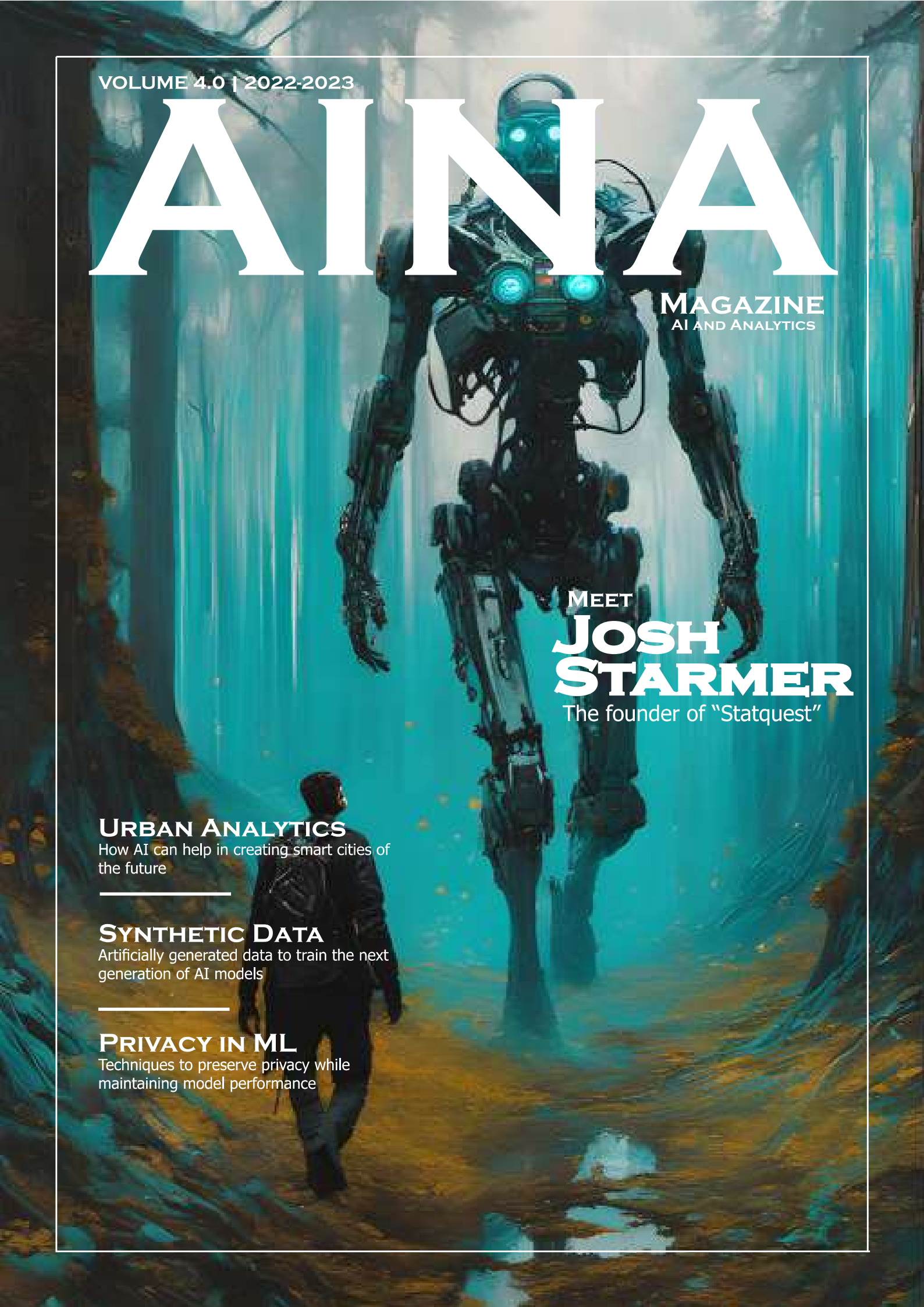
VOLUME 4.0 | 2022-2023

AINA

Magazine

AI and Analytics



Meet

Josh Starmer

The founder of "Statquest"

