SafeSim: Neuro-Symbolic Medical Text Simplification

Arunbh Yashaswi, Abhishek Rithik Origanti,

Matheshwara Annamalai Senthilkumar

Addressing Hallucination in Medical Text

SafeSim combines LLM fluency with symbolic verification.
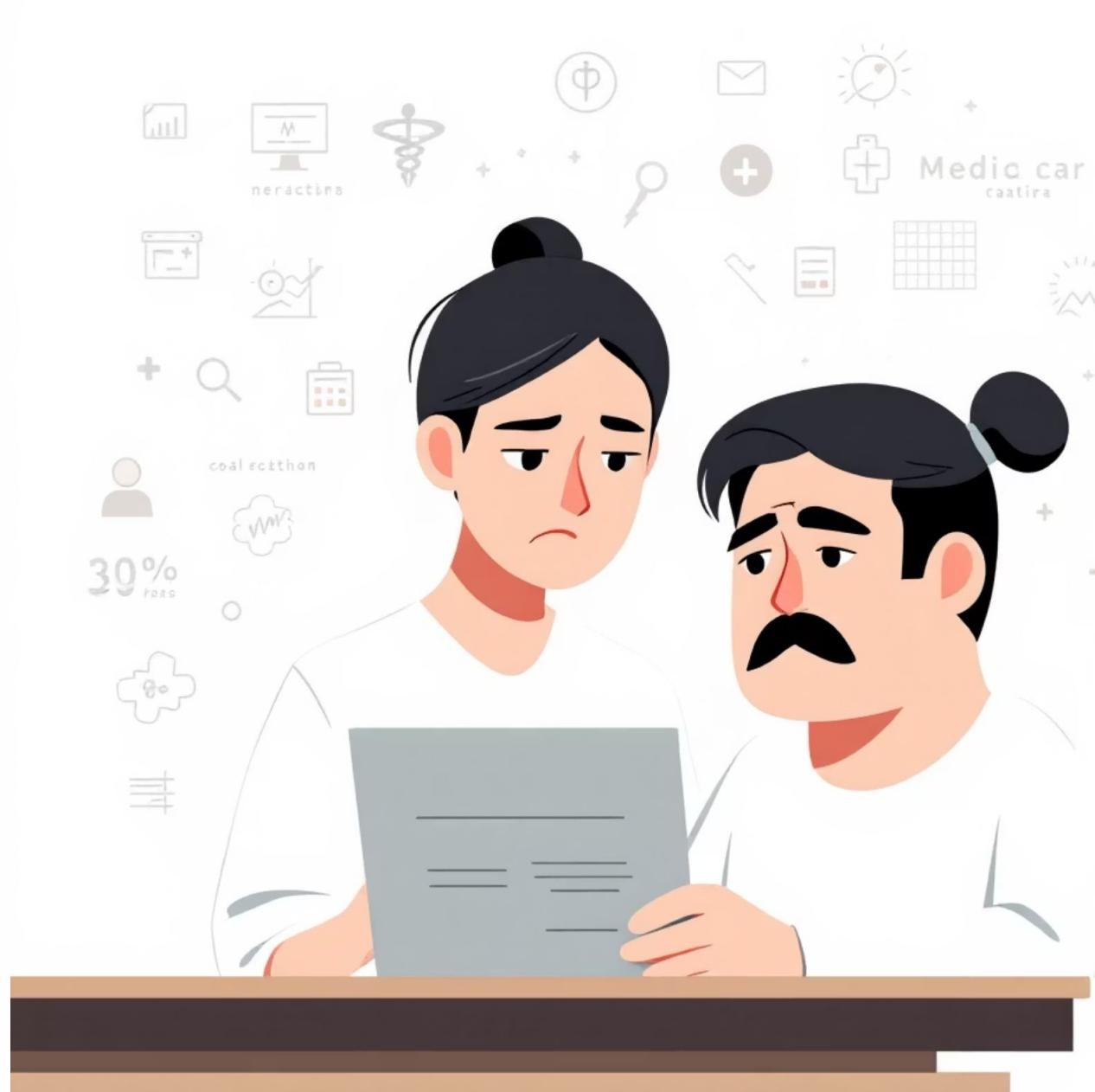
100% Entity Preservation
Ensures critical medical entities (dosages, vitals) are never altered.

