Book review

Review of Faking It: Artificial Intelligence in a Human World

by Toby Walsh
La Trobe University Press
in conjunction with Black Inc.
2023
pp. 1-234
ISBN: 9781760644826 (paperback)
ISBN 9781743823361 (ebook)
Review DOI 10.1108/BL-07-2024-231

Introduction

This is a review of Professor Toby Walsh's recently published book (2023), *Faking it: Artificial Intelligence in a Human World*. Professor Toby Walsh, UNSW Scientia Professor of artificial intelligence (AI), is the consummate blend of sceptical scientist and enthusiastic inventor–developer of AI who is able to extoll the potential virtues of AI and at the same time warn of possible threats to us humans.

Professor Walsh notes five broad themes or emerging trends from AI research and use: What's in a name: AI and cybernetics?; the new science of robotics; the pre-scientific practices of alchemy; the evolving science of human-AI sense-making; and enthusiasm, hype and fakery of emerging AI. I explore AI, digital records, knowledge management and collective intelligence in my recently published (July 2023) book, *Memory Curators and Memory Archivists in the Digital Memory Age*.

What's in a name: AI and cybernetics?

Fixing labels
Firstly, Professor Walsh observes and reasons that the term AI is a useful way of describing the concept. Setting the scene and context, Walsh reflects:

> [. . .] to my own surprise, too, I've come to believe that the name artificial intelligence is not a mistake, but a rather good description. That's because one of the key things about artificial intelligence, I now realise, is that it is artificial – that it is about imitating human intelligence. [. . .] By abstracting intelligence, we can hand over many tasks to machines (Walsh, 2023, pp. 9–10).

Walsh identifies the central issue of the AI–human intelligence debate by noting:

> [. . .] intelligence is the ability to extract information, learn from experience, adapt to the environment, understand, and reason about the world. So what, you should ask, is artificial intelligence? Most AI isn't embodied and situated in the world like our human intelligence, adapting and learning from the environment. How, then, can the intelligence of machines be identified and measured when it is so fundamentally different to human intelligence? (Walsh, 2023, p. 9).

Walsh offers a partial answer when writing:

AI is about getting computers to do tasks that humans require intelligence to do: the tetralogy of perceiving, reasoning, acting and learning. It requires intelligence to perceive the state of the world, to reason about those percepts, to act based on that perception and reasoning, and then to learn from this cycle of perception, reasoning and action. Before the 1950s, there weren't any computers around on which to experiment, so it was pretty hard to do any meaningful AI research (Walsh, 2023, p. 12).

Walsh observes early in his book:

I am confident, however, that the issues raised in this book will not be out of date. Indeed, I am sure they will be even more pressing. And this book will be even more useful, as a guide and a warning. We will, for example, be ever more deceived by fake AI and AI fakes. It is time, then, for concerned citizens to understand – and to act (Walsh, 2023, p. 7).

What's in a name?

Secondly, as described by Professor Walsh and noted by Associate Professor Catherine Ball from the Australian National University's (ANU) School of Cybernetics, AI is a very broad label covering a complexity of technologies, potential capabilities and diverse uses. For example, Ball describes the Machine Learning (ML) capability that:

[…] involves computers using data to discover how they can perform tasks without being explicitly programmed to do so. What many people think of when they hear "AI"—the ability of computers to think for themselves and program things without needing humans to enter the code —is actually ML (Ball, 2022, p. 30).

Following Walsh's comments Ball highlights the complexity by noting AI is part of complex adaptive systems that:

[…] include technological networks that are more dynamic and flexible than a food web in nature because there are "rules" that govern nature […] It may sound counterintuitive, but the rules in automated technologies and machine learning are so flexible that there are really no rules (Ball, 2022, pp. 42–43).

Luddites?

Thirdly, with AI's revolutionary social and economic potential comes turbulence and degrees of human concern to fear. Ball reasons when discussing the historical effects of technological change with the Luddites of northern England in the late eighteenth century the:

[…] new machinery itself was not really the target of people's anger, but destroying it was an easy way to upset the factory bosses when workers were fighting for their rights. Ned Ludd could not be blamed for this because he didn't actually exist. He was a made-up persona designed to frustrate the government. Ned Ludd was fake news (Ball, 2022, pp. 58–59).

At the broader level, Harvard social scientist Professor Spar reminds readers:

[…] machines taking over is in many ways only the updated version of the story humankind has been telling itself since our storytelling days began. We are a tool-making species, and a questioning one, probing to understand both the source of our life and its meaning (Spar, 2020, p. 287).

With a strong dose of realism, Spar notes that technology:

[…] is a tricky thing. It is our creation, a wholly human construction, yet developed now to a point where it could-might, probably will expand beyond our human ability either to comprehend

inner workings or to completely control them. We have built the pieces-millions of them, crafted by generations of us without a master plan for how those pieces should eventually interact and evolve (Spar, 2020, p. 288).

Walsh's observation about AI imitating leads to the complex issues of automation, cybernetics and robotics.

The new science of robotics

Meeting of minds?
Cybernetics appears to have emerged before AI. Walsh sets the context by adding cybernetics was emerging before:

[. . .] the famous Dartmouth conference in 1956, there was a remarkable series of conferences held between 1941 and 1960, organised by the Macy Foundation. The Macy conferences brought together a diverse group of anthropologists, biologists, computer scientists, doctors, ecologists, economists, engineers, linguists, mathematicians, philosophers, physicists, psychologists, social scientists and zoologists to explore intelligent systems.

