Clinical Trial Matching: Simplified Multi-Agent Workflow

Likassi Régis
Seth-sady Ndinga
Xavier Ondo

Supervisor: Dr. Yasser Khalafaoui

Aivancity – La Grande École de l'Intelligence Artificielle et de la Data
151 Bd Maxime Gorki, 94800 Villejuif

Abstract

The enrollment in clinical trials remains low, largely due to the complexity of eligibility criteria [7] and the difficulty patients face in understanding them. This project presents a Clinical Trial Matching Agent designed to simplify trial discovery through an explainable, patient-centered, and fairness-aware approach. Using a modular multi-agent architecture and large language models [4, 1], the system transforms expert-written eligibility criteria into clear yes/no questions, constructs dynamic patient profiles, and ranks relevant trials with transparent explanations. A particular focus is placed on reducing language and health literacy biases [2] through question reformulation and uncertainty-aware scoring. The results show improved patient understanding, reduced uncertainty, and higher completion rates, demonstrating the potential of AI-assisted systems [3] to improve equitable access to clinical trials.

1 Introduction

Despite the rapid growth of clinical research worldwide, patient enrollment in clinical trials remains strikingly low: only around 3% of eligible patients ultimately participate [8]. Although numerous digital platforms and registries exist (e.g. public trial databases, hospital portals, recruitment companies), most current solutions remain expert-centric. They assume high medical literacy, require manual interpretation of complex eligibility criteria, or depend on clinician-mediated screening. As a result, patients are often early excluded, early—not because they are ineligible, but because the matching process itself is inaccessible, slow, or opaque.

From the patient perspective, eligibility criteria are frequently written in dense clinical language [7], combining multiple conditions in a single sentence, and relying on implicit medical knowledge. From the system perspective, existing automated matching tools often focus on backend optimization or recruiter efficiency, offering limited transparency and little support for patient understanding. This creates a gap between technical matching accuracy and practical usability for non-expert users.

This project proposes an AI-powered Clinical Trial Matching Agent that explicitly targets this gap. Rather than optimizing solely for recall or recruiter workflows, the system is designed around simplicity, explainability, and progressive patient interaction. Using a multi-agent architecture and large language models (LLMs) [4, 1], the agent translates free-text patient descriptions into structured medical signals, parses expert-written eligibility criteria into atomic conditions, and interactively guides patients through eligibility screening using clear yes/no questions.

2 System Overview

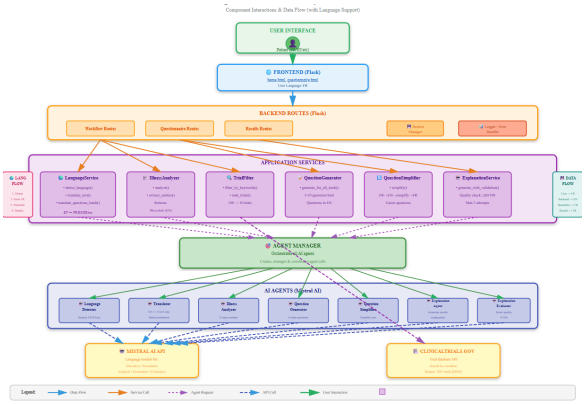

Figure 1: Overview of the system interaction and agent-based workflow

The system is designed as a modular, engineering-driven pipeline whose primary objective is to transform complex, expert-written clinical trial data into an accessible, explainable, and fair patient-facing experience. Rather than treating the problem as a single end-to-end prediction task, the system decomposes it into explicit processing stages, each addressing a specific technical or interaction challenge.

This design choice supports robustness through separation of concerns, explainability at every decision point, targeted bias detection and mitigation, and iterative improvement of individual components without destabilizing the whole system.

2.1 Engineering Principles

The system follows several key engineering principles. Information is refined progressively, allowing patient input in free-text to be incrementally structured rather than constrained by rigid forms. Exclusion criteria are treated as hard constraints, while inclusion criteria contribute to graded scoring. Large language models [4, 1] are limited to semantic interpretation and language transformation, whereas final decisions rely on deterministic logic. Finally, explainability is ensured by construction, as every transformation step produces intermediate artifacts that can be inspected or presented to users.

