AI for System Dynamics: Mapping Progress across Six Modelling Stages

Aisyah Dewi Muthiah[1], Hasan Turan[1,2], Sondoss El Sawah[1,2]

[1]School of Systems and Computing, UNSW Canberra, Australia
[2]Capability Systems Centre, UNSW Canberra, Australia

ABSTRACT

New opportunities emerge from the integration of Artificial Intelligence (AI) into the System Dynamics (SD) to improve its manual, traditional, and expert-driven modelling process. This paper presents a literature review of 10 studies that apply AI techniques across the six stages of SD modelling, ranging from problem identification and definition to the design of learning strategy and infrastructure. The results show that the current integration of AI is progressing in the stages of system conceptualization, with techniques such as causal extraction producing promising results. However, the last three stages remain under-explored. Different methods of performance evaluation, as well as limitations in traceability, are among the challenges that have been identified in the current literature. This paper identifies gaps in methodology and suggests the development of an integrated and interpretable AI-SD pipeline that includes proper testing of model behaviour and allows for engagement with stakeholders to produce reliable and efficient SD modelling practices.

Keywords: Artificial Intelligence, System Dynamics, AI supported Modelling, Large Language Models, Natural Language Processing

INTRODUCTION

System Dynamics (SD) is a modelling approach designed to analyze, understand, and simulate the behaviour of complex systems over time, represented by feedback loops, stocks, and flows (Forrester, 1971; Sterman, 2000; Ford, 2010). The model itself is built using structural elements such as stocks, flows, auxiliaries, and constants to represent behaviour within the system (Richardson 2011). SD has been applied to many domains, as it has proven to capture feedback-driven behaviour and dynamic complexity (Yasarcan, 2023). However, the SD process remains time and labour-intensive. As systems become more complex, the traditional modelling approach becomes challenging for modellers, especially novices, in terms of scalability and time to decision-making (Deutsch et al., 2024).

Current research in AI integration with SD modelling processes, such as the use of Natural Language Processing (NLP) and Machine Learning (ML), offer promising solutions to ease some of the challenges faced by traditional modelling processes. For example, AI can be used to extract causal relationships from textual data, automate the building of causal loops, or even simulate the model within AI interfaces. Despite these attempts, current studies still lack a systematic framework for assessing how AI techniques can support each stage of the SD modelling process. This paper addresses the gap by systematically reviewing the literature on AI-enhanced SD modelling processes across all the stages defined by Martinez-Moyano and Richardson (2013). The objective is to map current practices, identify gaps, and provide recommendations on integrating AI into the conventional SD workflow.

METHODOLOGY

This study conducts a literature review to investigate how modern AI techniques (e.g., ML, NLP, LLM) have been integrated into the six stages of SD modelling, as depicted in Figure 1. The literature was collected from a combination of peer-reviewed journals, conference proceedings, and high-quality preprint published between 2015 and 2025 with keywords such as "System Dynamics", "SD", "Artificial Intelligence", "AI", "Natural Language Processing", "Large Language Models", "Causal Extraction", "Causal Loop Diagram". The selection process itself depicted in Figure 2 adapting from PRISMA 2020 flow diagram with some adjustment.

The objective was to identify research that explicitly integrates AI into the SD model development rather than applying AI techniques to optimize the results post-modelling. The study must contribute to at least one stage of SD modelling processes, as depicted in Figure 1, and the AI technique is used to mimic, automate, or improve the traditional SD task. Studies that only apply AI to post-modelling tasks, such as non-AI automation, fuzzy systems, and expert rules, and focus on educational tools are exempt. Each paper will be identified in which stages they have been applying AI, determining the type of AI used, assessing the performance evaluation approach, validation mechanism, and limitations.

