Description for Collaborative LLM-Human Assistance Platform

GPT-4o

December 18, 2024

Overview

The platform enables efficient collaboration between humans and language models (LLMs) to solve complex problems. It features two primary user types:

- $\mathbf{h_1}$: The human user initiating a question.
- $\mathbf{h_2}$: The human user assisting the LLM in addressing specific gaps in knowledge.
- $\mathbf{a_1}$: The AI assistant (LLM) responsible for processing and responding to h_1.

Key Features

1. h_1-Facing Side

- A simple application interface for h_1 to submit questions to a_1.
- a_1 evaluates its confidence in answering the question. If necessary, it generates abstract sub-questions to address specific knowledge gaps.
- Sub-questions may be routed to h_2 for assistance.

2. h_2-Facing Side

- **Private Interaction Model:** Questions are served to h_2 one at a time in "answering mode," with no public display or direct interaction between users.
- **Moderation:** a_1 or a validation LLM evaluates h_2's contributions for relevance and appropriateness.
- **Motivation and Recognition:**
  - Integration with LLM providers to reward h_2 with usage credits for valuable contributions.
  - Leaderboards highlighting top contributors and their fields of expertise.
  - Creative visualizations, such as knowledge maps, to showcase activity and achievements.

- **Learning Opportunities:** h_2 can use the platform to encounter challenging STEM problems and improve their own understanding.

Workflow

1. h_1 submits a question.
2. a_1 evaluates its confidence in responding.
    - If confident, a_1 answers directly.
    - If not confident, a_1 generates abstract, impersonal sub-questions related to specific gaps in knowledge.
3. Sub-questions are routed to h_2 users for assistance.
4. h_2 provides focused inputs, which are validated by a_1 or a secondary LLM.
5. a_1 synthesizes the final answer and delivers it to h_1.
6. Feedback from h_1 impacts the ranking of both a_1 and h_2.

Focus and Constraints

- The platform restricts questions to abstract, impersonal topics, primarily in STEM fields.
- Filters and moderation remove contentious or offensive questions.
- Users are informed that their questions may be anonymized and viewed by another human (h_2).

Motivation and Sustainability

- Credits for LLMs provided as rewards.
- Gamification elements, such as leaderboards and badges, encourage engagement.
- h_2 can view how their input contributed to the final response, fostering a sense of accomplishment.
- The platform serves as both a problem-solving tool and a learning resource for users.