# Epidemic Modeling with Generative Agents

Ross Willams , Niyousha Hosseinichimeh , Aritra Majumdar , Navid Ghaffarzadegan

**摘要**

该研究提出了一种基于个体层面建模的新模式，以解决将人类复杂行为纳入流行病模型的重大挑战。通过在基于代理人的流行病模型中使用生成式人工智能，每个代理人都能通过连接到例如 ChatGPT 等大型语言模型来做出自己的推理和决策。通过各种模拟实验，该论文提供了令人信服的证据，证明生成性代理人能够模仿现实世界的人类行为，如生病和病例增加时的自我隔离。总的来说，这些代理人表现出类似于在最近的新冠疫情出现之后所观察到的疫情模式。此外，这些代理人成功地使疫情曲线趋于平缓。这项研究提供了一种模仿人类大脑、推理和决策的新途径，为改进动态系统建模的创造了新的可能性。

**关键词**：generative AI agent；epidemic modeling；computational social science

## 1 引言

社会中相互联系的个体组成的社会网络构成了当代世界的基石。与数学分析不同，计算机模拟为理解社会网络的形成和演化提供了新的途径。这对于社会科学家来说是一个基本的工具。早在 1996 年，就有一本名为《社会科学微观模拟》[1]的书，从社会科学的角度提供了关于模拟的宝贵见解。社会模拟涉及的领域非常广泛，如流行病、社交网络、交通网络的模拟。社会模拟可以通过两种形式实现：微观模拟[2][3]和宏观模拟[4][5][6]。在宏观模拟中，也称为基于系统的模拟，研究人员使用描述群体变化状态的方程对系统的动力学进行建模。相反，微观层次的模拟，或基于代理人的模拟，指研究人员使用人为制定的规则或参数化模型来描述与他人交互的个体 (简称代理人) 行为。

大型语言模型 ( Large Language Models，LLMs )[7][8][9][10][11][12]是深度学习领域的最新进展，其特点是使用了大量的神经层网络。这些模型在大量的文本语料上进行训练，获得了理解、生成和使用人类语言的惊人能力。

LLMs 在文本理解方面具备与人类水平相近的强大能力，因此研究人员已经开始[13][14][15][3]利用 LLMs 作为模拟复杂人类行为的代理人。首先，LLMs 具有感知和理解世界的能力，尽管它受限于文本形式描述的世界。其次，LLMs 能够利用推理技术设计和组织任务时间表，将任务要求和随之而来的奖励结合起来。在整个过程中，LLMs 有效地维护和更新记忆清单，采用基于人类推理模式的适当引导提示。最后，LLMs 能够生成与人类语言十分相似的文本，这些输出文本能够影响环境，并与其他代理人进行交互。人的行为是复杂多样的，在个体、社区、文化和情境中都是变化的。将这种复杂性整合到模型中，在表示和参数化方面均存在困难。因此，采用基于代理人的仿真模式，利用 LLMs 来模拟社会网络中的每一个体，从而捕获他们各自的行为和个体之间错综复杂的相互作用，具有重要的应用前景。

在这篇论文中，作者着重于流行病的实际场景 (比如 COVID-19)，提出了一个基于生成式代理人的流行病模型（GABM），与传统的 ABMs[此处引用] 不同，这种新的建模方式通过生成式人工智能（或 LLMs）赋予了每个代理人基于环境信息进行推理和决策的能力，而无需建模者确立任何决策规

则。如图 1 所示，这一新的建模模式能够引入社会中具有不同角色的代理人，从而捕捉多元化社会中个体对传染性病毒演变成流行病的反应。

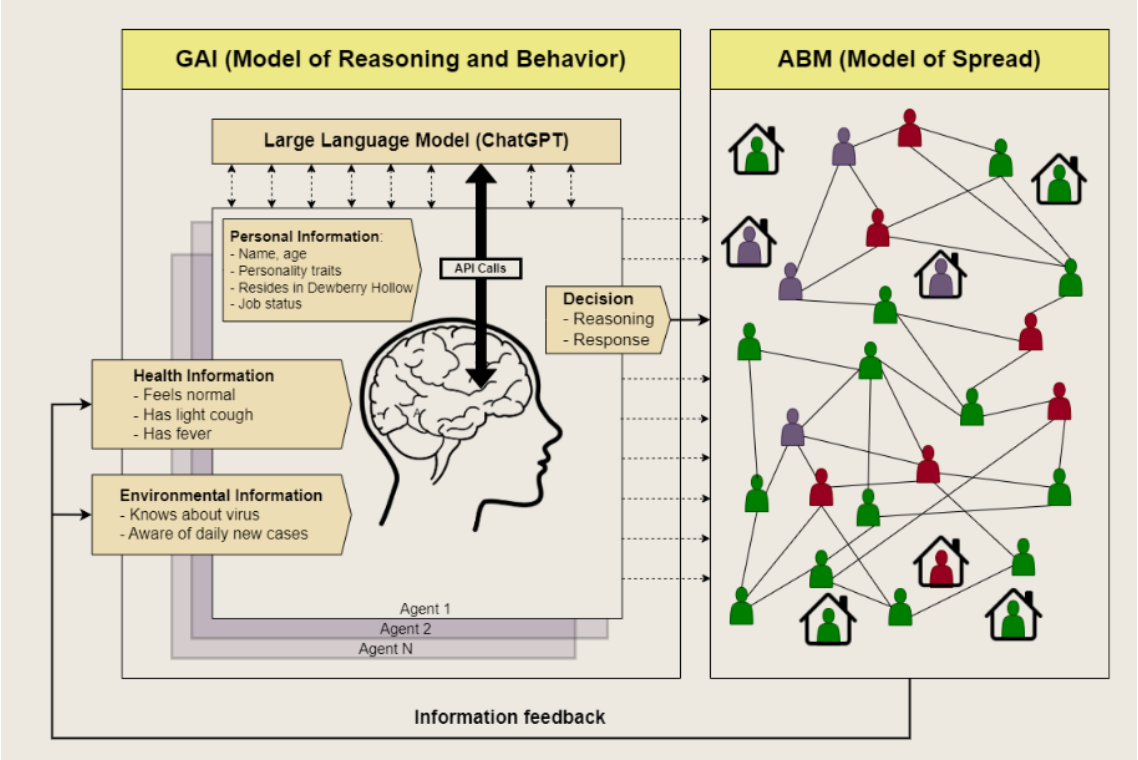



图 1: GABM Overview

## 2 相关工作

### 2.1 社会计算与模拟

根据[16]，“模拟意味着用合适的输入驱动一个系统的模型，并观察相应的输出”。社会模拟旨在模拟各种社会活动，其应用范围广泛[17]。由于驱动社会行为的内部机制无法被直接观察到，社会模拟的一个主要优势是它可以帮助社会科学家理解人类社会[18]的某些重要特征。通过采用能够合理地重现历史社会行为动态性质的模型，在某种程度上我们能够预测社会系统的未来