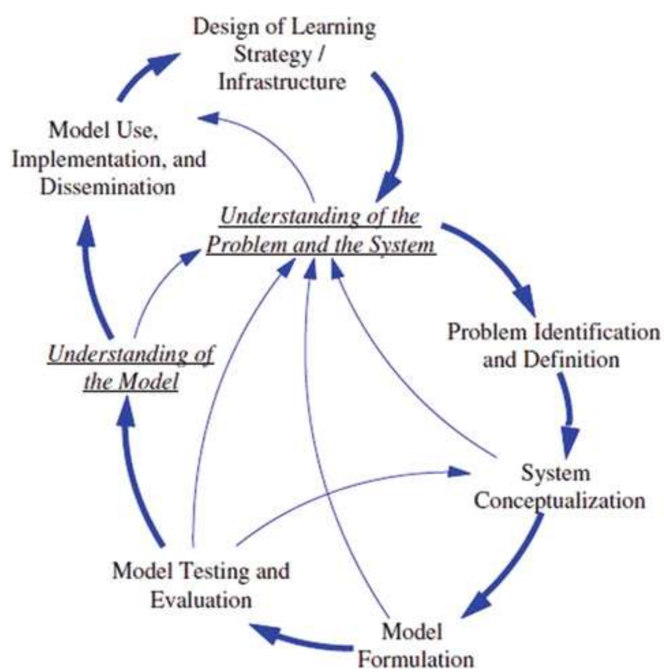

Figure 1. SD modelling approach (Martinez-Moyano and Richardson, 2013)

RESULTS AND ANALYSIS

According to the inclusion and exclusion criteria and the flow based on picture depicted in figure 2, thirteen validated studies were identified that integrate AI techniques across the SD modelling processes. These AI techniques include Natural Language Processing (NLP), Machine Learning (ML), and Large Language Models (LLM). From the table, it is evident that AI techniques have been increasingly adopted to support the traditional SD processes that rely heavily on stakeholders and experts, such as identifying and structuring causal relationships. However, stages such as model use and testing, as well as the design of learning strategy and infrastructure, remain heavily underexplored.

