Race to AI: the origins of artificial intelligence, from Turing to ChatGPT

In the winter of 1958, a 30-year-old psychologist named Frank Rosenblatt was en route from Cornell University to the Office of Naval Research in Washington DC when he stopped for coffee with a journalist. Rosenblatt had unveiled a remarkable invention that, in the nascent days of computing, created quite a stir. It was, he declared, "the first machine which is capable of having an original idea". Rosenblatt's brainchild was the Perceptron, a program inspired by human neurons that ran on a state-of-the-art computer: a five-tonne IBM mainframe the size of a wall. Feed the Perceptron a pile of punch cards and it could learn to distinguish those marked on the left from those marked on the right. Put aside for one moment the mundanity of the task; the machine was able to learn. Rosenblatt believed it was the dawn of a new era and the New Yorker evidently agreed. "It strikes us as the first serious rival to the human brain," the journalist wrote. Asked what the Perceptron could not do, Rosenblatt mentioned love, hope and despair. "Human nature, in short," he said. "If we don't understand the human sex drive, why should we expect a machine to?" The Perceptron was the first neural network, a rudimentary version of the profoundly more complex "deep" neural networks behind much of modern artificial intelligence (AI). But nearly 70 years on, there is still no serious rival to the human brain. "What we've got today are artificial parrots," says Prof Mark Girolami, chief scientist at the Alan Turing Institute, in London. "That in itself is a fantastic advance, it will give us great tools for the good of humanity, but let's not run away with ourselves." The history of AI, at least as written today, has no shortage of fathers. Many sired the same offspring. Rosenblatt is sometimes referred to as the father of deep learning, a title shared with three other men. Alan Turing, the wartime codebreaker at Bletchley Park and founder of computer science, is considered a father of AI. He was one of the first people to take seriously the idea that computers could think. In a 1948 report, Intelligent Machinery, Turing surveyed how machines may mimic intelligent behaviour. One route to a "thinking machine", he mused, was to replace a person's parts with machinery: cameras for eyes, microphones for ears, and "some sort of electronic brain". To find things out for itself, the machine "should be allowed to roam the countryside", Turing quipped. "The danger to the ordinary citizen would be serious," he noted, dismissing the idea as too slow and impractical. But many of Turing's ideas have stuck. Machines could learn just as children learn, he said, with help from rewards and punishments. Some machines could modify themselves by rewriting their own code. Today, machine learning, rewards and modifications are basic concepts in AI. As a means of marking progress towards thinking machines, Turing proposed the Imitation Game, commonly known as the Turing test, which rests on whether a human can discern whether a set of written exchanges come from a human or a machine. It's an ingenious test, but attempts to pass it have fuelled immense confusion. In one recent eyebrow raiser, researchers claimed to have passed the test with a chatbot that claimed to be a 13-year-old Ukrainian with a pet guinea

pig that squealed Beethoven's Ode to Joy. Turing made another hefty contribution to AI that is often overlooked, Girolami says. A declassified paper from the scientist's time at Bletchley Park reveals how he drew on a method called Bayesian statistics to decode encrypted messages. Word by word, Turing and his team used the statistics to answer questions such as: "What is the probability that this particular German word generated this encrypted set of letters?" A similar Bayesian approach now powers generative AI programs to produce essays, works of art and images of people that never existed. "There's been a whole parallel universe of activity on Bayesian statistics over the past 70 years that completely enabled the generative AI we see today, and we can trace that all the way back to Turing's work on encryption," says Girolami. The term "artificial intelligence" didn't appear until 1955. John McCarthy, a computer scientist at Dartmouth College, in New Hampshire, used the phrase in a proposal for a summer school. He was supremely optimistic about the prospects for progress. "We think that a significant advance can be made … if a carefully selected group of scientists work on it together for a summer," he wrote. "This is the postwar period," says Dr Jonnie Penn, an associate teaching professor of AI ethics at the University of Cambridge. "The US government had understood nuclear weapons to have won the war. So science and technology could not have been on a higher high." In the event, those gathered made negligible progress. Nevertheless, researchers threw themselves into a golden age of building programs and sensors that equipped computers to perceive and respond to their environments, to solve problems and plan tasks, and grapple with human language. Computerised robots carried out commands made in plain English on clunky cathode-ray tube monitors, while labs demonstrated robots that trundled around bumping into desks and filing cabinets. Speaking to Life magazine in 1970, the Massachusetts Institute of Technology's Marvin Minsky, a towering figure in AI, said that in three to eight years the world would have a machine with the general intelligence of an average human. It would be able to read Shakespeare, grease a car, tell jokes, play office politics and even have a fight. Within months, by teaching itself, its powers would be "incalculable". The bubble burst in the 1970s. In the UK, Sir James Lighthill, an eminent mathematician, wrote a scathing report on AI's meagre progress, triggering immediate funding cuts. The revival came with a new wave of scientists who saw knowledge as the solution to AI's woes. They aimed to code human expertise directly into computers. The most ambitious – though other words have been used – was Cyc. It was intended to possess all the knowledge that an educated person used in their daily life. That meant coding in the lot, but getting experts to explain how they made decisions – and coding the information into a computer – turned out to be far harder than scientists imagined. Twentieth-century AI did have notable successes, though. In 1997, IBM's Deep Blue beat the chess grandmaster, Garry Kasparov. The contest made global headlines, with Newsweek announcing "The Brain's Last Stand". During a game, Deep Blue scanned 200m positions a second and looked nearly 80 moves ahead. Recalling the contest, Kasparov said the machine "played like a god". Matthew Jones, a professor of history at Princeton University and co-author of the 2023 book How Data Happened, says: "It was, in some sense, a last great gasp of a more traditional mode of AI." Real-world problems are messier: rules are unclear, information is missing. A chess-playing AI cannot switch tasks to plan your day, clean the house or drive a car. "Chess isn't the best benchmark for artificial intelligence," says Prof Eleni Vasilaki, the head of machine learning at the University of Sheffield. Since Deep Blue, the most eye-catching leaps in AI have come from an entirely different approach, one that traces back to Rosenblatt and his card-sorting Perceptron. Simple, single-layered neural networks based on the Perceptron were not very useful: there were fundamental limits to what they could achieve. But researchers knew that multi-layered neural networks would be far more effective. What put them out of reach was a lack of computer power and any sense of how to train them. The breakthrough came in 1986 when researchers including Geoffrey Hinton at Carnegie Mellon University developed "backpropagation" as a way to teach the networks. Instead of single "neurons" communicating with their neighbours, entire layers could now speak to one another. Let's say you build a neural network to sort images of kittens from images of puppies. The image is fed in and processed by the network's different layers. Each layer looks at different features, perhaps edges and outlines or fur and faces, and sends outputs to the next layer. In the final layer, the neural network calculates a probability that the image is a cat or a dog. But let's say the network gets it wrong: Rover would never wear a bell around his neck! You can calculate the size of the error and work backwards through the network and adjust the weights of the neurons – essentially the strength of the network's connections – to reduce the error. The process is repeated over and over and is how the network learns. The breakthrough propelled neural networks back into the limelight, but once again researchers were thwarted by a lack of computing power and data. That changed in the 2000s with more powerful processors, particularly graphics processing units for video gaming, and vast amounts of data, thanks to an internet full of words, images and audio. Another step-change came in 2012 when scientists demonstrated that building "deep" neural networks – those with lots of layers – were immensely powerful. Hinton and others unveiled AlexNet, an eight-layered network with about 10,000 neurons, which trounced the opposition in the ImageNet challenge, an international competition that challenges AI to recognise images from a database of millions. "AlexNet was the first lesson that scale really matters," says Prof Mirella Lapata, an expert on natural language processing at the University of Edinburgh. "People used to think that if we could put the knowledge we know about a task into a computer, the computer would be able to do that task. But the thinking has shifted. Computation and scale are much more important than human knowledge." After AlexNet, achievements came thick and fast. Google's DeepMind, founded in 2010 on a mission to "solve intelligence", unveiled an algorithm that learned to play classic Atari games from scratch. It discovered, through trial and error, how to triumph at Breakout by smashing a channel through one side of the wall and sending the ball into the space behind. Another DeepMind algorithm, AlphaGo, beat the Go champion Lee

