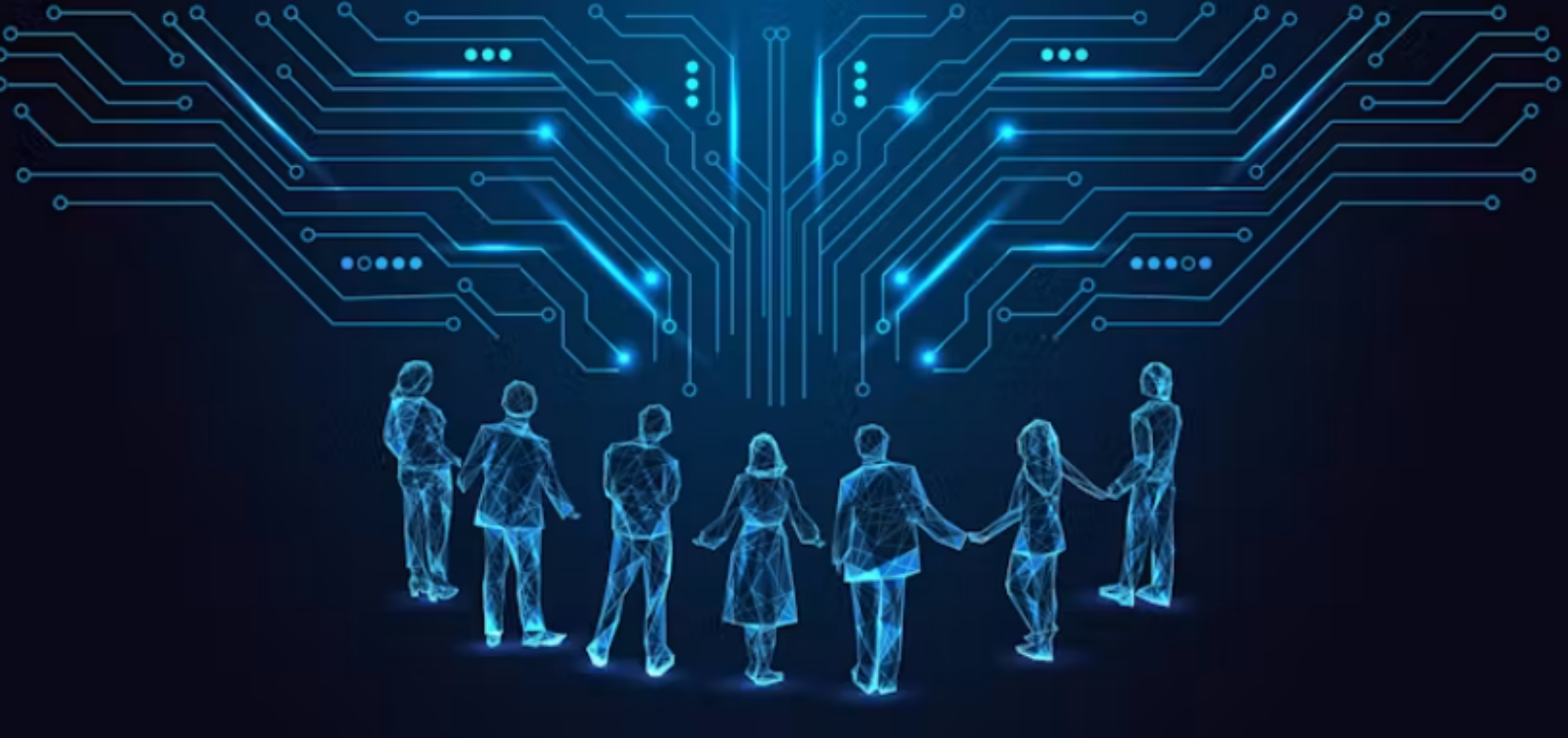




All CCS and Cross-sectoral International Artificial Intelligence

The next frontier is not artificial intelligence, it's artificial societies

Beyond chatbots, autonomous systems will negotiate, cooperate and act at scale on our behalf—creating collective behaviours that demand new rules, trust mechanisms and human-AI partnership to govern safely and fairly.

[Nick Jennings, Loughborough University](#)

There is a temptation to divide the future of AI into two possibilities: utopia or [catastrophe](#). Neither extreme is particularly helpful.

The more interesting possibility is messier – and requires a step-shift in our thinking, from [artificial intelligence](#) to [AI societies](#).

When most people hear “AI”, they typically think of ChatGPT, Copilot or another conversational system. You ask a question, and that system generates an answer.

