

OLD DOMINION UNIVERSITY

VMASC
OLD DOMINION UNIVERSITY

HOW CAN AI PROVIDE A SEMANTIC LAYER FOR COMPOSABILITY?

Jhon G. Botello[1,2], Virginia Zamponi[2], Jose J. padilla[2], Erika Frydenlund[2]

[1]Department of Computer Science

[2]Virginia Modeling Analysis and Simulation Center

Old Dominion University, Norfolk, VA, USA

jgbotello Storymodelers

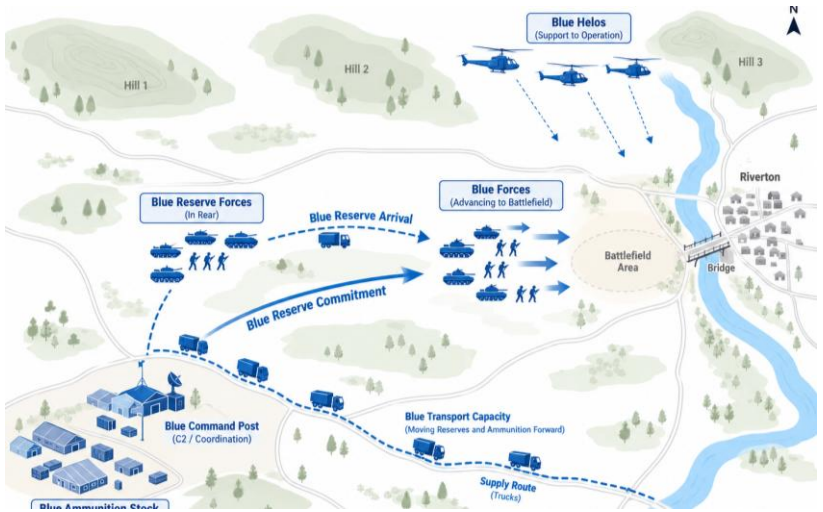
Modeling and Simulation for Autonomous Systems Conference (MESAS) 2026

Agenda

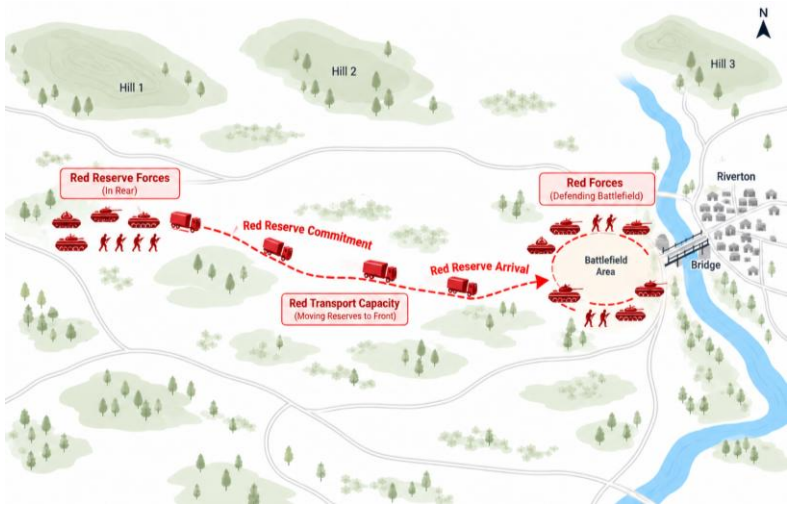
1. The Problem with Composition
2. How Can AI Help with Composition?
3. Challenges and Limitations

Models Represent Real-World Systems Through Abstractions Designed to Answer Specific Questions.

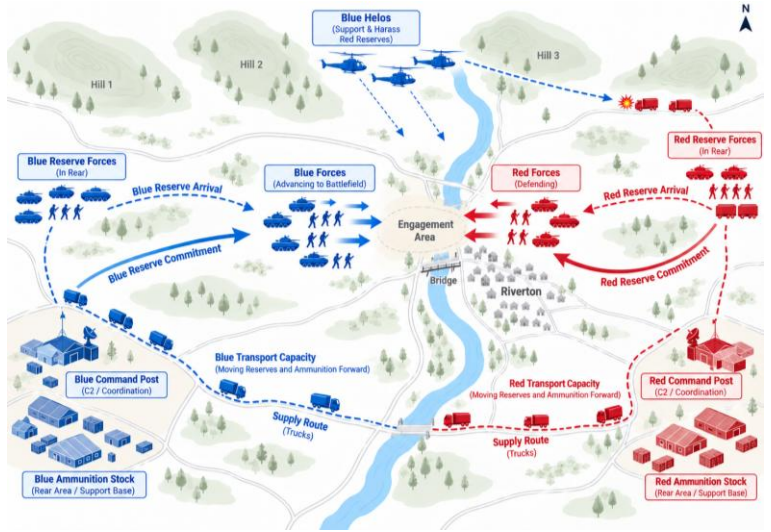
Modelers Also Use Models to Answer Questions They Were Not Originally Designed For, Driving Interest in Model Reuse and Composition.



Model A: Blue Forces Simulation

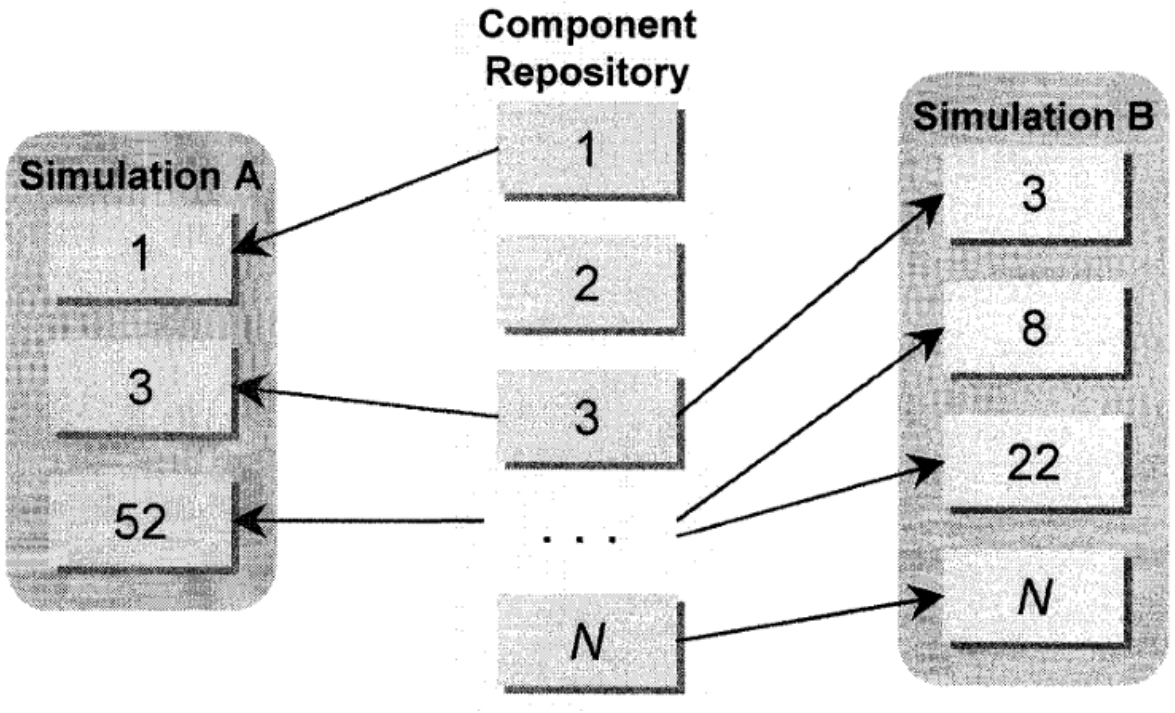


Model B: Red Forces Simulation



Model C: Blue-Red Engagement Simulation

Composition Is the Process of Combining Two or More Models to Form a Larger, Integrated Model. Composability Is the Ability of Models to Be Combined.