One of the participants at the Macy conferences was "Norbert Wiener, the father of what was at that time the emerging field of cybernetics" (Walsh, 2023, pp. 210–211).

Dartmouth Datafest, 1956
The subtext of hidden creativity emerges in the term AI when McCarthy:

[. . .] introduced the name to describe the topic of a seminal conference held at Dartmouth College in 1956. This meeting brought together many of the pioneers in artificial intelligence for the first time, and laid out a bold and visionary research agenda for the field (Walsh, 2023, pp. 11–12).

Walsh highlights the interesting history of the term AI by noting Alan Turing had used the AI in a paper:

[. . .] published six years before one of the other founders of the field, John McCarthy, coined the term. "I had to call it something," he wrote later, "so I called it 'Artificial Intelligence', and I had a vague feeling that I'd heard the phrase before, but in all these years I have never been able to track it down" (Walsh, 2023, pp. 11–12).

Walsh notes the unbridled optimism of the conference planners who funded by the Rockefeller Foundation claimed:

We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer (Walsh, 2023, pp. 11–12).

Histories of science need to delve into this era looking for more answers, possibly. Mustafa Suleyman and Michael Bhaskar are setting the scene for the next chapter.

Wiener and cybernetics
Before the genesis of AI, the associated discipline of cybernetics had issues with staking its technical claim itself. Walsh describes:

[…] a beautiful letter from Esther Potter, a director of the Library of Congress, to Dr Norbert Wiener, author of the seminal text Cybernetics, appealing for help in trying to classify his book. "We have read and reread reviews and explanations of the content of your book and have even tried to understand the content of the book itself," she wrote, "only to become more uncertain as to what field it falls into. I am appealing to you as the one person who should be able to tell us […] If we were not somewhat desperate about this particular problem, I should hesitate to bother you with it" (see https://tinyurl.com/WhatIsCybernetics.) (Walsh, 2023, p. 24).

The robots are already here, and more are coming!

Robotics is part of the co-evolving science of cybernetics. Walsh reasons that here is the key turning or tipping point. Robots need to make sense of their surroundings to apply their embodied skills and capabilities. Therefore, robots need to understand their environment and make sense. Robots are extensions of humanity (tools and technologies and systems). Robots, through people, sense, realise, make sense, learn and adapt.

Professor Toby Walsh's historical analysis of the effect of Asimov's famous three laws of robotics on the relationship between people and robotics reasons for the:

[…] last 80 years, Asimov's laws have largely been ignored by those actually involved in building AI and robotics. They have remained science fiction rather than science fact. However, in the last decade it has become clear that storm clouds have been brewing (Walsh, 2022, p. 125).

Walsh notes that robots:

[…] are often depicted as the very embodiment of artificial intelligence. That's not surprising. For a robot to act intelligently in the world, it needs AI. It needs to sense, reason, act and learn from an ever-changing world. Now, not all robots have AI. Some simply follow the same instructions repeatedly. These are the sort of robots you often find in factories, and usually they're in cages to protect humans from their repetitive, pre-programmed movements. But when robots are out in the real world, away from a controlled environment like the factory floor, they need some artificial intelligence (Walsh, 2023, p. 17).

Walsh's comments cover the essential feature of human-robotic interactivity including the cybernetics of designing and managing robots across social and economic environments.

Professor Pasquale's four laws of robotics

Pasquale explores the key issue of designing and controlling robots: the embodiment issue. Pasquale describes how the "philosopher Hubert L. Dreyfus developed theories of tacit knowledge to explain why expert systems performed so poorly.72 We know more than we can explain. Think of how difficult it would be to boil down your job into a series of 'if, then' statements" (Pasquale, 2020, p. 24). Essentially Pasquale asks key questions that need asking about robotic embodiment such as could:

[…] situations you encounter daily be recognized by a computer? Would possible responses to those situations also be readily articulated, evaluated, and ranked? If not (and I suspect for the vast majority of readers, this is the case), the difficulties of description (of job tasks) and translation (of human situations into computable code) (Pasquale, 2020, p. 24).

Walsh's comments on robots introduces the American legal academic Professor Frank Pasquale's four laws of robotics (McFadzean, 2023, pp. 66–69 and pp. 264–266). Pasquale's analysis and review are aimed at the designers and developers of robots rather than the robots. Governance is key to drafting and applying legislation and standards to designing robots. Pasquale sets the scene for his new laws of robotics by noting the emerging evolution between AI and intelligence augmented (IA). Pasquale's four laws (2020, pp. 4–12)

can be summarised as: 1. Roles: robots complement and supplement humans not replace (except for dangerous or dirty jobs); 2. Identity: when interacting with people robots must be identified; 3. Laws of Conflict: robots as weapon systems must follow the laws and rules of war and engagement; 4. Responsibility and Accountability: creators, controllers, managers and owners of robots must be identified and responsibilities defined and declared (McFadzean, 2023, pp. 67–69).

