Xtensyon.

CASE STUDY

Extracting Structured Data from Clinical Notes Without Breaking Workflows

June 18, 2025 • Xtensyon Labs • 8 min read

> *Clinicians write for humans, not databases. This case study covers a practical pipeline for extracting problem lists and medication details while keeping review steps simple and safe.*

TL;DR

- Keep humans in the loop for low-confidence fields; do not force automation.
- Normalize abbreviations and templates before model work.
- Measure accuracy by patient safety impact, not just F1 score.
- Make corrections feed back into the system as training data and rules.

Executive Summary

We implemented a note-to-structure pipeline for healthcare teams who needed consistent fields such as diagnoses, meds, and follow-up plans. The system combined template recognition, controlled extraction prompts, and a review UI for borderline cases. The workflow was designed around existing chart review habits so clinicians did not feel they were doing extra admin work.

Why It Matters

Clinical notes are rich, but messy. Pulling data out without care can create patient safety risk and erode trust. A careful approach focuses on a small set of fields that matter, adds confidence-based review, and keeps provenance so staff can verify where each extracted value came from.

What We Built

- A preprocessing layer that normalizes abbreviations, section headers, and templated phrases.
- An extraction pass for high-priority fields with strict output schemas and validation.
- A review queue for low-confidence values, with one-click accept or correction.
- A feedback loop that turns recurring corrections into rules and test cases.

Observed Outcomes

- Better completeness for medication lists compared with manual entry alone.
- Lower clinician frustration by review only when the system is unsure.
- Cleaner downstream analytics once confidence and provenance were stored per field.

Implementation Notes

- Start with a limited scope. One clinic, one note type, a handful of fields.
- Validate outputs with schema checks before saving anything.
- Keep an audit trail that shows the exact source sentence for each extracted value.
- Use shadow mode first so teams can compare results without operational risk.

Governance & Risk

- Treat extracted data as clinical documentation. Apply retention and access policies.
- Prevent model prompts from containing identifiers when not needed.
- Define who can override extracted fields and how corrections are recorded.

Release Checklist

- Do extracted fields include confidence and provenance?
- Is there a review queue for low-confidence outputs?
- Are schemas validated before persistence?
- Do corrections feed back into rules and tests?
- Is scope limited to fields with real operational value?

Conclusion

You can extract useful structure from notes without turning clinicians into data clerks. The key is narrow scope, clear review triggers, and traceable provenance for every field.

Keywords

healthcare clinical notes information extraction EHR data quality workflow