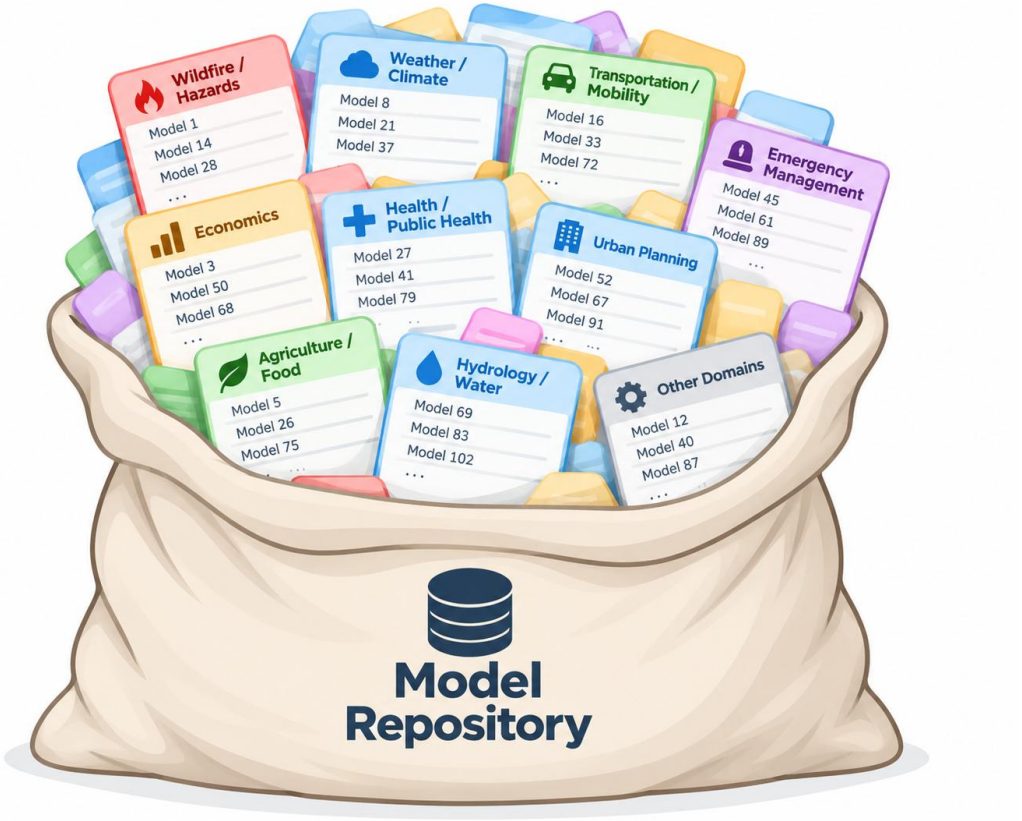
Notional Example of Composability. Source: [1]

There Are **Two** Main Challenges When It Comes to Model Composition:

1. **Searching** for Relevant Candidate Models to Combine.
2. **Combining** Models into a Coherent Simulation.

[1] Weisel, E. W. (2004). Models, Composability, and Validity [Old Dominion University Libraries]. https://doi.org/10.25777/43PVGS24

The Searching problem: Imagine the Complexity of Searching Among a Large Repository of Models



Inside the Bag:

1. Different Domains, Assumptions, Units, and Modeling Paradigms
2. Lot of Models to Search and Compare
3. Not All Models Are Relevant, but Several May Be Potential Candidates

The Combination Problem: Manually Analyzing Scenarios, Extracting Requirements, Assessing Composability, and Combining Models Is a Complex and Time-Consuming Process.

Given a Scenario:

Blue Forces advance toward an objective area defended by Red Forces. Both forces have units with different capabilities and movement speeds. Blue Forces move toward the objective, while Red Forces defend their positions. Terrain conditions and unit availability influence movement and engagement outcomes.

- Entities
- Attributes
- Variables
- Relationships

Modelers would need to identify the components required by the scenario, search for and retrieve relevant candidate models, and integrate them while reconciling differences in assumptions, units, and other model characteristics.

And a Modeling Question:

How Do Unit Capabilities, Terrain Conditions, and Force Availability Affect the Outcome of a Blue-Red Engagement?

System Dynamics (SD) Offers an Application Context to Study Model Reuse and Composition

What Do We Need to Build a System Dynamics Model?

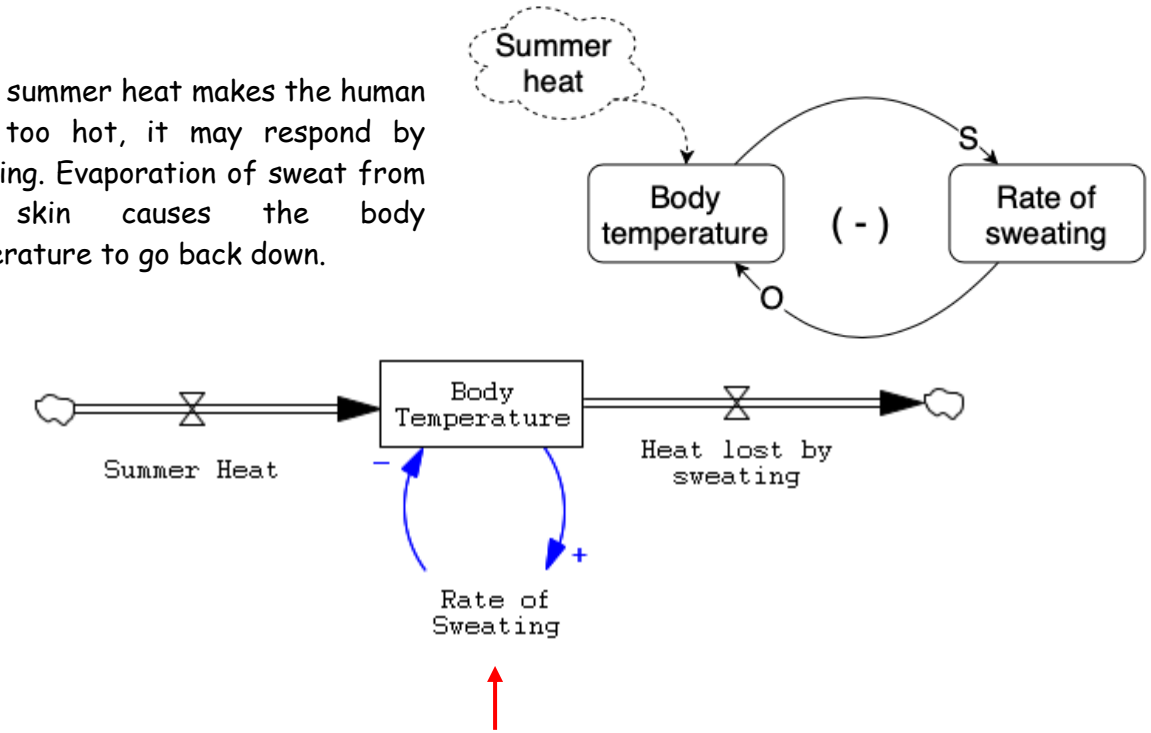
- System observation and understanding
- Key variable identification
- Causal relationships & feedback loops Identification (optional CLD Representation)
- Stock–flow & delays structure (SFD representation)
- Data and expert knowledge

