between the model and the referent maximizes alignment.

from errors in either the simulated or physical intervention. For instance, if the policy implementation in the referent encounters unforeseen obstacles that substantially differ from the initial plan, the source of the deviation may be traced back to the referent. Therefore, it is crucial to evaluate both simulated and physical intervention to determine whether any unexpected deviations arise from the model itself or from the referent. Rather than measuring 'how inaccurate' the policy model was, seeing all deviations as failure. This activity should be aimed at understanding and evaluating how well the model was leveraged and what can be improved going forward. By doing so, model developers and users alike can learn from previous applications, identifying what worked well and what did not. This learning process is vital for refining both policy-modeling and the strategies for applying them, fostering a continuous improvement cycle [12].

4 MODEL DOCUMENTATION

With the importance of evaluability in mind, we now turn toward identifying the crucial model documentation in M&S projects. While modelers are responsible for ensuring the availability of appropriate model documentation [26], the key question is what specific documentation should be provided upon the project's completion. The M&S literature typically describes three different types of model documentation: model specification, verification and validation (V&V) reports, and the reporting of results.

The first type is the model **Specification**, which describes the data-generating process. Understanding the model is crucial to determining whether the model's data-generating process is a sufficiently accurate representation of the real data-generating process [27]. This specification can be conveyed informally through, e.g., a conceptual or communicative models [11], via generic model specification protocols with the most prominent being the *Overview, Design concepts, and Details* (ODD) protocol [28] (and an increasing number of purpose-specific extensions of this protocol), by detailing the model's assumptions [29], as a formal construct using differential equations or first-order logic [30, 31], or, specific for the field of M&S, using the *Discrete Event System Specification* (DEVS) [32].

Recognizing the diverse foci of agent-based models, the varying origin of data (potentially promulgated by the adoption of KIDS modeling principles [33]), we can further observe a range of specification standards concerned with provenance aspects of the modeling process [34], including the RAT protocol [35] that is specifically aimed at systematically documenting the use of qualitative data at different stages of the modeling process. Transparency and guidance for the modeling process itself is provided by the EABSS approach [36], for instance, and efforts targeting the transparent documentation of the parameterization of agent-based models include the Characterization and Parameterization framework (CAP) [37]. Independent

of the presence or absence of model specifications with respect to underlying data and process, a common practice is to submit the model code to open online repositories, such as CoMSES or GitHub [1].

The second type of documentation concerns **Verification & Validation**. On the one hand, verification practices are aimed at ensuring the correctness of the implemented code [38]. On the other hand, validation seeks to maximize fit between model and referent. Validation entails various techniques for testing the model, ranging from informal to formal, to ensure a better fit with the referent. Informal techniques tend to rely heavily on human reasoning [11]. Examples of informal tests include face validation and Schruben's Turing test, where domain experts seek to distinguish shuffled reports of output between those generated by the simulated model and those generated by the real data-generating process [26]. More formal tests are conducted using computational testing techniques and mathematical proofs [11]. While V&V is often performed in the field of ABMS, it frequently suffers from underreporting [8, 9, 39, 40].

The third type of model documentation is **Reporting**. This documentation involves the scientific reporting of model results [29]. Reporting of results is crucial, as the produced artefacts (e.g., documents) enable stakeholders to independently analyze, reference, and utilize the model's results within the organization after the modelers leave [14]. Clear and accessible reporting has the potential to aid stakeholder's understandability and serve as effective instructions for use by clearly communicating a model's capabilities, limitations, and results [41]. This documentation often includes the scientific motivation for the included parameters, results of computational experiments, and outcomes of different experimental conditions. Consequently, this could take the form as a ranked list indicating the effectiveness of different policy interventions (*in-silico*) or as qualitative insights. However, while these ranked lists of simulated policy interventions most often possess internal validity (i.e., consistency of outcomes between experimental conditions [42]), it is important that the end-user organization and non-technical personnel are well informed about the underlying assumptions as well as their potential limitations regarding external validity [12]. External validity pertains to the extent to which the outcomes of a model can be generalized to the target system [42], reflecting how the ABMS community typically conceptualizes validation [11, 27].

