BenCourt: A Benchmark and Framework for Court Simulation using LLM-based Agents (Proposal)

Kaiyuan Zhang
DCST, Tsinghua University
Beijing,China
ky-zhang24@mails.tsinghua.edu.cn

Zhaoxi Li
DCST, Tsinghua University
Beijing,China
li-zx24@mails.tsinghua.edu.cn

Xuancheng Li
DCST, Tsinghua University
Beijing,China
lixuanch23@mails.tsinghua.edu.cn

Abstract

This is a preliminary proposal of a new framework: BenCourt, which aims to provide a testing and evaluation playground for court simulation using LLM-based agents. The full version is expected to be released in December. Github link: https://github.com/Miracle-2001/BenCourt.

1 Background

Court trail has became a useful method to maintain justice and social stability throughout the world. However, the scarce legal resources and huge legal cases put extreme pressure on lawyers, judges, and the whole judicial system. An explainable, portable and efficient autonomous court needs to be built. In other words, legal agents are expected to make their contributions.

After the release of Chatgpt 3.5[3] in late 2022, Large Language Models (LLMs) becomes an extremely hot topic in both research and application areas in the past 2 years. With the help of LLMs, researchers and engineers built and released multiple simulation systems based on LLM-based agents, such as "Stanford Town"[4], RecAgent[5], Agent Hospital[2], AgentCourt[1], etc.

Though primary and simple, these attempts indeed provide inspirations and hope for future designation on agent simulation.

In this paper, we proposed a newly defined benchmark and framework for building a more advanced court simulation system via LLM-based agent.

2 Definition

For a normal process of a court trail, 5 basic agents should be included, which are "Judge", "Plaintiff", "Plaintiff's Lawyer", "Defendant" and "Defendant's Lawyer". The abbreviation of these roles can be written as "JG","PT","PTL","DF", "DFL", respectively.

Each agents has its profiling and task. "PT" and "DF" should present the court and answer questions asked by the judge or lawyers. "PTL" and "DFL" should represent their parties and argue in the court. "JG" should conclude the trial and give the verdict. Besides these basic characters, other agents such as court clerk, witness may also participate in the trail.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

The goal of the court simulation is to give a proper instrument of judgment and make an explanation as clear as possible. Efficiency, portable and scalable are higher goals of these systems.

3 Related Work

In 2023, the "Stanford Town" project made a presentation of mutli-agents discussion and collaboration scenario. After that, RecAgent[5], Agent Hospital[2] are published aiming at simulating the areas of recommended system and hospital consult.

For, court simulation, recently a system called AgentCourt [1] was released and received much attention. It is the first court simulation system with 6 characters and each agent can maintain their role to some extent. However, further advancement such as using retreival augmented generation (RAG) technique or long context learning methods are inadequately discussed. Besides, more evaluation methods should be included to maintain a logical and proper trail process.

4 Proposed Method

4.1 Motivation and Basic Goals

As mentioned above, court simulation is valuable, which may save legal resources and provide strategies or recommendation to real world lawyers and judges. Considering the disadvantages of AgentCourt [1], we list our goals as followings:

First, a well-defined framework and playground should be established, which can accommodate at least 5 basic legal agents to simulate a complete trail process. Second, this system should be scalable and reliable. Functions such as agent adding, legal case editing should be provided. Third, proper evaluation methods has to be included for a better simulation result and explanation ability. Last but not least, some recently used multi-agent establishing and simulation methods should be covered.

4.2 Baselines

Besides AgentCourt [1], baselines should also include "single judge system" (only one judge to give out a trail result), "simple in-context-learning agent"(only use basic prompt command to guide each agent), and "supervised-finetuning agent"(use legal datasets to finetune each agent). Other baselines will be supplemented in the future.

4.3 Possible challenges and future planning

During the building process of the court simulation project, some possible challenges are listed as followings:

1. Multi-round conversations and long context retrieval ability. A real court process consists of many rounds of conversations, so agents should remember prior ones and give logical output.

2. Profile design and tasks understanding. Both judge and lawyer are sophisticated practitioner, so a well-defined profile and useful finetuning methods should be employed.

In the future, maybe we can aggregated more techniques such as RAG to make it a more advanced platform. Hope it can make contributions to the real world legal system.

References

[1] Guhong Chen, Liyang Fan, Zihan Gong, Nan Xie, Zixuan Li, Ziqiang Liu, Chengming Li, Qiang Qu, Shiwen Ni, and Min Yang. Agentcourt: Simulating court with adversarial evolvable lawyer agents. *arXiv preprint arXiv:2408.08089*, 2024.

[2] Junkai Li, Siyu Wang, Meng Zhang, Weitao Li, Yunghwei Lai, Xinhui Kang, Weizhi Ma, and Yang Liu. Agent hospital: A simulacrum of hospital with evolvable medical agents. *arXiv preprint arXiv:2405.02957*, 2024.

[3] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

[4] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22, 2023.

[5] Lei Wang, Jingsen Zhang, Xu Chen, Yankai Lin, Ruihua Song, Wayne Xin Zhao, and Ji-Rong Wen. Recagent: A novel simulation paradigm for recommender systems. *arXiv preprint arXiv:2306.02552*, 2023.