Maintains Readability
Simplifies complex medical jargon without compromising clarity.

Novel Approach
Outperforms neural baselines like BART and T5 in safety.

The Challenge: Patient Understanding

Complex medical jargon creates barriers, leading to poor adherence and adverse outcomes.

→ **Complex Jargon**

Discharge summaries and clinical notes are often difficult to understand.

→ **Safety Risks**

LLMs can hallucinate, changing "50mg" to "5mg" with life-threatening consequences.

→ **Need for Precision**

Automated simplification requires deterministic fact preservation.

SafeSim: A Neuro-Symbolic Framework

Ensuring deterministic fact preservation without sacrificing readability.

01

Symbolic Entity Extraction

Extracts critical entities (dosages, medications, vitals) using symbolic methods.

0
2
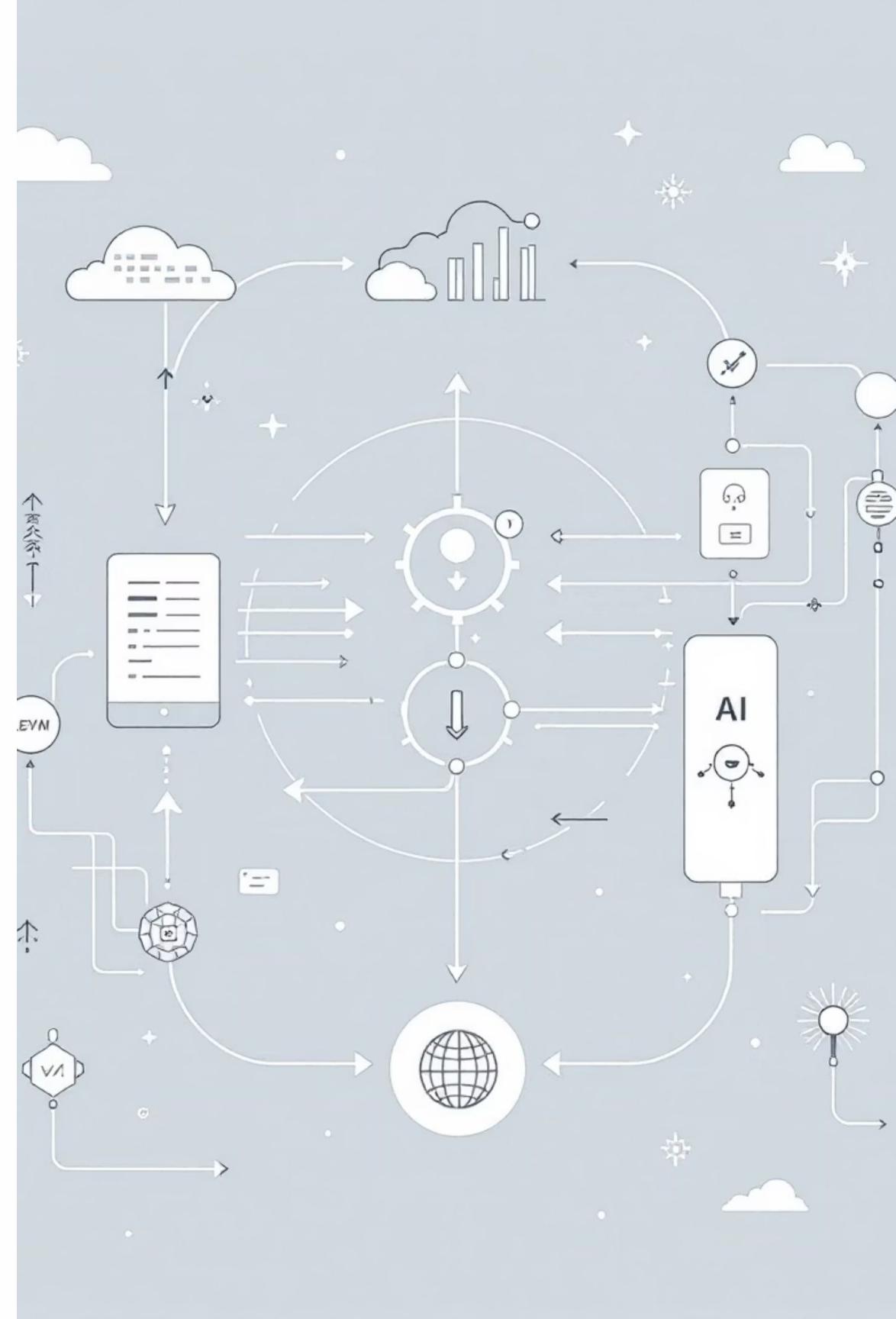
Neural Simplification

Simplifies text via LLMs, with extracted entities as explicit constraints.

0
3

Deterministic Verification

Verifies outputs using logic checks; failed verifications trigger refinement.

Evaluation Dimensions

SafeSim assessed across five key areas for robust performance.

System Design
Modular pipeline: NER, LLM rewriting, logic verification.

Method Comparison
SafeSim vs. BART and T5 baselines for safety and quality.

Data-Centric NLP
Performance across diverse medical specialties and terminology.

Error Analysis
Qualitative audit of failure modes, distinguishing critical safety violations.

Responsible NLP
Ethical implications, bias, and human-in-the-loop deployment.

Problem Formulation: Constrained Translation

Maximizing readability while ensuring strict safety constraints.

Source Sequence (S)

Complex medical tokens.

Target Sequence (T)

Simplified, readable text.

Critical Entities (ES)

