GRESHAM COLLEGE

12TH MAY 2020

Artificial Intelligence and Religion

Professor Yorick Wilks

"All the fear and superstition that existed once have been destroyed by the Machine" EM Forster The Machine stops, 1909.

"I am not a religious man, but I cannot help seeing every problem from a religious point of view." Wittgenstein

There are a number of tenuous links from AI to religion, some very old, and recently cults have appeared in the US dedicated to the worship of AI artefacts, if and when they become in some sense "super intelligent". After WWII, Karl Popper and IJ Good speculated on what it would be to have a machine that knew everything, with Good noting that the first ultra-intelligent machine would be "the last invention man need ever make". Laplace in the 19C had postulated a "demon" that knew the position and velocity of every particle in the universe: as he put it "if we conceive of a being whose faculties are so sharpened that he can follow every molecule in its course", though that level of knowledge of physics tends not to be what we think of "knowing all the facts there are". The World Wide Web is certainly pushing towards "knowing everything", although it does not know what it knows in any real sense; but the Semantic Web project AI is said to be working on that (as I discussed in a previous lecture).

Early cybernetics

One early technical landmark is Wiener's long essay *God and Golem*, also published just after WWII. Wiener was the founder of Cybernetics, the early form of AI, and the Golem was a medieval artificial creature with human-like properties supposedly created by a Prague rabbi. The essay is an assessment of the impact of the coming of intelligent machines on ethics and also on religion itself, particularly in view of man's acquisition of the god-like property of creating some form of himself. He asked what was to be our *image* of these new entities.

Early cyberneticists were more open to these speculations than their later, technically more successful, AI successors. In the UK, Stafford Beer developed a theory of "hylozoism"- that everything is alive - which is not far from the old doctrine of Spinoza (alias Pantheism) that everything is God. For Stafford Beer, the universe was a black box feedback system, and we were part of it, interacting with it. What separated off cybernetics totally from conventional science for Beer was that the universe was for him unknowable, and all we could do is to learn to accommodate to it, to adapt. This has resonances with both primitive animist theories and the far more sophisticated existentialist philosophy of Heidegger. This difference of attitude to knowledge on the part of cybernetics and traditional AI runs through this lecture and is close to what separated the ancient Gnostics off from other religions: the belief in the sufficiency and power of human knowledge.

As we shall see later, something of the flavour of early cybernetics has returned with the rise and technical success of Machine Learning in AI, which has also largely eclipsed the intervening "rationalist" period of the subject, where in the decades of the 60s-90s the investigation of intelligence in machines and humans was believed to be based on proofs, representations and logic. Cybernetics, on the other hand, was about learning, statistical mathematics, feedback and survival.

Traditional Theology: Knowing Everything and Beneficence

The relevance of this is that knowing everything, or omniscience, is what traditionally helps to define God in the western tradition, and so early AI can be seen as moving into that territory from, as it were, a different direction from cybernetics.

There is no doubt that the World Wide Web - as it is and with no limits to it knowledge in sight - does raise the possibility of an entity that effectively knows everything, and it is tempting to want to identify this with the 19C philosophical tradition of a "world consciousness" in Hegel and later in thinkers like Teilhard de Chardin. This was the idea that we, as the conscious part of the universe, perhaps the only part, are in some way making the universe both self-conscious as well as omniscient in the way God had traditionally been seen.

It is a separate and interesting question as to whether something that knows everything can be conscious in any sense like us. In the TV series *Years and Years*, a girl undergoes a brain enhancement that makes her feel conscious of everything everywhere, e. g. of every beggar in Peking. But can one be conscious of everything, since consciousness seems by its very nature selective and a matter of attention? If true, this raises the issue of how an omniscient being, god or machine, could be conscious at all.