尽管社会计算与模拟的应用前景广阔，但社会模拟是复杂的。最早的工作使用基于离散事件的模拟[4]或系统动力学[5][19][17]，通过一系列方程来近似描述系统的多个变量。这些早期的方法主要集中在准确预测变量，而不是阐明潜在的机制或因果关系。随后，受模拟在其他科学领域的快速发展和显著成功的启发，基于代理的模拟在社会模拟领域兴起。在这些模拟方法中，一个具有代表性的技术是使用元胞自动机[2]。这种方法最初建立了一个由众多个体组成的社会环境，随后制定了一套规则，规定个体如何相互作用并更新状态。基于代理的模拟可以看作是一种微观仿真，它通过描述明确定义的微观个体的行为来描述真实世界。因此，基于代理的模拟也被称为微观模拟。

近年来，随着机器学习和人工智能的迅速发展，基于代理的模拟仿真发生了显著的变化。一些代理开始由机器学习算法或日渐强大人工智能驱动。这些代理具有动态感知周围环境的能力，并表现出与人类行为近似的行为。个体模拟代理的快速发展不仅保持了传统模拟模式的有效性，并且带来了显著的改进。

### 2.2 基于大语言模型的模拟

近年来，GPT 系列[7][8]、PaLM 系列[9][10]、LLaMA[11]、GLM[12]等大型语言模型凭借其在理解和生成人类语言本文方面的强大能力，受到了广泛关注。

Aher 等人[13]进行了初步测试，发现 LLMs 能够重现一些经典的经济学、心理学和社会学实验的结果。Horton 等人[14]用基于 LLM 的代理人代替人类实验对象模拟经济行为，LLM 代理人 (或生成性代理人) 被赋予禀赋、信息、偏好等信息，并给予适当引导。使用 LLM 代理人的实验结果与原论文 (人工实验) 在质量上是相似的[20][21]。另一项研究[15]采用了基于 LLM 的众包方法，通过收集模拟现实人类的 LLM 的反馈结果，以进行社会科学计算的研究。

最近，Part 等人[3]基于电子游戏环境构建了一个由 25 个代理人组成的虚拟城镇，每个代理人均由 LLM 驱动，在虚拟城镇中，代理人可以计划和安排日常生活中要做的事情。虽然模拟完全基于生成范式，没有任何真实数据可以评估，但它提供了一些见解，即 LLMs 可以作为驱动代理人的强大工具。通过引导赋予每个代理人自己的身份和特征，促进代理人之间的沟通。值得注意的是，这种模拟完全在生成范式下进行，没有纳入任何真实世界的数据进行评估。尽管如此，这些发现为 LLMs 驱动智能体模拟的潜力提供了有价值的见解。

## 3 本文方法

### 3.1 本文方法概述

在这篇论文中，作者构建了一个由多位代理人组成的 small world，为了把人类行为的复杂性整合到模型中，作者提出了 GABM 模型，即在传统的基于代理的传播模型中引人入了生成式 AI 或 LLMs 来帮助代理人完成复杂的决策任务。如图 1所示，GABM 模型由两部分组成：GAI 和 ABM.GAI(Generative AI Model) 对每个代理人的推理和行为进行建模，ABM(Agent Based Model) 对流行病的传播进行建模。在每一时间节点，代理人由 GAI 模型结合环境信息进行推理和决策 (即是否出门与其他代理人进行交互)，代理人之间的交互由 ABM 建模。

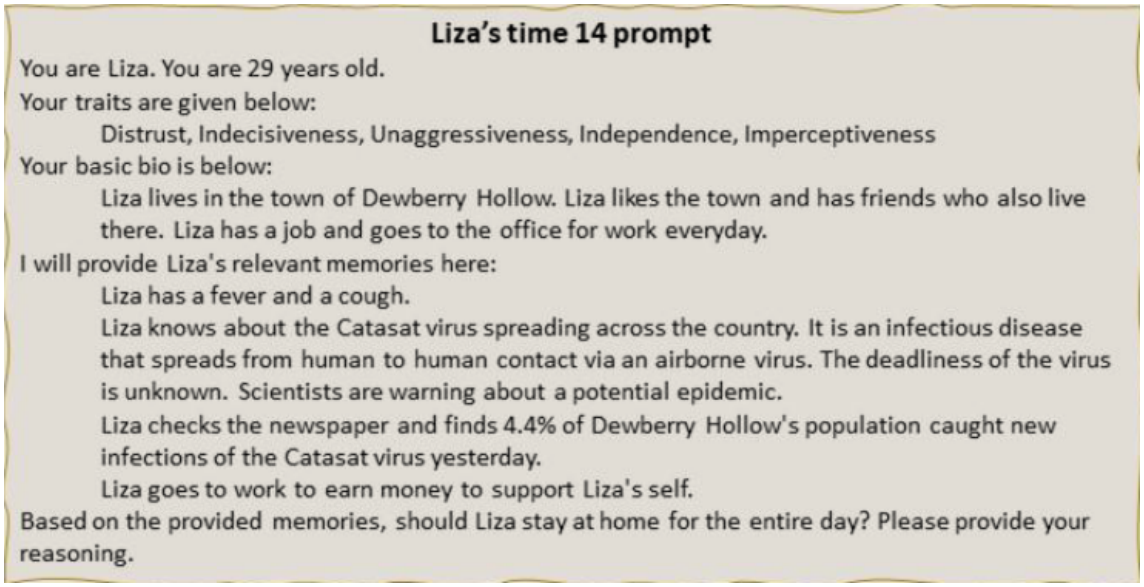



图 2: Prompt example

GAI 模型由 LLMs 驱动，每个代理人都有一个发送给 LLMs 的提示文本 (图 2)，包括代理人的名字、年龄、特征、相关记忆。提示文本还包括代理人的身体状况，城镇中患病个体的比率等信息。然后

