



If society is in denial or simply isn't paying attention, then the chance to shape artificial general intelligence could be lost forever. PHOTO: ISTOCK

A new type of AI will make trillions, reorder the world

Technology There are three arguments for taking progress towards artificial general intelligence more seriously, writes Kevin Roose.

Here are some things I believe about artificial intelligence: I believe that over the past several years, AI systems have started surpassing humans in a number of domains – maths, coding and medical diagnosis, just to name a few – and that they're getting better every day.

I believe that very soon – probably in 2026 or 2027, but possibly as soon as this year – one or more AI companies will claim they've created an artificial general intelligence, or AGI, which is usually defined as something like "a general-purpose AI system that can do almost all cognitive tasks a human can do".

I believe that when AGI is announced, there will be debates over definitions and arguments about whether or not it counts as "real" AGI, but that these mostly won't matter because the broader point – that we are losing our monopoly on human-level intelligence, and transitioning to a world with very powerful AI systems in it – will be true.

I believe that over the next decade, powerful AI will generate trillions of dollars in economic value and tilt the balance of political and military power towards the nations that control it – and that most governments and big corpor-

ations already view this as obvious, as evidenced by the huge sums of money they're spending to get there first.

I believe that most people and institutions are totally unprepared for the AI systems that exist today, let alone more powerful ones, and that there is no realistic plan at any level of government to mitigate the risks or capture the benefits of these systems.

I believe that hardened AI sceptics – who insist that the progress is all smoke and mirrors, and who dismiss AGI as a delusional fantasy – not only are wrong on the merits, but are giving people a false sense of security.

I believe that whether you think AGI will be great or terrible for humanity – and honestly, it may be too early to say – its arrival raises important economic, political and technological questions to which we currently have no answers.

I believe that the right time to start preparing for AGI is now.

This may all sound crazy. But I didn't arrive at these views as a starry-eyed futurist, an investor hyping my AI portfolio or a guy who took too many magic mushrooms and watched *Terminator 2*.

I arrived at them as a journalist who has spent a lot of time talking to the engineers building powerful AI systems, the investors funding it and the researchers studying its effects. And I've come to believe that what's happen-



OpenAI chief executive Sam Altman (left) and his counterpart at Google DeepMind, Demis Hassabis. PHOTOS: BLOMBERG

ing in AI right now is bigger than most people understand.

In San Francisco, where I'm based, the idea of AGI isn't fringe or exotic. People here talk about "feeling the AGI," and building smarter-than-human AI systems has become the explicit goal of some of Silicon Valley's biggest companies. Every week, I meet engineers and entrepreneurs working on AI who tell me that change – big change, world-shaking change, the kind of transformation we've never seen before – is just around the corner.

"Over the past year or two, what used to be called 'short timelines' (thinking

that AGI would probably be built this decade) has become a near-consensus," Miles Brundage, an independent AI policy researcher who left OpenAI last year, told me recently.

Outside the Bay Area, few people have even heard of AGI, let alone started planning for it. And in my industry, journalists who take AI progress seriously still risk getting mocked as gullible dupes or industry skills.

Honestly, I get the reaction. Even

though we now have AI systems contributing to Nobel Prize-winning breakthroughs, and even though 400 million people a week are using ChatGPT, a lot of the AI that people encounter in their daily lives is a nuisance. I sympathise with people who see AI slop plastered all over their Facebook feeds, or have a clumsy interaction with a customer service chatbot and think: *This is what's going to take over the world?*

I used to scoff at the idea, too. But I've come to believe that I was wrong. A few things have persuaded me to take AI progress more seriously.

The most disorienting thing about today's AI industry is that the people closest to the technology – the employees and executives of the leading AI labs – tend to be the most worried about how fast it's improving.

This is quite unusual. Back in 2010, when I was covering the rise of social media, nobody inside Twitter, Foursquare or Pinterest was warning that their apps could cause societal chaos. Mark Zuckerberg wasn't testing

Facebook to find evidence that it could be used to create novel bioweapons, or carry out autonomous cyberattacks.

But today, the people with the best information about AI progress – the people building powerful AI, who have access to more advanced systems than the public sees – are telling us that big change is near. The leading AI companies are preparing for AGI's arrival, and studying potentially scary properties of their models, such as whether they're capable of scheming and deception, in anticipation of their becoming more capable and autonomous.

Sam Altman, the chief executive of OpenAI, wrote that "systems that start to point to AGI are coming into view".

Demis Hassabis, the CEO of Google DeepMind, said AGI is probably "three to five years away".

Dario Amodei, the chief executive of Anthropic (who doesn't like the term AGI but agrees with the general principle), told me last month that he believed we were a year or two away from having "a very large number of AI systems that are much smarter than humans at almost everything".

Maybe we should discount these predictions. After all, AI executives stand to profit from inflated AGI hype, and might have incentives to exaggerate.

But lots of independent experts – including Geoffrey Hinton and Yoshua Bengio, two of the world's most influential AI researchers, and Ben Buchanan, who was the Biden administration's top AI expert – are saying similar things. So are a host of other prominent economists, mathematicians and national security officials.