Example

Narrative

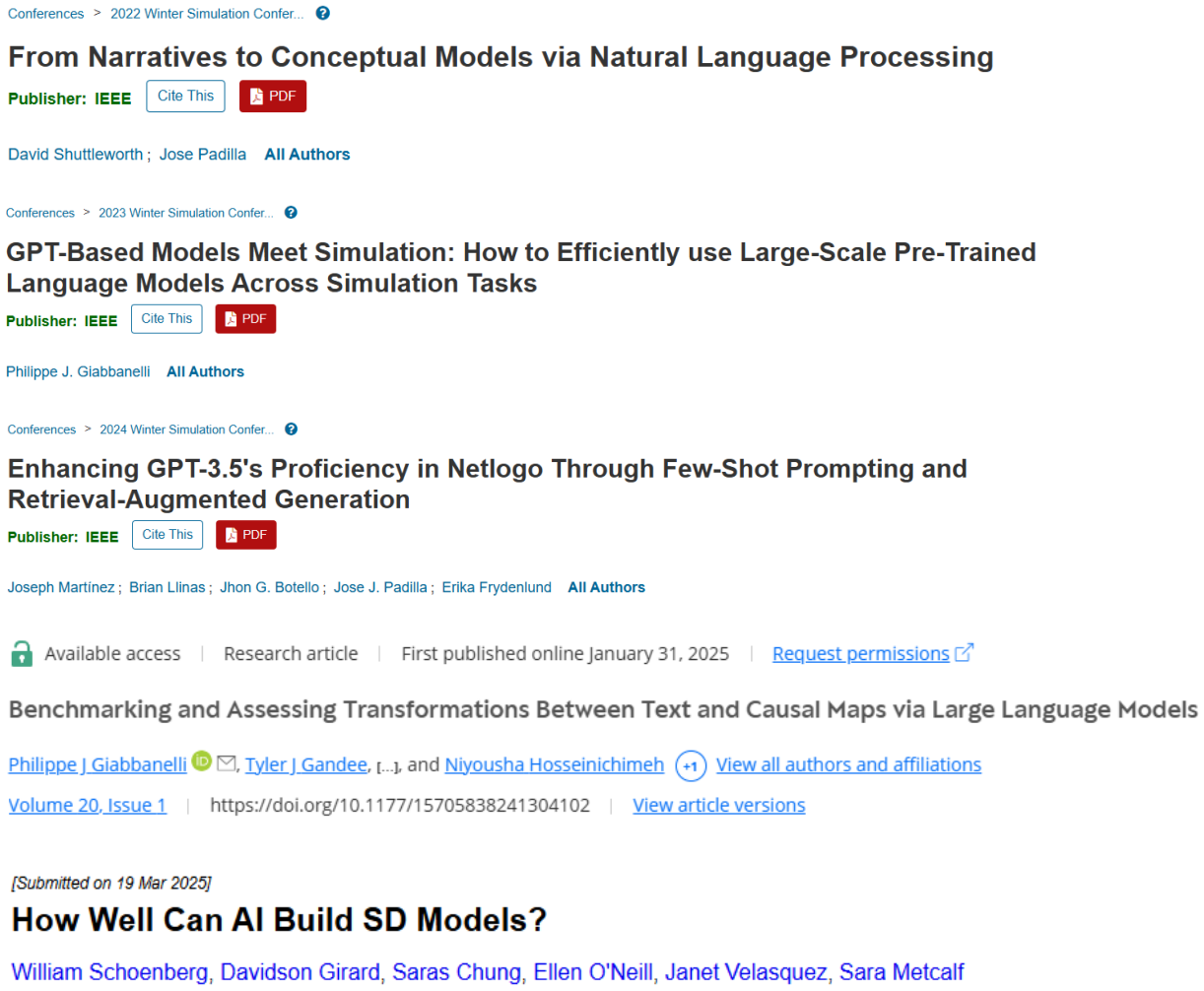
When summer heat makes the human body too hot, it may respond by sweating. Evaporation of sweat from the skin causes the body temperature to go back down.

Corresponding causal loop diagram



We Focus on SFD Representations

Can AI Help With Model Composition?



Conferences > 2022 Winter Simulation Confer...

From Narratives to Conceptual Models via Natural Language Processing

Publisher: IEEE | Cite This | PDF

David Shuttleworth ; Jose Padilla All Authors

Conferences > 2023 Winter Simulation Confer...

GPT-Based Models Meet Simulation: How to Efficiently use Large-Scale Pre-Trained Language Models Across Simulation Tasks

Publisher: IEEE | Cite This | PDF

Philippe J. Giabbanelli All Authors

Conferences > 2024 Winter Simulation Confer...

Enhancing GPT-3.5's Proficiency in Netlogo Through Few-Shot Prompting and Retrieval-Augmented Generation

Publisher: IEEE | Cite This | PDF

Joseph Martínez ; Brian Llinas ; Jhon G. Botello ; Jose J. Padilla ; Erika Frydenlund All Authors

Available access | Research article | First published online January 31, 2025 | Request permissions

Benchmarking and Assessing Transformations Between Text and Causal Maps via Large Language Models

Philippe J Giabbanelli, Tyler J Gandee, [...], and Niyousha Hosseinichimeh +1 View all authors and affiliations

Volume 20, Issue 1 | https://doi.org/10.1177/15705838241304102 | View article versions

[Submitted on 19 Mar 2025]

How Well Can AI Build SD Models?

William Schoenberg, Davidson Girard, Saras Chung, Ellen O'Neill, Janet Velasquez, Sara Metcalf

Research at the intersection of **AI and M&S** has largely focused on **building models from scratch**, while **model reuse and composition** have received comparatively **less attention.**