The evolving AI-IA balancing trend is set in the context between utopian or dystopian visions and ideas about robots and humanity. Pasquale reasons that the "future of automation in the workplace—and well beyond—will hinge on millions of small decisions about how to develop AI" (Pasquale, 2020, p. 14). Pasquale summarises the situation as when "robotics revolutionized manufacturing by replacing assembly-line workers [. . .] many business experts want similar technological advances to take over more complex work, from medicine to the military" (Pasquale, 2020, pp. 13–14).

Pasquale observes that many journalists and commentators, focusing on:

[. . .] managerialist enthusiasms, have discussed "robot lawyers" and "robot doctors" as if they are already here [. . .] To the extent that technology transforms professions, it has tended to do so via IA, not AI. [. . .] The question now for innovation policy is where to sustain this predominance of IA, and where to promote AI. This is a problem we must confront sector by sector, rather than hoping to impose a one-size-fits-all model of technological advance (Pasquale, 2020, p. 14).

The essential question of this evolution is:

How far should machines be entrusted to take over tasks once performed by humans? What is gained and lost when they do so? What is the optimal mix of robotic and human interaction? And how do various rules—from codes of professional ethics to insurance policies to statutes—influence the scope and pace of robotization in our daily life? (Pasquale, 2020, p. 14).

Sadly, as Walsh suggests, Wiener, McCarthy and Minsky missed the fantastic opportunity to merge cybernetics, robotics and AI in 1956. Is Walsh asking is it too late to right this wrong by exploring alchemy?

When thinking about designing robots, Walsh cautions us about the uncanny valley issue of robot design and planning. Japanese robotics professor Masahiro Mori wrote about the uncanny valley in 1970. Mori reasoned initially "small improvements to the design of a robot make it seem more human-like. [. . .] as the robot is made even more human-like, our brains begin discounting any residual differences and fill in the gaps. We rapidly climb the heights of indistinguishability, towards total realism" (Walsh, 2023, p. 75).

Records for robots?

Following the themes of Professor Walsh's book, robots must become reflexive therefore learning through vital feedback (loops). Reflexivity is about sensing, learning and adapting to new stimuli. Robots could be defined and described as cybernetic systems. Taking the logic further, robots could be crudely labelled as business systems, making products or providing services, privately, corporate, governmental or non-government sectors. Therefore, robotics could be seen as part of a business' recordkeeping informatics system. AI (software, sensors and creative components) are part of robotic sensemaking and must be recorded for the feedback loops of a system. The operations of robots must be financially, managerially and legally accountable. Someone must be responsible for a robot's operations, actions and reactions.

Recordkeeping informatics is the study and science of what information is created and shared when, where, by whom and how. Essentially informatics applies systems analysis.

Recordkeeping systems are composed of four separate yet interactive, interdependent parts or sub-systems:

(1) people (users, managers, administrators, technical support, developers);
(2) business rules (policies, procedures and workflows [automated, people, integrated]);
(3) network (architecture of the connected networks, what is linked to which system); and
(4) technologies (business, communications, corporate-management and recordkeeping) (McFadzean, 2023, p. 114).

Grant, Upward, Reed, Oliver, Evans, Poole and Donnelly put forward a strong business case for recordkeeping to become data management (McFadzean, 2023, p. 124, pp. 148–149, pp. 154–156, p. 195, p. 244, p. 259 and p. 265). Informatics supplies the data and information to build workflows that become operational business processes that manage the access to corporate memory (Upward *et al.*, 2018, pp. 74–76, pp. 88–89, pp. 223–224 and pp. 278–283). This creates the conditions for reflexivity.

Upward, Reed, Oliver and Evans describe the three skillsets of the three-part Monash model of digital registrar-systems analyst-feedback specialist based on reflexivity (McFadzean, 2023, p. 120). The Monash model works through the science of recordkeeping informatics. Specifically, the roles would conduct the following tasks: the digital registrar is an expert in business processes and interpreting the metadata schema. Metadata are assertive statements about the origin, ownership, purpose and history of information and the record; a systems analyst is the expert who understands, supports and uses the record as an electronic system (the parts that byte). Analysts create trust in systems as systemic records; the auditor examines and tests the system and curates data by applying data to business processes. McFadzean summarises the descriptions in summary tables (McFadzean, 2023, pp. 59–60, pp. 121–150).

The pre-scientific practices of alchemy?
Thirdly, Walsh then introduces alchemy. Counterintuitively, alchemy could be seen as the activity between speculation, conjecture, serendipity and insight. The place where ideas can emerge as proto-hypotheses. Walsh sets the scene for alchemy by noting ironically the comparison between alchemy and AI:

> [. . .] was proposed by [Hubert] Dreyfus (1965) as a condemnation of artificial intelligence, but its aptness need not imply his negative evaluation. Some work can be criticized on the grounds of being enslaved to (and making too many claims about) the goal of creating gold (intelligence) from base materials (computers). But nevertheless, it was the practical experience and curiosity of the alchemists which provided the wealth of data from which a scientific theory of chemistry could be developed (Walsh, 2023, p. 22).

