

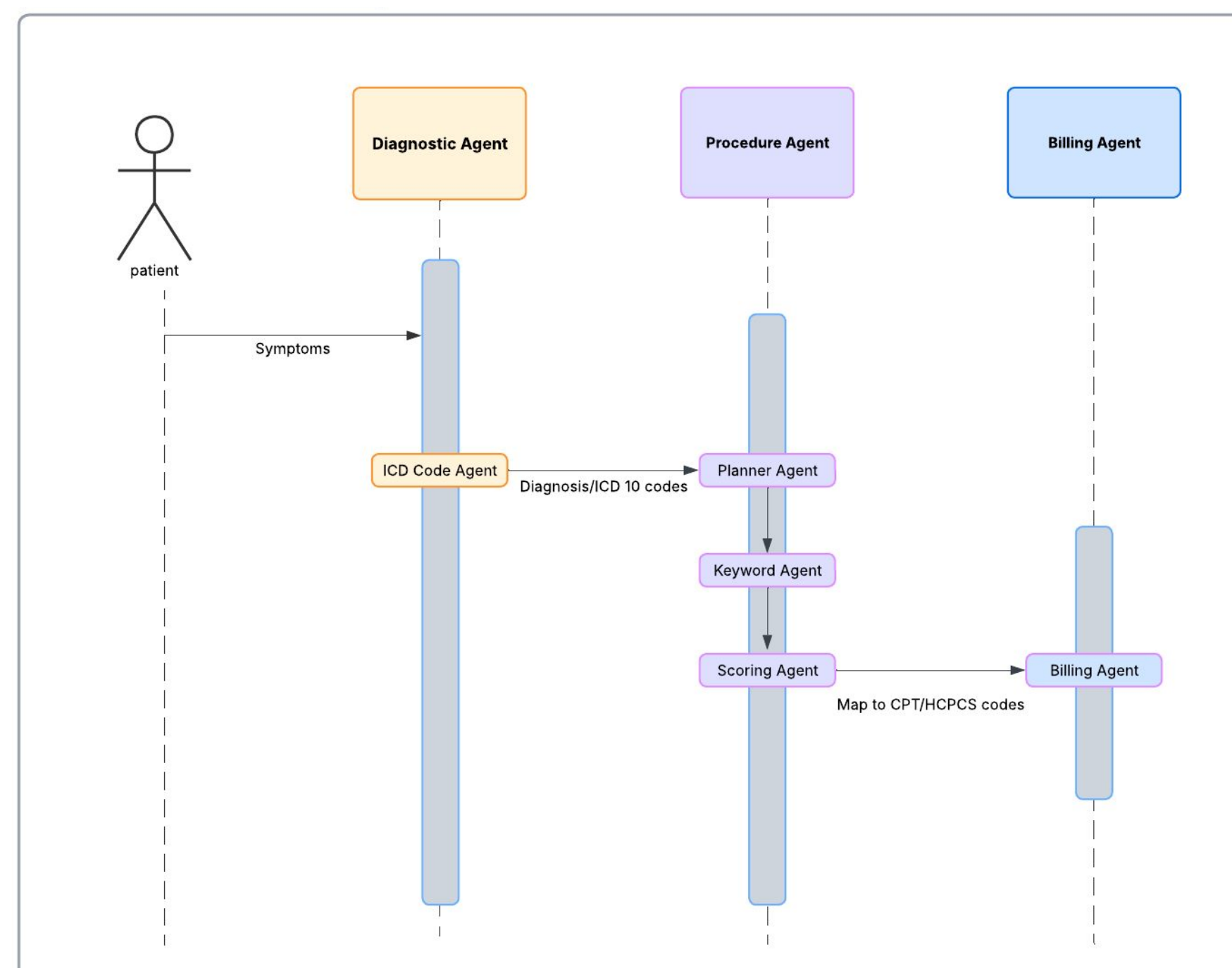
Making the Cost of Care Visible: A Multi-Agent Simulation of the Healthcare Administrative Pipeline

Motivation

Patients enter the healthcare system with symptoms but **little insight into how those symptoms will translate into diagnoses, procedures, billing codes, or insurance claim outcomes**. Existing clinical decision-support tools do not model this full administrative pipeline, leaving patients effectively “blind” to the financial consequences of their care.

Research Goal

The aim of this project is to build a modular multi-agent system that **simulates the end-to-end journey from symptoms to diagnosis, procedure selection, medical coding, and billing**. By allowing uncertainty to accumulate across each step – mirroring real hospital and insurance workflows – we aim to evaluate how predictable, transparent, and aligned these outcomes are from the patient’s perspective.



Methodology

We built a multi-agent pipeline that **models the full clinical and billing workflow**, from a patient’s symptom description to the final estimated out-of-pocket cost. The system is divided into three coordinated agents, each producing structure outputs that feed into the next stage.

- Diagnostic Agent** – Symptoms → ICD-10 Diagnoses
 - Patients provide free-text symptoms and basic demographics. An LLM diagnostic agent infers likely conditions and assigns ~~Text~~ ICD-10 diagnosis codes. The output is a ranked list of diagnoses with confidence scores and rationales.
- Procedure Agent** – Diagnoses → ICD-10-PCS → HCPCS/CPT Codes
 - A planning agent generates high-level “procedure intents” specific to the patient case. Keyword extraction and lexical search identify matching ICD-10-PCS procedure codes, which are then scored by a clinical reasoning agent. Each recommended ICD-10-PCS procedure is mapped to HCPCS/CPT billing codes using a hybrid approach combining word-bank filtering and LLM-guided selection.
- Billing Agent** – HCPCS/CPT Code → Cost Estimation
 - The billing agent receives the HCPCS/CPT codes along with insurance parameters (deductible, Medigap, locality, etc.). Using Medicare fee schedules, it computes allowed amounts, deductible application, coinsurance, and the patient’s total projected out-of-pocket cost.

Together, the three agents form an end-to-end pipeline that follow how symptoms translate into procedures that ultimately determine a patient’s financial burden.

Results

Diagnostic Agent

- Evaluated on 212 symptom-diagnosis pairs
- Exact Match Accuracy:
 - Top 1: 23.1%
 - Top 3: 28.8%
- Semantic Alignment:
 - Average cosine similarity: 0.58 (Top 1), 0.68 (Top 3)
 - Semantic accuracy at threshold of 0.8: 26% (Top 1), 34% (Top 3)
- ICD-10 Coding:
 - Correct or semantically relevant ICD-10 codes in 198/212 cases
 - 62% showed meaningful ICD-disease alignment
- Misclassifications largely occur between clinically similar conditions (e.g. eczema and psoriasis)
- No nonsensical predictions → stable clinical reasoning

Procedure Agent

- Tested on 95 ICD-10-PCS codes using an externally validated CPT/HCPCS mapping dataset
- Retrieval-augmented mapping achieved Top 3 accuracy of 74%
- Incorporating lexical retrieval and structured code catalogs substantially improves mapping reliability vs. pure LLM prediction

Limitations

- CPT codes are copyrighted by the AMA – don’t have full mapping of ICD-10-PCS codes to CPT
- Only have cost data for Medicare fees. Private insurance rates can vary greatly.