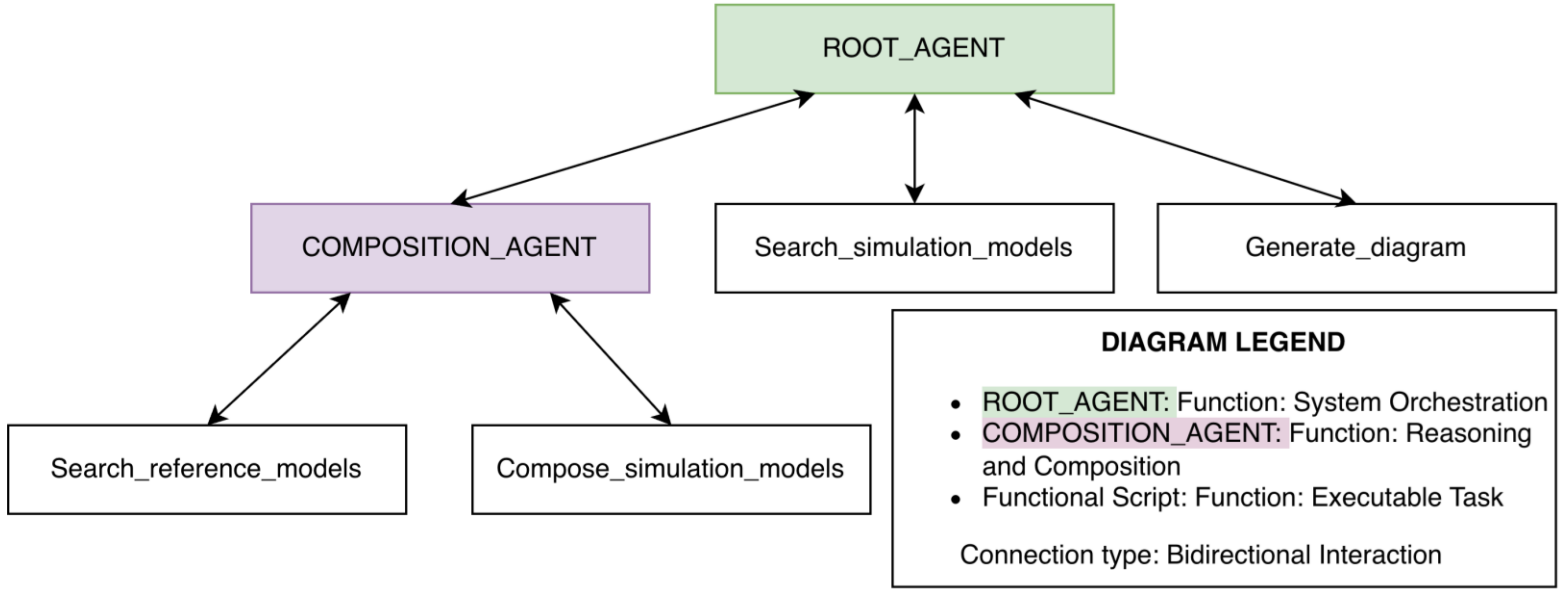
AI-Driven Model Composition Raises Several questions and Challenges

- How can models be indexed and retrieved when AI systems have not been designed for them?
- What does a modeling question look like?
- How much does AI know about the topic and method that we are trying to explore?
- What limitations may LLMs have when processing information from large datasets/models?

We Built an Agentic AI system Aiming to Help in the modeling process by Reusing Models and Combining Them



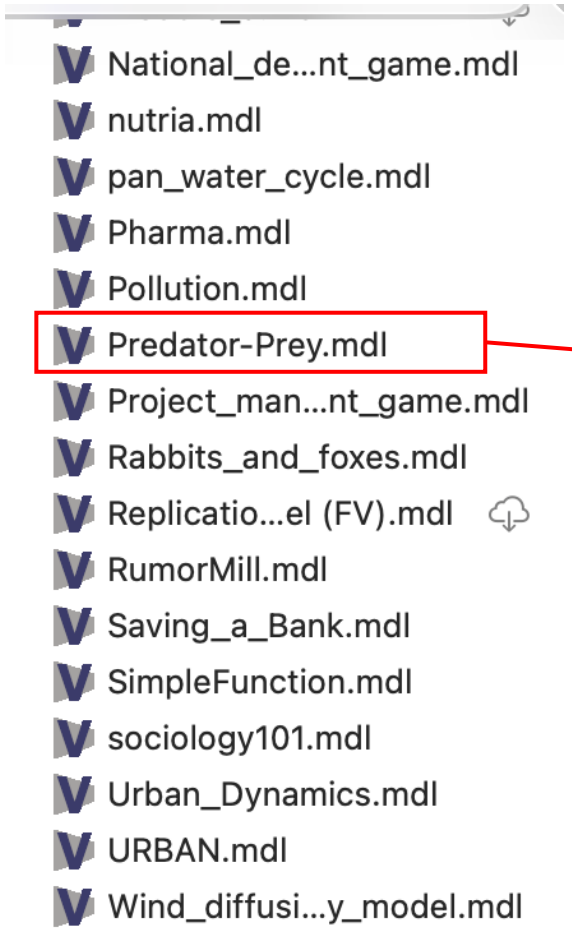
Workflow for Developing Our Agentic AI System



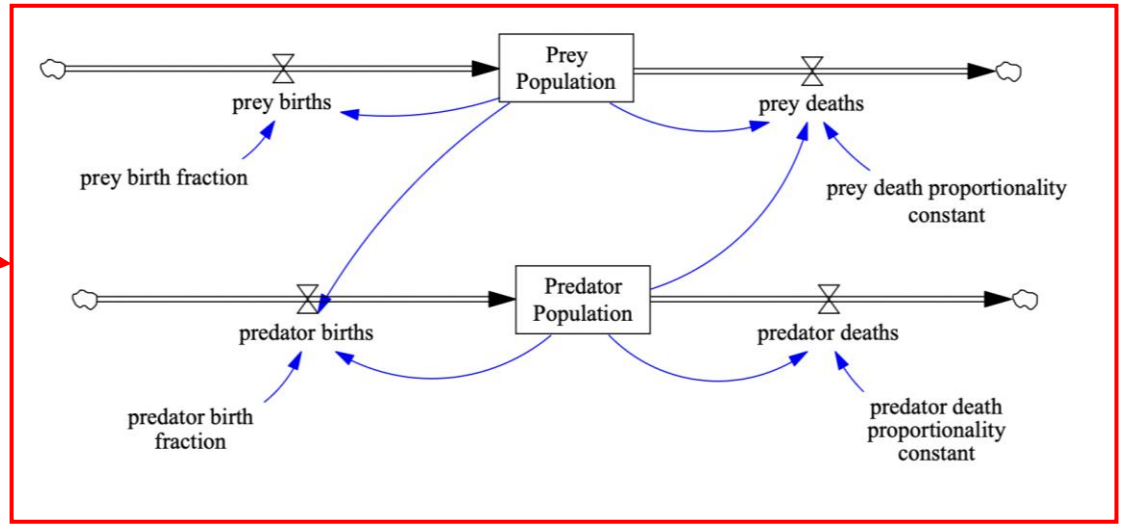
Architecture of the Proposed Agentic AI System

How Did We Make an AI System to Search and Retrieve Models?