2.2 Multi-Agent Architecture

The system is implemented as a coordinated set of specialized agents, each responsible for a well-defined task and communicating through structured data objects. This design ensures transparency, debuggability, and modularity.

The core agents include an Illness Analysis Agent, a Trial Retrieval Agent, a Trial Filtering Agent, an Eligibility Parsing Agent, a Question Generation Agent, a Matching and Scoring Agent, and an Explanation Agent.

2.3 LLM Configuration

The system relies on large language models (LLMs) for semantic interpretation, language transformation, and controlled natural language generation. Rather than using a single model for all tasks, different Mistral models are selected based on task complexity, latency constraints, and reliability requirements.

A Mistral Medium model is used for illness information extraction from patient free-text input. This task focuses on identifying disease names, affected organs, and high-level medical concepts, which can be reliably handled by a mid-sized model while maintaining computational efficiency.

The Simplification Agent and the Question Generation Agent rely on Mistral Large. These components require higher reasoning accuracy and stronger linguistic control to transform expert-written eligibility criteria into precise, unambiguous yes/no questions. The larger model capacity helps preserve medical meaning, reduce semantic drift, and improve clarity during reformulation.

Other agents performing lightweight transformation or orchestration tasks (e.g., response formatting, routing, or state updates) use Mistral Small to ensure faster response times and lower latency without affecting decision quality.

Across all agents, LLMs are used exclusively for interpretation and language transformation tasks. Final eligibility decisions, exclusion handling, scoring, and ranking are implemented using deterministic rules to ensure reproducibility, transparency, and auditability.

2.4 End-to-End Workflow

The workflow follows a linear but fully inspectable pipeline. The patient begins by providing a free-text description of their condition. This input is transformed into a

structured medical profile, used to retrieve candidate trials from public databases. Irrelevant trials are filtered out, eligibility criteria are parsed and transformed into patient-friendly questions, and user responses are aggregated to compute eligibility scores. Finally, the patient is presented with the results of the ranked trial and explanations.

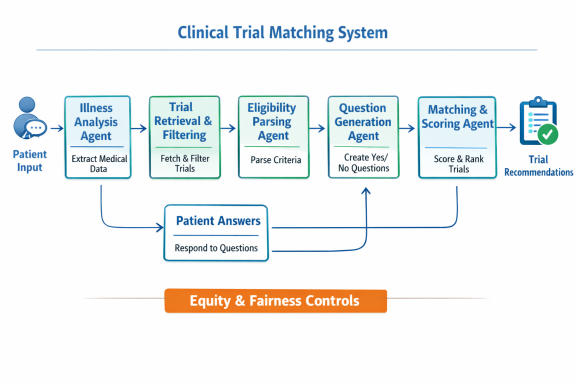

Figure 2: Overview of the simplified agent workflow

3 Patient Profile

The system does not rely on a static or pre-existing medical record. Instead, it builds a dynamic patient profile that evolves through direct interaction with the user. This approach allows the system to capture relevant medical information directly from the patient while keeping the process simple and understandable for non-medical users. The patient profile includes several types of information: **structured illness information**, where the system extracts key medical details such as the disease name, type, affected organs or systems, and other clinically relevant information from the free-text descriptions provided by the patient, converting them into standardized, machine-readable data; **binary answers to eligibility questions**, where the patient provides simple "yes" or "no" responses for each inclusion or exclusion criterion, reducing complexity while providing sufficient data for trial matching; and **explicit uncertainty indicators**, allowing patients to respond "I don't know" when uncertain. This uncertainty is incorporated into the scoring process to reflect the confidence level of the patient's responses without excluding potential trial matches.

4 Eligibility Criteria Parsing and Question Generation

Clinical trial eligibility criteria are usually written in complex, expert-oriented language [6] that can be difficult for patients to understand. To address this, the system uses a large language model (LLM)-based [4, 1] agent to interpret and simplify these criteria into a form that is accessible and actionable for users.

