

CS224V Homework 1: Finding the Patient

Agentic Literature Research for Rare-Disease Diagnosis

Stanford University

Due: 2026-10-06 23:59 PDT

1 Overview

You are given a cohort of ten patients, each described only by a list of clinical findings. Exactly one of them has the rare disease assigned to your pair. Your job is to identify that patient and to justify the choice with evidence that someone else can check.

Some of these diseases are missing from the phenotype knowledge base you would normally match against. For example, ReNU syndrome, was first described in 2024 and is absent from databases built before then. Matching the target patient against such a database would not return “unknown”. It returns a specific, wrong diagnosis. The information needed to fix this exists, but only in the published literature.

The central goal of this assignment is to understand **how literature research systems reach their conclusions, where they fail, and how to investigate those failures**. You should be able to explain how the system found its evidence, interpreted it, and used it to produce a ranking.

You will start with retrieval augmented generation (RAG). Retrieving relevant passages and asking a model to answer those passage.

Then you will then work through **GRILL** and inspect its intermediate outputs to understand what each contributes.

GRILL searches and reads the literature while maintaining a *belief ledger*: hypotheses, evidence for and against them, and unresolved questions. You will trace how it moves from an initial question to competing hypotheses and chooses what to investigate next.

Most of the infrastructure is provided. We already pre-compute all the LLM calls for you. The assignment asks you to inspect each stage and perform error analysis. It is graded on how well you explain why components work or fail.

You will use the notebook to complete the homework. All the information you need to complete the assignment is already included.

[Link to Google Colab Notebook](#)

Please pick the disease you want to work with here: [Disease Assignments](#)

Before you begin, please make a copy of the notebook and remember to save your work regularly.

After you run the final cell, the notebook will generate a `todo_answers.json` file. Please submit this file along with your completed notebook.

Assignment Structure: The assignment follows the pipeline from the problem to the diagnosis:

- **Section 2: Background:** rare-disease diagnosis as a knowledge-base problem, and setup.

- **Section 3: Establishing Baseline:** for the assigned disease and its cohort, run some baselines experiments (Part 1-3).
- **Section 4: GRILL:** how a research agent's beliefs are formed and verified (Part 4).
- **Section 5: Patient Matching:** We provide the grill report to the LLM and check the results (Part 5).
- **Section 6: Error Analysis:** You will look at the trajectory of the GRILL agent and fine failure points (Part 6).
- **Section 7: Submission** (Part 7).

Grading. There are 12 TODOs worth 53 marks in total (Table 1). The marks reward explanation: a submission that diagnoses correctly without saying why scores below one that fails and explains precisely where. Every quantitative claim must cite a notebook cell or a `file.md:line` locator, and your report must name at least one error in the pipeline that you found through deliberate inspection.

Table 1: Assignment tasks and marks

TODO	Task	Marks
1.1	Evaluate web search result	2
1.2	Compare answer with and without retrieved web result	4
3.1	Patient matching results	5
3.2	Differentiating factors in results	5
4.1	Analyzing refuted hypotheses	5
4.2	Compare GRILL vs RAG vs direct LLM call	4
4.3	Guessing the Patient	1
5.1	Repeated runs of GRILL	6
5.2	Differentiating factors in results of GRILL	6
5.3	Ranking difference between GRILL and RAG	4
5.4	Guess the patient Part 2	1
6.1	Error analysis	10