在每个时间步骤，代理人需要决定是否整天留在家中以及他们的理由 (图 3)。对于那些决定离开自己家的代理人，ABM 模型将每个代理人单独与数量等于模型接触率的其他代理人进行交互，这就可能导致疾病在易感人群和受感染人群之间的传播。一旦所有代理人交互完成，时间步长就会增加 1，代理人的健康状态就会更新。

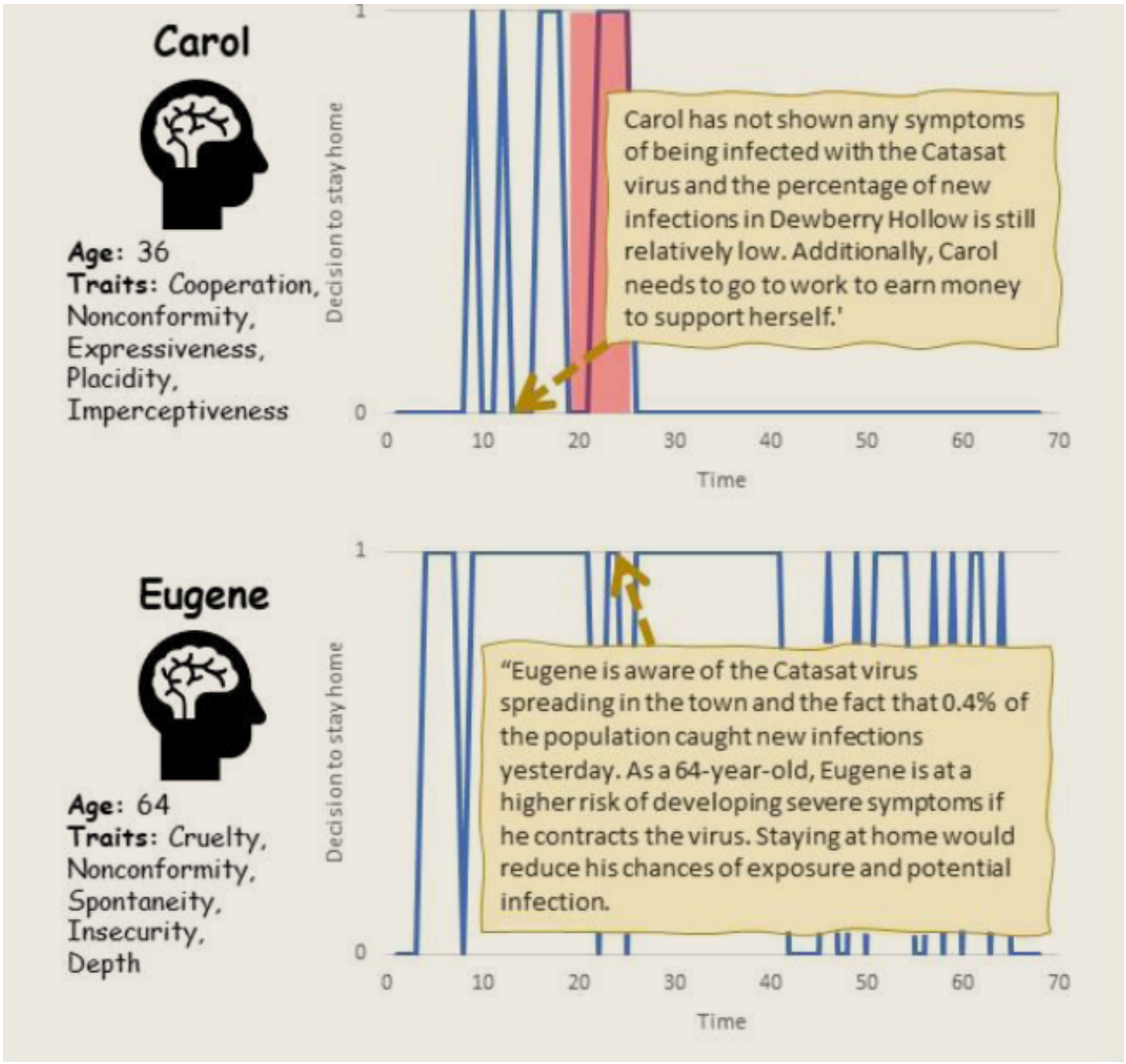



图 3: 示例：代理人特征、反应以及推理

最终，研究者统计了模拟实验的结果，其中包括每日新增病例、历史累计病例、当前累计病例、以及人口流动率等指标。在实验结果中，他发现模拟世界中出现一些与现实人类社会相似的特征。比如代理人能够通过了解疫情信息，结合自身记忆决定是否自我隔离，并且通过这种方式延缓疫情的发展，平缓疫情曲线。此外，个体层面的分析显示了代理人决策的变化以及对有关病毒的信息的不同反应，接近于真实世界的情况。

# 4 复现细节

## 4.1 与已有开源代码对比

复现的代码参考了 Github 项目 ($https://github.com/bear96/GABM-Epidemic$), 具体包括以下文件：

- $citizen.py$
- $eval.py$
- $main.py$
- $utils.py$
- $world.py$

在复现这篇论文的过程中我沿用了 ABM 传播模型，主要对 GAI 模型进行了修改。

这篇论文中提出的 GABM 模型存在一个显著的缺陷，即相对于资源密集、成本高、运行时间较长。截至 2023 年 6 月，一个由 1000 个代理人组成的流行病传播模型在配备 32GB RAM 的 CPU 上的运行时间可超过 90 小时，每次程序运行的成本约为 20 美元。这主要是由于对 OpenAI 服务器的数以万计的应用程序接口 (API) 调用导致的，同时还有其他硬件和电力成本。另外该论文缺乏与其他大语言模型效果的对比，缺少基准方法。

“然而，随着未来几年中大型语言模型（LLMs）在成本和速度上的进一步提升，GABM 的计算成本和运行时间有望减少。随着本地运行的 LLMs 逐渐成为可能，我们已经看到了一些希望”。作者在论文中提到了本地运行的 LLMs 可能带来的运行时间与成本上的改进。为进一步探索本地 LLMs 的潜力，并作为这篇论文的基准方法，我在这篇论文的基础上做出了如下改进:

### 4.1.1 部署本地运行的 LLM

由于频繁调用 OpenAI 服务器不仅大幅增加了成本，还导致运行时间的延长。为了探索轻量级的本地部署的大语言模型（LLM）是否能够胜任 ChatGPT 的任务，我进行了在本地部署开源大语言模型 Vicuna 的实验。在 2023 年上半年，UC 伯克利学者与 CMU、斯坦福等大学联合推出了一个在 LLaMA 基础上微调的模型，被俗称为小羊驼。Vicuna 目前已经发布了两个版本：Vicuna-13B 和 Vicuna-7B。Vicuna-13B 在 OpenAI ChatGPT 和 Google Bard 的质量方面达到了 90% 以上，在大多数情况下优于 LLaMA 和 Stanford Alpaca 等其他模型。本地部署的 LLM 开源在 github 项目 ($https://github.com/lm-sys/FastChat$). 由于实验资源有限，我选择部署参数规模为 70 亿的 Vicuna-7B-v1.5 作为这篇论文的基准方法，并探索部署在本地上的轻量级大语言模型的潜力。

分别将 ChatGPT 与 Vicuna 驱动 GAI 模型时收集的数据进行对比，研究 Vicuna 能否结合上下文信息进行合理推测，在多大程度上能够完成 ChatGPT 的任务，同时对比统计运行时间与成本的改善。

### 4.1.2 提示设计 (Prompt Engineering)

”Prompt” 对于大型语言模型（如 GPT 系列）的输入文本，通常被认为是影响模型生成回答质量的重要因素之一。透过巧妙设计的 prompt，可以引导模型产生更准确、更相关或更符合特定语境的回答。适当设计的 prompt 可以使模型更加准确地理解用户的意图，从而提高生成回答的质量。然而，过于模糊或不清晰的 prompt 可能导致模型产生不符合期望的回答。因此，在使用大型语言模型时，prompt 的设计成为调整和优化模型输出的关键策略之一。

该论文的提示文本 (prompt) 是基于 ChatGPT 的特性设计，并不适用于 vicuna-7B-v1.5。vicuna 的整体性能不如 ChatGPT，如果把论文设计的 prompt 原封不动的发给 vicuna，vicuna 将很难理解 prompt 的内容并做出有效回答。通过大量调试与查阅资料，我设计了内容更为明确，更容易被 vicuna 理解的 prompt 如图 4。

```
You are {self.name}. You are {self.age} years old.
Your basic bio is below:
{self.name} lives in the town of Dewberry Hollow. {self.name} likes the town and has friends who also live there. {self.name} has a job and goes to the office for work everyday.
Your traits are given below:
{self.traits}
I will provide {self.name}'s relevant memories here:
{self.name}'s health condition:{self.get_health_string()}
{self.name} knows about the Catasat virus spreading across the country. It is an infectious disease that spreads from human to human contact via an airborne virus. The deadliness of the virus is unknown. Scientists are warning about a potential epidemic.
{self.name} checks the newspaper and finds that {(self.model.day_infected_is_4[self.model.schedule.steps]*100)/self.model.population: .1f}% of Dewberry Hollow's population caught new infections of the Catasat virus yesterday.
{self.name} goes to work to earn money to survive.
Based on 1.your traits 2.health condition 3.daily infection rate in the town 4. work need, should {self.name} stay at home for the entire day? Please provide your reasoning.
please pay atention!
"Yes" means stay at home,"No" means not stay at home but go to work instead,your "Response" must align with "Reasoning". This is crucial.
If the answer is "Yes," please state your reasoning as "Reasoning: [explanation]."
If the answer is "No," please state your reasoning as "Reasoning: [explanation]."
The format should be as follow:
Reasoning:
Response:
Example response format:
Reasoning: {self.name} is tired.
Response: Yes
It is important to provide "Response" in a single word.
```

图 4: 改进的 prompt

按照图 4 的 prompt 格式，vicuna 能够综合考虑多种因素并给出合理且清晰的回答，避免了回答与推理的不一致或忽略一些重要信息。这里发送给 Vicuna 的提示文本是系统消息，与用户消息相比，系统消息增强了 ChatGPT 在长时间和会话中保持自己扮演的角色以及遵守相关规则的能力。

提示文本的第一部分给出代理人的姓名、年龄、特征来灌输代理人的人格。此外还介绍了代理人的基本生活状况。第二部分向代理人提供其健康状况以及病毒以及的相关信息来帮助代理人做推理决策。第三部分询问代理人是否应该待在家中进行自我隔离，指定其需要结合的因素，规范其回答及推理的格式。

## 4.2 实验环境搭建

### 4.2.1 硬件环境

- CPU: Platinum 8260 CPU @2.30GHz(2*48 线程)
- GPU: RTX 2080Ti 12G*2
- Memory: 128GB
- OS: Linux

### 4.2.2 安装依赖包

- mesa
- names-dataset
- openai
- matplotlib
- numpy
- pandas
- tqdm

### 4.2.3 部署本地 LLM

先从 github 克隆仓库, 用下面的命令

- $git\ clone\ https://github.com/lm-sys/FastChat.git$
- $cd\ FastChat$

安装本地 LLM 需要的依赖包，使用以下命令。如果第二条指令报错，可能是当前 pip 版本过旧第一条指令需要正确执行。

- $pip3\ install\ --upgrade\ pip$
- $pip3\ install\ -e\ ".[model\_worker, webui]"$

需要调用 API 接口来与本地部署的 LLM 进行通信，新建三个窗口，依次使用下面的命令

- $python3\ -m\ fastchat.serve.controller$ 启动模型控制器

如网络出现问题，无法远程下载模型权重，可以手动缓存至本地，再修改 model-path 为本地路径

- $python3\ -m\ fastchat.serve.model\_worker\ --model-path\ lmsys/vicuna-7b-v1.5$ 加载模型，注意 Vicuna-7B 需要至少 14G 的显存

- $python3\ -m\ fastchat.serve.openai_api_server\ --host\ localhost\ --port$ 8000 启动 API 服务器，端口 8000