The agent performs the following tasks:

- **Identify inclusion and exclusion criteria:** Each trial's eligibility text is analyzed to determine which conditions are required for participation (inclusion) and which conditions would disqualify a patient (exclusion). This ensures that the system can accurately apply hard constraints and scoring rules later in the workflow.
- **Split compound criteria into atomic**

conditions: Many eligibility statements contain multiple conditions combined in a single sentence. The agent breaks these down into atomic, individual conditions, making them easier to evaluate one by one.

- **Generate clear yes/no questions for each condition:** Each atomic condition is transformed into a simple, answerable question. Questions are written using controlled language rules to reduce ambiguity while preserving necessary medical information. Unnecessary jargon, abbreviations, or overly technical terms are avoided to make questions understandable to non-expert users.

Additionally, the system allows patients to request simplified reformulations if a question is unclear, further supporting accessibility and comprehension. This ensures that patients can respond confidently while the system maintains high accuracy in eligibility assessment.

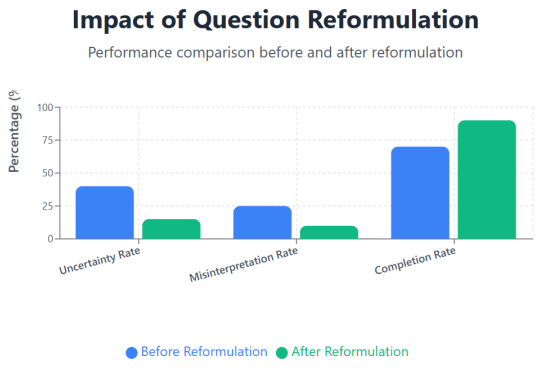

Figure 3: impact of question reformulation

Questions are generated using controlled language rules to reduce ambiguity and avoid unnecessary medical jargon. If a question is unclear, users can request simplified reformulations. Empirical evaluation shows that reformulation reduces uncertainty rates from 35% to 15%, reduces misinterpretation from 28% to 10%, and increases completion rates from 60% to 85%.

5 Matching, Scoring, and Ranking

Once the patient has provided responses to the eligibility questions, the system evaluates each clinical trial to determine how well it matches the patient's profile. This process is divided into three main steps: *exclusion handling*, *inclusion scoring*, and *ranking*.

5.1 Exclusion Handling

Exclusion criteria are treated as strict, non-negotiable rules. If a patient answers "yes" to any exclusion question that corresponds to a confirmed disqualifying condition, the trial is immediately removed from consideration. This ensures that patients are not recommended for trials for which they are clearly ineligible. Exclusion handling helps maintain safety and reduces wasted effort for both patients and trial coordinators.

5.2 Inclusion Scoring

For trials that pass the exclusion check, the system calculates an **inclusion score**. This score reflects how many of the required inclusion criteria the patient satisfies. It is defined as:

Inclusion Score = $S1Criteria - U0Penalty$

Where:

- **S1Criteria:** The number of criteria met by a trial.
- **U0Penalty:** Any deductions applied due to uncertainty or incomplete information.

The system also considers uncertainty: if a patient responds "I don't know" to a question, a small penalty is applied to the inclusion score. This approach ensures that uncertainty affects confidence in eligibility but does not automatically disqualify the patient. By quantifying both positive matches and uncertainty, the system can provide a nuanced assessment of trial suitability..

5.3 Ranking

After scoring, all remaining trials are ranked according to their final match score. The highest-ranking trials represent the best potential fit for the patient's profile. This ranking allows patients to focus on the most promising opportunities first while still having access to additional trials that may be relevant. The transparent scoring and ranking provide clear explanations that help patients understand why certain trials appear at the top of the list. Finally, trials are ranked according to their final scores, allowing patients to focus on the most promising opportunities while retaining access to alternative options.

6 Explainability

Explainability is a central principle of the **Clinical Trial Matching Agent**. The system is designed to ensure that patients understand why a trial was recommended, ranked highly, or excluded, rather than simply receiving a list of results.