For some alchemy is a pejorative and for others it is exciting, such as for Eric Horvitz. Walsh notes:

> [. . .] many have compared the field of AI to medieval alchemy. Rather than attempting to turn base metals into gold, the ambition of artificial intelligence is to turn simple computation into intelligence. Walsh documents how Eric Horvitz, chief scientific officer of Microsoft Research and a past president of the Association for the Advancement of Artificial Intelligence, told The New York Times in 2017: "Right now, what we are doing is not a science but a kind of alchemy." I checked with Eric and he stands by this observation today. He remains "intrigued, curious, and optimistic that there are deeper insights and principles to be uncovered" (Walsh, 2023, pp. 21–22). Horvitz's curiosity supports Walsh's sense that AI researchers are "at the stage of pouring

together different combinations of substances and seeing what happens, not yet having developed satisfactory theories" (Walsh, 2023, p. 22).

Ida Lovelace and dazzling diversity

Through references to creativity, Walsh introduces the critical power of human creativity and insight by narrating the significant scientific and historical contribution of Ada Lovelace to AI. Ada Lovelace was the daughter of the great British poet Lord Byron of the nineteenth century Romantic era. As a gifted mathematician Lovelace:

[. . .] worked alongside Charles Babbage in his unsuccessful attempts to build a mechanical computer. [. . .] she wrote what is generally considered to be the first computer program for that mechanical computer. [. . .] was also the first to recognise that, while computers manipulate 0s and 1s, those 0s and 1s could represent things besides numbers. [. . .] could represent dots in a beautiful picture, or the musical notes from a melodic symphony (Walsh, 2023, p. 94).

Lovelace recognised the dazzling diversity of data and the potentially limitless scale that could be applied.

Walsh's comment on Newcomen's steam engine highlights the surprises of research by observing that when:

[. . .] the steam engine was invented by Thomas Newcomen, no one worried about the exhaust gases. And yet the steam engine launched the industrial age, and that has led us inexorably to the climate emergency that we face today. Similarly, when Einstein invented his theory of general relativity, no one predicted that it would give us a global positioning system. It's equally hard to imagine the less direct effects of artificial intelligence (Walsh, 2023, p. 38).

This important theme of Walsh's book is explored in Antony Funnell's ABC RN science podcast Future Tense. Antony Funnell, the presenter of one of the ABC RN's science program, Future Tense, interviews Professor Iwan Rhys Morus, Historian and Author at Aberystwyth University in Wales, Professor John van Reenen, an Economist at London School of Economics and Massachusetts Institute of Technology and Professor Jay Bhattacharya, Professor of Medicine, Economics and Health, Stanford University about trends in scientific research.

Funnell describes a fascinating comparison about how novel ideas can take a while to show the benefits. Funnell notes that the concept of CRISPR:

[. . .] the idea how bacteria fend off attacks from viruses [. . .] now just looking at the bacteria stores in genetic material the history of attacks it has received from viruses [. . .] that person wasn't planning on gene editing software [. . .] what happened is people used these ideas over decades to amend it or change it [. . .] but it's the set of novel ideas that accumulate one after another that led to the gene editing breakthrough that CRISPR actually represents (Funnell, 12 March 2023).

The confusion created by alchemy is exacerbated by forces propelling organisational climate change. Seven forces are driving business organisational climate: personal compiling to curating to blogging; expression explosion: the new literacies; rhizomic networking cultures and emerging localism; reinventing new information professionals; digital-creative economies and big data; people power: tacit personal knowing through memory to agile remembering; King Lud in the digital-memory age: deskilling, upskilling, reskilling and appropriability (McFadzean, 2023, 12–26). These forces may be ripe for sensible AI support to lessen the negative effects and to identify and develop the positive influences on organisational climate change.

The evolving science of human-AI sense-making

The meeting of the minds?
Walsh reviews five key differences between human cognition and AI computational learning. Currently, humans and computers learn very differently which highlights the gap between people and AI. Firstly, Walsh reminds us that machine learning algorithms "typically require thousands of examples to recognise a single concept. Humans, on the other hand, can learn from a single example." Machine learning "transfers poorly outside the training set. Humans are marvellous at applying their learning to new domains" (Walsh, 2023, p. 195).

Power of people filtering
Secondly, machine learning for data collecting while able to collect massive volumes of diverse data very rapidly has filtering problems. Walsh describes part of the human-machine central dichotomy when our senses:

> [. . .] generate megabits of information every second, but out conscious brains can only process tens or hundreds of those bits. A smartphone or camera shooting high-definition video also captures megabits of information every second, and quickly struggles to buffer all that data. Fortunately, we don't need all that data. We only really need to hear the screech of tyres and the honk of the horn over all the other noise of the city.

Walsh notes that our human "reticular activating system performs the vital function of filtering the immense torrent of sensory information coming into our brains down to something more manageable" (Walsh, 2023, p. 159).

Artificial intelligence neural networks
Thirdly, Walsh highlights another critical difference between machine learning and our brains by noting that neural networks of machine learning:

> [. . .] are only loosely related to biological neural networks. [. . .] the human brain doesn't use back propagation, the weight-updating algorithm which is at the centre of deep learning. The human brain is asynchronous. Neural networks are not. The human brain has a complex interconnected topology. Neural networks typically have a simple layered structure (Walsh, 2023, p. 196).

Walsh reasons that:

> [. . .] deep learning methods like those at the heart of AlexNet were quickly shown to be successful in other domains, such as speech recognition and natural language processing. Deep learning has been successfully applied to domains from drug design to climate science, from material inspection to board games. AlexNet itself was built on research into neural networks dating back to the 1960s (Walsh, 2023, p. 30).