On the other hand, in the 17C Leibniz famously argued that, in terms of his monads that composed everything in the world, God was the supreme monad that was aware of the contents of all other monads at once; it was the only fully conscious thing. His was the only monad that was aware of many things at once or, in another metaphor Leibniz used, that could see things from every point of view at once. Some theologians have argued from this that God could not, on such a view, be conscious in the way we are, because consciousness, as we noted, requires some limitation, some deprivation of access. So, if computers really can do more than one thing at once "at their topmost level" of operation---and it seems obvious that a big computer can now do many things at once---- then it must follow that computers, *in so far as they could be conscious, now or in the future*, could be a stage further on than humans towards the condition of Leibniz's God monad, since we cannot be conscious of more than one thing at once. This is, of course, a very weak and slightly circular argument, and many would still deny that machines can be conscious, even in principle, a notion I argued in favour of in a previous lecture.

A related question is whether an artificial "superintelligence", were one to emerge, would be well-disposed towards us. Religious traditions historically assume that creations (i.e. us vis-a-via a creator or a Superintelligence vis-a-via us) are well disposed to those who made them. If that is true, we need not fear what we may create in the future with AI. Bostrom has recently in his book "Superintelligence" argued that such a creature might well destroy us all, but I would counter that the tradition of created things and their benevolence towards their creator suggests otherwise.

Romantic visions of machines as perfect: (a) making human-like things

Making human-like things artificially is a tradition that goes back to Ovid's story of Pygmalion making the statue Galatea come alive, a story that reappears, via Shaw's play *Pygmalion*, in our own time as *My Fair Lady*. We noted the Golem of Prague and in 19C there was, of course, Frankenstein's monster. In 19C German romantic thought, Kleist argued that marionettes and puppets were in some ways more perfect than real humans, in part because they were not conscious, with all the problems that state brings. John Gray has taken this observation as the starting point for an investigation of free will, and raises the possibility that having conscious choice is drawback. It remains a lingering idea on the edge of AI that, like Kleist's marionette, the artificial can be more perfect than human is and, in that sense, more god-like. Many Eastern religious traditions also see consciousness as something to be transcended in meditation, not indulged or celebrated. In the 1920s JD Bernal was one of a group of British eugenicist scientists who thought that with human augmentation and ingenuity : "Consciousness itself might vanish in a humanity that had become completely etherialised".

Romantic visions of machines as perfect: (b) augmenting humans

The last quotation is a transition from futurist discourse about creating humans to that of adapting or evolving existing humans (or some of them anyway) so that they become immortal or superintelligent themselves. Groups or individuals who believe this are now called *Transhumanists* and have close links back to the Greek Gnostics and then 18C rationalism and humanism, moving on to the worship of adapted humans or "Supermen" to use Nietzsche's term. Transhumanists and superintelligencers only differ in the source of the supreme intelligence, artificial or ourselves transformed, not the goal itself. This is not a new idea: one of Schubert's most famous songs is set to a poem by the 19C writer Mueller and contains the line "Will kein Gott auf Erden sein, sind wir selber Götter"---if a god won't come down to earth, we will be gods ourselves!

The practical elements for this augmentation are surgery, exoskeletons fitted on to our bodies to aid human work, as well as the huge amount of current state effort to build "perfect soldiers" with drugs and prostheses, a tradition going back to hashish-fueled Ottoman armies of the Middle Ages. Kurzweil, who was not only an AI originator, but the modern source of the notion of the *Singularity moment* foreseen by IJ Good long before: the moment of the emergence of the superintelligent machine. For Kurzweil, this notion remains connected to Transhumanism, and he believes a future machine-human fusion will remain human and benign, whereas in Bostrom's dystopian future the superintelligence is wholly alien and hostile.

Romantic visions of machines as perfect: (c) God-machines and AI religions

Making gods, like all traditions, is ancient, back to Aaron's Golden Calf, the Baal Gods of Iron, as well as the later Roman Emperors after death. Bostrom's emergent Superintelligence certainly has god-like qualities, and it is interesting to note here his complex and almost certainly false argument that this emergent superintelligence will be unique since it will kill all its competitors. His god-like AI world ruler is thoroughly monotheist. The reaction against this tradition is very strong in monotheistic religions: Islam, Judaism and reformed Christianity, which tend to be hostile any notion of 'graven images", which is to say entities created in the image of humans or gods. Unsurprisingly perhaps, this spread to Marxism, which was almost a fourth Abrahamic religion, and Marx's fear of fetishisation, in his case of commodities/ IN *Capital*, he writes "The fetishism of commodities." The form of wood, for instance, is altered by making a table out of it. Yet for all that the table remains an ordinary thing, wood. But, as soon as it steps forth as a commodity, it is changed into something transcendent."