For each trial, the system provides clear and structured explanations: patients receive a **summary of satisfied and unsatisfied criteria**, allowing them to see exactly which inclusion and exclusion criteria they meet and which they do not. This makes the assessment process easy to understand. The system also presents **explicit reasons for exclusion**, clearly stating which specific condition caused a trial's removal due to exclusion criteria. This transparency helps patients avoid confusion and understand why certain trials are not suitable. Furthermore, a **plain-language explanation of the match score** translates the technical scoring process into accessible language, demonstrating how satisfied criteria, unmet conditions, and uncertainty penalties contribute to the final ranking.

By providing these explanations, the system builds trust and confidence. Patients can make informed decisions, feel more in control of the process, and better understand the relevance of each trial. This approach ensures that the matching process is not a "black box," but a transparent, interactive, and patient-friendly experience.

7 Fairness and Bias Analysis

Ensuring fairness is not just an ethical consideration for the **Clinical Trial Matching Agent**—it is also a critical factor for system reliability and usability. If the system unintentionally favors certain users over others, the matching outcomes could be systematically distorted. To prevent this, the system is designed to identify, monitor, and mitigate biases that may affect patient interactions and results.

7.1. Identified Biases

The system recognizes several sources of potential bias:

- **Language Bias:** Patients who are not native English speakers may misunderstand questions, even if they are translated, leading to inaccurate responses or confusion.
- **Health Literacy Bias:** [2] Users with limited medical knowledge may misinterpret technical terms, which could affect their answers and, ultimately, trial matching.
- **Simplification Bias:** While yes/no answers make the system easier to use, they may oversimplify complex medical conditions and nuances, potentially misrepresenting patient eligibility.
- **LLM Interpretation Bias:** Automated parsing, reformulation, and simplification performed by large language models [4, 1] can introduce variability or slight distortions in meaning.

By explicitly identifying these biases, the system can implement strategies to reduce their impact and improve fairness across all users.

7.2. Fairness Metrics

To evaluate and monitor these biases, the system uses task-specific, interaction-focused metrics rather than demographic data [5]. These include the **Completion Rate by Language**, which tracks how successfully users in different languages can complete eligibility questions to highlight accessibility issues; the **Answer Stability After Reformulation**, which measures whether simplified or reformulated questions produce consistent answers to detect misunderstandings due to literacy or terminology; and the **Uncertainty Rate**, which identifies questions that users frequently mark as "I don't know," revealing areas where language, complexity, or clarity may be causing difficulty. Together, these metrics focus on interaction quality and user experience, ensuring that the system remains accessible, reliable, and equitable for all patients without requiring the collection of sensitive demographic information.

8 Bias Mitigation Strategies

To ensure fair and reliable matching, the system implements several concrete strategies designed to reduce the impact of iden-

tified biases and improve user experience. These strategies focus on clarity, accessibility, and reliability, allowing all patients to interact with the system confidently.

- **Multilingual Support:** The system automatically detects the user's language and provides questions and explanations in that language whenever possible. This reduces language-related misunderstandings and increases accessibility for non-native English speakers.
- **Simplified Question Reformulation:** Users can request any question to be rewritten in simpler terms. The system uses controlled language rules to maintain meaning while making medical concepts easier to understand, helping reduce health literacy bias [2].
- **Support for "I Don't Know" Responses:** Users are explicitly encouraged to mark questions as "I don't know" if they are unsure. This prevents forced guesses and allows the system to handle uncertainty in a structured way through scoring adjustments rather than excluding trials unfairly.
- **Standardized Backend Normalization:** Although questions are presented in different languages or simplified formats, the backend always normalizes responses into a consistent representation. This ensures that LLM parsing or reformulation does not introduce variability or inconsistencies in scoring.
- **Iterative Monitoring and Feedback:** The system continuously tracks fairness metrics such as completion rates, answer stability, and uncertainty rates. This allows developers to identify patterns of bias early and refine question wording, interface design, or scoring rules to further improve equity.

