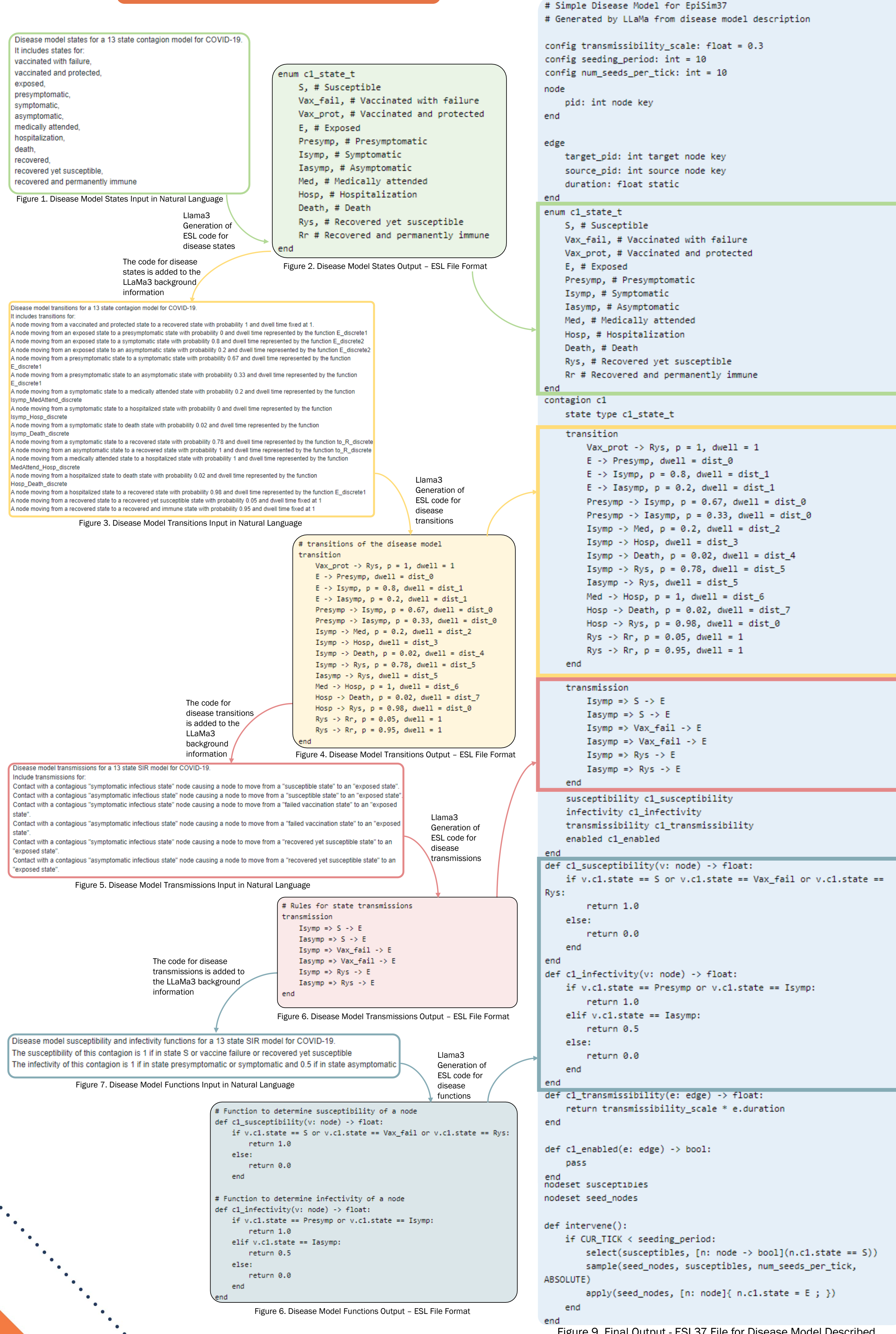


EpiGen: Generating HPC Epidemic Simulators using Natural Language by leveraging Large Language Models

Libby Trainum
Mentor: Parantapa Bhattacharya

- Creating efficient high-performance computing (HPC) epidemic simulators can be difficult for researchers
- Large language models (LLM) are good at parsing complex system descriptions from natural language
- Episim37 is framework for writing HPC epidemic simulators that run on top of social contact networks
- EpiGen is a system for creating HPC epidemic simulators using Episim37 and LLMs
- EpiGen takes in natural language descriptions of compartmental disease model and generates an Episim37 simulator using the ESL37 domain-specific language
- Use of natural language reduces technical barriers for researchers for creating efficient epidemic simulations and accelerates the setup phase of epidemiological research

Simulator Generation Process



Episim37

- High performance simulation framework for modeling epidemics on top of social contact networks using custom domain-specific language
- Significant advancement over its predecessor, EpiHiper, with ~30x efficiency and ~15x fewer computational resources (for whole US simulations)
- Allows comprehensive analysis and modeling of disease spread
- Easily adaptable to a wide range of epidemic scenarios and hypotheses
- Powerful tool for researchers and public health officials

ESL37 Domain-Specific Language

- ESL37 is a domain-specific language for specifying epidemic simulations for Episim37
- Describes 2 components, the social contact network and the contagion model
- Each node in the social contact network represents a member of a population
- Edges in the social contact network represent interactions between nodes where disease transmission can occur
- Disease states represent various phases of disease progression
- Transitions define rules for moving from one disease state
- Transmissions define disease spread due to contact between nodes
- Allows for the definition of interventions, like seeding infections or the simulating of public health responses

Large Language Models

- Large language models leverage deep learning techniques
- For this work, we used LLaMa3 and Gemma
- EpiGen translates natural language descriptions into ESL37 using LLaMa3
- Provides an incremental method to generate the disease simulation
- Simplifies setup and adjustment of epidemic models
- Architecture and training on example files ensure accuracy
- Further training and refinement for broader array of diseases and scenarios is possible—ongoing learning process improves utility and precision over time
- Integration makes simulation file generation efficient, accessible, and responsive