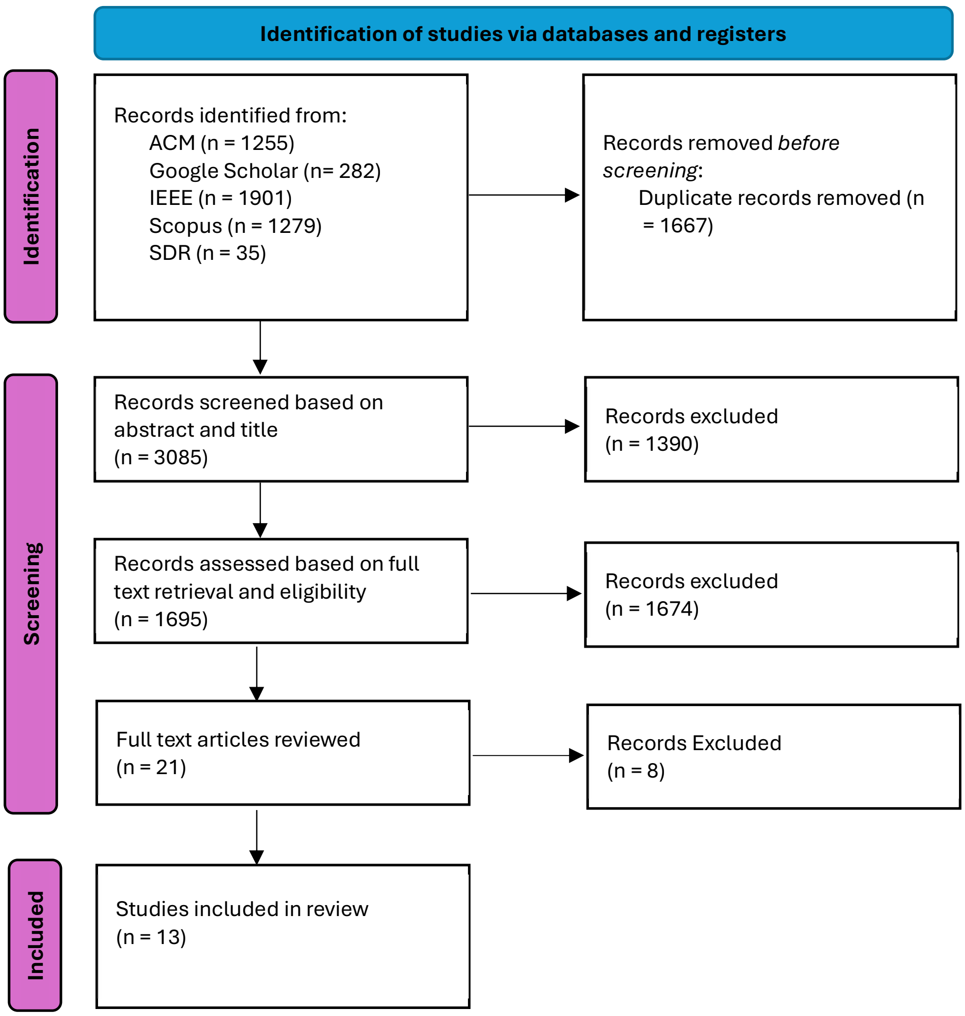

Figure 2. The selection process of the included literature

Problem identification and system conceptualization are the early and foundational stages of SD modelling to understand system boundaries, select relevant variables and relationships, and construct the causal map. Traditionally, this process heavily relies on engagement with experts or stakeholders (Sterman 2000). Among the reviewed studies, none explicitly support the problem identification stage, as they often assume a predefined problem context. Akhavan & Jalali (2024) shown an exception as it gives a preliminary insight on how ChatGPT can be used as a support through stages 1-4 of SD modelling including problem identification. In which it able to refine research question, exploring problem background, as well as suggesting system boundaries. All based on the input from the user.

Significant contribution was observed in the system conceptualization stage, particularly through the use of qualitative data, including automatic causal extraction and the generation of causal loop diagrams. For the use case scenario in this stage Akhavan & Jalali (2024) and Jalali & Akhavan (2024) provide an insight on how to use Chat GPT for variable and relation extraction as well as polarity checking based on user input and interview. The interview case study shows that the results generated by Chat GPT comparable to the one produced manually by the expert. Bakker (2023) compares the use of traditional NLP and LLM for extracting causal entity and relationships from text with an emphasis on textual preprocessing and sentence classification to build a knowledge graph while Veldhuis et al (2024) used BERT and GPT for causal sentences extractions. Those two studies provide early insight for conceptual modelling. Based on these two studies, Hosseinichimeh et al. (2024) propose a GPT-based SD bot with structured prompts to extract variables and causal links and produce a Causal Loop Diagram (CLD). This study shows the ability to mimic human-generated CLD with reasoning chains and relevant text that support the claims. It also allows novice modellers to be confident whether the identifying variables, relations, and polarity are made up or not. Liu and Keith (2025) have similar work on automating the building of CLD by using curated prompting that can be applied across different texts, and the results show that for simple structures, the prompt can produce similar quality to CLD built by experts. Schoenberg et al. (2025) proposed "SD-AI" in attempt to improve the SD Bot pipeline with its multi LLM benchmarking framework, highlighting GPT-4.5 's superior performance.

On more advanced context, Giabbanelli et al. (2025) benchmarks multiple LLMs for their ability to transform text to causal maps and the other way around, proposing different metrics as a form of assessment like accuracy, coverage, and quality of causal inferences. Valdivia Cabrera et al. (2025) study utilised SBERT-based semantic clustering to merge multiple CLD's in attempt to enable scalable stakeholders' integration.

