From Free-Text Follow-up Notes to Structured Actions

Extracting clinical instructions and timeframes Models in a controlled mini-world.

Motivating Use Case

The Need

Doctors write free-text instructions (e.g., "Do blood test in 2 weeks"). Systems need to know **what** action and **when** it's due for every patient.

The Value

Structured data allows for automatic compliance checks ("Did the patient do the test?"), overdue alerts, and population health dashboards.

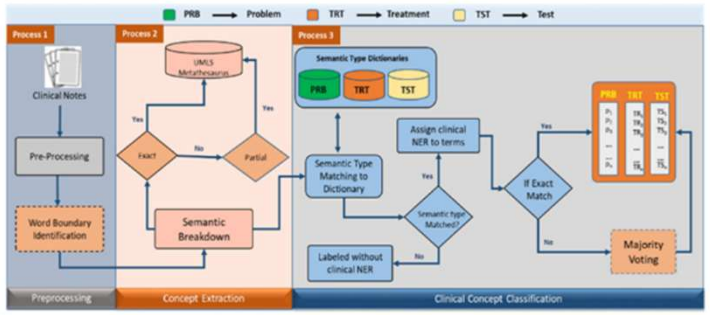
Current State

Instructions remain locked in text notes. Verification requires manual chart review. Existing tools often ignore these crucial instructional nuances.

Project Task: Instruction Extraction

- **Input:** A synthetic clinical note containing symptoms, history, and 1–3 natural language instructions (e.g., "Come back in 4 weeks").
- **Output:** For a fixed library of actions, predict:
  - Present (True/False)
  - Timeframe (Days)
- **Novelty:** Focus specifically on the "Action + Timeframe" extraction task using a controlled "mini-world" with perfect ground truth.

Conceptual illustration of extracting structured data from clinical notes



Synthetic Data Generation

Closed Instruction Library

We define a fixed set of actions (CBC, X-Ray, Glucose Log) to constrain the problem space.

Generative Process

We sample instructions first, then ask an LLM to "weave" them into a realistic clinical narrative. This ensures we have **perfect ground truth** labels for training and evaluation.

Why Synthetic?

It eliminates privacy concerns and removes human annotation error, allowing us to focus purely on the extraction capability of the NLP models.

Models & Methods: Baseline vs. Pipeline

(A) Baseline Approach

Free-form JSON List

The model reads the note and simply lists whatever actions it finds.

- "List instructions found..."
- Risk of hallucination or omission.
- Requires post-processing to map back to the library.

(B) Structured Pipeline

Per-Code Checklist

The model is forced to make a binary decision for every code in the library.

- "For code CHEST_XRAY: Is it present?"
- Explicit decision making.
- Produces a complete, structured matrix directly.

Example Extraction

Synthetic Clinical Note

"The patient presents for an outpatient visit with 3 weeks of intermittent headaches… Past medical history is notable for well-controlled hypertension… Assessment: primary headache disorder is most likely…

Plan: advise nonpharmacologic measures… ***Please schedule a follow-up visit with the same physician in 4 weeks*** *to reassess symptoms and review any test results."*

Structured Output

GROUND TRUTH

```
codes: ['FOLLOW_UP_GP']
time: {'FOLLOW_UP_GP': 28}
```

MODEL PREDICTIONS

Baseline ✓ 28 Days

Pipeline ✓ 28 Days

Both models correctly identified the action and converted "4 weeks" to 28 days.

Metrics & KPIs

1. Instruction Detection

Did the model find the right action? Measured via **Accuracy** and **F1-Score** on the binary present label.

2. Timeframe Quality

If the action was found, did the model understand "2 weeks" = 14 days? Measured via **Mean Squared Error (MSE)**:

$$\mathrm{MSE} = \frac{1}{N} \sum_{i=1}^{N} \left(y_{\mathrm{pred}} - y_{\mathrm{true}} \right)^2$$

3. Comparison

Does the structured pipeline offer more consistent detection and accurate timeframes compared to the baseline?