5 PREVALENCE OF BEST PRACTICES

Having identified three types of model documentation in the previous section, we now explore how frequently best practices associated with this documentation are performed in applied projects. A prior literature review examined the use of ABMS in policy-making [9]. Using this data, we assess the prevalence of best practices across these documentation types. 34 publications were examined with respect to the prevalence of best practices related to these three types of model documentation in projects involving various policy actors. In this study, a model specification was considered documented if the publication or its supplementary materials include a model protocol (such as ODD or variants thereof) or if the model has been uploaded to an online repository. For validation, four distinct model testing techniques were considered: face validation, quantitative validation, sensitivity analysis, and robustness checks. Model reporting were recorded if the modelers reported their results to the end-user organization in written format.

The Venn diagrams in Figure 2 illustrate the overall distribution of best practices related to the three distinct sets – Validation, Specification, and Reporting – within the superset of model practices, denoted as $MP = \mathcal{P}(\{V,S,R\})$. Condition 1 (C1), the Venn diagram on the left-hand side, represents a less restrictive condition, including models that only perform face validation and whether the model was documented using a model protocol **or** uploaded to an online repository. Only 5.9% of policy models can be found in the universe U outside of the union of the sets $(V \cup S \cup R)^c$, representing the publications that do not report performing any kind of model practices. It is directly observable that good practices tend to accompany other good practices and that validation appears to be the most reported practice. This is evidenced by the clustering of data points in the intersections surrounding the validation set. In fact, the validation set com-

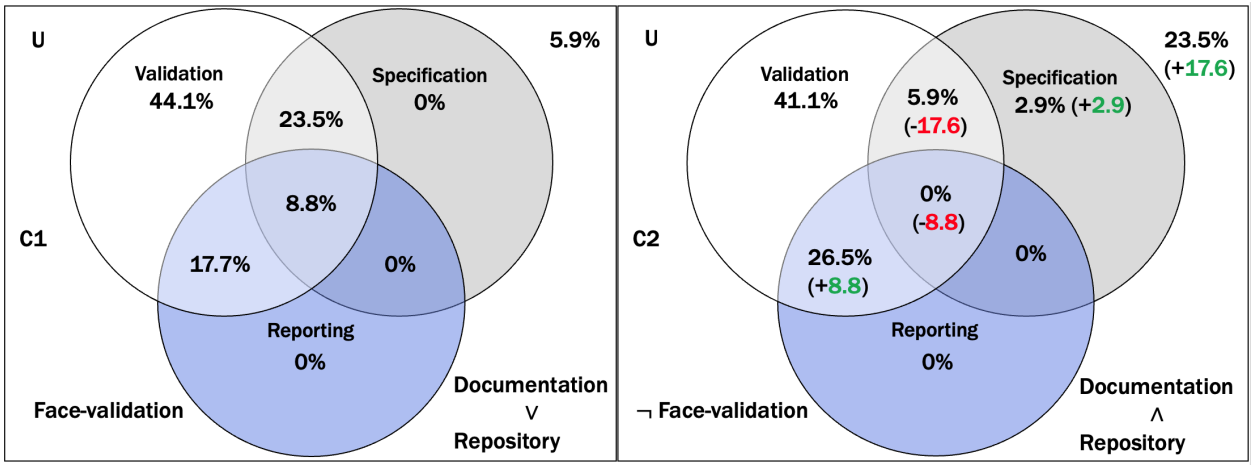

Figure 2: Best practices for the three sets of model documentation are evaluated under two conditions: C1 (less stringent) and C2 (more stringent). Changes between conditions are highlighted in red for decreases and green for increases.

prises 94.1% of the data in C1. Another insight is that relatively few publications report performing all three types of best practices $(V \cap S \cap R) = 8.8\%$.

In Condition 2 (C2), the Venn diagram on the right-hand side applies more stringent criteria for inclusion in the specification and validation sets. The specification set requires both access to code through an online repository **and** including a model protocol. For the validation set, it excludes informal model testing i.e., face validation. In this diagram, the more stringent interpretation results in an additional 17.6% from $(V \cap S)$ and 8.8% from $(V \cap S \cap R)$ being pushed outside of any set, totaling 23.5% in $(V \cup S \cup R)^c$. Unfortunately, the stricter inclusion criteria of C2 removes all data points from the inner intersection of the Venn diagram $(V \cap S \cap R)$. Interestingly, the data in intersection $(V \cap R)$ of C2 remains robust to this operation. In fact, the data in $(V \cap S \cap R = 8.8\%)$ from C1 is pushed into this area, resulting in an increase, because it no longer meets the criteria for inclusion in set S. Hypothetically, this could indicate that modelers view model testing as an integral component of reporting the outcomes of policy experiments to end-user organizations prior to application. In summary, this data indicates that applied policy-modeling suffers from a deficit in reporting or implementing best practices prior to application.