Urban Analytics

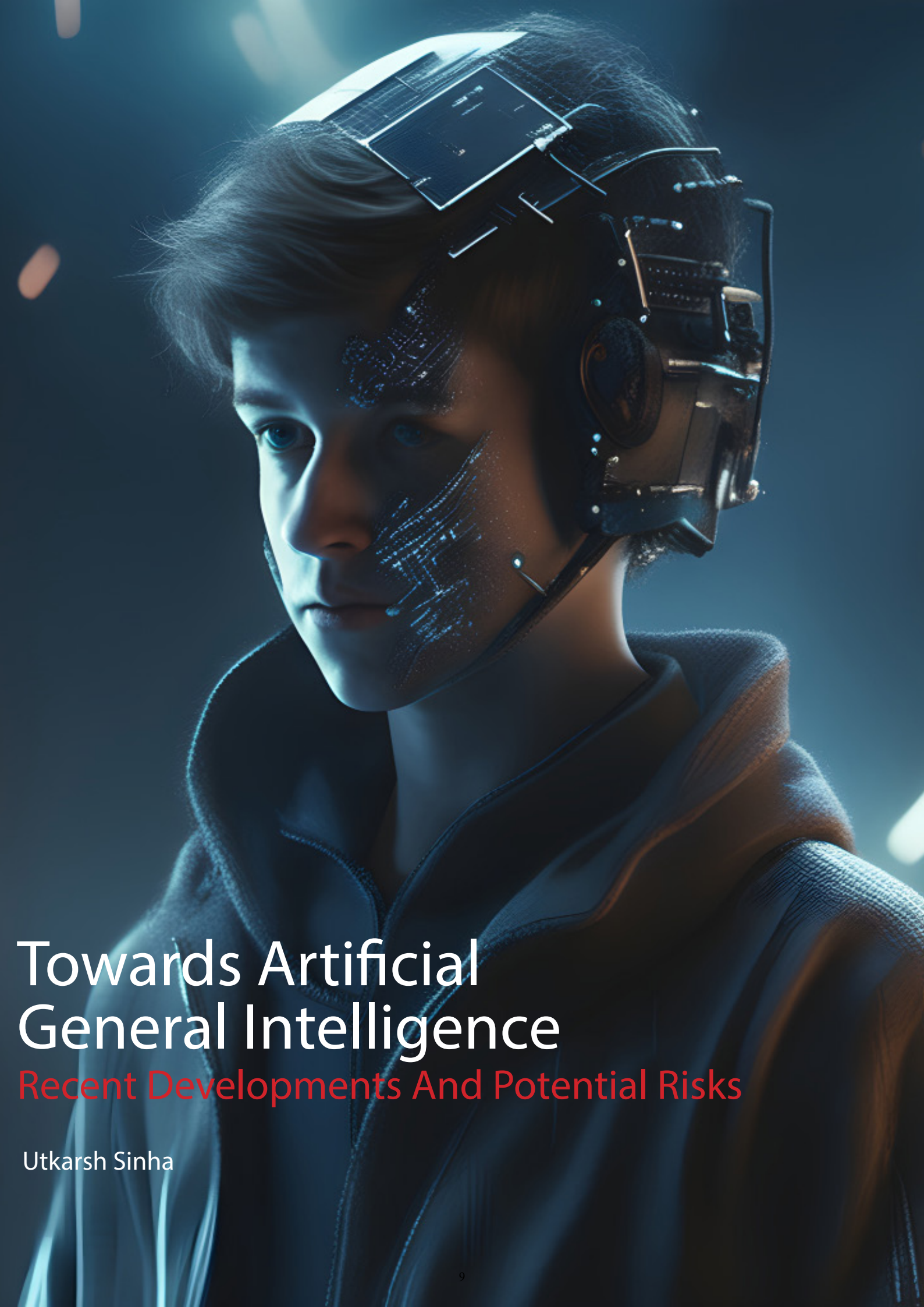
How AI can help in creating smart cities of the future