The model formulation stage involves the building of a mathematical model and stock and flow structures, as well as preparing the simulation process. In terms of modern AI techniques, two studies have been trying to incorporate them into this stage. Du Plooy and Oosthuizen (2023) build, refine, and simulate the SD model using ChatGPT 4. Although the results still need to be validated by experts, their work shows promising results in GPT-assisted model development. On the other hand, Gadewadikar and Marshall (2024) specifically target the parameter estimation activity within the model formulation by using several machines learning techniques, including support vector machines (SVMs), artificial neural networks (ANNs), and random forests (RF). This study offers a scalable approach to tuning model parameters using AI despite concerns about generalizability and interpretability.

Besides Du Plooy and Oosthuizen's (2023) study that integrates AI in the stage of model testing and evaluation by simulating the model within the GPT 4, Hu et al. (2025) introduce a more advanced level by embedding SD model into Chat GPT 4 and Chat GPT

4o. This preliminary work enables simulation in conversational interface eventhough it still requires human validation but shows a possibility of reducing technical barriers for novice users.

Table 1. AI supported SD papers

Author	Year	AI Method	AI Role	SD Stages Supported	Key Contributions
Valdivia Cabrera et al.	2025	NLP (SBERT)	NLP-based variable extraction and clustering	Stage 2: System Conceptualization	Used Sentence-BERT to merge 13 participatory CLDs by clustering semantically similar variables. Achieved a high F1-score for automatic merging. Supported stakeholder CLD integration at scale.
HB Taramsari et al.	2024	LLaMA-2 (LLM) + NLP technique (LDA)	LLM for variable (cause of accidents) identification and categorisation	Stage 2: Conceptualization	Extracting variables from unstructured aerospace incident reports, linked to cascading failures to support expert-led CLD generation
Schoenberg,B et al.	2025	11 LLMs (incl GPT-4.5, GPT-4o, o1)	Extract variable, causal relations and polarity in one pass prompting	Stage 2: Conceptualization	Developed "sd-ai" to generate CLD from given input, benchmarking several LLMs. Gpt-4.5-preview achieved highest total score on causal translation + instruction conformance.
Giabbanelli et al.	2025	GPT	LLM transformation from maps to text and text to maps	Stage 2: Conceptualization	Providing open datasets, formats, and tools to support and assess map-to-text and text-to-map tasks
Bo Hu	2025	ChatGPT-4 and ChatGPT-4o	Run and Modify SD simulation in ChatGPT	Stage 3–4: Model Formulation & Testing	ChatPySD, embedding SD models into the ChatGPT 4 interface through PySD in order to enable natural language-based simulation, scenario testing, and parameter tuning without coding expertise
C. du Plooy & R. Oosthuizen	2023	GPT-4	Develop and run the SD model and simulation in Chat GPT	Stage 3–4: Model Formulation & Testing	Assesses GPT-4's ability to iteratively build, refine, and translate SD models into Python with human guidance, showing strong simulation accuracy but limited error detection without expert input
Liu & Keith	2025	GPT-3.5	Variable, causal relation, polarity identification and CLD building	Stage 2: Conceptualization	Generated CLDs from user-defined dynamic hypotheses using GPT-4 with structured prompting. The two-stage approach has the best results.
Hosseinichimeh et al.	2024	GP- 4 Turbo	Variable, causal relation, polarity identification and CLD building	Stage 2: Conceptualization	Created an SD Bot that builds CLDs iteratively from user inputs and provides reasoning and relevant text as well. Validated via expert review and comparison to reference maps.

Author	Year	AI Method	AI Role	SD Stages Supported	Key Contributions
Bakker et al.	2023	GPT-3.5	Extracting entities and causal relations to create knowledge graph	Stage 2: System Conceptualization	Compared traditional NLP and GPT for causal extraction. Found LLM performed better but needed structured feedback to correct the loop logic. A prototype allows feedback integration.
Jalali & Akhavan	2024	ChatGPT 4	Variable, causal relation, polarity identification and CLD building	Stage 2: System Conceptualization	Used GPT-3.5 to code 10 interview transcripts. Showed comparable thematic coding accuracy to human coders in the SD context. Emphasised the cognitive support role.
Akhavan & Jalali	2024	ChatGPT 4	Refine research question and explore problem background, map to text conceptualisation refinement, provide feedback on model equations and parameter values, develop code, and evaluate and see the simulation model	Stage 1-4	Demonstrated how LLM can be use to support SD modelling and Identified misuse risks of LLMs in SD modeling stages
Veldhuis et al.	2024	NLP (Bert based model and GPT 3.5)	causal elationship extraction	Stage 2: System Conceptualization	Demonstrated how NLP models can extract causal sentences from text to support early SD model development
Gadewadikar et al.	2024	ANNs models, SVMs models, and RF models	AI for parameter estimation	Stage 3: Model Formulation	Applied ML-based estimation for SD parameters, resulting in the simulation giving similar results to real-world data