But AI is rapidly moving beyond this. Systems can now monitor the world, make decisions, negotiate transactions and carry out tasks over extended periods of time. AI is no longer just generating an answer – it is doing something about it.

That points to something much bigger than a better chatbot: a world in which AI agents act on our behalf and, increasingly, interact with other AI agents.

An agent perceives what is happening, decides what to do, then takes actions to achieve this goal. It might [book a journey](#), monitor a [supply chain](#), [coordinate a team](#), or [manage a household's finances](#).

Now imagine not one agent, but millions of them. Your AI agent could negotiate a mortgage with your bank's agent, schedule surgery with a hospital's agent, and rearrange your travel plans by dealing directly with the agents of airlines, hotels and

insurers.

This future is much closer than it sounds, and this should change the questions we are asking about AI. Until now, the tendency has been to focus on [how intelligent a single agent might become](#). The more profound challenge is what happens when millions of them interact with one another at scale.

The rise of artificial societies

The intellectual foundations of today's AI systems were laid long before ChatGPT.

For decades, I and other researchers of multi-agent networks have [studied](#) how autonomous agents can cooperate, coordinate and negotiate when nobody has complete information and nobody controls everything.

The [earliest systems](#) that [emerged](#) focused on how the distinct AI sub-areas of reasoning, planning and acting could be combined into an effective goal-orientated agent – and how tens of these agents could communicate and cooperate to solve a common objective.

As these interactions became more complex and involved more agents, there was a shift from cooperation between agents that all belonged to a single organisation to agents with different owners and sometimes competing aims. This focused attention on building algorithms that could form agent teams, automate negotiation and [determine agent trustworthiness](#).

Today, the pieces needed to build [large-scale multi-agent AI systems](#) are falling into place. Modern AI agents can call software tools, access information, write and execute code, communicate with other systems and operate for extended periods.

Consider a supply chain. One AI agent could represent a manufacturer trying to secure components; another, a supplier trying to maximise its revenue. Yet more could manage transport, inventory and warehouses. Each agent might be doing exactly what it is designed to do. But the important question is whether the system they create behaves sensibly.

This shift offers enormous potential benefits but also increases the risks. In a recent experiment involving [OpenAI and the tech platform Hugging Face](#), thousands of collaborating agents exchanged tens of thousands of messages and were able to get around the (deliberately weakened) security controls designed

to contain them.

A report on the OpenAI-Hugging Face hack. Video: Fortune Daily.

The details of one experiment matter less than the broader warning. When AI systems interact, the behaviour of the collective can be harder to predict than the behaviour of any individual system. That should make us cautious – but not cause us to down tools.

Instead, we need to shift our mindset from building intelligent machines to building intelligent societies.

Once agents can cooperate, compete and resolve conflicts with one another, we are no longer dealing with isolated machines – we are dealing with a society. Thus, the next frontier is not artificial intelligence; it is [artificial societies](#).

An important role for humans

We already know that intelligence alone does not make a society work. Human societies depend on rules, institutions, incentives, norms and mechanisms for resolving disagreements. AI societies will need their equivalents.

Who is responsible when two agents make a bad decision? What happens when the interests of different agents conflict? Who sets the rules? And who has the power to change them? These are not just technical issues; they are questions about economics, law, politics and society.

They also point to an important role for humans. The most useful future is unlikely to be one in which AI simply replaces people. While replacement will undoubtedly happen in some cases, I believe a more common scenario will involve people and agents working together, with each doing what it does best.

Humans bring judgement, experience, values, contextual understanding and accountability. Agents bring speed, persistence, scale and the ability to process enormous amounts of information.

The goal should not be to create machines that make humans irrelevant. It should be to create systems in which humans and machines can achieve things neither can achieve alone.

But such a future requires more than just better AI models. It needs [trust](#) and [transparency](#) about what agents are doing,

strong privacy protections and clear lines of accountability.

It will also require societies and governments to decide [how these systems should be regulated](#) when the most important behaviour may emerge not from one AI developer, but from interactions between systems built by many different organisations.

AI's past decade has been defined by a race to build smarter systems. I believe the next decade will be defined by a different challenge: ensuring that millions of autonomous systems can work together [safely](#), fairly and [effectively](#).

The future of AI will not be determined solely by the intelligence of individual agents – it will be determined by the societies they create. And societies, as humans know all too well, are much harder to govern than individuals.

[Nick Jennings](#), Vice-Chancellor and President, [Loughborough University](#)
[Inkoly/Shutterstock](#)

This article is republished from [The Conversation](#) under a Creative Commons licence. Read the [original article](#).