High-risk categories: Dosage, Medication, Frequency.

Safety Constraint (C)

Semantic invariance between S and T for critical entities.

SafeSim uses a **neuro-symbolic approach** to simplify medical discharge summaries while guaranteeing the preservation of critical facts (medications, dosages, vitals).

📄 Input: Medical Text

Paste discharge summary or clinical note:

> Give 10 units insulin subcutaneously b.i.d before meals.

🔄 Simplify Text

✨ Output: Simplified Text

✅ **SAFE**

Verification Score: 100%

Simplified text:

> Give 10 units insulin under the skin twice a day before meals.

> 🔍 Detailed Analysis

🔍 Detailed Analysis

Extracted Entities

Give 10 units insulin subcu b.i.d before meals.

Entity List:

10 units (DOSAGE)

insulin (MEDICATION)

b.i.d (FREQUENCY)

Verification Detai

```
{
  "is_safe" : true,
  "missing_entities"
  "modified_entities
  "warnings" : []
  "score" : 1
}
```

🏥 SafeSim: Safe Medical Text Simplification

SafeSim uses a **neuro-symbolic approach** to simplify medical discharge summaries while guaranteeing the preservation of critical facts (medications, dosages, vitals).

📄 Input: Medical Text

Paste discharge summary or clinical note:

Administer 10 units insulin subcutaneously b.i.d. before meals. Check blood glucose q.i.d. Target range 80-120 mg/dL.

🔄 Simplify Text

✨ Output: Simplified Text

⚠️ **NEEDS REVIEW**

Verification Score: 80%

Simplified text:

give 10 units insulin under the skin twice a day. before meals. Check blood glucose four times a day. Target range 80-120 mg/dL.

⚠️ Safety Alerts

⚠️ SAFETY ALERT: Critical entity 'Administer' not found in simplified text!

▽ 🔍 Detailed Analysis

Extracted Entities

Administer 10 units insulin subcutaneously b.i.d. before meals. Check blood glucose q.i.d. Target range 80- 120 mg /dL.