Fourthly, processes are another critical difference between artificial and human intelligence. Computers process massive volumes at great speed and significantly, never forget. People:

> [. . .] by comparison, work at much slower biological speeds, performing hundreds or at most thousands of operations per second. There is no doubt the human brain is an impressively parallel machine, with billions of neurons working together simultaneously. [. . .] But it is clear that human brains and computers have very different architectures with which to achieve intelligence. It is unsurprising that intelligence is different when the underlying architecture is very different (Walsh, 2023, pp. 144–145).

Artificial intelligence is learning how we speak!
A major theme of Walsh's book is the key and fundamental role of communication in human-computer interaction. Walsh reasons that:

AI is needed to power those interactions: first to understand your speech, and then to speak back to you. This is why artificially intelligent virtual assistants are a big part of the future of human–computer interaction. But this opens up a host of problems, especially when these virtual assistants are designed to fool you into thinking they're real (Walsh, 2023, p. 71).

West and Allen support Walsh's analysis noting the key translating and communicating function of voice intelligence (McFadzean, 2023, 178–179). Turow (2021, p. 238) describes how the Web has led to an:

[. . .] unprecedented ability to gather enormous amounts of information about individuals [. . .] following their activities on the web, and [. . .] the ability to reach people through a variety of digital platforms— advertising on websites, ads on Google and Bing search engines, email, social media such as Facebook and Twitter (McFadzean, 2023, pp. 183–184).

Incredible power of who we are: human consciousness
Fifthly, is almost the most significant, our individual human consciousness. Walsh highlights the immensity of this difference by reasoning that each person's consciousness can be described as "your awareness of your thoughts, memories, feelings, sensations and environment." Walsh asks how "do the billions of neurons in your brain – or, indeed, the trillions of cells in your body – act or feel as one? How do you feel you? What is the experience you have of being conscious?" [. . .] It is your perception of yourself and the world around you. Of course, this is subjective and unique to you. I have no way of directly knowing what it is like to be the conscious you (Walsh, 2023, 155–156). Walsh notes that currently "no coherent scientific theory exists to explain the experience that you had this morning when you opened your eyes. Consciousness remains one of the greatest mysteries left to science" (Walsh, 2023, p. 160).

Walsh unravels consciousness by reasoning that as "humans, we can choose between good and bad, and we must rely on some sort of morality to make a good choice. But if machines don't have free will – and computers are arguably the most deterministic devices we have built – how can they be moral? And is consciousness somehow connected?" (Walsh, 2023, p. 155). Walsh quotes the Australian David Chalmers who reasons consciousness "poses the most baffling problems in the science of the mind. There is nothing that we know more intimately than conscious experience, but there is nothing that is harder to explain. All sorts of mental phenomena have yielded to scientific investigation in recent years, but consciousness has stubbornly resisted" (Walsh, 2023, pp. 159–160).

Enthusiasm or the fakery and hype of emerging AI

Powerful simplicity of common sense
Fifthly, and finally, Walsh describes the current era as one of hype and fakery. Walsh cautions that researching, developing, applying and testing AI has seven issues to confront. Firstly, the:

[. . .] fundamental problem is that we have very little idea how to program such common sense into machines. And without it, they will at best be "idiot savants" – superhuman at a few narrow tasks, but lacking in our all round general intelligence. And building machine savants can be dangerous. If someone is good at faking one intelligent task, we are likely to trust them on others. But with AI, trust could easily be a perilous mistake (Walsh, 2023, p. 202).

Computers are incredibly powerful yet ironically fragile. Walsh notes this is a "dangerous difference, since we are constantly being caught out by machines falling over on tasks that would not defeat humans" (Walsh, 2023, p. 145).

Humanity's aces: social intelligence and tacit knowledge
Two, computers lack social intelligence. Social intelligence could be grounded on collective individual personal knowing. Personal knowing could also be labelled tacit knowledge or implicit memory. Through social intelligence people "empathise with other people because we share the same biology. We can guess how someone will react because it will likely be similar to how we would react" (Walsh, 2023, p. 148). The under-rated significance of social intelligence is clearly described and assessed by Dr Richard Baldwin. Baldwin notes social intelligence (read human general intelligence) allows communities and teams to share and collaborate through the theory of the mind and empathy. Conversely, AI uses narrow, algorithmic intelligence (McFadzean, 2023, pp. 63–65).

Science journalist and communicator Annie Murphy Paul explains that (McFadzean, 2023, pp. 161–162) "listening—and telling—is at the heart of one more way we can use social interaction to enhance our thinking: through the exchanging of stories" through the work of five researchers: Hugo Mercier, Christopher Myers, Brian Uzzi, and the coauthors Andy Clark and David Chalmers.

Paul notes Hugo Mercier describes how people are:

> [...] "natural born arguers" who can deliberately deploy that innate capacity to correct the mistakes, clarify our thinking, and reach sound decisions. The key is to approach the act of arguing with the aim not of winning at all costs but of reaching the truth through a vigorous process of advancing claims and evaluating counter—claims [...] when we make the best case for our own position while granting the points lodged against it (Paul, 2021, p. 204).