6 ATTRIBUTES ENABLED BY DIFFERENT TYPES OF MODEL DOCUMENTATION

Given the lack of best practices prior to application, we propose a framework for understanding the utility of various types of documentation and their combinations in applied projects. Evidence based policy-making aims to integrate the scientific method into the political decision-making process [43]. This integration requires distinct elements from both scientific and political frameworks to ensure adherence to democratic principles while respecting scientific practices. Scientifically, appropriate model documentation provides transparency, facilitates the reproducibility of model behavior, and enables the tracing of the origins of data, models, and code, ultimately supporting the peer-review process. Politically, the model documentation provides an evidentiary foundation serving as a motivation for the policy prescription. Building on this thinking, we examine how various aspects of evidence-based policy-making can be facilitated by specific forms of model documentation, and what implications this has for modelers and policy actors. Leveraging the same set-theoretical operationalization as in Figure 2, we identify four distinct attributes emerging at the intersections of the three types of model documentation in Figure 3, drawing on literature and examples.

The implication being that the collective value of various types of model documentation exceeds the sum of their individual parts, as different combinations of documentation enables different attributes.

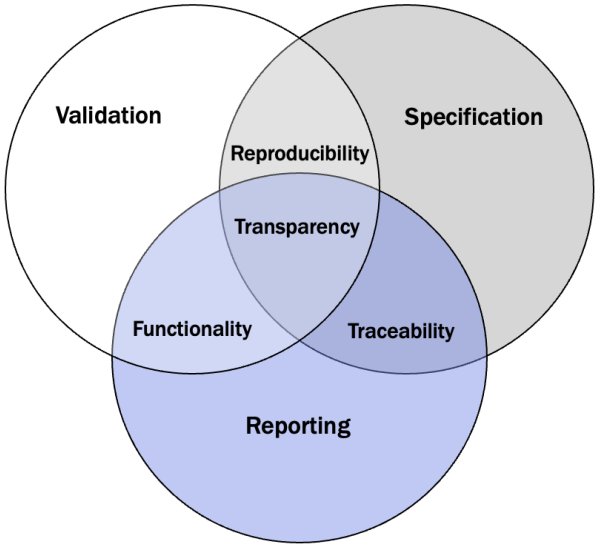

Figure 3: This application framework illustrates the attributes that emerge from combining different types of model documentation.

The first attribute is *Reproducibility*, sitting at the intersection of validation and specification. This should permit a complete understanding of the model's data-generation process and the methods used to test the model [28]. Theoretically, this should allow for the re-implementation of the model while maintaining similar levels of model fit during validation. This is a fundamental aspect for modelers, as reproducibility aligns with scientific practices and supports peer review. Between validation and reporting lies *functionality*; the results of model testing, along with written model results, indicating the intended functionality of the model. This intersection serves to establish internal validity i.e., consistency of outcomes between experimental conditions [42]. Thus, this attribute effectively supports policy actors with information about the model's application. This reporting format was applied by leading epidemiologists from the Imperial College of London during the COVID-19 pandemic, which influenced policymakers worldwide [44]. However, this work was later criticized for lacking transparency about the model's inner workings, as the model specification was not disclosed [1]. This objection indicates that documentation from the specification is also necessary to assess the external validity of the model.

Traceability lies in the intersection of specification and reporting, meaning that the model specification and the description of the data-generation process producing the results were accessible. In software engineering, traceability refers to a software's capability to link any uniquely identifiable artifact to another while sustaining these linkages over time. This allows engineers to answer questions concerning the software product and its development process [45]. Democratically, traceability ensures that any policy prescription based on the model's results could be traced back to its original specification, making it important for policy actors. This enables citizens to understand the reasoning behind political decisions and evaluate the efficacy of the political system, ultimately fostering accountability [46]. The model specification is also crucial to assess the construct validity of the model. Construct validity relates to appropriate operationalization of measurable constructs ('time', 'money' and 'weight' are examples exhibiting high construct validity) [42]. Construct validity is also required to achieve external validity, which was the fundamental reservation related to the Imperial College of London's COVID-19 model, as the model specification was missing [1]. Thus, without the specification and reporting documents, it is difficult to establish the scientific

relevance of the included parameters, assess their construct validity, and determine whether the model is appropriately specified and accurately represents the real data-generating process. At the intersection of all sets lies *transparency*, enabling evaluability, supporting the assessment of external validity, and facilitating quality assurance before application and evaluation after implementation. This, in turn, reduces obstacles to the deployment of policy models while upholding democratic legitimacy in the policy-making process, particularly following the introduction of the AI Act.