We Collected 40 SD Models From Educational Platforms in Vensim (.mdl) Format



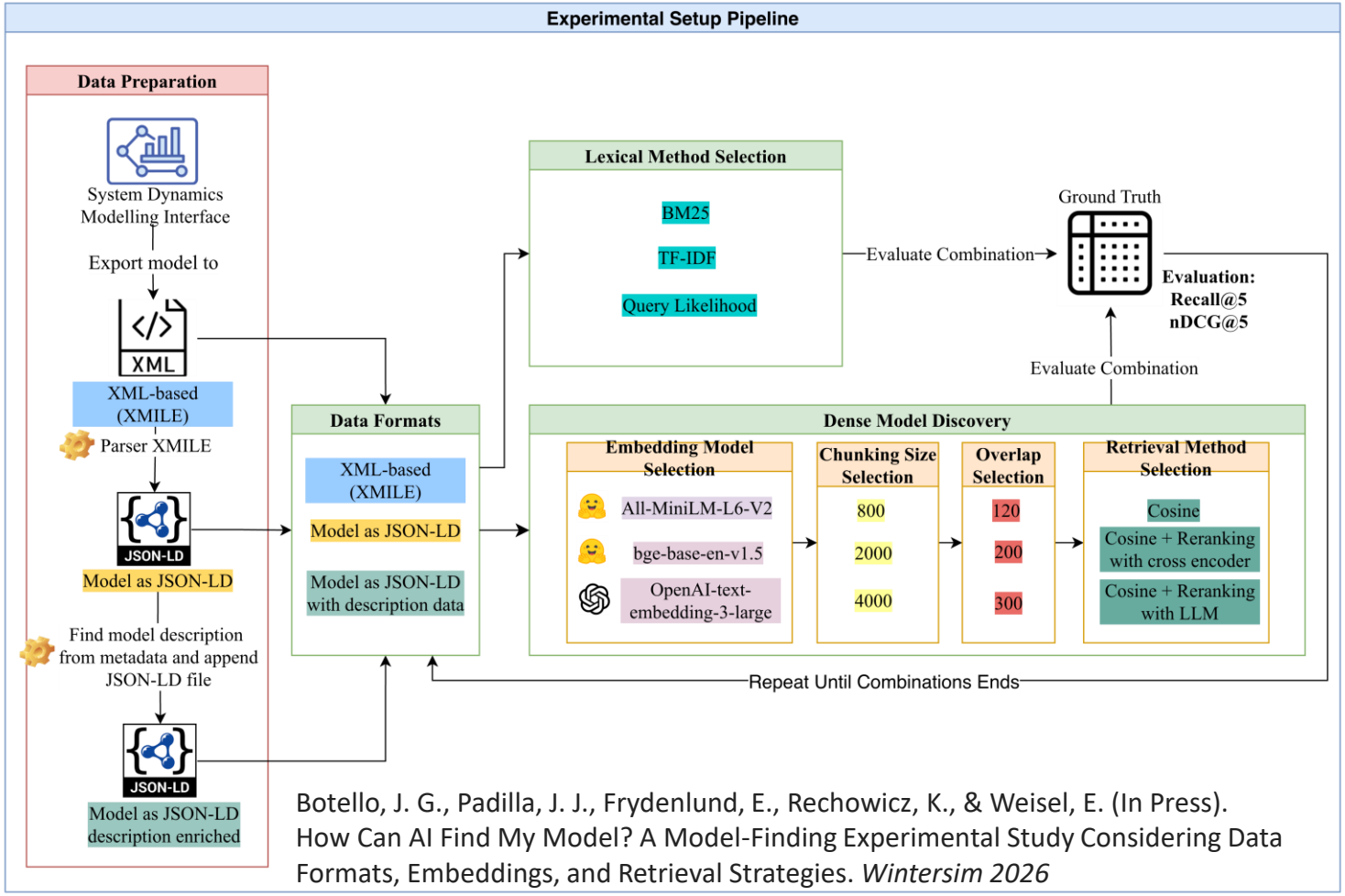
Subset of SD Models



We converted each .mdl file into its corresponding .xmile file. XMILE is an XML-based standard format designed to support interoperability among System Dynamics (SD) platforms.

```
<?xml version="1.0" encoding="utf-8" ?>
<xmile version="1.0" xmlns="http://docs.oasis-open.org/xmile
    <header>
        <product version="1.0" lang="en">Vensim</product>
        <vendor>Ventana Systems, Inc.</vendor>
        <created>
        </created>
        <modified>
        </modified>
        <name>
        </name>
        <caption>
        </caption>
    </header>
    <sim_specs method="RK4" time_units="seasons">
        <start> 0</start>
        <stop> 12</stop>
        <dt> 0.03125</dt>
    </sim_specs>
    <model>
        <variables>
            <stock name="Predator_Population">
                <units></units>
                <doc></doc>
                    <eqn>
                    15
                    </eqn>
                    <inflow>
                    predator births
                    </inflow>
                    <outflow>
                    predator deaths
                    </outflow>
            </stock>
            <stock name="Prey_Population">
                <units></units>
                <doc></doc>
                    <eqn>
                    100
                    </eqn>
                    <inflow>
                    prey births
                    </inflow>
                    <outflow>
                    prey deaths
                    </outflow>
            </stock>
            <aux name="predator_births">
```

In Our Preliminary Work, We Conducted a Series of Experiments to Evaluate Different Data formats, Indexing Methods, Embedding Models and Configurations, and Ranking Strategies After Retrieval.



Botello, J. G., Padilla, J. J., Frydenlund, E., Rechowicz, K., & Weisel, E. (In Press). How Can AI Find My Model? A Model-Finding Experimental Study Considering Data Formats, Embeddings, and Retrieval Strategies. *Wintersim 2026*

Human-Curated Ground Truth (15 Questions With Ranked Candidates)


```
{
  "query_type": "direct",
  "query": "Which simulation models represent population or resource dynamics using explicit stock-flow
  structures with processes such as growth, depletion, consumption, harvesting, or competition?",
  "relevant": ["Fish_bank", "Goats_and_grass", "nutria"],
  "graded_relevance": {
    "Fish_bank": 3,
    "Goats_and_grass": 2,
    "nutria": 1
```

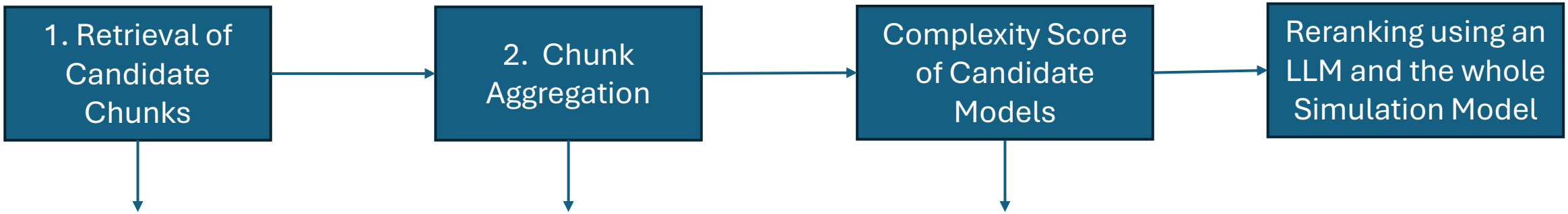

Results

Query Type	R@5	nDCG@5
Direct	0.59	0.63
Abstract	0.29	0.22
Ambiguous	0.17	0.20

Results from preliminary Work (Best Configuration is showed)
Json-LD + OpenAIEmbedding + 4000 + 300 + LLM

- **Recall@5 (R@5):** Measures the proportion of relevant models retrieved among the top 5 results.
- **nDCG@5:** Measures how well relevant models are ranked within the top 5 results, giving higher scores when more relevant models appear first.

