

# Pretty Good AI – Engineering Challenge

## Overview

Build a voice bot that calls our test line and has conversations with our AI agent. Your bot will act as a "patient" testing our system—finding bugs, evaluating quality, and stress-testing edge cases.

This challenge tests what we actually care about: can you build something that works, reason through ambiguous problems, and ship?

## The Task

Build an automated voice bot that:

1. Calls our test number: **805-439-8008**
2. Simulates realistic patient scenarios (scheduling, refills, questions, etc.)
3. Records and transcribes the conversations
4. Identifies bugs or quality issues in our agent's responses

## Requirements

### Deliverables

Item	Description
Working code	Your voice bot, written in Python
README	How to set up and run your bot
Architecture doc	1-2 paragraphs explaining how your system works and why you made key design choices
Call transcripts	Minimum 10 calls. Include both sides of each conversation.
Bug report	Document any issues you found—incorrect responses, hallucinations, failures to understand, awkward phrasing, etc.

<b>Loom video</b>	5-7 min walkthrough of your approach, what you built, and how you used AI tools
-------------------	---

## Code Standards

- Clean, readable code with reasonable structure
- Clear instructions to run (ideally a single command after setup)
- Document any API keys or environment variables needed (don't commit secrets)

## Call Scenarios to Test

Your calls should cover a variety of scenarios. Ideas:

- Simple appointment scheduling
- Rescheduling or canceling
- Medication refill requests
- Questions about office hours, locations, insurance

You're testing our AI—be creative about finding its limits.

## Time & Cost Expectations

- **Expected time:** 6-12 hours
- **API costs:** Expect \$10-20 for telephony, STT/TTS, and LLM usage

**Cost reimbursement:** We'll reimburse up to \$25 in API costs for submissions that demonstrate genuine effort. Include screenshots of your usage/billing with your submission.

## Submission

1. Create a public GitHub repo with your solution
2. Email us with:
  - Link to your repo
  - Link to your Loom walkthrough
  - Start your reply with: "I built it."

## Evaluation Criteria

What we're looking for	What we're NOT looking for
Working code that makes real calls	Perfect code or over-engineering
Clear thinking in your architecture doc	Fancy diagrams

Useful bugs found with good descriptions	Nitpicks about punctuation
Evidence you iterated and problem-solved	One-shot copy-paste from AI
Clean enough code to read	Production-grade infrastructure

## Questions?

If something is unclear or you hit a blocker, email [kevin@prettygoodai.co](mailto:kevin@prettygoodai.co). Asking good questions is a positive signal, not a negative one.

---

**Good luck. Show us how you build.**