This framework naturally lends itself for an interpretation of the division of responsibilities between modelers and policy actors in applied policy-modeling projects. By ensuring that all necessary model documentation is delivered to the 'end-user' organization, ownership of the solution can be transferred from modelers to policy actors. If the public organization deems the solution appropriate for the problem, considering the most current knowledge, risks, and other evidence, accreditation may be granted (either to a specific part of the model or to the model as a whole), allowing the model to be applied [12]. Optimally, reusing as much project documentation as possible – such as the model protocol, verification and validation plan, and reporting documents for future publication – could significantly reduce effort while also ensuring that initial project goals have been met [12, 41]. This approach effectively separates model development from model application. This would place the onus on modelers to be responsible for the development and to illustrate the intended use of the model, while policy actors become responsible for the application of the model. This does not suggest that policy actors should be excluded from the development of the model or that modelers cannot assist in its application. Rather, it serves to clarify responsibilities, ensuring that policy actors benefit from science-based solutions while simultaneously protecting modelers from personally shouldering any legal liability.

7 DISCUSSION

As AI continues to advance, particularly large language models (LLMs), determining what constitutes appropriate model documentation becomes increasingly important. It is foreseeable that modelers will (at least partially) 'outsource' much of the documentation work to generative LLMs, leading to the trade-off of conceivably providing more comprehensive documentation than humans might (i.e., are all aspects of the model captured), while bearing the risk of 'hallucinating' about model attributes, or reflecting model attributes at varying levels of detail or granularity (i.e., are all algorithms/execution cycles, spelled out in sufficiently detailed form). Similarly, when reviewing verification and validation of models, the strongest possible form of testing in the form of formal proofs [11] counteracts a key underlying promise of rich agent-based models, their ability to reflect empirical reality in great detail by recognizing the underlying complexity and diversity of evidence (see the KIDS approach [33]), but can no longer be reduced to equation systems. In this context 'rich' refers to the contrast to simple conceptual agent-based models that can in fact be reduced to equation-based models.

Addressing these documentation challenges, accreditation offers benefits by explicitly assigning responsibility – shifting it from an unspecific 'set of shoulders' to well-defined roles accountable for adhering to best practices in Validation, Specification, and Reporting. To this end, an association of distinctive responsibility to archetypical roles such as 'modeler', 'domain expert', 'policy expert', 'user' (to offer an exposition) and associated accountability (e.g., the modeler's responsibility for verification) would drive ownership for the respective areas and inadvertently lead to increased methodological rigor for policy model development, and potentially elevate areas in which responsibilities are potentially overlapping (e.g., shared responsibility for Specification). However, the specification of such roles, as well as the association to specific attributes of the framework (i.e., sets and intersections) offers grounds for further development, ideally building on sound empirical support.

In light of this work, developing this or other frameworks to guide documentation standards are crucial to ensure that evaluability and consistency are maintained. Furthermore, assuming that the applied utility of a

policy model will always guide its testing and accreditation suggests that the effort required for model testing and documentation will increase with model complexity and the novelty of leveraged techniques, while potentially increasing its utility as a decision-support tool. Accordingly, the design of future documentation standards must be carefully tailored to ensure they support, rather than hinder, innovative solutions. With this in mind, we welcome any modifications to further develop this framework, or altogether different proposals, that could guide the community's way forward in application of *Responsible ABMS*. Future research based on this framework could focus on defining qualitative aspects for achieving sufficiency of documentation and beyond, enabling a scalar representation that could indicate levels of 'applicability'. The adoption of the EU AI Act should be seen as a catalyst for change in this regard.

8 CONCLUSION

While it is still uncertain to which extent ABMS will fall under the regulation of the AI Act, we have argued that operating under the assumption that ABMS will be subject to its regulatory scope is the most sound approach as case law and secondary legislation develops. Even if a subset of ABMS falls outside this regulation, it will still be subject to the new norms of transparency, which are increasingly pervasive. Thus, compliance with the AI Act, whether as a formal requirement or social expectation, is warranted. As the first AI law having global implications, with many others in the pipeline, it could also serve as a vital test case for future demands on the ABMS community.

