

每天，结交新朋友，邀请对方出去参加聚会，并协调在合适的时间一起出现在聚会上。我们通过消融证明了我们的智能体架构的组成部分——观察、计划和反射——每一个都对智能体行为的可信度做出了重要贡献。通过将大型语言模型与计算交互代理融合，本工作引入了架构和交互模式，以实现可信的人类行为模拟。

## CCS的概念

- 以人为中心的计算→交互系统和工具;
- 计算方法→自然语言处理。

## 关键字

## 1介绍

我们怎样才能创造一个互动的人工社会来反映可信的人类行为呢？从《模拟人生》等沙盒游戏到认知模型[23]和虚拟环境[10, 59]等应用，四十多年来，研究人员和实践者一直在设想可以作为人类行为可信代理的计算代理。在这些愿景中，计算驱动的代理与他们过去的经验一致，并对他们的环境采取可信的行动。这种对人类行为的模拟可以用现实的社会现象填充虚拟空间和社区[27, 80]，训练人们如何处理困难的人际关系[44, 52, 94]，测试社会科学理论[12, 46]，为理论和可用性测试制作人类处理器模型[23, 39, 51]，为无处不在的计算应用程序提供动力[31]和社交机器人[10, 14]，并支持不可玩的游戏角色[59]。[85]能够在一个开放的世界中处理复杂的人际关系。然而，人类行为的空间是巨大而复杂的[85, 108]。尽管大型语言模型[18]在单个时间点模拟人类行为方面取得了显著进展[39, 80]，但确保长期一致性的全通用智能体更适合于管理不断增长的记忆（随着时间的推移，新的交互、冲突和事件出现和消退）的架构，同时处理多个智能体之间展开的级联社会动态。成功需要一种方法，可以在很长一段时间内检索相关事件和互动，反思这些记忆以概括和得出更高层次的推论，并应用推理来创建在当前和代理行为的长期弧中有意义的计划和反应。在本文中，我们引入了生成代理-代理绘制生成模型来模拟可信的人类行为-和

证明他们能产生可信的个人和突发群体行为的模拟。生成型智能体对自身、其他智能体及其环境做出各种各样的推断；他们制定日常计划，反映他们的特点和经验，执行这些计划，做出反应，并在适当的时候重新计划；当最终用户改变他们的环境或用自然语言命令他们时，他们会做出反应。例如，生成型智能体在看到自己的早餐在燃烧时关掉炉子，如果有人在浴室外面等着，当遇到另一个想要交谈的智能体时停下来聊天

一个充满生成主体的社会以新兴的社会动态为特征，在这个社会动态中，新的关系形成，信息扩散，主体之间的协调出现。为了实现生成智能体，我们描述了一个智能体架构，它存储、综合和应用相关的记忆，使用一个大型语言模型来生成可信的行为。我们的架构包括三个主要部分。第一个是记忆流，这是一个长期记忆模块，用自然语言记录了智能体经历的全面列表。记忆检索模型结合了相关性、近代性和重要性，以显示为代理的实时行为提供信息所需的记录。第二种是反射，随着时间的推移，它将记忆合成为更高层次的推论，使智能体能够得出关于自己和他人的结论，从而更好地指导自己的行为。第三是计划，将这些结论和当前环境转化为高层次的行动计划，然后递归地转化为行动和反应的详细行为。这些反思和计划被反馈到记忆流中，以影响代理的未来行为。这种架构建议应用于多个领域，从角色扮演和社交原型到虚拟世界和游戏。在社会角色扮演场景中（例如，面试准备），用户可以安全地排练困难的、充满冲突的对话。当对社交平台进行原型设计时，设计师可以超越暂时的人物角色，对动态的、复杂的交互进行原型设计，从而展开时间。在本文中，我们关注的是受《模拟人生》（the Sims 2）等游戏的启发，创造一个小型的、互动的代理社会的能力

通过将我们的架构连接到ChatGPT大型语言模型[77]，我们在游戏中展示了一个由25个智能体组成的社会。最终用户可以观察这些代理并与之交互。例如，如果终端用户或开发者希望小镇举办游戏中的情人节派对，传统的游戏环境需要手动编写数十个角色的行为脚本。我们证明，对于生成智能体，简单地告诉一个智能体她想举办一个聚会就足够了。尽管有许多潜在的失败点——派对策划者必须记住邀请其他代理人参加聚会，参与者必须记住邀请，记住的人必须决定实际出现，还有更多——我们的代理人成功了。他们把派对的事传开了，然后

当提到参与行动或去某个地方的生成代理时，这是可读性的简写，而不是暗示它们参与了类似人类的代理。我们的代理人的行为，类似于迪士尼动画人物，旨在创造一种可信度，但他们并不意味着真正的代理。可以在以下链接查看生成代理社会的实际模拟演示：[https://reverie.herokuapp.com/UIST\\_Demo/](https://reverie.herokuapp.com/UIST_Demo/)。模拟代码的公共存储库位于这里：[https://github.com/joonspk-research/generative\\_agents](https://github.com/joonspk-research/generative_agents)