Myers' research covers how workers identify and share critical workplace information informally and through immediate, shared yet ironically distinctive experiences rather than the conventional, traditional training and set course processes. This "tacit knowledge" generically comprises essential need-to-know information about workplace workflows and processes (often as shortcuts, workarounds and heuristics as rules of thumb that make businesses work informally and efficiently though not always necessarily legally or safely). Myers speculates that this information, labelled "tacit knowledge", is identified as needed, and the real value of this knowledge is that it is expressed in context and finely, specifically tailored to meet the information needs for an immediate situation (Paul, 2021, pp. 206–209).

Uzzi, a professor of management at Northwestern University, draws the theme of Mercier's research further by observing "that the process of knowledge creation has fundamentally changed." More broadly, Uzzi notes, "Almost everything that human beings do today, in terms of generation of value, is no longer done by individuals. It's done by teams" (Paul, 2021, p. 231). Uzzi is one of many scholars exploring the emerging study of collective intelligence. Paul (2021, p. 231) recommends:

> [...] "error pruning" by team members in a specific sequence of actions in response to our teammates' contributions: we should acknowledge, repeat, rephrase, and elaborate on what other group members say. Studies show that engaging in this kind of communication elicits more complete and comprehensive information. It re-exposes the entire group to the information that was shared initially, improving group members' understanding of and memory for that information [...] it may seem cumbersome or redundant, research suggest that this kind of enhanced communication is part of what makes expert teamwork so effective (McFadzean, 2023, pp. 162–163).

Fourth and finally, Paul observes that research of Clark and Chalmers into the extended mind space describes experts as having:

> [...] learned how best to marshal and apply extra neural resources to the task before them [...] actually do more experimenting, more testing, and more backtracking than beginners [...] more apt

than novices to make skilful use of their bodies, of physical space, and of relationships with others [...] are less likely to use their heads and more inclined to extend their minds (Paul, 2021, p. 16).

Part of Paul's answer about how human and AI currently interacts and may coevolve can be found in the early history of AI (McFadzean, 2023, p. 163). This could also be called Consciousness.

Fakery and evolving artificial intelligence

Walsh's analysis of fakery and the value of social intelligence to essential collaboration and cooperation for humans by highlighting one of the key flashpoints between human and AI: how to define, create and sustain trust. Walsh describes how the unfortunate synergy between imagery and deep learning has encouraged the making and distributing of video and audio manipulated online vblogs known idiomatically as deep fakes. Put simply, deep fakes focus human attention as "we are social animals and mostly prefer to interact with people rather than computers. Deep fakes are thus an engaging means of getting our attention" (Walsh, 2023, p. 74).

Deep fakes

Walsh then correctly challenges deep fakes by noting deep fakes have "the potential to become weapons of mass persuasion. [...] could use the same AI technologies pioneered by Cambridge Analytica to tailor the conversation to the different political sensitivities of every voter" (Walsh, 2023, p. 79). Walsh reasons that one of the strengths of social intelligence is our human Achilles heel, people:

[...] are easily fooled by such fakes, despite the fact that the capacity to recognise voices and faces well is an important human skill. Our ability to cooperate depends on our ability to remember and recognise voices and faces. Indeed, there are regions of the brain – the superior temporal gyrus for voices, and the fusiform gyrus for faces – devoted to these tasks. But deep fakes are already good enough to fool these specialised parts of our brains (Walsh, 2023, p. 75).

ChatGTP

ChatGPT now enters the shadowy, grey space of what is fakery for text? Walsh cautions that as ChatGPT or similar software becomes expansively accessible and:

[...] more sophisticated, there are concerns about the potential misuse of this technology. It could be used to spread misinformation or to generate fake news at an unprecedented scale. The creators of ChatGPT have already taken steps to address these concerns by limiting the chatbot's ability to generate content related to sensitive topics like politics or religion (Walsh, 2023, p. 50).

Walsh offers a reason for his caution by reminding readers that like all sources, ChatGPT has problems with biases. Walsh observes historically:

[...] despite OpenAI having put a lot of effort into trying to prevent the system from saying anything offensive or controversial. Indeed, this was one of the most important ways in which ChatGPT improved upon GPT-3. Since GPT-3 was trained on a large chunk of the internet, and since there's a lot of offensive and controversial content to be found on the internet, GPT-3 will sometimes output text that is offensive or controversial. It will, for example, happily agree with an interlocutor that climate change is a conspiracy, and that black lives don't matter (Walsh, 2023, pp. 55–56).

Privacy, transparency and intellectual property rights
The fourth issue concerns the hidden and plain-sighted issue of privacy. The business model known colloquially as Surveillance Capitalism has also been labelled "'stalker economics' [. . .] with strong incentives to ignore your privacy. [. . .] AI algorithms are processing all the personalised data that the tech companies are recording" (Walsh, 2023, p. 181). Fifth, despite AI companies highlighting "transparency as a key principle in the responsible deployment of artificial intelligence [. . .] transparency seems to be on the decline. This is driven, in part, by the commercial pressures of winning the AI race" (Walsh, 2023, p. 189). The sixth issue covers:

> [. . .] intellectual property law. I've already discussed how it is challenging copyright law, but it is also challenging other types of intellectual property, such as patents. Patent law is based on the assumption that inventors are human. Courts around the world are wrestling with how to cope with patent applications that name an AI system as the inventor (Walsh, 2023, pp. 211–212).