This paper has aimed to establish practices that enable the assessment of ABMS in terms of transparency and reliability – key aspects for their effective use in policy decision-making. To this end, we have sought to identify different types of model documentation and the prevalence of best practices related to these forms of documentation in applied policy models. The main takeaways from this analysis indicate that best practices from applied policy models are rather low, but that good practices tend to be exercised in concert. While the applied use of policy models dates back to at least the early 2000s, policy-modeling remains in an exploratory phase with few established guidelines. This could hypothetically be attributed to the dispersed and tool-focused nature of policy-modeling, which is applied across various policy areas and fields of research, leading to a lack of a comprehensive overview. However, recent work has improved this knowledge base, particularly by highlighting the lack of evaluability in applied policy models. With the introduction of the AI Act, which demands greater transparency, additional pressure is now placed on model documentation. It is up to us in the modeling community to collectively shape the future of policy-modeling and the broader application of ABMS.

To this end, we have proposed a framework for conceptualizing how various types of ABMS documentation can enable essential attributes in an applied setting. We believe this is a valuable contribution that could constructively guide future efforts. This framework highlights various user/developer attributes that emerge, and validity types that can be assessed, from different combinations of documentation for modelers and policy actors. The combination of documentation from all three sets – **Specification**, **Validation**, and **Reporting** – ensures transparency and facilitates evaluability. However, this framework is neither fully developed nor without challenges. Although certain aspects are trivial to operationalize (e.g., "counting" complementary forms of documentation, such as text and code), others bear inherent challenges. Highlighting some of those, let us consider the risk of underdocumentation not only as a matter of quantity, but also quality. While coverage of these three sets of model documentation would seem necessary for any evidence-based policy-making, the qualitative aspects of these documents could provide valuable guidance to modelers on what should be considered 'sufficient documentation'.

The introduction of the EU AI Act has sparked important discussions around the application of AI, offering a valuable opportunity to leverage this momentum to improve model documentation, thereby increasing transparency, accountability, and the credibility of ABMS. Proactively developing and establishing guidelines will allow us to stay ahead of current and future requirements introduced by AI regulation.

ACKNOWLEDGMENTS

This work was partly supported by the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation.

REFERENCES

[1] F. Squazzoni, J. G. Polhill, B. Edmonds, P. Ahrweiler, P. Antosz, G. Scholz, E. Chappin, M. Borit, H. Verhagen, F. Giardini *et al.*, "Computational models that matter during a global pandemic outbreak: A call to action," *Journal of Artificial Societies and Social Simulation*, vol. 23, no. 2, p. 10, 2020.

[2] R. Axelrod, "Advancing the art of simulation in the social sciences," in *Simulating social phenomena*. Springer, 1997, pp. 21–40.

[3] N. Gilbert, P. Ahrweiler, P. Barbrook-Johnson, K. P. Narasimhan, and H. Wilkinson, "Computational modelling of public policy: Reflections on practice," *Journal of Artificial Societies and Social Simulation*, vol. 21, no. 1, 2018.

[4] M. Calder, C. Craig, D. Culley, R. De Cani, C. A. Donnelly, R. Douglas, B. Edmonds, J. Gascoigne, N. Gilbert, C. Hargrove *et al.*, "Computational modelling for decision-making: where, why, what, who and how," *Royal Society open science*, vol. 5, no. 6, p. 172096, 2018.

[5] B. Edmonds and L. ní Aodha, "Using agent-based modelling to inform policy–what could possibly go wrong?" in *Multi-Agent-Based Simulation XIX: 19th International Workshop, MABS 2018, Stockholm, Sweden, July 14, 2018, Revised Selected Papers 19*. Springer, 2019, pp. 1–16.

[6] A. Tolk, T. Clemen, N. Gilbert, and C. M. Macal, "How can we provide better simulation-based policy support?" in *Annual Modeling and Simulation Conference*. IEEE, 2022, pp. 188–198.

[7] A. Tolk, J. A. Richkus, F. L. Shults, and W. J. Wildman, "Computational decision support for sociotechnical awareness of land-use planning under complexity—a dam resilience planning case study," *Land*, vol. 12, no. 5, p. 952, 2023.

[8] F. Lorig, E. Johansson, and P. Davidsson, "Agent-based social simulation of the covid-19 pandemic: A systematic review," *Journal of Artificial Societies and Social Simulation*, vol. 24, no. 3, 2021.