Entity List:

Administer (MEDICATION)

10 units (DOSAGE)

insulin (MEDICATION)

b.i.d. (FREQUENCY)

q.i.d. (FREQUENCY)

Verification Details

```
▼ {
    "is_safe" : true
  ▶ "missing_entities" : []
  ▶ "modified_entities" : []
  ▶ "warnings" : []
    "score" : 1
}
```

Model Information

Backend: rule-based
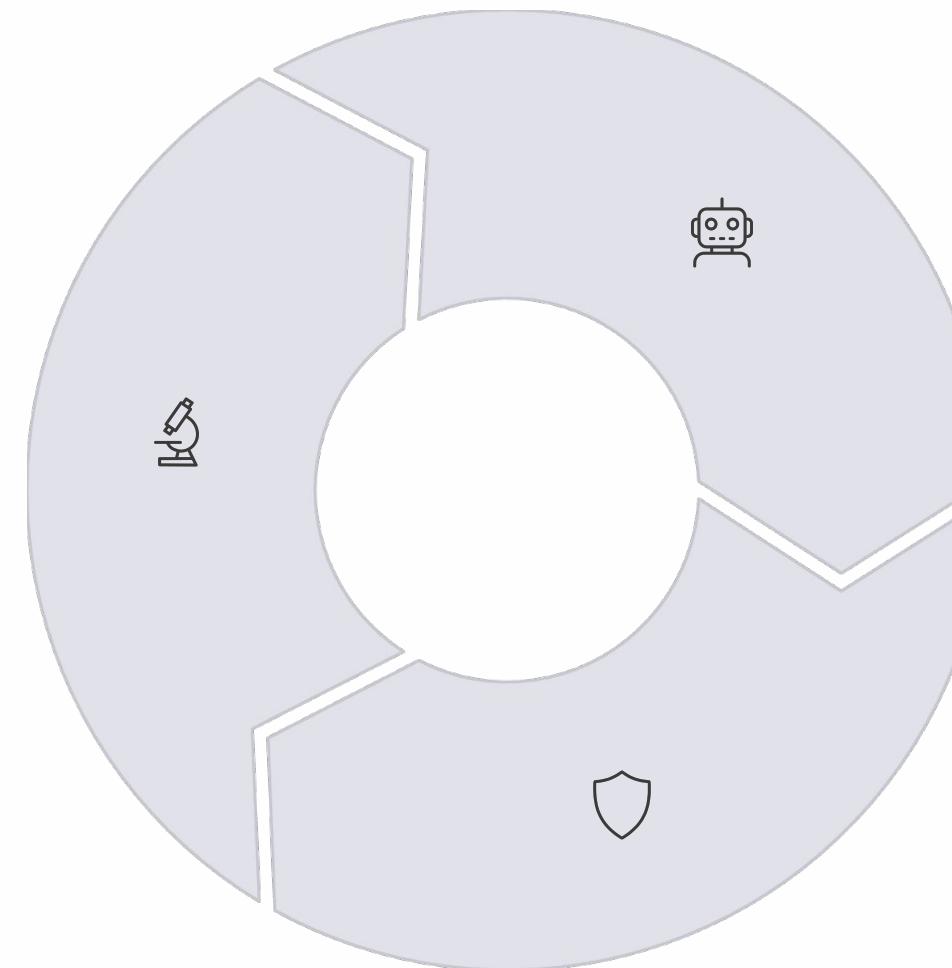
Strictness: high

Try SafeSim Demo

Let's Start!

SafeSim Architecture: Verify and Refine

A neuro-symbolic pipeline for robust medical text simplification.