Our Preliminary Work Informed the Design of Our Agentic AI System While Highlighting Areas for Improvement, Such as the Need to Improve Our Retrieval Strategy



$$sim(q,d) = \frac{\boldsymbol{q} \cdot \boldsymbol{d}}{\|\boldsymbol{q}\|\|\boldsymbol{d}\|}$$

Given a query $\boldsymbol{q}$ and a chunk $\boldsymbol{d}$, relevance is computed using cosine similarity, where $\boldsymbol{q}$ and $\boldsymbol{d}$ denote the embedding vectors of the query and chunk, respectively

$$S_D = \frac{\sum_{i=1}^{m} \alpha^{i-1} s_i}{\sum_{i=1}^{m} \alpha^{i-1}}$$

where $\mathbf{S_D}$ is the overall retrieval score for simulation model **D**, $\mathbf{s_i}$ is the cosine similarity of the **i-th** highest-ranked chunk associated with that model, $\mathbf{m}$ is the number of retained chunks for the model, with $\mathbf{m \leq 3}$, and $\boldsymbol{\alpha = 0.7}$ is a decay factor that gives progressively greater importance to higher-ranked chunks.

$$C_j = 3N_{\{s,j\}} + 2N_{\{f,j\}} + N_{\{a,j\}} + 0.5N_{\{e,j\}} + 0.25N_{\{c,j\}}$$

$$P_j = 100\left(1 - \frac{C_j - C_{min}}{C_{max} - C_{min}}\right)$$

where $\mathbf{C_j}$ is the structural complexity score of candidate model **j**, and $\mathbf{N_{\{s,j\}}}$, $\mathbf{N_{\{f,j\}}}$, $\mathbf{N_{\{a,j\}}}$, $\mathbf{N_{\{e,j\}}}$, $\mathbf{N_{\{c,j\}}}$ correspond to the numbers of stocks, flows, auxiliaries, equations, and estimated structural connections, respectively.

$\mathbf{C_j}$ is transformed into a relative simplicity score P_j.

We Used the Same Ground Truth Developed in Our Preliminary Work to Evaluate Model Search and Retrieval.

Query Type	R@5	nDCG@5
Direct	0.59	0.63
Abstract	0.29	0.22
Ambiguous	0.17	0.20

Results from preliminary Work

Query Type	Mean R@5	Mean nDCG@5
Direct	0.80 +21%	0.72 +9%
Abstract	0.27 -2%	0.32 +10%
Ambiguous	0.15 -2%	0.39 +19%

Results from Current Work (Same Configuration From Preliminary Work Was Used)

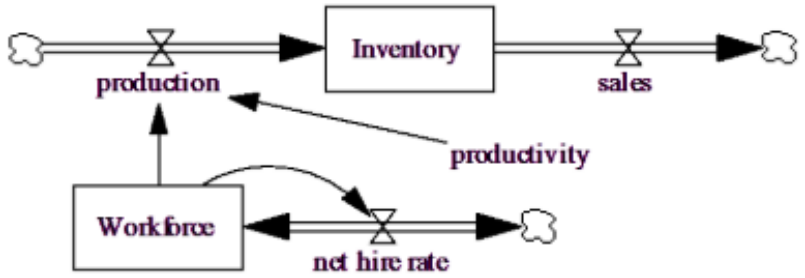
Insights

- **Results Highlights The Importance of the Quality of Question and Context**
- **Adding Descriptions to Models May Improve Retrieval**

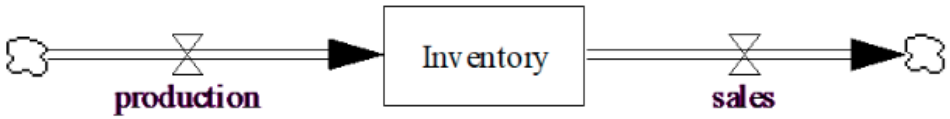
How Did We Assess Model Composition?

We Evaluated AI-Driven Model Composition Through Two Case Studies with Increasing Reasoning Complexity