Finally, some policymakers recognise that the Silicon Valley business models need to be regulated to avoid the creation or exploitive expansion of various digital monopolies.

The European Union's artificial intelligence act 2023
Walsh draws attention to the "Artificial Intelligence Act is a proposed new European law on artificial intelligence. This would be the first law on AI by a major regulator anywhere" (Walsh, 2023, p. 212). Since publication the European Union passed the Artificial Intelligence Act on 14 June 2023. Significantly, the European Union's Artificial Intelligence Act (2023):

> [. . .] divides AI applications into three risk categories. [. . .] an unacceptable risk, such as government-run social credit scoring of the type used in China; high-risk applications, such as a CV scanning tool that ranks job applicants; these are legal, subject to specific requirements. The third category consists of any other AI applications, and these are left largely unregulated (Walsh, 2023, pp. 212–213).

Ghosh's warning about algorithmic age
Walsh's comments supporting the need for algorithmic transparency aligns with Ghosh's highlighting the alchemy of algorithms that are used by data scientists to create inferences (McFadzean, 2023, pp. 65–66). Ghosh reasons that humanity has:

> [. . .] left the information age and are now in the algorithmic age. It is an entirely novel commercial landscape, in which the business model dictates that the value of our information is not the details themselves but rather the amalgamation, inference development, and automated algorithmic application of the information (Ghosh, 2020, pp. 69–70).

Smith supports Walsh and Ghosh by noting algorithmic software programs seeking to curate the data "have become so opaque that it is comparable to medieval pre-science" (Smith, 2020, pp. 177–178). Smith warns that flowing from the alchemy of algorithms, these:

> [. . .] now search for their own representational atoms, and self-organize based on information they glean from big data. That is what has made AI commercially viable, but to give these algorithms sufficient flexibility in finding their own atoms, we have had to give them scale and structure that is intractably complex. Yet we expect them to deliver generalizations that are effective, through emergence (Smith, 2017, pp. 177–178).

Explainable AI or XAI
As an antidote to the inferential profiling noted by Ghosh through "black box" algorithms, Ball (2022, pp. 29–30) reasons the:

[. . .] answer is to smash open the "black boxes" of the systems we are using (and feeding). To turn a black box of secrecy into a glass box of openness, we need to employ a methodology called "explainable AI" (XAI). XAI is already kicking off in the USA, where health insurance companies have been required, through case law, to reveal their data management and analysis algorithms and processes.

Australia's Artificial Intelligence Ethics Framework 2019
Returning to the earlier theme, recordkeepers are like alchemists who transform data into valuable records by applying XAI. As one of the AI world's leaders the Australian Commonwealth Department of Industry, Science and Resources published Australia's Artificial Intelligence Ethics Framework on 7 November 2019. The Framework is founded on eight key principles: AI systems should benefit society; respect for human rights and diversity; inclusive and accessible; protect privacy and security; reliable and safe. The last three are vital to AI-human interaction. AI should be: transparent and explainable; contestable and accountable. Contestable is fundamental to the other seven because to be contested all parties should be able to understand and make sense of AI (Australian Commonwealth Department of Industry, Science and Resources 7 November 2019, Australia's Artificial Intelligence Ethics Framework, www.industry.gov.au/publications/australias-artificial-intelligence-ethics-frame work/australias-ai-ethics-principles). This describes and explains XAI.

Microrecords and recordkeeping metadata
XAI turns data into records by adding recordkeeping metadata as assertive records about the data. This allows the data to be identified, described and verified when registered as a business record under a business' governance regime (McFadzean, 2023, pp. 20–21, p. 65, p. 155 and p. 200). I labelled this data as microrecords. Describing the data for the record, pardon the pun, would include responses to five questions key to recordkeeping informatics. International standard ISO 15489 is based on informatics. ISO 15489 authenticates authoritative records. ISO 15489 is key to the principles of explainability and contestability. Essentially, data morphing to microrecords, is evidence of human activity and therefore a form of historical evidence of past activity. The methods of historical enquiry and disciplines that study humanity can serve data analysis well. The five questions at the end of this article are questions from historical analysis and assist in making AI explainable.

This information about the data are metadata which can be defined as assertive records identifying the data. These five questions form the metadata identify the data collectively. These assertive statement, as metadata makes the data XAI:
Purpose (what is the question?)
1. Why is the data being collected (history of idea or request, reason, purpose, business case, vendor or internal business generated request or idea, evidence based research, professional identities and experiences of the data scientists, governance officers, executives and data management recordkeepers)?
The source (searching for evidence)
2. What data is being collected (the plain language, jargon-free description of the algorithm/s and learning data set/s, the business case explaining what is required and why, source or origin, history of the data, single or multiple data sets, age and geography of the data)?
Provenance (collecting the evidence)
3. How was the data collected (source/s, sensors, ownership and equity, governance and integrity, how does it comply with jurisdictional legislation and regulation, corporate governance standards [security of the data system]?

Provenance (collecting the evidence)

4. When and where was the data collected (times and places the data was collected or transferred, history of earlier data sets and how these may have changed, geospatial mapping and geographic maps of data)?