Synthetic Data

Artificially generated data to train the next generation of AI models

Privacy in ML

Techniques to preserve privacy while maintaining model performance

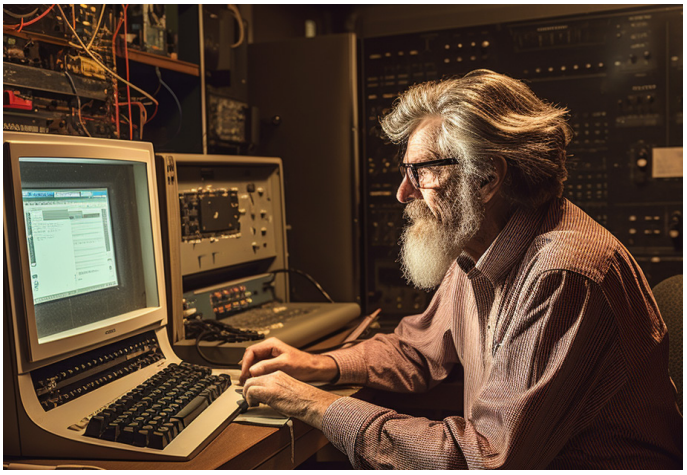


Towards Artificial General Intelligence

Recent Developments And Potential Risks

Utkarsh Sinha

In the history of evolution of life on earth, the emergence of humans represents a fleeting moment in a vast expanse of time. Yet, in the short amount of time we have been on the planet, we have managed to become the dominant species. The primary reason for this rapid ascent of the food chain is our intelligence. Given this fact, it is not surprising that we have always been fascinated by how intelligence works, and if it is something we can create or enhance. There have been several examples of fictitious creatures representing "artificial intelligence" throughout our history - from the myth of the Homunculus residing inside our body to modern-day stories like The Terminator. In the quest to develop artificial intelligence, the esteemed Professor John McCarthy initiated the modern version of this field as a research project at Stanford University in 1956. He believed that human intelligence isn't something that humans can never understand.



"I don't see that human intelligence is something that humans can never understand."
- Prof John McCarthy

The scientific community has worked relentlessly in the decades since then, and we have made major strides in the development of artificial intelligence. Today, we might be on the cusp of achieving the holy grail of AI – developing an "Artificial General Intelligence" (AGI). A machine which can perform any cognitive task we can conceive of - at or above human level. While current AI applications excel at specific functions, they lack general intelligence and cannot adapt to different tasks. For example, AI systems may outperform humans in one task but be outperformed by a child in another. Unlike AI, humans possess general intelligence, allowing them to perform a wide range of tasks and learn with fewer experiences. AGI aims to replicate this capability, enabling machines to learn and apply knowledge across various domains.

There is no consensus on when AGI will be realized. Estimates range from near-future to 100 years later to never, depending on the source. In any case, the development of such a machine will have profound impacts on our society – a society where cognitive and physical labour is abundant. It could create a self-improving machine that might lead to an intelligence explosion. The resulting artificial superintelligence (ASI) could either spell doom for our society or create a post-scarcity civilisation that would make present-day look like the Middle Ages in comparison. Borrowing the words of author James Barrat, the first true AGI may very well be "our final invention".

We will look at what form such an AGI might take and when based on the developments happening in the field at present. Would it be a completely digital entity, possess a physical body or interface with the human brain and enhance our intelligence? Or would it require some entirely new kind of approach that is not yet mainstream? How dangerous could it be in the hands of rogue actors, and what can governments do to mitigate this risk? Such questions, that existed only in the realm of science fiction just a few years ago, are becoming fundamentally important now. Let's try and explore these questions, starting with a look at the state of the art in the development of AGI.

"Of course, machines can't think as people do. A machine is different from a person, hence they think differently. The interesting question is, just because something is thinking differently from you, does that mean it's not thinking?"

– The Imitation Game

An Intelligent Digital Assistant

Imagine a future with computers that can take commands in conversational language and use the tools we use today to accomplish the task described to them. Indecisive about dinner? Snap a pic of your fridge for AI-generated recipes. Seeking a used cot on OLX? Your AI assistant will shortlist options, contacts owners, and schedules meetings. Office tasks like updating Salesforce or handling code documentation will become a breeze with one command to your AI office assistant. Imagine a visually impaired person walking down the aisle of a supermarket. The AI assistant could be their eyes, taking the video stream of the aisle on their mobile and directing them to the correct section of the mart to get what they need.

Microsoft launched Windows co-pilot: A natural language interface for all office tools

These capabilities may not be constrained to the realm of imagination much longer. Companies like Adept AI are already working on models that can use tools like Photoshop, Excel, Tableau etc. and work with you through a natural language interface. Microsoft recently launched Windows co-pilot, which aims to provide a natural language interface to all Microsoft Office tools. "BeMyEyes", a mobile app that connects visually impaired users to volunteers, is testing a beta version of its virtual assistant. Such applications show us the enormous positive impact that AGI can have on society - from freeing us up from the drudgery of work to helping members of society regain a sense of independence

It is important to note that these AI assistants show the capability of "multi-modality". They can see and identify objects in the real world, converse with humans in natural language, and understand and generate sound. The recent merging of the three modes of interaction is a big step towards achieving AGI. This, however, leads us to an open question in AI – can a true AGI be created without it ever interacting with the physical world? Some scientists believe so, and they are working towards what is known as "AI Embodiment".