在调用 API 时需要在代码中指定 api 服务器地址

- $openai.api\_base = "http://localhost:8000/v1"$

### 4.3 界面分析与使用说明

启动模型控制器后的界面如图 5 所示。

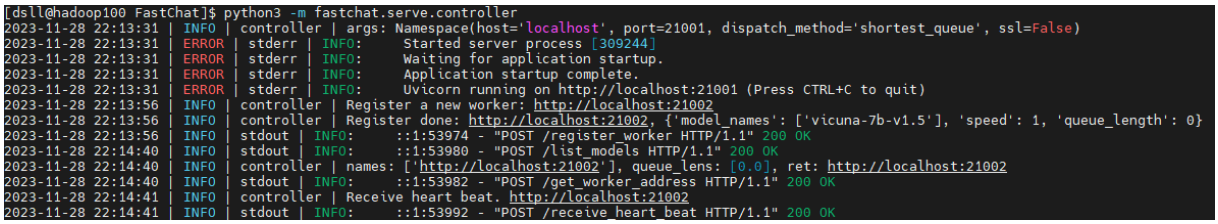



图 5: controller 界面

加载模型后的界面如图 6 所示，

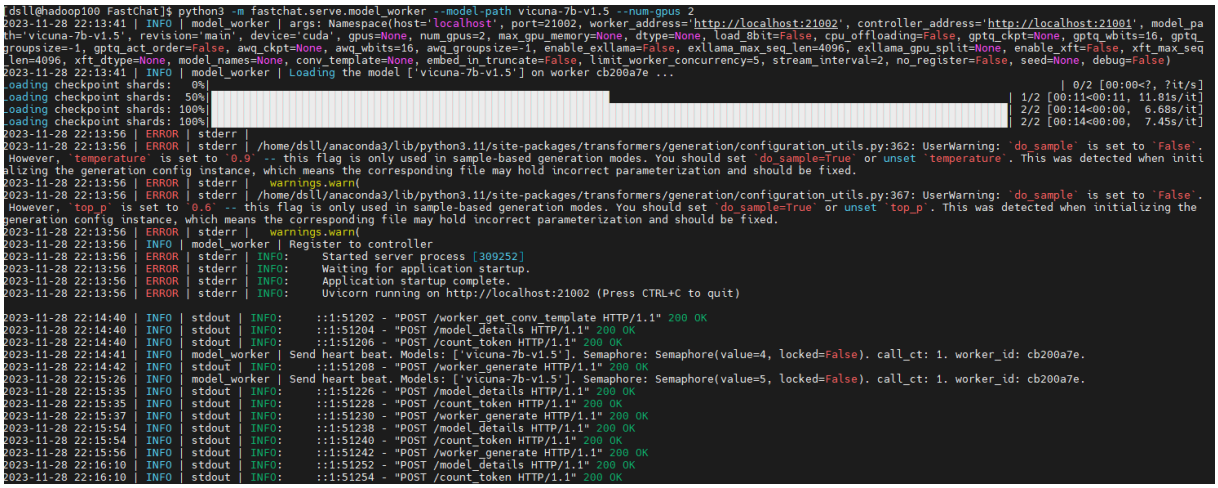



图 6: 加载大语言模型界面

启动 api 服务器后的界面如图 7 所示

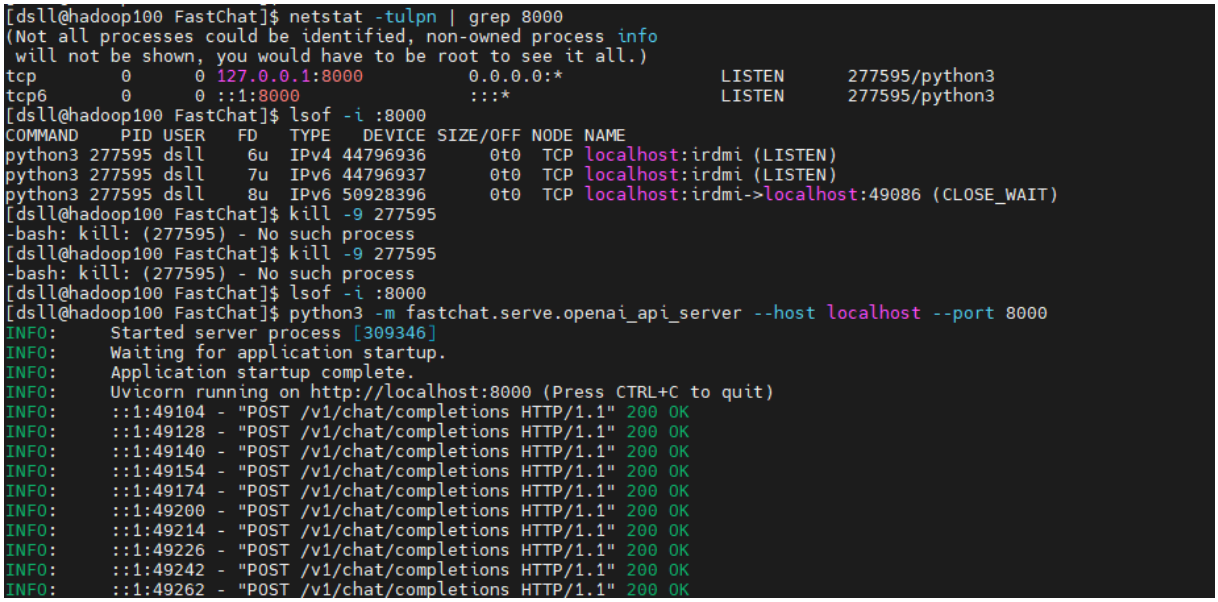



图 7: API Server 界面

再新建一个窗口，进入 GABM 目录，用命令 $python3\ main.py$ 运行 GABM 模型可选的参数如下：

- $--contact\_rate$ 接触率，默认 5
- $--infection\_rate$ 感染率，默认 0.1
- $--no\_init\_healthy$ 初始健康人群数，默认 98
- $--no\_init\_infect$ 初始感染人群数，默认 2
- $--no\_days$ 模拟进行的天数，默认 50
- $--time\_to\_heal$ 感染后需要的痊愈天数，默认 6
- $--no\_of\_runs$ 运行次数，默认 1

