

Intellect

Brain Games

OLIVER ROEDER

Keeping up with the ChatGPT neighbours

John Lin is a small-town pharmacist who takes great pride in his work. He enjoys helping people and making it easy for his customers to get their medicine. He lives with his wife, a college professor, and their son, a student of music theory. He is active in local politics and friendly with his neighbour. John Lin thinks his neighbour is a kind man. And he loves his family very much.

John Lin is also a chatbot, a souped-up version of ChatGPT. He “lives” in a digital hamlet called Smallville alongside two dozen other chatbots. He and his neighbours were placed there recently by a team of computer scientists from Stanford and Google to create a society with “believable human behaviour”. The modern generative AI movement has moved through text, image, audio and video. It is now attempting community.

“The main idea here is to ask ourselves, ‘Can we now create human-like agents that can populate an open-world setting?’” Joon Sung Park, a computer science PhD student at Stanford, told me. He’s the lead author of a recent paper describing the project. “If we poke at the right angle, we can retrieve believable human behaviour in a very narrow setting.”

Aesthetically, Smallville is a cute, cartoonish place inspired by *The Sims* video games. It has houses, shops, a café, a bar, a park, a college and a dorm. The buildings have rooms and the rooms have appliances and so on. It is inhabited by these AI



agents, which are augmented, crucially, with “memory streams” – long lists of past events and reflections. This allows them to consider and act upon their personal histories and motivations, far beyond the limited “context window” of the standard chatbot. These “generative agents” are aware of themselves, their fellow residents, the local establishments and their daily schedules. In other words, they are set loose to live their small-town lives.

From these lives emerges behaviour beyond the dictates of any Stanford programmer: the agents diffuse information, form relationships and co-ordinate. They wake, cook breakfast, go to work and school, shop, socialise, run for mayor and discuss mayoral candidates. And so Smallville’s social network complicates and densifies all on its own. The town enriches itself.

In one case, Smallville residents threw a Valentine’s Day party. They autonomously spread invitations, asked each other out on dates (Maria invited Klaus, her secret crush; he accepted!) and co-ordinated their arrivals on the day. They even decorated for the occasion, organised by café owner Isabella, though they were never instructed to decorate.

Practically, the hope is that “believable” agents like these could soon populate and enrich video games as realistic non-player characters, facilitate human-like role-playing and serve as “rehearsal spaces for interpersonal communication” – preparatory arenas for job interviews, say, or other difficult interactions (an idea explored

recently, using real-life subjects, by comedian Nathan Fielder in *The Rehearsal*).

Examined life in Smallville unfolded for just two days but, as technology improves and costs go down, Park hopes to populate larger AI towns that thrive for long periods – a thousand residents over multiple years. He also thinks John, Isabella and their friends could one day be valuable to social scientists. Much easier, of course, to experiment on and analyse generative agents than human beings. (And as Baudrillard wrote, “The simulacrum is never what hides the truth... The simulacrum is true.”)

Broadly, Park seeks to revive in Smallville a couple of academic fields that he sees as floundering. One is creating machines that believably simulate human behaviour, an arduous task that dates to the very earliest days of AI. The other is agent-based modelling, which aims to understand complex systems by simulating their component members.

But, as in any real small town, not everything was hunky-dory in Smallville. Tom knew he was supposed to chat with Isabella at the party, but was also unsure the party even existed. Isabella hallucinated Sam’s mayoral campaign announcement. Yuriko incorrectly believed her neighbour to be the 18th-century economist Adam Smith.

And some residents appeared to develop drinking problems, going to the local bar each day at noon. The researchers interpreted this as an error

in their system until they learnt that some humans do indeed go to the bar everyday at noon.

“I think there is something quite interesting and deep about that,” said Park. “This concept of what is believable, and what is actually human, is a very hard thing to define.” Residents of any small town or large city can testify to that.

The Smallville code was recently made open-source and placed on GitHub so anyone can find their own algorithmic town – an attractive prospect for any aspiring digital godhead. But Park also emphasised the now-familiar guard rails and warnings that ought to accompany AI: steering agents away from inappropriate behaviour, for exam-

ple, transparency about the underlying code and reminding users that John Lin is not sentient after all, despite his love for his family.

“The obvious concern here is if people form parasocial relationships with these agents, and that ends up replacing their actual human-to-human relationships,” said Park.

I also asked Park if, while building Smallville, he considered the possibility that someone else, his own creator, had built a simulation that he was now living in. “It’s a question that I never really bother thinking about,” he said. “Like, how would you test it?”

.....
 Oliver Roeder is a senior data journalist and author of “Seven Games: A Human History” (WW Norton)

The Smallville code was made open-source – an attractive prospect for any digital godhead

WORKLIFE
 By
AMY HWANG



“I LIKE FALL BECAUSE EVERYTHING GOES BACK TO THE NEW NORMAL”