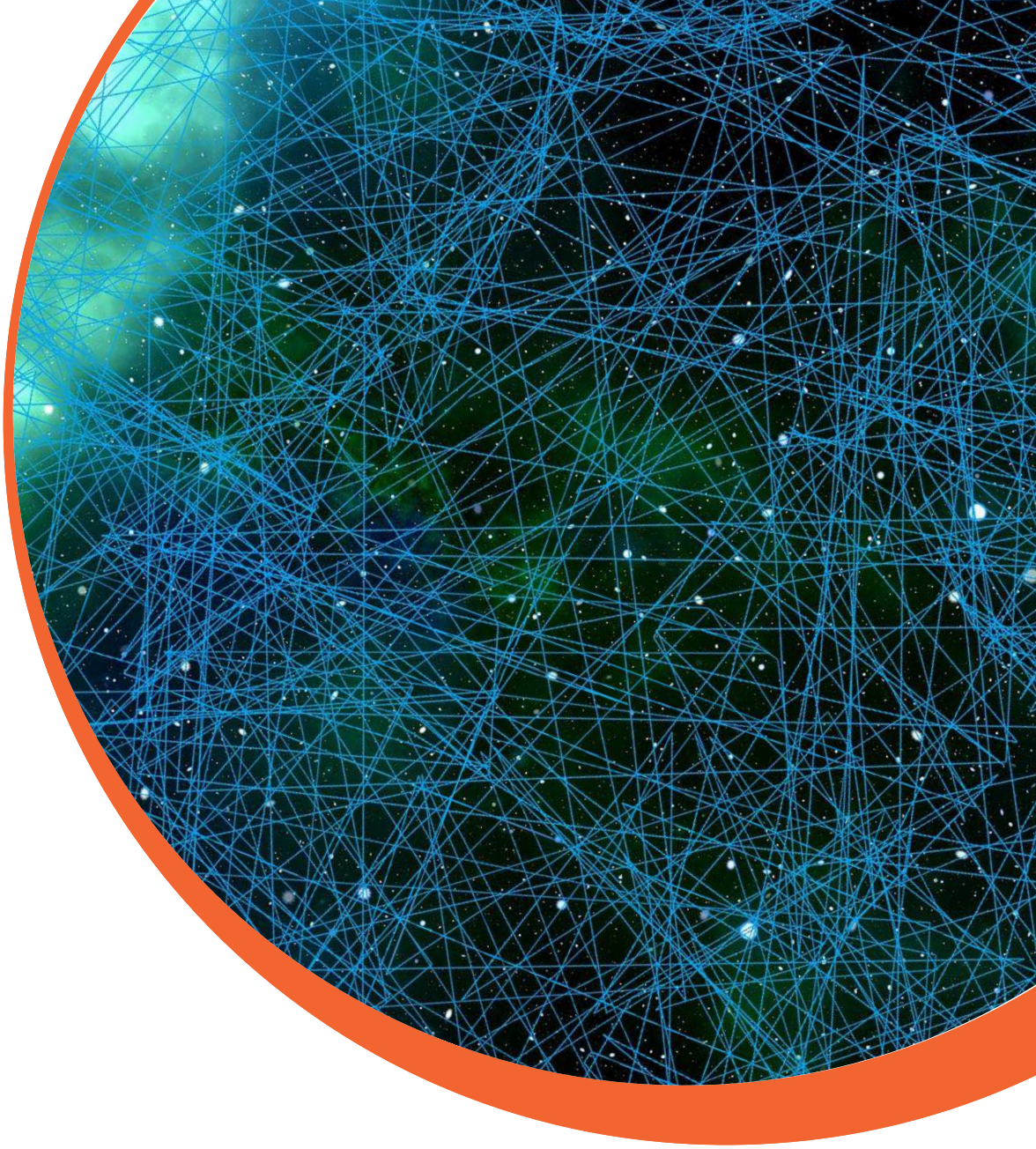

Thinking Critically About **AI**

Discussion paper by

Calum Samuelson

Jubilee Centre
BIBLICAL THINKING FOR PUBLIC LIFE

Thinking Critically About AI

This article acts as a second installment in our ongoing thinking about AI and robotics. It seeks primarily to provide clarity and encouragement in the midst of dialogue that seems increasingly confused, contradictory, inaccessible, and provocative. It also aims to generate dialogue and inspire critical thinking. It is neither comprehensive nor exhaustive.

'AI' as currently used is a relatively poor and misleading term. Berkeley Prof. **Michael Jordan** explains: 'Most of what is being called 'AI' today, particularly in the public sphere, is what has been called 'Machine Learning' (ML) for the past several decades.' It is true that ML itself is increasingly being implemented in tangible ways throughout society, but it has been in use for quite a long time (e.g. the Apollo Spaceships in the 1960s) and without explicit mention (millions were listening to ML-driven Pandora Radio in the early 2000s).