Inevitably, AI has already spawned explicit religious cults of worshipping, yet to be created, AI intelligences and it is hard to know how seriously their creators take them. In the USA, Levandowski has founded what purports to be an AI religion, set up to worship a superintelligence, soon to be manifested.

Automating religious practice

Religious machines go back to Tibetan prayer wheels turned by streams and medieval automata celebrating Easter. A wide range of Apps are now available from the trivial to the potentially interesting to support religious practices: the home chat Companions ALEXA and SIRI can now recite prayers on demand. In Japan, MINDAR is a Buddhist priest in robot form that conducts prayers, blessings and ceremonies. In East Asia there is undoubtedly more of this activity and a Japanese AI professor has expressed the view that, in the Buddhist tradition, there is not seen to be the same firm distinction between the animate and inanimate—there are spirits in everything, a sentiment we already located in the West in Leibniz and Stafford Beer. As Minoru Asada, President of the Robotics Society of Japan, puts it: "In Japan we believe all inanimate objects have a soul, so a metal robot is no different from a human in that respect, there are less boundaries between humans and objects."

Such Apps would only be of serious religious interest if they were able to perform religious rites like confession or saying Mass, and it is hard to imagine any church validating an automaton to do any of those things at present. But the notion is worth exploring if one believes that there is nothing *in principle* a human can do that a machine cannot. One can well imagine a machine confessor by extending the notion of a machine psychiatrist, of which there have been many incarnations over fifty years, or perhaps even a robot sports coach. One can conceive an automated Companion, one that has already been implemented to a limited degree, that knows its subject's habits and life in great detail—possibly gleaned from years of conversation.

That kind of Companion would be essentially a copy of its subject but not at all like the proposed "uploaded brain" of the Transhumanists, but working at the level of language rather than neurons. It is not hard to see how such an entity might also function as an ethical advisor or a confessor based on its detailed knowledge of the person it had learned from, or copied. AI has now largely shifted to machine learning as its principal paradigm, which has resulted in programs whose operations and reasons for action are often obscure, even to those who programmed them, based as they are on sometimes inscrutable neural nets rather than rules. One can then argue that human motivations are, from the point of view of outside observers, no more or less obscure than those of these most advanced machines, and we are now at an interesting point where discovering the reasons behind the actions of humans and machines may be in much the same position. It is this fact that represents the return of some of the attitudes of long-discarded cybernetics, and which now renders plausible the possibility of an advisor on human action and conduct, since it can have the same status as efforts to explain the actions of machines.

Yuval Harari has argued, however, that having a machine like this that "understands us better than ourselves " can be seen as the end of the long tradition of individualist rationalism on which our western assumptions about religion and politics rest: the belief that the individual has a free choice as to his or her actions and beliefs. He would almost certainly say that an ethical companion/confessor makes nonsense of that, although many might find it perfectly acceptable that someone close to them knows better than they do what they want and why they act as they do. Arranged marriages are said work on just that principle.

Conclusion: the uniqueness of the human?

In his famous paper on whether machines could think, Turing considered and dismissed a range of objections, including one that God had given man a soul and that was involved in thought, hence machines and animals could not have a soul or think. Turing dismissed this since it "implies a serious restriction of the omnipotence of the Almighty … should we not believe that He has freedom to confer a soul on an elephant if He sees fit?". Turing could not take any of this very seriously, but it does echo a long tradition in the west that man is *Imago Dei*, the image of God, and is therefore unique in being god-like. As we have seen, that has proved an attractive possibility to many thinkers beyond the confines of religion itself. We have also touched on the more Eastern tradition that does not see so strong a demarcation between the human and non-human, and that accepts that there may be nothing in principle that a human can do and a machine cannot, a tradition I would subscribe to myself. On that view, whatever godlikeness or distinctive human likeness are, they could pop up elsewhere.