No other studies extend the contribution of AI integration within SD modelling stages beyond the model formulation and testing. While automation accelerates the modelling processes, some tools are good enough for integration within the decision-making processes. The challenge persists in terms of ensuring usability, transparency, and stakeholder engagement.

METHODOLOGICAL GAP AND CHALLENGE

Despite promising progress in integrating AI into SD, it remains fragmented and incomplete. None of the studies implemented an end-to-end SD modelling pipeline. Most of them focused on key activities within a single stage while heavily relying on human input for validation, refinement, and interpretation. For example, Liu and Keith (2025) and Hosseinichimeh et al. (2024) demonstrate how AI can aid in generating CLD; however, users must still validate the output. This highlights the need for AI-human collaboration workflows rather than full automation.

User feedback mechanisms are also rare within the studies; only Bakker et al. (2023) integrate interactive and user-in-the-loop modelling to iteratively review and annotate AI-suggested variables and links, Schoenberg et al. (2025) enable incremental edit through prompt based interactions and structured output. Both system promote collaborative modelling framework

Other than that, the evaluation approach across the studies varies widely and is underspecified. Veldhuis et al. (2024) apply F1, precision, recall, and completeness scores to assess model quality, Hosseinichimeh et al. (2024) apply link and loop match, Schoenberg et al. (2025) and Giabbanelli et al. (2025) employ more detailed metrics but in a different form. Most studies apply visual inspection without a systematic evaluation framework. This raised concerns regarding scalability, generalizability, and trust in the results. Additionally, no study mentions model saturation, a critical concept to determine whether the model is sufficient enough before moving on to the next stage (Tomaia-Cotisel, A. et al., 2023). This raises concerns about whether an AI-supported model is considered complete.

Lastly, only Hosseinichimeh et al. (2024) provide reasoning and traceability, while most other studies remain black boxes. This undermines the usage of AI in collaborative modelling and learning contexts.

In summary, although AI tools are making SD modelling more efficient. There is a clear need for integrative, explainable, and user-centred solutions that support SD modelling through all cycles and enable feedback, provide robust evaluation, and offer transparency on how outputs are derived (Bakker et al., 2023; Hosseinichimeh et al., 2024; Giabbanelli et al., 2025; Liu and Keith, 2025; Schoenberg et al., 2025; Veldhuis et al., 2024).

CONCLUSION

This review demonstrates that while there is significant progress in SD modelling supported by AI, especially in system conceptualization and model formulation, the integration across the SD lifecycle remains limited. Most studies fall short of enabling transparent, traceable, and interactive pipelines. The lack of an evaluation framework, model saturation assessment, and user feedback mechanism are challenges present in current studies. These challenges limit the usability of AI tools to support novice modellers. Future research should focus on developing AI systems that support end-to-end modelling with human-the-loop features, transparent reasoning path, and iterative validation mechanisms that mimic the dynamic and collaborative nature of traditional SD modelling workflows

REFERENCES

Akhavan, A. and Jalali, M.S., 2024. Generative AI and simulation modeling: how should you (not) use large language models like ChatGPT. *System Dynamics Review*, *40*(3), p.e1773.

