

Saad Ahmed Qadeer  
**EXECUTIVE SUMMARY**

Automated Medical Transcript Processing & ICD-10 Extraction

---

## PROJECT OVERVIEW

This initiative addresses the operational bottleneck of converting unstructured medical narratives into structured billing data. By leveraging Natural Language Processing (NLP), the system automates the extraction of patient demographics and treatment plans, mapping them directly to standardized ICD-10 codes for insurance billing.

---

## THE BUSINESS CHALLENGE

Manual review of clinical transcripts is labor-intensive, costly, and susceptible to human error. These inefficiencies delay claim submissions and increase the risk of denials due to coding inconsistencies. As patient volumes grow, manual extraction becomes a non-scalable barrier to revenue cycle optimization.

---

## TECHNICAL SOLUTION

The solution is a Python-based pipeline integrating the OpenAI GPT-4o-mini API to process text in two distinct stages:

- **Structured Extraction:** The model parses narrative text to identify key entities (e.g., patient age, specific treatments) using function calling to enforce a consistent schema.
- **Code Mapping:** Extracted treatments are automatically correlated with accurate ICD-10 billing codes, leveraging the model's embedded medical knowledge.

---

## KEY RESULTS & IMPACT

- **Operational Efficiency:** Reduces processing time from hours to approximately 1-2 seconds per transcript.
- **Cost Reduction:** Operates at less than \$0.01 per transcript, significantly undercutting manual labor costs.
- **Data Integrity:** Delivers high-accuracy, structured CSV output compatible with existing Electronic Health Record (EHR) and billing systems.

---

## CONCLUSION

This project delivers a production-ready automation tool that transforms clinical documentation into actionable billing data. It offers immediate ROI through reduced administrative overhead and improved claim accuracy, providing a scalable foundation for future healthcare data analytics.