Internship Proposal

TITOLO:

Counterfactual What-If Sepsis Advisor

DESCRIPTION:

Sepsis is one of the leading causes of death in intensive care units and is characterized by a dysregulated host response to an infection that leads to organ dysfunction.

According to Sepsis-3, sepsis is defined as:

"infection + organ dysfunction (SOFA increase ≥ 2)."

Early diagnosis is essential to improve clinical outcomes, and in recent years Artificial Intelligence has shown increasing potential in identifying patients at risk.

Having the possibility to perform counterfactual predictions would allow clinicians to explore how a patient's trajectory might evolve under different clinical decisions or recommendations, providing valuable support for personalized treatment strategies and decision-making.

The objective of this project is to build a proof-of-concept system to address the gap. The system allows a clinician to type a hypothetical intervention in plain language – "what if I increase the vasopressor dose?", "What if I delay antibiotics?" – and receive three things in return: a data-driven prediction of how the patient's trajectory would change over the next 6 hours, a symbolic check of that prediction against formal clinical rules, and a natural-language explanation that is explicitly grounded in both.

The architecture combines three components that together instantiate a neuro-symbolic AI pipeline. A sequence model trained on MIMIC-IV handles the trajectory forecasting: it takes the patient's recent clinical history and the hypothetical intervention as input and predicts key outcomes at t+6h. A Logic Tensor Network (LTN) acts as a symbolic safety gate: it evaluates the predicted trajectory against clinical rules derived from sepsis guidelines and returns a satisfaction score for each, flagging physiologically dangerous trajectories before they reach the clinician. MedGemma serves as the language interface for at both ends of the pipeline: it parses the clinician's free-text query into a structured intervention specification, and at the other end it generates the narrative explanation conditioned on the LSTM output and the LTN constraint scores.

OBJECTIVES

1. Train a forecast model on MIMIC-IV that predicts at t+6h for sepsis patients given a hypothetical intervention

2. Build an LTN constraint verifier encoding logical rules that reliably detects clinically unsafe predicted trajectories
3. Demonstrate that grounded MedGemma prompting reduces clinically unsafe statements vs. ungrounded prompting
4. Deliver a working demo of the full pipeline on at least three end-to-end patient scenarios

WORK PLAN

Phase 1 – Data and LSTM forecaster

1. **Access to MIMIC-IV data**: complete PhysioNet credentialing and data use agreements
2. **Cohort extraction**: filter sepsis-3 patients and export hourly time series
3. **Exploratory data analysis**: explore variable distributions, missing value patterns, temporal coverage
4. **Feature engineering**: build 12-hour input windows, split dataset in train/validation/test
5. **LSTM training and evaluation**

Phase 2 – LTN constraint verifier

1. **Clinical rule study**: study sepsis-3 definition and read Surviving Sepsis Campaign guidelines. Extract threshold values for logical rules
2. **Logical rule implementation**: implement shock evidence, organ dysfunction, antibiotic urgency, treatment guidelines, etc., as predicates objects and combine them using logical operators
3. **LTN training and evaluation**: run LTN verifier and compute accuracy, false positive rate, false negative rate
4. **Pipeline integration**: integrate LSTM forecasting with LTN verifier

Phase 3 – MedGemma integration and experiment

1. **MedGemma setup**
2. **Template design**: write structured evidence template (system prompt + patient state + predicted trajectory + LTN scores)
3. **Scenario selection**: select a subset of patients from MIMIC-IV covering: early shock, late shock, AKI, delayed antibiotics, etc.
4. **Output generation**: for each scenario generate three outputs: a) grounded, b) LSTM-only, c) ungrounded

EXPECTED RESULT

The expected outcome is a clinician-interactive prototype capable of receiving a free-text what-if intervention query, predicting the patient's trajectory under that hypothetical, and producing a grounded recommendation supported by formal constraint verification and natural-language explanation anchored in both the trajectory evidence and the symbolic rule that was checked.

The project's main contribution lies not only in improving the safety of LLM-generated clinical

narratives, but also in defining a more robust and explainable method for counterfactual reasoning in the ICU.

TECHNOLOGIES INVOLVED

- Python
- Dataset open source (e.g., MIMIC-IV)
- LLMs, deep learning models (e.g., LSTM, Transformer), neuro-symbolic frameworks (e.g., LTNs)

RESOURCES

- MIMIC-IV
- Sepsis-3 definition
- Logic Tensor Network paper – PyTorch implementation
- Surviving Sepsis Campaign guidelines

Contact:

Prof. **Francesco Zanichelli** (francesco.zanichelli@unipr.it)
Dott. **Fabrizio De Santis** (fabrizio.desantis@unipr.it)

Distributed Systems Group, Dipartimento di Ingegneria e Architettura