



Auckland Unitarian Church  
Love beyond belief



## The Age of Machine Learning — What Kind of Society Do We Want?

Speaker: John Maindonald  
Service leader: Shirin Caldwell  
Music: Jaime Taylor

10<sup>th</sup> November 2024

John Maindonald © 10 November 2024

Today, no business of any size uses paper based accounting. The move from paper based systems started in the late 1950s when large corporations started using computer mainframes, as they were called, for basic operations. In the 1980s personal computers made computer-based accounting systems widely available.

The chatbots that have attracted wide attention in the past several years are a very different technology from accounting systems. In January of last year, Ted Zorn spoke about "[Finding our humanity in a technologized world](#)", looking in particular at the abilities of 2024 chatbots akin to ChatGPT. Chatbots per se are not all that new. [The Eliza chatbot](#) came on the scene in the mid 1960s. One could ask it "What can you say that will make me happy?", and it would come back with a sympathetic response. ChatGPT is of course a vastly more powerful beast, using information that it gathers from the internet for answers that usually appear plausible.

Ted asked ChatGPT to "Write a sermon for the Auckland Unitarian church of around 1000 words focusing on the most prominent themes in Unitarian church sermons", and then used the chatbot's response for the first part of his talk. The response included a reference to principles 1 and 7, noting that "As Unitarian Universalists, we believe in the inherent worth and dignity of every person" and "We also hold a respect for the interdependent web of all existence." It went on to say that "it's up to us to use technology in a way that enhances our humanity, rather than diminishing it." Two of the other principles that are very pertinent to today's discussion, not mentioned by ChatGPT are the third that speak of "Justice, equity and

compassion in human relations”, and the sixth that speaks of “The goal of world community with peace, liberty, and justice for all”.

ChatGPT is one of several chatbots that are available to try on the web. Others include Microsoft Copilot, Google Gemini, and Perplexity. The term Artificial Intelligence (or AI) has come into common use in the past decade as a generic name for this and other computer technologies that are now in use. What constitutes AI is not well-defined. What is in mind are powerful computational systems that, with advances in the power of computers and in the code they run, have extended their reach dramatically in the recent past. As I will illustrate, even with an older and more conventional computer code, things can go seriously wrong. Too often, managers who buy such systems are not adequately informed on what they are taking on. Their training fitted them for a different time and place.

Perplexity had no problem when I asked it “What are the implications of AI for the human future?” I have selected key sentences from its response:

The implications of AI for the human future are vast and multifaceted. Here are some key areas to consider:

1. **Economic Impact:** *AI has the potential to transform industries, automate jobs, and increase productivity. While it may create new job opportunities, there’s also concern about job displacement and the need for reskilling.*
2. **Healthcare Advances:** *AI can enhance diagnostics, personalize treatment plans, and streamline operations in healthcare, leading to better patient outcomes and more efficient systems. **JM: But check that systems really do what is claimed.***
3. **Ethics and Governance:** *The rise of AI raises important ethical questions about bias, privacy, and accountability. Establishing frameworks for responsible AI development and usage will be crucial.*
4. **Social Dynamics:** *AI could alter social interactions and relationships, impacting everything from communication to privacy. There’s also the potential for increased surveillance and erosion of personal freedoms. **JM: AI systems are altering social interactions, in concerning ways.***
5. **Environmental Sustainability:** *AI can contribute to environmental monitoring, resource management, and energy efficiency, aiding in efforts to combat climate change. **JM: AI systems are very energy hungry, ~2% of global electricity use in 2020.***
6. **Education:** *AI has the potential to personalize learning experiences, making education more accessible and tailored to individual needs, but it also raises questions about data security and equity.*

Ethics and Governance, and Social Dynamics, are where I want to focus attention. There’s first of all the issue of putting too much trust in these systems.

Banking systems have changed dramatically in the past 50 years. One unfortunate result is that the new online systems have given scammers a field day, with losses steadily increasing year by year. Those who are not very tech savvy are of course the most vulnerable. In a talk he gave in Wellington recently, Andy Prow who is chair of NZTech's Digital Safety Board argued that banks are not doing nearly enough to protect customers from scammers, noting specific steps they could and should be required to take.

In 1999 the British Post Office installed a new accounting system that replaced its former paper-based systems, for use by the subpostmasters in its more than 11,000 post office branches. The Post Office had the authority to bring its own prosecutions, independently of the police. Faults in the software had the result that more than 900 subpostmasters were wrongly convicted of theft, fraud and false accounting. It took until 2020 to get an inquiry into what had gone wrong. Their convictions are still in process of being overturned, with the cost of reparations is likely to be of the order of 1.4 billion pounds. There have been devastating effects for the subpostmasters involved —many had their livelihoods ruined, some went to prison, and four committed suicide. The story is told in the TV series “Mr Bates versus the Post Office” that was released at the beginning of this year. After his Post Office contract was terminated in 2003 over a false shortfall, Mr Bates sought out other subpostmasters in the same position and went on to found the Justice For Subpostmasters Alliance in 2009, leading finally to the 2020 inquiry.

See <https://www.personneltoday.com/hr/the-post-office-horizon-scandal-an-explainer/> for more about this.

How did those who supplied the software, the post office management, and the courts, allow these horrors to extend over two decades. It seems unbelievable.