AI Embodiment

Embodiment is the idea that an intelligent agent has a physical body that can interact with its environment. There is no consensus among researchers on whether embodiment is indispensable for AGI. Embodiment enables the AI to learn from sensory inputs and allows it to explore and experiment with physical objects in the real world. In a sense, we already have a limited version of embodied AI with self-driving cars. However, an embodied AGI must have greater capabilities than navigating roads. It should be able to interact with humans in their natural language and be able to perform tasks it was not specifically trained for. The embodied AI doesn't need to have a humanoid form either. Think R2-D2 and BB-8 from Star Wars or the robot assistants from Interstellar. They could be low-powered robots with designs suited for specific tasks.

Embodied AI could be of great use to humans in space exploration

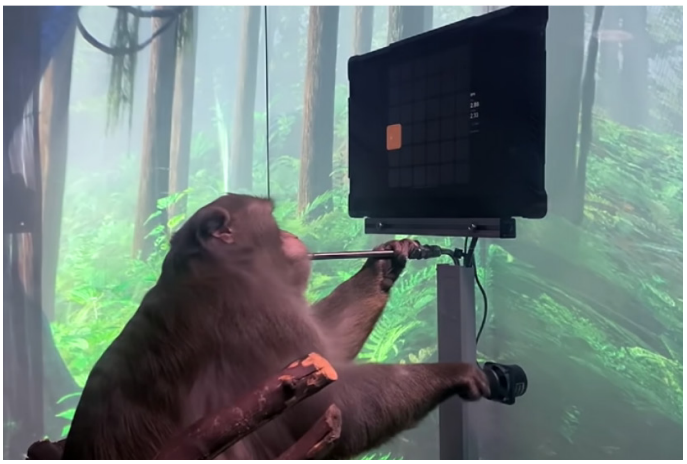
The humanoid robotics market is projected to reach $17.3 billion by 2027, driving intense competition for developing advanced humanoid robots. Google introduced the Palm-E (E for Embodied) model in March 2023, capable of solving diverse tasks across different robots and modalities. Around the same time, Open AI invested $23.5 million in 1X, focusing on androids collaborating with humans. Elon Musk announced Tesla Optimus at Tesla's inaugural AI day in August 2021.

Ultimately, the likelihood of AGI being realized as a humanoid robot appears high, given the increasing investments by tech giants in embodied AI systems. Whether it will be the final form remains uncertain and can only be determined with time. There is, however, a third possibility – that of humans merging with AI through a "brain-computer interface".

Brain-Computer Interface

A brain-computer interface (BCI) is a direct communication pathway between the brain's electrical activity and an external device, most commonly a computer or robotic limb. A BCI can translate a person's brain activity into external responses or directives, such as controlling a prosthetic limb with their thoughts. BCIs can also be used for augmenting or repairing human cognitive functions.

There have been successful experiments on Parkinson's patients that regained mobility through stimulation from a brain implant. Experiments on rats have shown the potential of implants being able to augment memory and transfer information using a brain-to-brain interface. Companies like Neuralink and Synchron are planning to start human trials of their implants.



Monkey playing video games with its mind using Neuralink's brain computer interface

However, these current applications only scratch the surface of what is possible. In the future, brain-computer interfaces could allow humans to leverage the strengths of digital computing, surpassing the capabilities of the unaugmented human brain. Today, smartphones act as extensions of ourselves, but limited data transmission between brains and mobiles hinders their full potential. Overcoming this limitation could enable direct access to machine intelligence, resulting in a collaborative "superintelligence" surpassing any previous cognitive abilities.

Would the creation of such a superintelligence be possible? It is too early to tell. The technology needed for humans to merge with AI might still be decades away. As we approach the development of AGI, the "misalignment problem" arises: its goals may not align with humanity's, posing an existential threat. BCIs can serve as a hedge, potentially granting us more control over AI by merging with it.

The approaches discussed till now assume the development of AGI in the form of a self-learning software. However, some scientists are working towards emulating the human brain through specialized hardware, in the nascent but promising new field of "Neuromorphic Computing".