Sedol at the Chinese board game. The firm has since released AlphaFold. Having learned how protein shapes relate to their chemical makeup, it predicted 3D structures for 200m more, accounting for nearly every protein known to science. The structures are now driving a new wave of medical science. Copious headlines flowed from the deep learning revolution, but these now look like ripples ahead of the tidal wave unleashed by generative AI. The powerful new tools, exemplified by OpenAI's ChaptGPT, released in 2022, are named for their proficiency at generation: essays, poems, job-application letters, artworks, movies, classical music. The engine at the heart of generative AI is known as a transformer. Developed by Google researchers, originally to improve translation, it was described in a 2017 paper whose title, Attention Is All You Need, riffs on a Beatles hit. Even its creators seem to have underestimated the impact it would have. Llion Jones, a co-author of the paper, and the one responsible for its title, has since left Google to co-found a new company, Sakana AI. Speaking from his office in Tokyo where he was cooking up a new transformer experiment, he reflected on the paper's reception. "We did think we were creating something quite general, it wasn't designed to do translation specifically. But I don't think we ever thought it would be this general, that it would take over," he says. "Almost everything runs on transformers now." Before the transformer, AI-driven translators typically learned language by processing sentences one word after another. The approach has its drawbacks. Processing words in sequence is slow, and it doesn't work well for long sentences: by the time the last words are reached, the first have been forgotten. The transformer solves these problems with help from a process called attention. It allows the network to process all the words in a sentence at once, and understand each word in the context of those around it. OpenAI's GPT – standing for "generative pre-trained transformer" – and similar large-language models can churn out lengthy and fluent, if not always wholly reliable, passages of text. Trained on enormous amounts of data, including most of the text on the internet, they learn features of language that eluded previous algorithms. Perhaps most striking, and most exciting, is that transformers can turn their hand to a broad array of tasks. Once it has learned the features of the data it is fed – music, video, images and speech – it can be prompted to create more. Rather than needing different neural networks to process different media, the transformer can handle the lot. "This is a step-change. It's a genuine technological watershed moment," says Michael Wooldridge, a professor of computer science at the University of Oxford and author of The Road to Conscious Machines. "Clearly Google didn't spot the potential. I find it hard to believe they would have released the paper if they'd understood it was going to be the most consequential AI development we've seen so far." Wooldridge sees applications in CCTV, with transformer networks spotting crimes as they take place. "We'll go into a world of generative AI where Elvis and Buddy Holly come back from the dead. Where, if you're a fan of the original Star Trek series, generative AI will create as many episodes as you like, with voices that sound like William Shatner and Leonard Nimoy. You won't be able to tell the difference." But revolution comes at a cost. Training models like ChatGPT requires enormous amounts of computing power and the carbon emissions are hefty. Generative AI has "put us on a collision course with the climate crisis", says Penn. "Instead of over-engineering our society to run on AI all the time, every day, let's use it for what is useful, and not waste our time where it's not."