Case Study 1: Structural Recompositing



Model

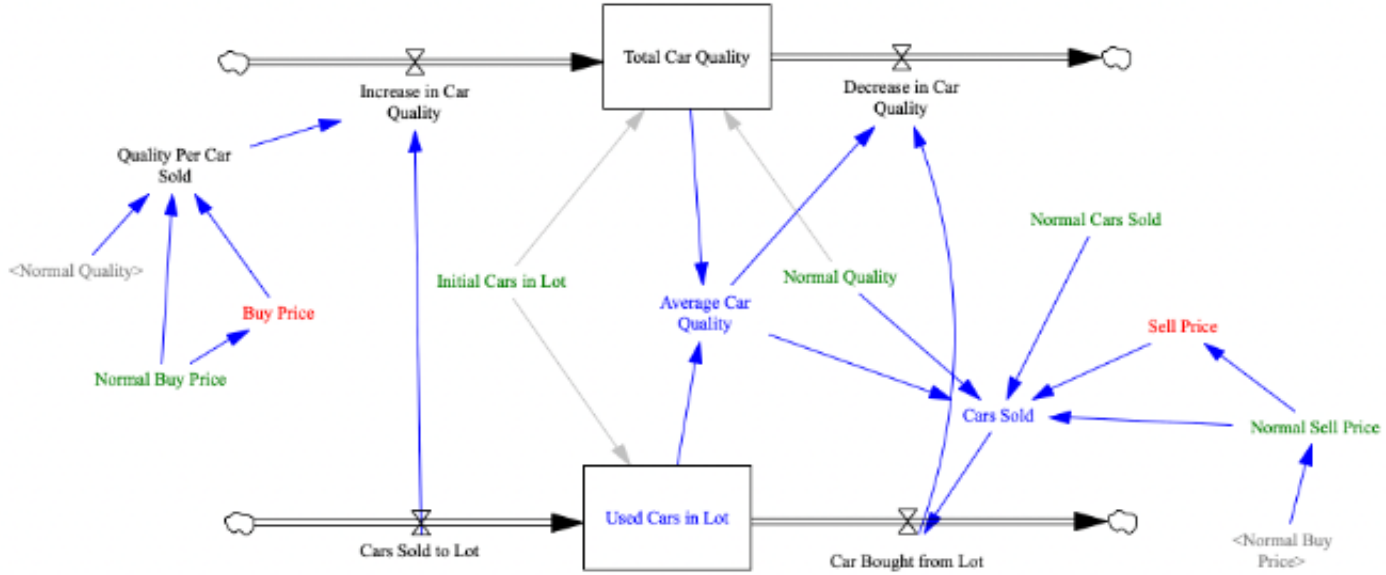


Module 1



Module 2

Case Study 2: Cross-Model Composition



Module 1



Module 2

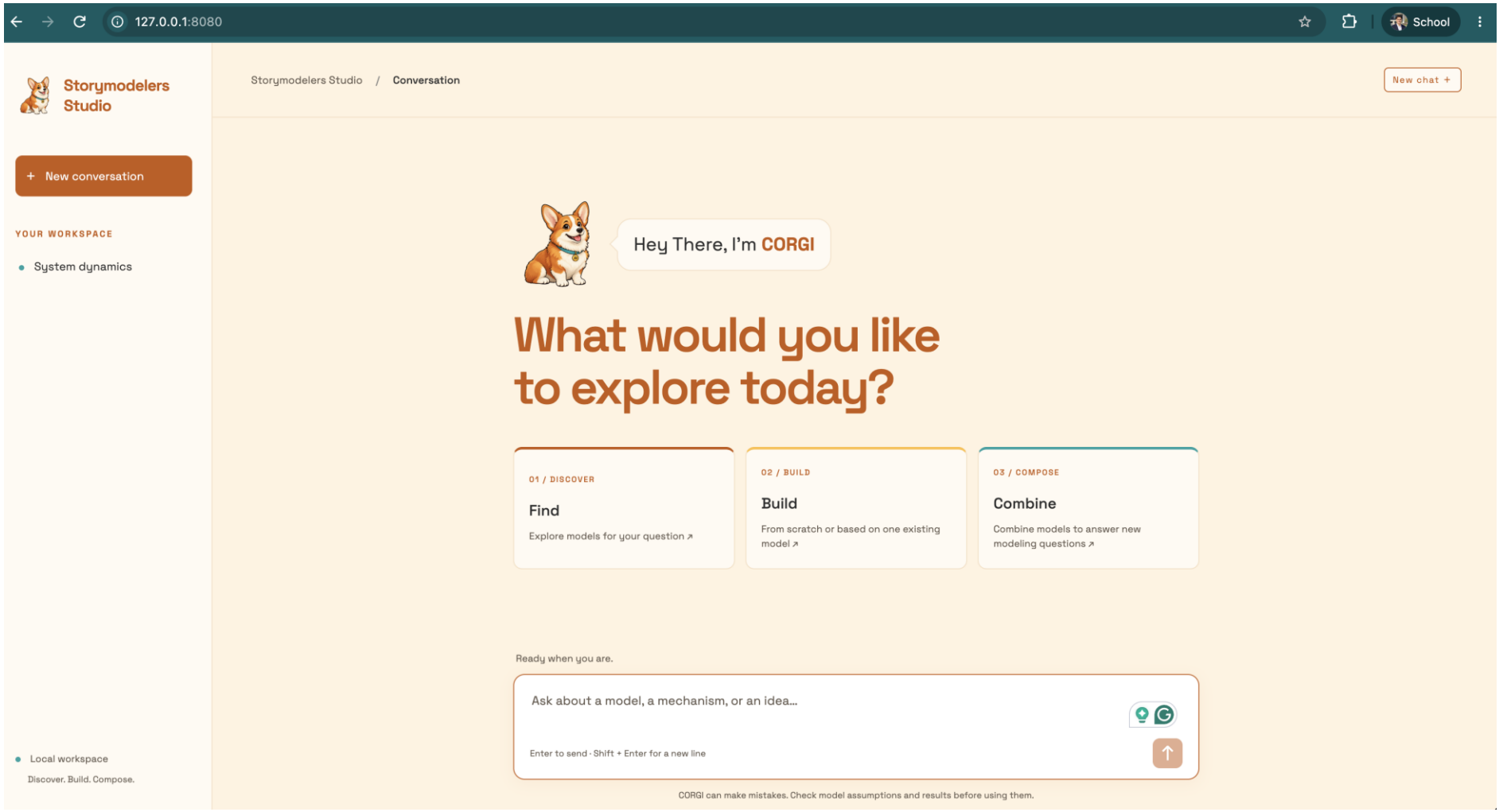
We Designed a Structured Prompt Using Chain-of-Thought (CoT) Prompting Strategies to Guide the Composition Agent in Breaking Down the Task into Multiple Steps.

1. Interpret the composition objective
2. Assess compatibility
3. Propose connections between models
4. Document provenance and adaptations
5. Record assumptions and unresolved incompatibilities,
6. Incorporate external evidence when provided
7. Generate the composed model as schema-constrained JSON-LD
Generate Visualization via .mdl File

Composition Results

Dimensions	Qualitative Assessment	Case study 1	Case Study 2	Case Study 2.1
Structural Composition	Yes / Partially / No	Yes	Partially	Full
Semantic Consistency	Yes / Partially / No	Yes	Partially	Full
Modeling Question Coverage	Full / Partial / None	Full	Full	Full
Unit Consistency	% passing unit check	100%	100%	100%
Equation Correctness	% correct equations	100%	88.9%	96.7%
Simulation Executability	Success / Failure	Success	Failure	Success

We Developed an Interface of Our Agentic AI systems to Continued Working on Scaling Up Our Approach to Consider More Realistic Model Composition Scenarios.



We Tested Our System Using a Realistic Combat Scenario Developed by [2]

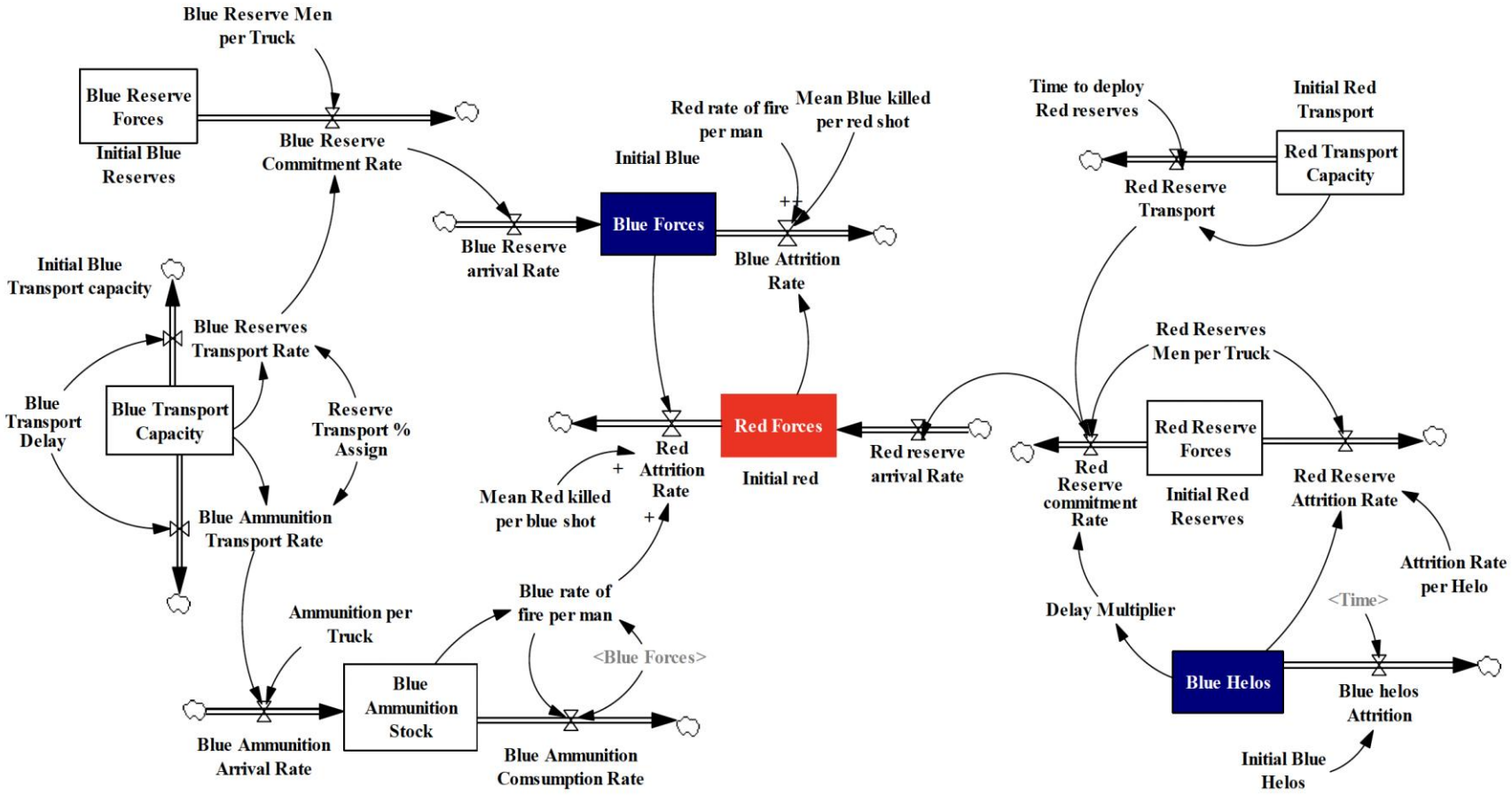


Figure 14. Combat model

[2] Bull, O. M. (2010). System Dynamics Applied to Combat Models (Lanchester Laws). Universidad Andres Bello. https://proceedings.systemdynamics.org/2010/proceed/papers/P1004.pdf

We Modularized the Model by Dividing It into Smaller Subcomponents and Used the Scenario Narrative Provided by [2] to Search for, Retrieve, and Combine the Components.