Timeline and Form of AGI Arrival

The timeline for the arrival of the first true AGI is difficult to predict, with estimates ranging from as early as 2033 to as late as 2200, but most experts agree it will eventually happen. We saw last year that an increase in model size led to a breakthrough in performance of GPT 3.5 compared to GPT 3. Another 2 or 3 breakthroughs of similar magnitude may well lead us to achieving AGI in the form of a digital assistant, but the timing of such breakthroughs remain uncertain. Also, the ability of these chat bots to learn entirely new information by looking at new text hasn't been clearly demonstrated yet. Some argue that these

"One of the skills of AGI is not any particular milestone, but the meta skill of learning to figure things out. It can go and decide to get good at whatever you need."

– Sam Altman

bots are merely stochastic parrots imitating the patterns and language of the data they were trained on.

Embodied AGI is comparatively more difficult to build. The humanoid robots of today have limited generalizability in terms of the tasks they can perform. The potential job displacement in blue-collar work due to embodied AGI may also encourage governments to deter companies from making such an AGI on a mass scale. Hence, the development of AGI in this form is likely to take longer than in its digital form. Brain-computer interfaces have potential to become AGI, but their immediate use case is focused on curing neurological disorders rather than enhancing healthy individuals. The slow adoption and potential health risks may make this technology the least likely scenario for achieving AGI, although this approach gives us the best chance of achieving AI alignment. Other novel approaches, like emulating the hardware of the whole brain (neuromorphic AI), are still in their early stages and have a longer way to go to reach AGI compared to the previously discussed approaches.

Moving Forward: AGI Societal Impact and Risk Mitigation

Recently, there has been an active discourse among researchers and tech CEOs regarding the societal impact of AGI and ways to mitigate its risks. Proponents of AGI like Andrew Ng and Yan LeCun highlighted its positive contribution. On the other hand, notable figures like Elon Musk and Steve Wozniak have warned about its risks, calling for a pause on "Giant AI experiments". Geoffrey Hinton expressed concerns about the dangers of AI and subsequently left Google. Open AI and DeepMind published articles addressing existential risks and the need for governance and alignment of superintelligence. Bill Gates also posted on his blog comparing the development of AI to the development of the Graphic User Interface in its importance.

The pace of AGI development is accelerating, with advancements becoming more and more frequent. This is giving rise to risks like copyright infringement, learning from biased data, and helping rogue actors in cyberattacks and weapon development. These risks demand serious discussion and proactive mitigation. As we strive for AGI, important questions arise: Can big tech companies prioritize ethics and safety? Will the goals of a superintelligent being align with humanity's? How can AGI development be effectively regulated?

Big tech's proactive engagement with the government is promising, but uncertainty persists about regulations keeping up with AI advancements. Historical cases, such as banning of CFCs and nuclear power regulation after accidents, show that governments are not proactive in technology regulation. Artificial General Intelligence may not allow for such delayed regulation.

"Digital Superintelligence may be the biggest existential threat humanity has ever faced"

- Elon Musk

Open AI Charter: "OpenAI's mission is to ensure that artificial general intelligence (AGI), by which we mean highly autonomous systems that outperform humans at most economically valuable work benefits all of humanity."

The development of AGI can be compared to the development of nuclear weapons. The country that develops it first will be at a decisive advantage. However, other countries will soon follow suit. The possible military capabilities developed as a result of AGI may be comparable to other weapons of mass destruction in its power. This warrants an agency like the IAEA to regulate AI at an international level. In fact, Open AI CEO Sam Altman has made such a suggestion himself.

Ultimately, like the Industrial Revolution granted us physical power beyond what the human body was capable of, AGI may grant us mental abilities beyond that which our minds are capable of. The industrial revolution was responsible for a quantum leap in the growth rate of the world economy. A similar jump in the growth rate of our economy may lead to a level of prosperity beyond anything we can imagine. Simultaneously, it may also lead to an increase in our destructive capabilities. Hence, we may be standing at an inflection point in human history. Our actions at this time might be remembered for generations in the future if we get it right. If we get it wrong, however, there might not be any future generations left to remember us.

The Prime Minister Rishi Sunak meets with Demis Hassabis, CEO DeepMind, Dario Amodei, CEO Anthropic, and Sam Altman, CEO OpenAI, in 10 Downing Street