[9] M. Belfrage, F. Lorig, and P. Davidsson, "Simulating change: A systematic literature review of agent-based models for policy-making," in *Annual Modeling and Simulation Conference*. IEEE, 2024, pp. 1–13.

[10] M. Belfrage, *Agent-based Social Simulation & Policy-Modelling: Facilitating Realistic and Credible Decision-making Support*. Licentiate Thesis: Malmö University Press, 2025.

[11] O. Balci, "Validation, verification, and testing techniques throughout the life cycle of a simulation study," *Annals of operations research*, vol. 53, pp. 121–173, 1994.

[12] M. Belfrage, E. Johansson, F. Lorig, and P. Davidsson, "[in]credible models–verification, validation & accreditation of agent-based models to support policy-making," *JASSS: Journal of Artificial Societies and Social Simulation*, vol. 27, no. 4, 2024.

[13] M. Cocco, G. Cultrera, A. Amato, T. Braun, A. Cerase, L. Margheriti, A. Bonaccorso, M. Demartin, P. M. De Martini, F. Galadini *et al.*, "The l'aquila trial," *Geological Society, London, Special Publications*, vol. 419, no. 1, pp. 43–55, 2015.

[14] A. Voinov and F. Bousquet, "Modelling with stakeholders," *Environmental modelling & software*, vol. 25, no. 11, pp. 1268–1281, 2010.

[15] L. G. Morra-Imas, L. G. Morra, and R. C. Rist, *The road to results: Designing and conducting effective development evaluations*. World Bank Publications, 2009.

[16] S. Barakat, F. Hardman, D. Connolly, V. Sundaram, and S. A. Zyck, "Programme review & evaluability study (PRES) – UNICEF's education in emergencies & post-crisis transition (EEPCT) programme," 2010.

[17] European Commission, "Regulatory framework on artificial intelligence," Digital Strategy, December 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[18] A. Bradford, *The Brussels Effect: How the European Union Rules the World*. Oxford, UK: Oxford University Press, 2020.

[19] "Regulation (EU) 2024/1689 of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts," Article 3, 2024, available at: https://eur-lex.europa.eu/.

[20] R. Matheus, M. Janssen, and T. Janowski, "Design principles for creating digital transparency in government," *Government Information Quarterly*, vol. 38, no. 1, p. 101550, 2021.

[21] K. Fink, "Opening the government's black boxes: freedom of information and algorithmic accountability," *Information, Communication & Society*, vol. 21, no. 10, pp. 1453–1471, 2018.

[22] OECD, *Glossary of Key Terms in Evaluation and Results-Based Management for Sustainable Development (Second Edition)*. Paris: OECD Publishing, 2023. [Online]. Available: https://doi.org/10.1787/632da462-en-fr-es

[23] B. S. Onggo, L. Yilmaz, F. Klügl, T. Terano, and C. M. Macal, "Credible agent-based simulation – an illusion or only a step away?" in *Winter Simulation Conference*. IEEE, 2019, pp. 273–284.

[24] U.S. Coast Guard, "Verification, validation and accreditation (vv&a) of models and simulations (m&s)," https://media.defense.gov/2017/Mar/13/2001710628/-1/-1/0/CI_5200_40.PDF, 2006, [Online; accessed 17-Aug-2023].

[25] B. Edmonds, "Different modelling purposes," *Simulating social complexity: A handbook*, pp. 39–58, 2017.

[26] L. W. Schruben, "Establishing the credibility of simulations," *Simulation*, vol. 34, no. 3, pp. 101–105, 1980.

[27] P. Windrum, G. Fagiolo, and A. Moneta, "Empirical validation of agent-based models: Alternatives and prospects," *Journal of Artificial Societies and Social Simulation*, vol. 10, no. 2, p. 8, 2007.

[28] V. Grimm, U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback, "The odd protocol: a review and first update," *Ecological modelling*, vol. 221, no. 23, pp. 2760–2768, 2010.

[29] A. M. Law, "How to build valid and credible simulation models," in *2022 Winter Simulation Conference (WSC)*. IEEE, 2022, pp. 1283–1295.

[30] V. Goranko, *Logic as a tool: a guide to formal logical reasoning*. John Wiley & Sons, 2016.