GABM 运行界面如图 8 所示

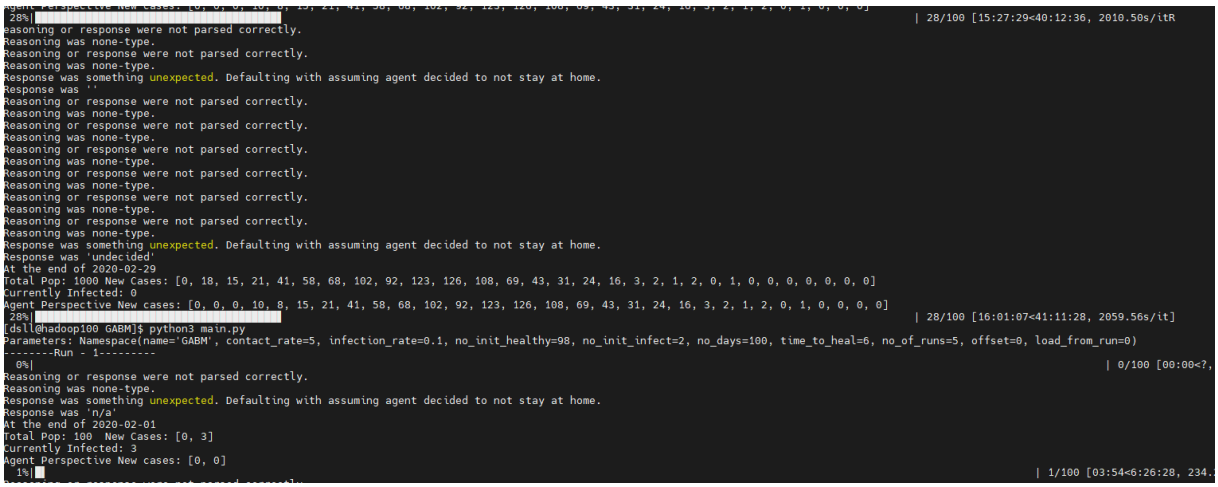


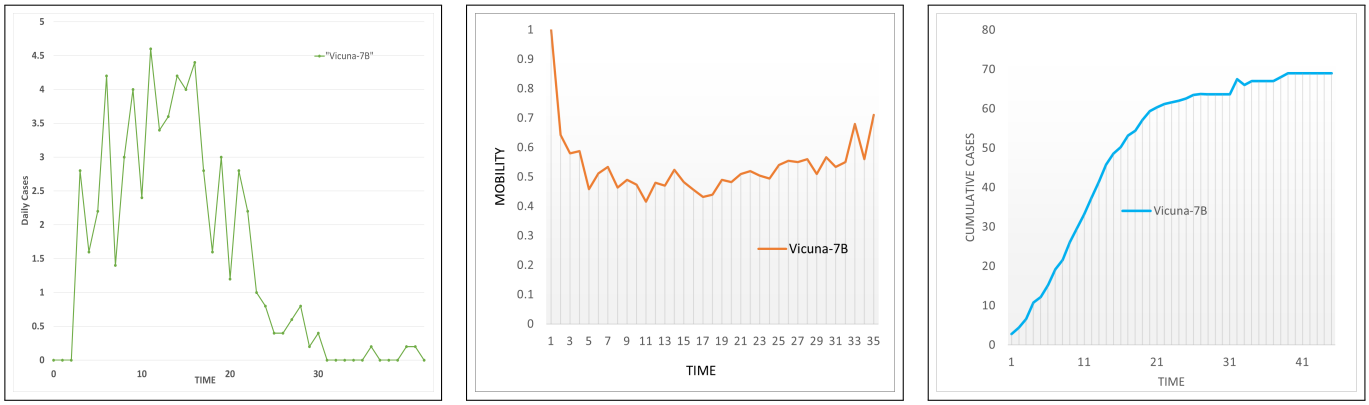

图 8: GABM 运行界面

### 4.4 创新点

该论文提出的 GABM 模型是资源密集的模型，由于需要对 OPEN AI 服务接口进行频繁的调用，其花费成本高，运行时间长。作者在论文中提到“然而，随着未来几年中大型语言模型（LLMs）在成本和速度上的进一步提升，GABM 的计算成本和运行时间有望减少。随着本地运行的 LLMs 逐渐成为可能，我们已经看到了一些希望”。由此，在实验资源有限的情况下，我在本地部署了性能最优的开源轻量级 LLM，即 vicuna-7B-v1.5。通过统计对比本地 LLM 驱动 GAI 模型时的实验数据，探索本地部署 LLM 的潜力，对运行成本和时间的改进，此外作为该论文的基准方法。

## 5 实验结果分析

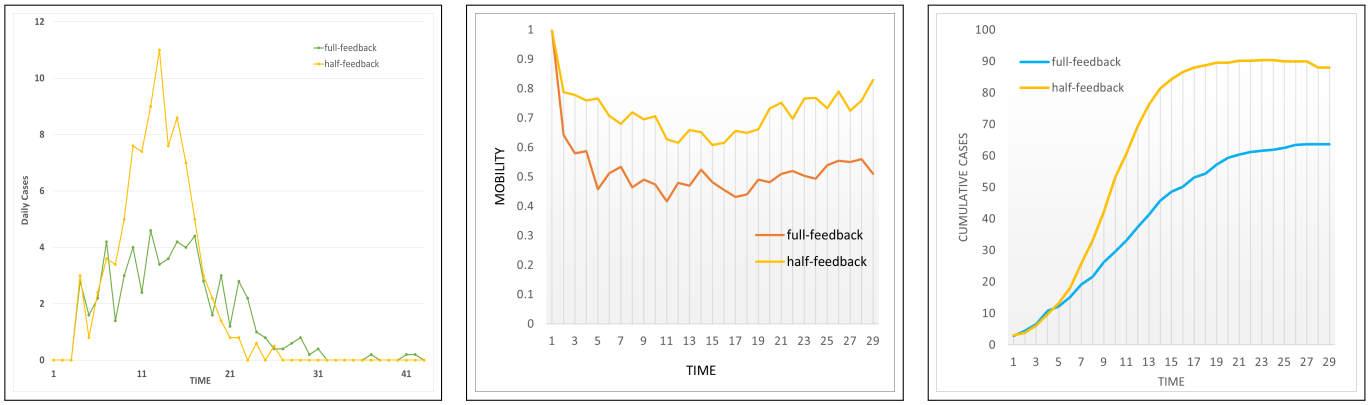
首先给出 Vicuna 驱动 GAI 模型时，全反馈情况下 (指代理人能够通过阅读新闻等方式获取疫情的详细信息) 统计的每日新增病例 Daily cases 以及人口流动率 Mobility。在小镇人口为 100,500 与 1000 的实验设置下，图 9a Daily new cases 表明小镇的代理人能够结合自身情况 (性格特征和健康状况) 以及疫情状况，对当前疫情局势做出及时反应。具体而言当疫情状况比较严重，例如每日新增病例数较高时，代理人可能会进行自我隔离，避免与其他人接触，图 9b 统计的小镇中代理人的流动性反应了这一现象。图 9c 统计的累计病例表明，通过代理人的自我隔离，小镇的疫情曲线趋于平缓，疫情得到了有效控制。