Bakker, R.M., Blok, D., de Boer, M.H., Kalkman, G.J., Veldhuis, G.A. and van Waas, R.P., 2023. Exploring knowledge extraction techniques for system dynamics modelling: Comparative analysis and considerations. BNAIC/BENELEARN 2023.

Deutsch, A.R., Frerichs, L., Perry, M. and Jalali, M.S., 2024. Participatory modeling for high complexity, multi-system issues: challenges and recommendations for balancing qualitative understanding and quantitative questions. System Dynamics Review, 40(4), p.e1765.

du Plooy, C. and Oosthuizen, R., 2023. AI usefulness in systems modelling and simulation: GPT-4 application. South African Journal of Industrial Engineering, 34(3), pp.286-303.

Ford, A., 2010. Modeling the environment. Washington, DC: Island Press.

Forrester, J.W., 1971. Counterintuitive behavior of social systems. Theory and Decision, 2(2), pp.109-140.

Gadewadikar, J. and Marshall, J., 2024. A methodology for parameter estimation in system dynamics models using artificial intelligence. Systems Engineering, 27(2), pp.253-266.

Giabbanelli, P.J., Gandee, T.J., Agrawal, A. and Hosseinichimeh, N., 2025. Benchmarking and assessing transformations between text and causal maps via large language models. Applied Ontology, p.15705838241304102.

Hosseinichimeh, N., Majumdar, A., Williams, R. and Ghaffarzadegan, N., 2024. From text to map: a system dynamics bot for constructing causal loop diagrams. System Dynamics Review, 40(3), p.e1782.

Hu, B., 2025. ChatPySD: Embedding and simulating system dynamics models in ChatGPT-4. System Dynamics Review, 41(1), p.e1797.

Jalali, M.S. and Akhavan, A., 2024. Integrating AI language models in qualitative research: Replicating interview data analysis with ChatGPT. System Dynamics Review, 40(3), p.e1772.

Liu, N.Y.G. and Keith, D.R., 2025. Leveraging large language models for automated causal loop diagram generation: Enhancing system dynamics modeling through curated prompting techniques. arXiv preprint arXiv:2503.21798.

Martinez-Moyano, I.J. and Richardson, G.P., 2013. Best practices in system dynamics modeling. System Dynamics Review, 29(2), pp.102-123.

Richardson, G.P., 2011. Reflections on the foundations of system dynamics. System Dynamics Review, 27(3), pp.219-243.

Schoenberg, W., Girard, D., Chung, S., O'Neill, E., Velasquez, J. and Metcalf, S., 2025. How well can AI build SD models?. arXiv preprint arXiv:2503.15580.

Sterman, J.D., 2000. Business dynamics: Systems thinking and modeling for a complex world. Boston: Irwin/McGraw-Hill.

Taramsari, H.B., Rao, B., Nilchiani, R. and Lipizzi, C., 2024, March. Identification of variables impacting cascading failures in aerospace systems: A natural language processing approach. In Conference on Systems Engineering Research (pp. 413-427). Cham: Springer Nature Switzerland.

Tomoaia-Cotisel, A., Allen, S.D., Kim, H., Andersen, D.F., Qureshi, N. and Chalabi, Z., 2024. Are we there yet? Saturation analysis as a foundation for confidence in system dynamics modeling, applied to a conceptualization process using qualitative data. System Dynamics Review, 40(4), p.e1781.

Valdivia Cabrera, M., Johnstone, M., Hayward, J., Bolton, K.A. and Creighton, D., 2025. Integration of large-scale community-developed causal loop diagrams: a Natural Language Processing approach to merging factors based on semantic similarity. BMC Public Health, 25(1), pp.1-9.

Veldhuis, G.A., Blok, D., de Boer, M.H., Kalkman, G.J., Bakker, R.M. and van Waas, R.P., 2024. From text to model: Leveraging natural language processing for system dynamics model development. System Dynamics Review, 40(3), p.e1780.

Yasarcan, H., 2023. The four main elements of dynamic complexity. System Dynamics Review, 39(2), pp.171-179.