Purpose, source and provenance.

5. By whom, of whom and for whom was the data created, captured, collected and transferred (source/s, governance of authority [author and owner equity] [old archival principle of chain of custody], conditions of access [transfer, use, storage and sharing], is the data safe and ethical to use, what are the risks and possible consequences)?

The five identity questions that can convert data to microrecords become critically important when applied to three themes for creating, capturing and applying generative AI. Firstly, generative AI can be used for creating and applying synthetic data (replicating real data created from clinical trials, laboratory tests, field observations, test platforms, data collection systems or software programs) to system modeling. This covers renewable energy from a variety of sources and networks, production, distribution, storage and use, biotechnology (medical research, agricultural production), epidemic and pandemic research and climatology (extreme or latent weather events or trends).

Secondly, generative AI can be used to manage robotic systems focusing on navigation, close motion planning (individual, teams, fleets or swarms), including the mapping of roles and processes to govern human-robot collaboration. Thirdly, planning includes computer-aided designing and process mapping developed through sequential prediction modeling. This third theme extends to many forms of modeling using synthetic data to map and evolve movement and logistics including military deployments, managing infrastructure and traffic flows (rail, maritime, road and aviation, goods and services) and movements and migrations of people and exchanges of ideas, data and information.

Recordkeeping informatics consists of information culture (business rules, workflows, roles and responsibilities), metadata (as assertive identity records), access (terms and conditions to protect and make records accessible) and continuum thinking (mapping the movement of information in records through time and systems as it is re-used into forms of business information) (McFadzean, 2023, p. 157).

Professor Toby Walsh and artificial intelligence 2023

The last word should belong to Professor Walsh who clearly sums the essential issue in an earlier book:

> [. . .] machines will always and only ever be machines. They do not have our moral compass. Only humans can be held accountable for their decisions. [. . .] Machines aren't, and probably won't ever be, conscious of their actions. [. . .] If we can't build moral machines, it follows that we should limit the decisions that we hand over to machines. But here's the rub: we're not going to be able to avoid handing over some of our decisions to machines (Walsh, 2022, p. 222).

Almost as an ending rather than in the text Walsh partially answers the Minsky-McCarthy and Weiner debate reflecting:

> We should therefore not be distracted by the immediate and flashy present. There is not enough thoughtful discussion about the more distant AI future, when machines are much more capable. What, for instance, will the impact be on the workplace when artificial intelligence matches or exceeds human intelligence? How should we start educating kids today for the AI-enabled jobs of the second half of this century? And what will the impact of AI be on science? Could it perhaps help us accelerate the rate of scientific discovery? (Walsh, 2023, p. 37).

Andrew James McFadzean
Independent Scholar, Canberra, Australia

References

Ball, C. (2022), *Converge. A Futurist's Insights into the Potential of Our World as Technology and Humanity Collide*, Major Street Publishing, Melbourne Victoria.

McFadzean, A. (2023), *Memory Curators and Memory Archivists in the Digital Memory Age*, Cambridge Scholars Publishing, Newcastle.

Pasquale, F. (2020), *New Laws of Robotics. Defending Human Expertise in the Age of AI*, The Belknap Press of Harvard University Press, Cambridge, MA.

Smith, R.E. (2020), *Rage against the Machine. The Prejudice of Algorithms and How to Stop the Internet Making Bigots of Us All*, Bloomsbury Business, London.

Spar, D.L. (2020), *Work Mate Marry Love. How Machines Shape Our Human Destiny*, Picador. Farrar, Strauss and Giroux, New York, NY.

Upward, F., Reed, B., Oliver, G. and Evans, J. (2018), *Recordkeeping Informatics for a Networked Age*, Monash University Publishing, Melbourne.

Walsh, T. (2022), *Machines Behaving Badly. The Morality of AI*, La Trobe University Press with Black Inc. Publishing, Collingwood, Victoria.

Walsh, T. (2023), *Faking It: Artificial Intelligence in a Human World*, La Trobe University Press, Melbourne, Victoria.

Further reading

Bell, G. *ABC Boyer Lecture 2017: Fast, Smart and Connected: What is It to Be Human, and Australian, in a Digital World? Public Lecture Podcast Sound Recording*, Australian Broadcasting Corporation, Sydney, available at: http://hdl.handle.net/1885/130314

Grant, R. (2017), "Recordkeeping and research data management: a review of perspectives", *Records Management Journal*, Vol. 27 No. 2, pp. 159-174.

Oliver, G. and Harvey, R. (2016), *Digital Curation*, 2nd ed. American Library Association. Neal-Schuman, Chicago.

Reed, B., Oliver, G., Upward, F. and Evans, J. (2018), "Multiple rights in records: the role of recordkeeping informatics", in Brown, C. (Ed.), *Archival Futures*, Facet, London, pp. 99-116.

Roberts, J. (2015), *A Very Short, Fairly Interesting and Reasonably Cheap Book about Knowledge Management*, SAGE Publications, London.

Suleyman, M. and Bhaskar, M. (2023), *The Coming Wave. AI, Power and the 21st Century's Greatest Dilemma*, Random House, Penguin.

Tett, G. (2021), *AnthroVision. How Anthropology Can Explain Business and Life*, Penguin Random House, London.