Neural Simplification

LLM backend (GPT-4, Claude, BART) with entity injection.
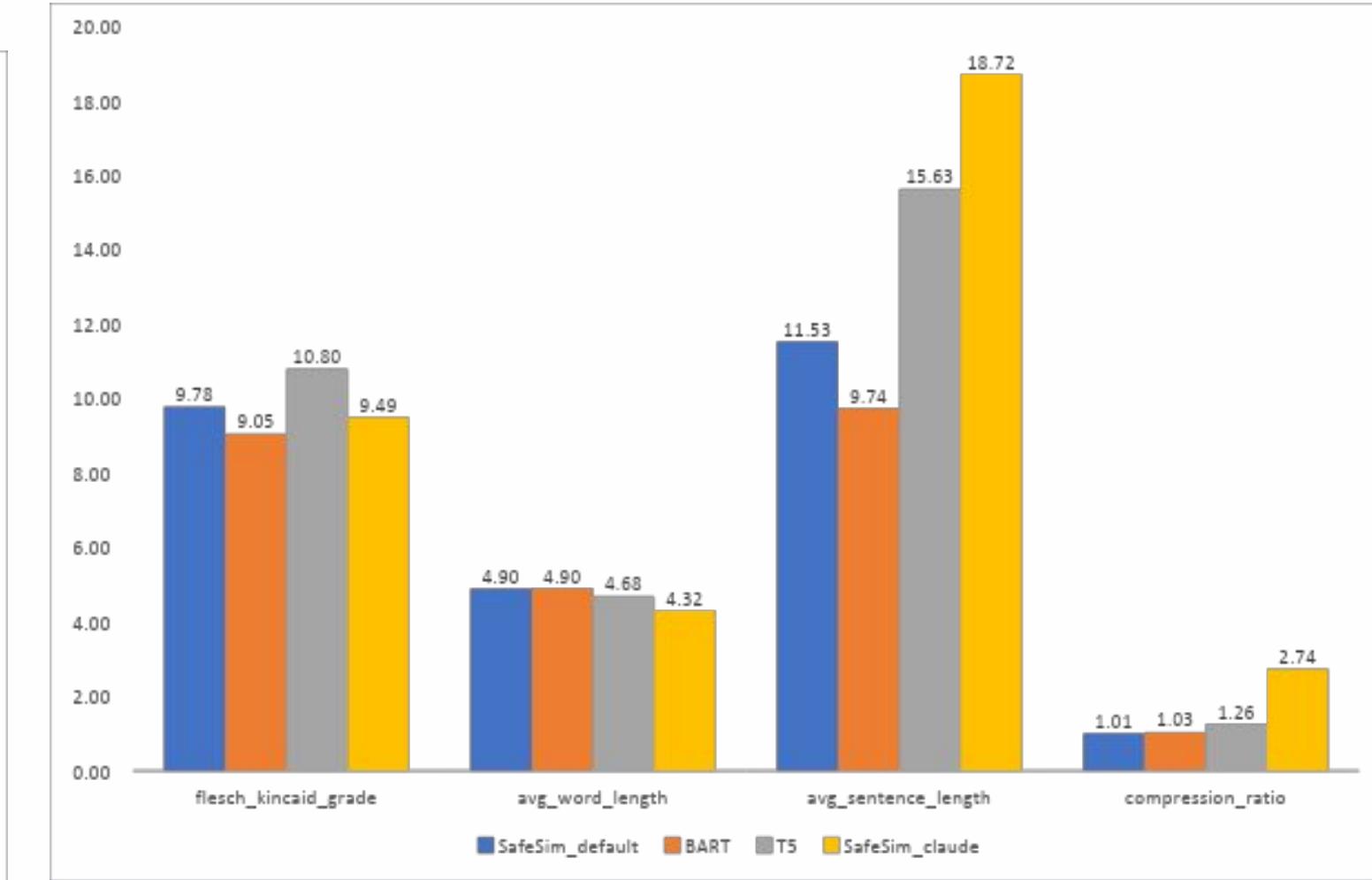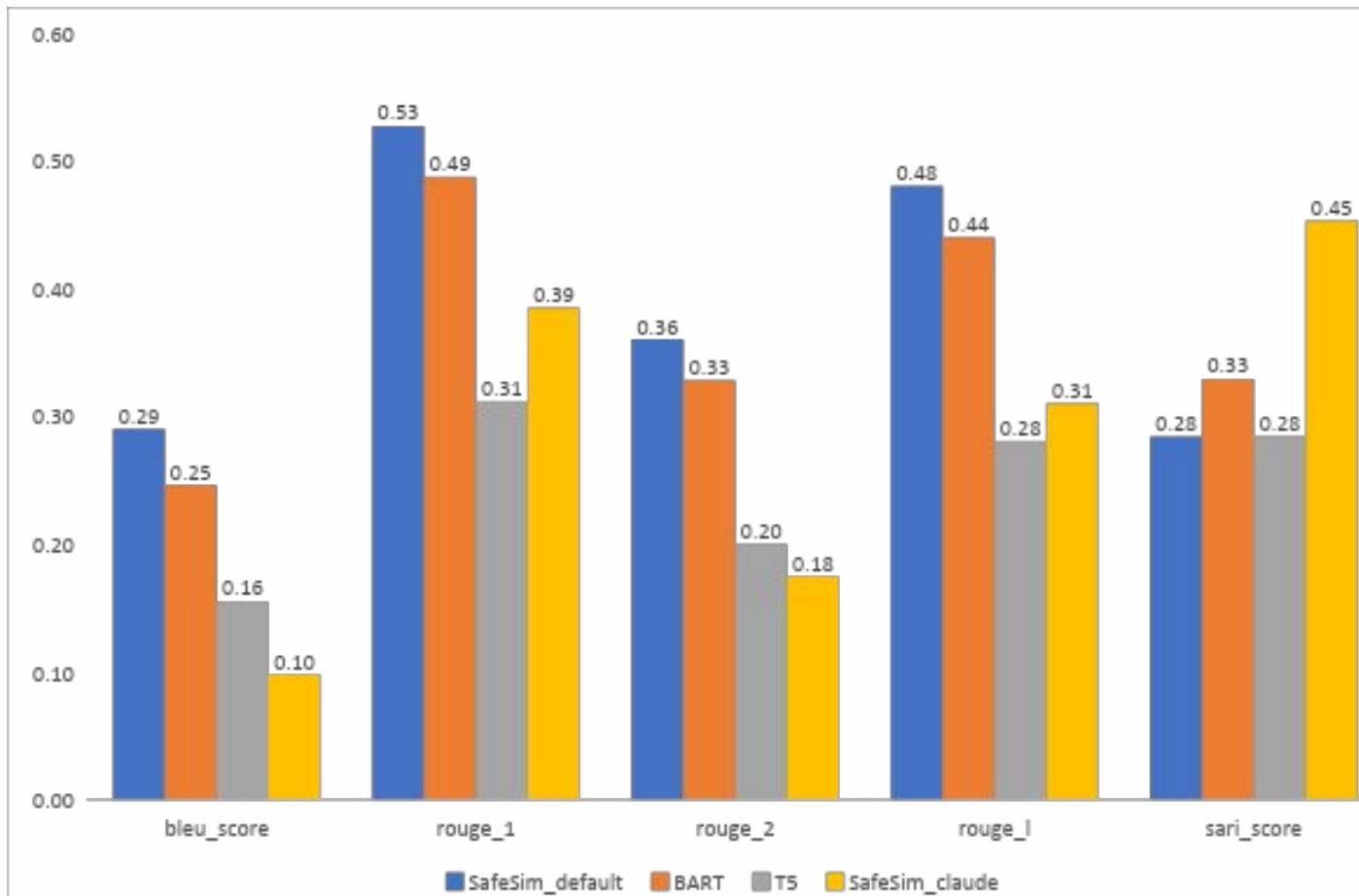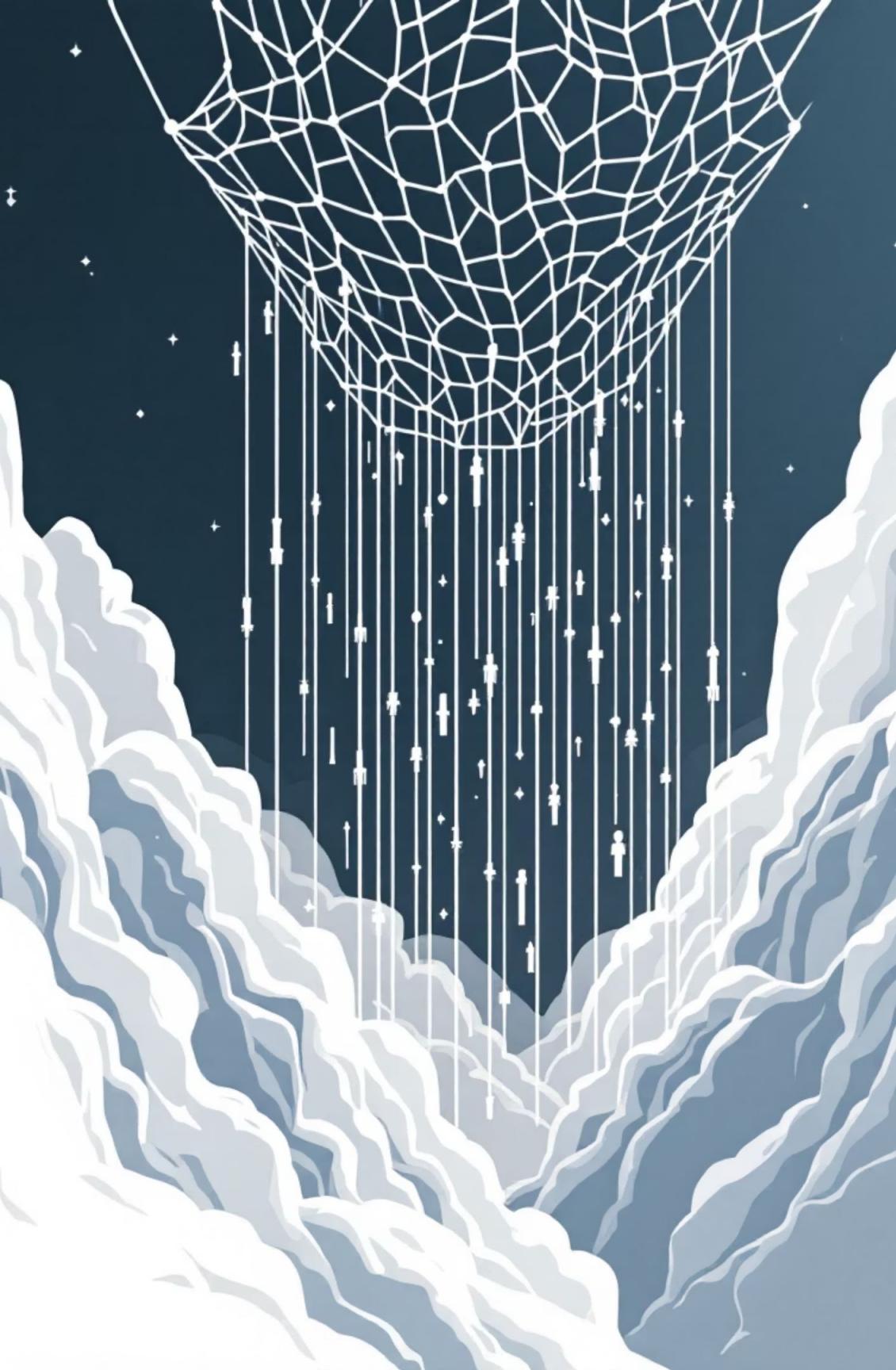
Symbolic Entity Extraction

Scispacy NER + Regex for structured and unstructured entities.

Deterministic Verification

Re-extract entities, set comparison, semantic relaxation, regeneration on failure.

Performance Metrics: SafeSim vs. Baselines

Key Findings & Discussion

SafeSim's neuro-symbolic approach ensures safety and effective simplification.

Simplicity-Overlap Trade-off

High SARI, low BLEU indicates aggressive, effective simplification, not just copying.

Neuro-Symbolic Safety

100% Entity & Dosage Preservation, eliminating stochastic risk of pure neural models.

Structural Readability

Merges fragmented shorthand into coherent sentences, improving patient comprehension.

Conclusion & Future Work

SafeSim solves hallucination in medical text simplification, setting a new standard.

Key Achievements

- 100% Entity Preservation Rate (EPR)
- Clinically safe LLM-based simplification
- Effective rewriting, not just text copying

Future Directions

- Expand entity coverage (e.g., conditional dosing)
- Multilingual support for global accessibility
- End-to-end optimization with RL feedback

THANK YOU