(a) Daily Cases (b) Mobility (c) Cumulative Cases

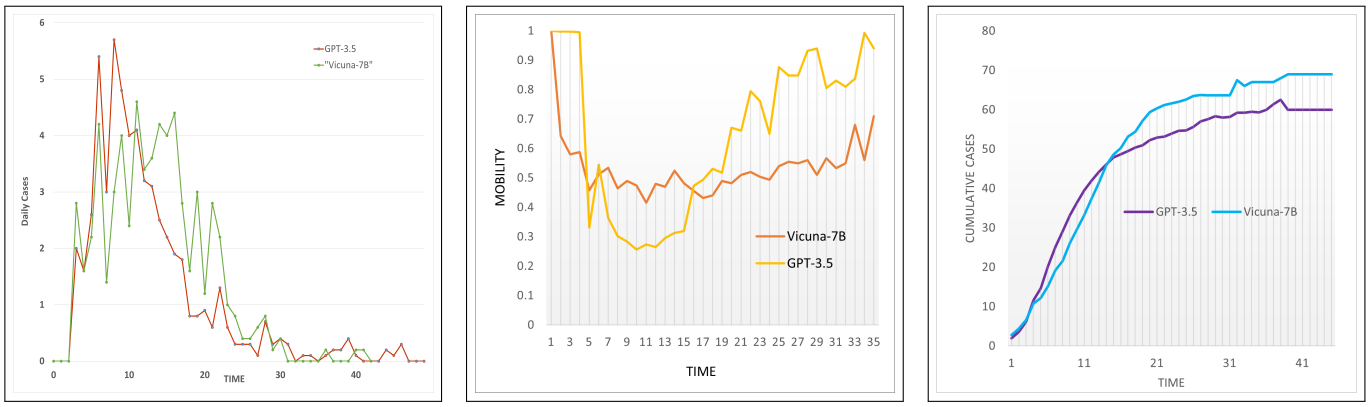
图 9: 100 Agents in town

接着我统计了半反馈与全反馈的实验数据来做对比实验。半反馈指代理人知道自己的基本信息以及自己的健康状况，但不知道小镇的疫情信息。图 10a 表明当代理人不了解小镇的疫情信息时，疫情每日新增病例较多，图 10b 表明半反馈的条件下多数代理人选择出门工作，图 10c 说明疫情没有得到有效控制，几乎 90% 的小镇人口感染了病毒。对比实验反应出疫情信息对代理人的推理决策十分重要，这与现实世界的情况也是符合的。




(a) Daily Cases (b) Mobility (c) Cumulative Cases

图 10: half and full feedback comparision



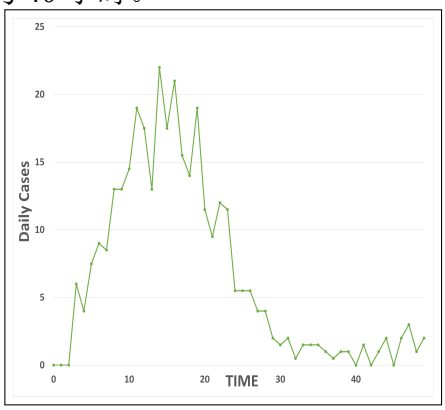

(a) Daily Cases (b) Mobility (c) Cumulative Cases

图 11: Vinuca-7B and GPT-3.5 feedback comparision

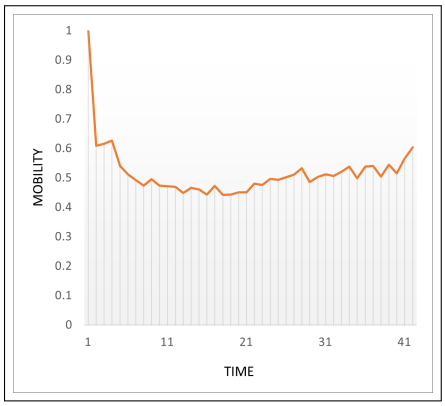
图 11 对比了 GPT-3.5 与 Vicuna-7B 驱动的 GAI 模型的实验数据,Vicuna 的曲线与 GPT-3.5 具有相同的趋势，且两者曲线比较贴合。Vicuna 虽然是本地部署的、参数只有七十亿的轻量级 LLM，但通过提示文本 (prompt) 设计以及实验配置，Vicuna 驱动的代理人依然可以结合各种环境与自身因素做出推理与决策，决定是否自我隔离，最终平缓整个小镇的疫情曲线，有效模拟真实世界的情况。GPT-3.5

能够处理的上下文容量比 Vicuna 多，能综合各项因素进行推理的能力也比 Vicuna 强。Vicuna 不能完全替代 GPT-3.5，但仍可以作为一种低成本的替代方案，

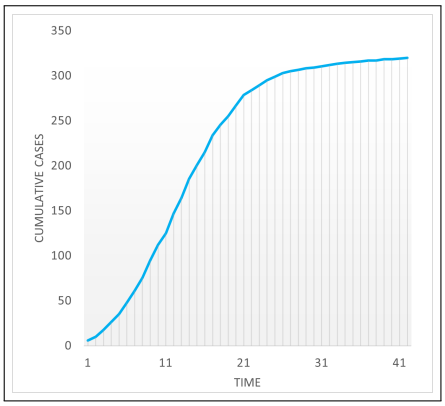
接着增大模型规模，给出小镇人口分别为 500,1000 的设置下的统计数据。他们在 Daily Cases、Mobility、Cumulative Cases 指标上具有相同的趋势。当小镇人口为 1000 时，进行一次模拟的时间约为 40 小时。