This is by no means the only example of its kind, but it is an especially serious one. Organizations that supply software, whether for accounting or for what is now being called AI, need to be held to high standards. The EU has been a leader with a set of standards that were mandated last August 1st , to come into force in three years time.

Accurate data is very often an issue. Modern aircraft are heavily reliant on automated guidance systems. In a race to compete with a new more fuel efficient Airbus 320 design that was released in 2010, Boeing immediately started work on a new 737 design, with larger engines that were moved up under the wing and slightly forward, changing the aerodynamic characteristics. One result was that the nose had a tendency to pitch upward under certain conditions when the plane was under manual control. An add-on software system was therefore installed, relying on just one of the two angle of attack sensors, that was designed to come on automatically when the upward pitching was detected. Boeing obtained permission from Federal Aviation Authority to omit mention of the add-on software system from the instruction manual for pilots, arguing that it was for all practical purposes failsafe. Problem was — the software was unable to detect when the information it was getting from the one angle of attack sensor on which it relied was faulty. Two crashes, one in 2018 and one in

2019, led to the death of 346 passengers. These sorts of problems arise all over the place. Mechanisms for detecting when the data is flawed and allowing control to pass over to human experts are essential.

See <https://www.seattletimes.com/seattle-news/times-watchdog/the-inside-story-of-mcas-how-boeings-737-max-system-gained-power-and-lost-safeguards/> for more about this.

I used a chatbot to get a perspective on AI, and it gave a pretty reasonable answer. But, when it comes to detail where text that is on the web has to be carefully read and understood, chatbots can get it badly wrong. Thus, a court reporter who checked what the Copilot chatbot would say about articles he had written, found himself identified as having committed crimes on which he had reported. A search on the web for “chatbot nonsense” will find more such examples. The extent to which chatbots are being promoted as a way for firms to provide customer guidance on websites is worrying. In 2022, a chatbot on Air Canada’s chatbot promised Jake Moffatt that he could book a full-fare flight for his grandmother’s funeral and then apply for a bereavement fare after the fact. A tribunal sided with Moffatt, rejecting the argument that he should have gone to the link provided by the chatbot, where he would have seen the correct policy. This set an important precedent.

A NZ government Chatbot website, [<https://govgpt.govt.nz/>] currently directed at small-business facing ministries and organisations, went online a few weeks ago. One can type in questions, and it will respond. Users are warned:

- May produce answers that are incorrect, biased, or not referenced in source material.
- Does not provide advice, nor represent the New Zealand government.

Can the government escape thus easily?

Those who manage systems for interacting online with banks should by now have a clear understanding of the issues, even if those issues are not as well handled as they should be. Chatbots that are used for providing advice to customers are a new challenge. Managers are unlikely to have been trained to cut through the hype of AI salespeople, and to carefully assess the risks as well as the benefits that go with investing in what may be very expensive systems. They are told “If competitors install this AI system with its competitive advantages before you do, you will miss out.” [A study published in mid-August by American think tank RAND](#) noted that by some estimates 80% of AI projects fail, more than double the rate for non-AI projects.

The Australian government has recently put out a set of draft standards. Of particular interest are the ‘guardrails’ (as they are called) that accompany them. For now at least these are voluntary, but they do set a standard to which those who sell AI systems can be asked to commit. [<https://www.industry.gov.au/publications/voluntary-ai-safety-standard/10-guardrails>]

The guardrails ask organisations to commit to:

- understanding the specific factors and attributes of their use of AI systems
- meaningfully engaging with stakeholders
- performing appropriately detailed risk and impact assessments
- undertaking testing
- adopting appropriate controls and actions so their AI deployment is safe and responsible.

Eight associated principles are:



Human, societal and environment wellbeing: 1,2,5,6,7



Reliability and safety: 2



Human-centred values: 1,2,6



Transparency and explainability: 2



Fairness: 2,6



Contestability: 2



Privacy protection and security: 6,7



Accountability: 2

There is a much in common, both with the seven principles, and with the newly agreed values statement.

<b>1</b>	The inherent worth and dignity of every person
<b>2</b>	Justice, equity and compassion in human relations
<b>3</b>	Acceptance of one another and encouragement to spiritual growth in our congregations
<b>4</b>	A free and responsible search for truth and meaning
<b>5</b>	The right of conscience & the use of the democratic process within our congregations & in society at large
<b>6</b>	The goal of world community with peace, liberty, and justice for all
<b>7</b>	Respect for the interdependent web of all existence of which we are a part



If only such principles could be applied, not just in relation to AI, but in our society more generally. Unitarians and UUs are called by the standards that we have set for ourselves to be awake to the changes taking place in our societies, and to do what we can to help direct them in positive directions. As Forrest Church said in the reading that we listened to earlier “a 21st-century theology needs nothing more and requires nothing less than a new Universalism”, one that will draw a fractured world together. It is indeed “up to us to use technology in a way that enhances our humanity, rather than diminishing it.” Where do we start? Beyond making ourselves informed, there is no easy answer.

For further reading, some of the links, above and below, may be useful. The Auckland and Wellington public libraries both have several hundred books on AI. One that I recommend is “[Moral AI And How We Get There](#)”, by Jana Borg and two others. The authors are a philosopher, a data scientist, and a computer scientist.

---

## Meditation / Conversation starter

- What can we do to educate ourselves on the risks and opportunities of the world into which we are moving?
- Are there steps, however small, that we can as individuals and as a community take that will help push changes in a positive direction?