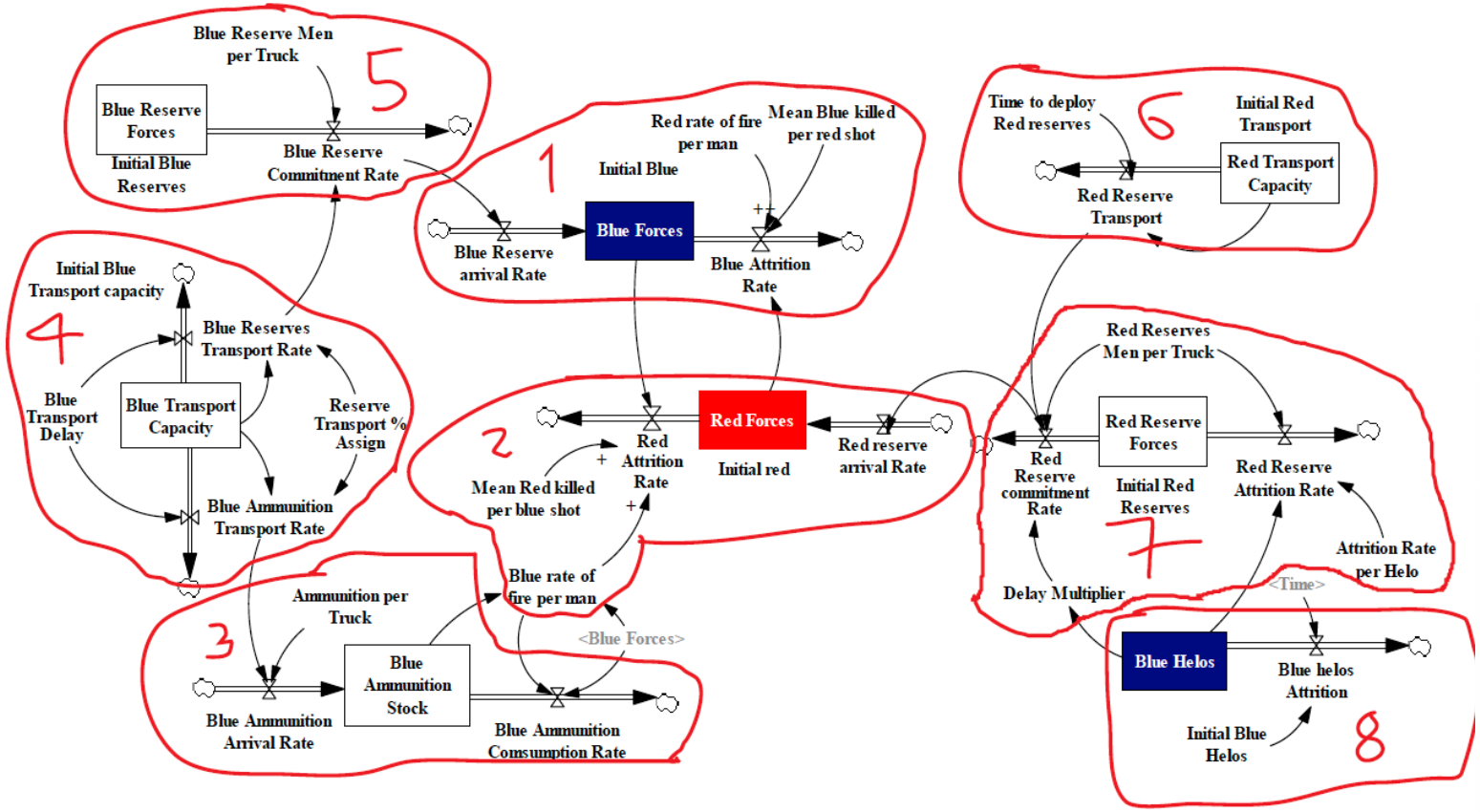


Figure 14. Combat model

A scenario for a simple land combat model

We will simulate the combat between two opposing forces, the Blue and the Red [Coyle, 1996]. The Blue forces consists of 1200 initial men in the front line, with their complete ammunition stock; 2000 soldiers in Reserve, with their ammunition ready to be transported in 30 trucks to be shared between reserve forces and ammunition. It should be noted that the Blue Reserve forces only have a limited transport capability, to share among the reserve soldiers and their ammunition. Also the Blue forces have 20 armored helicopters to be employed mainly to attack the Red reinforcements, in order to deter and to make difficult their advance to the front line.

The Red forces are only 1000 initial men in the front line of the battle (i.e. 200 soldiers less than the Blue combatant), and 2000 soldiers on reserve, to be committed at the request of the Tactical Commander. The Red Reserves have sufficient transport and ammunition for sustaining the whole combat.

We will be testing the following tactical decision:

The Red Tactical Commander will take advantage of his transport facilities and will commit the reserves as soon as the battle starts, accepting the heavy casualties that may be caused by the action of the Blue helicopters.

The Blue Tactical Commander is confident about the 20% difference of the initial forces for providing a winning edge, and the action of the helicopters for difficulty the Red reinforcement. Also been conscious of his limited transport capability will commit his reserves only if the Blue forces are matched by the Red ones, i.e. at a Red/Blue Ratio=1. In this event, he will give the orders to commit the reserves leaving to the Reserves Commander and the Logistic Commander to deal with the Transport Commander for sending to the front line the soldiers and their ammunition.

The following causal diagram shows this scenario. In the center is the Lanchester model previously explained with their corresponding additions to consider the reinforcements, the transport of men and ammunitions in the Blue side (in the Red side the ammunitions transport is omitted as no new information will provide), and the action of the Blue Helicopters against the Red reserves.

[2] Bull, O. M. (2010). System Dynamics Applied to Combat Models (Lanchester Laws). Universidad Andres Bello. https://proceedings.systemdynamics.org/2010/proceed/papers/P1004.pdf

Final Insights

- The system partially identifies the required modules. However, it proposes the missing components during composition.
- The system demonstrates knowledge of System Dynamics concepts and model structures.
- A key challenge lies in how model information and requirements are presented to the AI system.
- Empirical results highlight that AI can accelerate the modeling process, but human-in-the-loop involvement is always required, as AI and humans may interpret things differently.
- A potential path is to work on a prompt template to improve how we ask modeling questions.
- Future work includes augmenting datasets and assessing scalability.

Thank You