[31] J. M. Epstein and R. Axtell, *Growing artificial societies: social science from the bottom up*. Brookings Institution Press, 1996.

[32] G. Wainer and C. R. Martin, "Defining devs models using the cadmium toolkit," in *2022 Winter Simulation Conference (WSC)*. IEEE, 2022, pp. 1356–1370.

[33] B. Edmonds and S. Moss, "From kiss to kids – an 'anti-simplistic' modelling approach," in *Multi-Agent and Multi-Agent-Based Simulation*, P. Davidsson, B. Logan, and K. Takadama, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 130–144.

[34] E. Pignotti, G. Polhill, and P. Edwards, "Using provenance to analyse agent-based simulations," in *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, 2013, pp. 319–322.

[35] E. C.-B. Sebastian Achter, Melania Borit and P.-O. Siebers, "Rat-rs: a reporting standard for improving the documentation of data use in agent-based modelling," *International Journal of Social Research Methodology*, vol. 25, no. 4, pp. 517–540, 2022.

[36] P.-O. Siebers and F. Klügl, *What Software Engineering Has to Offer to Agent-Based Social Simulation*. Cham: Springer International Publishing, 2017, pp. 81–117.

[37] A. Smajgl and O. Barreteau, "Framing options for characterising and parameterising human agents in empirical abm," *Environmental Modelling Software*, vol. 93, pp. 29–41, 2017.

[38] C. A. Fossett, D. Harrison, H. Weintrob, and S. I. Gass, "An assessment procedure for simulation models: a case study," *Operations Research*, vol. 39, no. 5, pp. 710–723, 1991.

[39] S. Balbi and C. Giupponi, "Reviewing agent-based modelling of socio-ecosystems: a methodology for the analysis of climate change adaptation and sustainability," *University Ca'Foscari of Venice, Dept. of Economics Research Paper Series*, no. 15_09, 2009.

[40] J. Schulze, B. Müller, J. Groeneveld, and V. Grimm, "Agent-based modelling of social-ecological systems: achievements, challenges, and a way forward," *Journal of Artificial Societies and Social Simulation*, vol. 20, no. 2, 2017.

[41] M. Belfrage, F. Lorig, and P. Davidsson, "Making sense of collaborative challenges in agent-based modelling for policy-making," in *2nd Workshop on Agent-based Modeling and Policy-Making (AMPM 2022)*. CEUR-WS Vol-3420, 2022.

[42] T. D. Cook, D. T. Campbell, and W. Shadish, "Experimental and quasi-experimental designs for generalized causal inference," 2002.

[43] P. Cairney, "The politics of evidence-based policy making," 2016.

[44] N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg *et al.*, *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand*. Imperial College London London, 2020, vol. 16.

[45] J. Cleland-Huang, O. C. Gotel, J. Huffman Hayes, P. Mäder, and A. Zisman, "Software traceability: trends and future directions," in *Future of software engineering proceedings*, 2014, pp. 55–69.

[46] V. A. Schmidt, "Democracy and legitimacy in the european union revisited: Input, output and 'throughput'," *Political studies*, vol. 61, no. 1, pp. 2–22, 2013.

AUTHOR BIOGRAPHIES

MICHAEL BELFRAGE holds a MSc in computational social science with focus on social networks. Currently, he is a PhD-candidate in computer science at Malmö University, Sweden. His research interests include complex systems, agent-based simulation and computational politics. michael.belfrage@mau.se.

FABIAN LORIG holds a PhD in information systems research with focus on agent-based simulation. He is an assistant professor at Malmö University, Sweden and his research interests include the use of agent-technology for decision support with applications in logistics, transportation, and healthcare. fabian.lorig@mau.se.

CHRISTOPHER FRANTZ holds a PhD in information science with focus on agent-based modelling and simulation. He is an associate professor at the Norwegian University of Science and Technology (NTNU). His research interests include institutional science, policy analysis as well as agent-based modelling. christopher.frantz@ntnu.no.

JASON TUCKER holds a PhD in Social and Policy Sciences with a focus on global governance. He is a Researcher at the Institute for Futures Studies, Sweden. His research interests include artificial intelligence, healthcare, policy and global politics. jason.tucker@iffs.se.

PAUL DAVIDSSON is a professor in computer science at Malmö University, Sweden, as well as the director of the Sustainable Digitalisation Research Centre. His research interests include the theory and application multi-agent systems, artificial intelligence, and Internet of Things. paul.davidsson@mau.se.