To be fair, some experts doubt that AGI is imminent. But even if you ignore everyone who works at AI companies, or has a vested stake in the outcome, there are still enough credible independent voices with short AGI timelines that we should take them seriously.

To me, just as persuasive as expert opinion is the evidence that today's AI systems are improving quickly, in ways that are fairly obvious to anyone who uses them.

In 2022, when OpenAI released ChatGPT, the leading AI models struggled with basic arithmetic, frequently failed at complex reasoning problems and often "hallucinated", or made up nonexistent facts. Chatbots from that era could do impressive things with the right prompting, but you'd never use one for anything critically important.

Today's AI models are much better. Now, specialised models are putting up medallist-level scores on the International Math Olympiad, and general-purpose models have become so good at complex problem-solving that we've had to create new, harder tests to measure their capabilities. Hallucinations and factual mistakes still happen, but they're rarer on newer models. And many businesses now trust AI models enough to build them into core, customer-facing functions.

(*The New York Times* has sued OpenAI and its partner, Microsoft, accusing them of copyright infringement of news content related to AI systems. OpenAI and Microsoft have denied the claims.)

Some of the improvement is a function of scale. In AI, bigger models, trained using more data and processing power, tend to produce better results, and today's leading models are significantly bigger than their predecessors. But it also stems from breakthroughs that AI researchers have made in recent years – most notably, the advent of "reasoning" models, which are built to take an additional computational step before giving a response.

Reasoning models, which include OpenAI's o1 and DeepSeek's R1, are trained to work through complex problems, and are built using reinforcement learning – a technique that was used to teach AI to play the board game Go at a superhuman level. They appear to be succeeding at things that tripped up previous models. (Just one example: GPT-4o, a standard model released by OpenAI, scored 9 per cent on AIME 2024, a set of extremely hard competi-

If you really want to grasp how much better AI has gotten ... talk to a programmer.

tion math problems; ol, a reasoning model that OpenAI released several months later, scored 74 per cent on the same test.)

As these tools improve, they are becoming useful for many kinds of white-collar knowledge work. My *Times* colleague Ezra Klein recently wrote that the outputs of ChatGPT's Deep Research, a premium feature that produces complex analytical briefs, were "at least the median" of the human researchers he'd worked with.

I've also found many uses for AI tools in my work. I don't use AI to write my columns, but I use it for lots of other things – preparing for interviews, summarising research papers, building personalised apps to help me with administrative tasks. None of this was possible a few years ago. And I find it implausible that anyone who uses these systems regularly for serious work could conclude that they've hit a plateau.

If you really want to grasp how much better AI has become recently, talk to a programmer. A year or two ago, AI coding tools existed, but were aimed more at speeding up human coders than at replacing them. Today, software engineers tell me that AI does most of the actual coding for them, and that they increasingly feel that their job is to supervise the AI systems.

Jared Friedman, a partner at Y Combinator, a startup accelerator, recently said a quarter of the accelerator's current batch of startups were using AI to write nearly all their code.

"A year ago, they would've built their product from scratch – but now 95 per cent of it is built by an AI," he said.

In the spirit of epistemic humility, I should say that I, and many others, could be wrong about our timelines.

Maybe AI progress will hit a bottleneck we weren't expecting – an energy shortage that prevents AI companies from building bigger data centres, or limited access to the powerful chips used to train AI models. Maybe today's model architectures and training techniques can't take us all the way to AGI, and more breakthroughs are needed.

But even if AGI arrives a decade later than I expect – in 2036, rather than 2026 – I believe we should start preparing for it now.

Most of the advice I've heard for how institutions should prepare for AGI boils down to things we should be doing anyway: modernising our energy infrastructure, hardening our cybersecurity defences, speeding up the approval pipeline for AI-designed drugs, writing regulations to prevent the most serious AI harms, teaching AI literacy in schools and prioritising social and emotional development over soon-to-be-obsolete technical skills. These are all sensible ideas, with or without AGI.

Some tech leaders worry that premature fears about AGI will cause us to regulate AI too aggressively. But the Trump administration has signalled that it wants to speed up AI development, not slow it down. And enough money is being spent to create the next generation of AI models – hundreds of billions of dollars, with more on the way – that it seems unlikely that leading AI companies will pump the brakes voluntarily.

I don't worry about individuals overpreparing for AGI, either. A bigger risk, I think, is that most people won't realise that powerful AI is here until it's staring them in the face – eliminating their job, ensnaring them in a scam, harming them or someone they love. This is, roughly, what happened during the social media era, when we failed to recognise the risks of tools such as Facebook and Twitter until they were too big and entrenched to change.

That's why I believe in taking the possibility of AGI seriously now, even if we don't know exactly when it will arrive or precisely what form it will take.

If we're in denial – or if we're simply not paying attention – we could lose the chance to shape this technology when it matters most. **AFR**

THE NEW YORK TIMES.