The function normally proposed for human uniqueness is language, which we seem to possess and nothing else in the universe does. Its god-likeness can be felt in the opening words of John's Gospel: "In the beginning was the Word", a sentiment no one fully understands after 2000 years. Neil Lawrence, now Professor of Machine Learning at Cambridge and a former colleague, has recently proposed an account of human-machine difference in terms of information transfer rates. He argues that machines can output data at millions of bits a second but humans, using language, can only manage a handful, by reason of their speech anatomy and physiology. This, he argues, has caused us during evolution to develop language which can only be understood on the basis of enormous stores of background knowledge, which is what enables us to send only such very sparse signals to each other. He has further arguments, not relevant here, that these complex mental models then become the basis of consciousness. His argument is

reminscent of Minsky, the founder of AI at MIT, and his early emphasis on the role of mental models which, he went on in a 2013 article in the *Jerusalem Post* to suggest, would one say lead to an AI notion of a soul which he defined as "The word we use for each person's idea of what they are and why".

I think Lawrence gives an interesting new twist to the old claim that language distinguishes humans, and it seems to me a consequence that intelligent machines, to be human- or god-like, will eventually have to communicate with us in a language like ours and they will then be in some kind of equality with us. It makes no sense to talk of something understanding language *better than humans do* since humans *own* language and define what it is to understand it. That view is of course, close to the "Gnostic" one, quite different from more Eastern and cybernetic views that we touched on earlier, in which language and knowledge are not central to intelligence, and humans are not distinct from other forms of intelligence.

John Wisdom the professor of metaphysics at Cambridge in the 1960s, and on the fringe of the then paramount analytic tradition in philosophy, used to spend some weeks of lectures on the analysis of just what was meant by a sentence like "Prussia invaded France in 1870", a historical truth. Land masses cannot invade so the statement was metaphorical, or more accurately metonymic. But who exactly did invade and does it matter for the statement to be understood and judged as to its truth? One can imagine, in the spirit of Lawrence, access to some vast stream of data covering the movement of millions of individual Prussian soldiers. But even with that data, one would almost certainly not, as Danto argued and we noted earlier, have been able to extract the proposition that the sentence above expresses, and which is its significance or meaning. This memory of old-fashioned linguistic analysis captures for me the difference between meaning and data, and what it is that is unique to humans and what machines will need to have to be human-like, let alone god-like. Gray and Yuval Harari both express, in their different ways, their disquiet with the modern Gnostic tradition----of which Transhumanism, or what Lawrence calls "religion for nerds", is the most extreme expression: the tradition of total knowledge supremacy, as opposed to something closer to the cybernetic tradition in which the universe remains mysterious and beyond any full understanding.

Further Reading

If reading only one item on this list, I would recommend the short, lucid, book by John Gray, which touches on many of the themes of this lecture.

Beer, Stafford, J. 1966. "Cybernetics and the knowledge of God", *The Month*, 34.

Foerst, A. 2004. *God in the Machine: What Robots Teach Us About Humanity and God*. New York: Dutton.

Gray, J. 2015. *The soul of the marionette.* London: Penguin Books.

Herzfeld, N. L. 2002. "In Our Image: Artificial Intelligence and the Human Spirit", *Theology and the Sciences*, ed. Kevin J. Sharpe. Minneapolis: Fortress Press.

Johnson, B. 2008. *Persons and Things.* Cambridge, MA: Harvard University Press.

Kurzweil, R.. 1999. *The Age of Spiritual Machines: How We Will Live, Work and Think in the New Age of Intelligent Machines.* New York: Viking, 1999. Reprint, London: Phoenix, 1999.

Mayor, A. 2018. *Gods and Robots: Myths, machines and ancient dreams of technology*. Princeton: Princeton University Press.

Robertson, J. 2018. *Robo Sapiens Japanicus.* Oakland, CA: University of California Press.

Rosenfeld, . A. 1966. "Religion and the Robot: Impact of Artificial Intelligence on Religious Anthropologies." *Tradition: A Journal of Orthodox Thought* 8.

Turkle, S. 1984. *The Second Self: Computers and the Human Spirit*. New York: Simon and Schuster.

Wiener. N. 1964. *God and Golem: a comment on certain points where cybernetics impinges on religion*. Cambridge, MA: MIT Press.