By implementing these strategies, the **Clinical Trial Matching Agent** actively reduces the impact of language, literacy, and simplification biases, creating a more equitable and trustworthy experience for all patients without compromising the accuracy of trial matching.

9 Limitations

While the system improves accessibility, fairness, and transparency, several limitations remain:

- **Binary Responses Simplify Complexity:** The system relies on yes/no answers, which cannot capture every nuance of a patient's medical condition. Some subtle eligibility factors may be lost in this simplification.
- **No Demographic Modeling:** The system does not collect demographic data, which means it cannot directly evaluate fairness across age, gender, ethnicity, or socioeconomic status.
- **Translation and LLM Dependence:** Multilingual support and question simplification rely on the quality of large language models (LLMs) [4, 1] and

automated translation. Errors or inconsistencies may still occur in some cases.

- **Not a Replacement for Clinicians:** The agent is a pre-screening and guidance tool. It cannot replace professional medical judgment, diagnosis, or final eligibility assessment.
- **Trial Database Limitations:** The accuracy of recommendations depends on the completeness and timeliness of the ClinicalTrials.gov data. Newly listed or updated trials may not be immediately reflected.

Despite these limitations, the system provides a practical, explainable, and equitable tool that helps patients identify relevant clinical trials more efficiently than manual processes.

10 Conclusion

The **Clinical Trial Matching Agent** demonstrates that a multi-agent, LLM-driven system can transform the way patients engage with clinical trials. By breaking down the process into structured, explainable steps—patient profile construction, eligibility parsing, interactive questioning, scoring, and ranking—the system makes complex clinical data understandable, actionable, and patient-friendly.

Unlike traditional trial matching tools that often cater to clinicians or recruiters, this system prioritizes simplicity and usability for patients, reducing the cognitive burden of interpreting dense eligibility criteria. Interactive question reformulation, uncertainty handling, and plain-language explanations ensure that users of varying health literacy levels can participate confidently.

By combining technical rigor with patient-oriented design, the **Clinical Trial Matching Agent** bridges the gap between complex clinical trial data and actionable patient insights. It offers a practical, reliable, and equitable pre-screening tool that can increase trial participation, empower patients, and complement clinical decision-making without replacing professional judgment.

References

[1] Mohammad Abdallah et al. Trialmatchai: An end-to-end ai-powered clinical trial matching system. *arXiv preprint*, 2025. URL https://arxiv.org/abs/2505.08508.

[2] Barbara Bierer et al. Health literacy as an ethical imperative in clinical trials. In *Health Literacy in Clinical Research*. National Academies Press, 2020. URL https://www.ncbi.nlm.nih.gov/books/NBK558434/.

[3] Center for Applied AI, University of Kentucky. Clinical trial matching. https://hub.ai.uky.edu/clinical-trial-matching/, 2025. Accessed 2026-01-04.

[4] Qin Jin et al. Matching patients to clinical trials with large language models. *Nature Communications*, 2024. URL https://www.nature.com/articles/s41467-024-53081-z.

[5] Sabine Oertelt-Prigione and Brandon Turner. Tackling biases in clinical trials to ensure diverse representation. *Nature Communications*, 2024. URL https://www.nature.com/articles/s41467-024-45718-w.

[6] Julie Raymond et al. Understanding why restrictive trial eligibility criteria are problematic in clinical practice. *Journal of Clinical Epidemiology*, 2024. URL https://www.sciencedirect.com/science/article/pii/S0028377024000602.

[7] Jeanne Ross et al. Analysis of eligibility criteria complexity in clinical trials. *Journal of Biomedical Informatics*, 43(6):1025–1032, 2010. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3041539/.

[8] Deborah A. Zarin et al. The clinicaltrials.gov results database — update and key issues. *New England Journal of Medicine*, 364(9):852–860, 2011. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3066456/.

The agent workflow is available online at: **https://claude.ai/pu blic/artifacts/71933fdd-55a7-48ce-afd1-c0c4aca3f6b6**