(a) Daily Cases

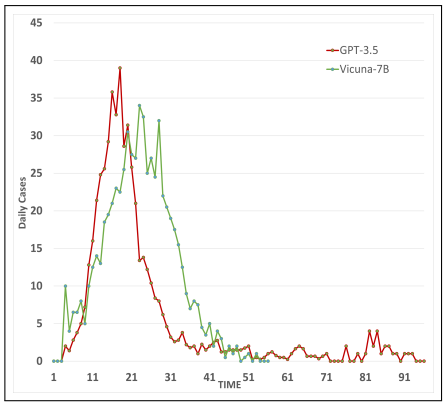



(b) Mobility

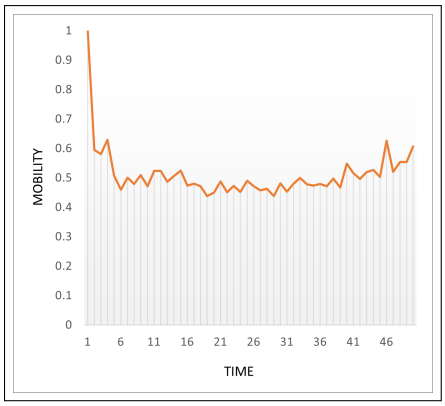



(c) Cumulative Cases

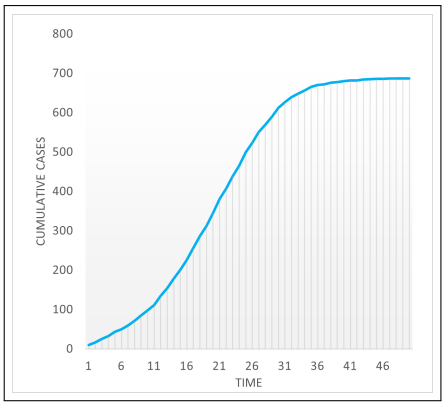
图 12: 500 Agents in town




(a) Daily Cases




(b) Mobility




(c) Cumulative Cases

图 13: 1000 Agents in town

## 6 总结与展望

本次实验复现了一种新的流行病建模方法 GABM，通过利用生成式人工智能，特别是 LLM，将人类的复杂行为融入传统的传播模型。该基于个体层面的建模模式中，每个代理人结合多种上下文信息 (如性格特征、健康状况、患病情况等) 以及 LLM 中自有的常识进行推理决策。LLM 不是为每个代理人指定决策规则，评估相应参数值得建模器，而是赋予每个代理人推理能力。原论文证明在没有外部施加决策规则等限制条件的情况下，生成式代理人可以做出与真实世界中人类行为一致的决策。

作者指出 GABM 存在运行成本高，耗时长的问题。截至 2023 年 6 月，一个由 1000 个代理人组成的流行病传播模型在配备 32GB RAM 的 CPU 上的运行时间可超过 90 小时，每次程序运行的成本约为 20 美元。这主要是由于对 OpenAI 服务器的数以万计的应用程序接口 (API) 调用导致的。作者在展望中提及了本地 LLMs 可能带来的改变，基于这一想法，在复现的过程中，我部署了本地运行的轻量级的 LLM(Vicuna-7B-v1.5) 替代 ChatGPT-3.5 来驱动 GAI 模型，赋予代理人推理能力。

从统计的实验数据来看，Vicuna 的在每日新增病例 Daily Cases、人口流动率 Mobility、累积病例 Cumulative Cases 等指标上表现出与 ChatGPT-3.5 相同的趋势，且数据曲线比较贴近。说明本地运行的

LLM(Vicuna) 在一定程度上可以胜任 ChatGPT-3.5 的工作，同时运行成本是可以忽略不计的电力、硬件成本。但 ChatGPT 能结合的上下文信息更多，综合复杂信息进行推理的能力也更强，本地部署的 LLMs 不能完全替代 GPT-3.5，但仍可以作为一种低成本的解决方案。

本次实验也存在不足之处，由于实验资源的限制，无法在本地部署能力更强、更接近 ChatGPT 的 Vicuna-13B 来驱动 GAI 模型。Vicuna-7B 不时的会做出明显错误的判断，有时推理和决策也不一致，为实验结果带来一定的误差。相信随着开源大模型的不断发展，基于生成式代理人的社交网络建模的成本能够逐渐减少，运行时间能够缩短，以推动更大规模、更深入的研究的开展。

## 参考文献


[1] TROITZSCH K G. Social science microsimulation[M]. Springer Science & Business Media, 1996.

[2] CHOPARD B, DROZ M. Cellular automata[J]. Modelling of Physical, 1998: 6-13.

[3] PARK J S, O’BRIEN J C, CAI C J, et al. Generative agents: Interactive simulacra of human behavior [J]. 2023. eprint: arXiv:2304.03442.

[4] KOLESAR P, WALKER W E. A simulation model of police patrol operations: program description[M]. New York City Rand Institute, 1975.

[5] MEADOWS D L, BEHRENS W W, et Al. Dynamics of growth in a finite world[M]. Wright-Allen Press Cambridge, 1974.

[6] MARSH L C, SCOVILL M. Title of the Conference Paper[C]//In NBER Workshop on Policy Analysis with Social Security Research Files. 1978: 15-17.

[7] BROWN T, MANN B, et Al. Language models are few-shot learners[C]//Advances in neural information processing systems. 2020: 33:1877-1901.

[8] OpenAI. Gpt-4 technical report[R]. 2023.

[9] Language modeling with PATHWAYS P S. Aakanksha Chowdhery and Sharan Narang and et al[J]. 2022. eprint: arXiv:2204.02311.

[10] ANIL R, et Al. Palm 2 technical report[R]. 2023.

[11] TOUVRON H, LAVRIL T, et Al. Llama: Open and efficient foundation language models[J]. 2023. eprint: arXiv:2302.13971.

[12] ZENG A, LIU X, et Al. Glm-130b: An open bilingual pre-trained model[J]. 2022. eprint: arXiv:2210.02414.

[13] AHER G V, ARRIAGA R I, et Al. Using large language models to simulate multiple humans and replicate human subject studies[C]//In International Conference on Machine Learning. 2023: 337-371.

[14] HORTON J J. Large language models as simulated economic agents: What can we learn from homo silicus?[R]. Technical report, National Bureau of Economic Research, 2023.

[15] HÄMÄLÄINEN P, TAVAST M, et Al. Evaluating large language models in generating synthetic hci research data: a case study[C]//In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023: 1-19.

[16] BRATLEY P, FOX B L, et Al. A guide to simulation[M]. 1987.

[17] GILBERT N, TROITZSCH K. Simulation for the social scientist[M]. McGraw-Hill Education (UK), 2005.

[18] AXELROD R. Advancing the art of simulation in the social sciences[J]. In Simulating social phenomena.Springer, 1997: 21-40.

[19] FORRESTER J W. ystem dynamics and the lessons of 35 years[J]. In A systems-based approach to policymaking.Springer, 1993: 199-240.

[20] SAMUELSON W, ZECKHAUSER R. Status quo bias in decision making[J]. Journal of risk and uncertainty, 1988: 1:7-59.

[21] CHARNESS G, RABIN M. Understanding social preferences with simple tests[J]. The quarterly journal of economics, 2002: 117(3):817-869.