Two important shifts can help us unpack some of the ambiguity surrounding 'AI': **shifts in terminology and shifts in perception.**

1) Shifts in Terminology: The term 'Artificial Intelligence' has gone through several stages of 'rebranding'. Originally (1950s), 'Artificial Intelligence' referred to the 'the heady aspiration of realizing in software and hardware an *entity* possessing human-level intelligence' [emphasis added].[1] Jordan calls this 'human-imitative AI' (others might call this AGI or 'Artificial General Intelligence'[2]). Today, the term 'AI' is more liberally applied to relatively 'low-level intelligence' developments in fields ranging from engineering to statistics, and to an even broader range of technologies in popular discussions.

2) Shifts in Perception: In a way, 'AI' is a paradoxical and moving target, a phenomenon suggested by the so-called 'AI effect'. In the words of **Rodney Brooks**, 'Every time we figure out a piece of it, it stops being magical; we say, "Oh, that's just a computation".' Alternatively, **Douglas Hofstadter** famously said, 'AI is whatever hasn't been done yet.' The usual example given is the defeat of Garry Kasparov by **IBM's Deep Blue** in 1997, which critics then retrospectively quipped didn't actually demonstrate 'intelligence'. Potentially, this means that what we popularly agree is 'AI' today may not be viewed as such in 20 years.

---

[1] https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7.

[2] Dialogue often includes three levels of AI: 1) **ANI** (Artificial Narrow Intelligence), which is what we have currently; 2) **AGI** (Artificial General Intelligence - when computer intelligence becomes as good or better than that of a human being across the board), which if/when achieved may signal the so-called 'Singularity'; 3) **ASI** (Artificial Super Intelligence), which is when 'things get scary' (AI becomes *much* smarter than humans, in every way).

Another factor affecting ambiguous conceptions of 'AI' relates to the way that entrepreneurs, investors, film makers, novelists, marketers, and journalists—seeking attention or profit—have *exaggerated or misrepresented the actual abilities and/or characteristics of AI*. Along these lines, it is important to note that some of the most impressive 'demonstrations' of 'AI' have, in fact, been staged.[3]

At the moment, *no single person has the final say on these issues*. In the midst of ongoing conversations, it is worth reminding ourselves that we can still be reasonably critical of people who are more qualified than we are in certain areas. If a speaker begins a talk about AI by casually mentioning that they work with 'philosophers, mathematicians, and computer scientists', we can fairly observe the absence of *neuroscientists, psychologists,* and *sociologists* in the list, and decide to suspend judgment until sounding these other views as well. This need not mean we are claiming to be better or smarter; it simply indicates an unwillingness to be swayed by simplistic appeals to authority. One would be completely justified in trusting Stephen Hawking's word about black holes, but equally unjustified in trusting his opinion on the provenance of a rare medieval manuscript.

'Expert' statements about AI should not be accepted unequivocally.

Ultimately, whilst no statements about AI should be accepted unequivocally, most at the frontlines of this field would resonate with the opening quotation from Prof. Michael Jordan—which involves much more than a mere penchant for semantics. Indeed, AI terminology influences the *public's perception*, which is largely characterised by fear of some type of radical takeover or displacement due to the **'exponential growth of AI'**. It is to this question that we now turn.

The Exponential Question

Exponential growth is a big deal because it makes it difficult for accurate predictions about the future based upon the past to be made. Taken generally as a frame of mind, the implications of this concept undergird much of the urgency in current dialogue around 'AI'. **A key issue here involves processing/computing speed.**

The main driver behind discussion about exponential growth is '**Moore's Law**' (the observation that computing power doubles every 18 months - which, of course, is not an actual law of physics).[4] However, it appears that Moore's Law is significantly slowing down,

[3] This usually involves 'staged conversations and teams of puppet-masters pulling strings from behind the curtain', as explored in 'How real is that Atlas robot video?' A great recent example is when Jimmy Fallen 'met' Sophia.

[4] Moore's Law has not been exactly as consistent or as accurate as some flippantly imply. The number of transistors has roughly doubled every 18 months rather than two years (which was itself a revision of an earlier prediction) and there have been several periods where growth has been faster or slower than this.

and many believe it will soon come to an end.[5] This is because transistors can only be made so small. Chip manufacturers have long used *nanometers*[6] as the metric for their transistors. The smallest currently in production are 7nm, although 5nm prototypes have already been created by multiple companies. For perspective, the A11 Bionic chip in the iPhone X features 10nm transistors. The problem is that transistors will soon be only one order of magnitude larger than silicon atoms themselves (0.2nm) which is likely to lead to an occurrence known as **quantum tunneling**, meaning that electrons jump across the boundaries set up for them. At this point, a transition to some type of **quantum computing** may be required—but a smooth transition is not inevitable, and certainly not at the current 'exponential growth rate' of Moore's Law. Ergo, **if the rate of growth in computing power plateaus or even simply slows down, many of the more extreme predictions about AI will be revealed as too ambitious.**

This is certainly not to deny that rapid technological growth is currently taking place and will likely continue to do so for the foreseeable future. But it is vital to debunk false narratives about AI *inevitably* spiraling out of control in an 'intelligence explosion'.[7] Obviously, there are other factors involved in this growth besides just processing speed.[8] Luke Dormehl rightly observes: 'Even if Moore's Law was (sic) to end tomorrow, optimizing today's software would still provide years, if not decades, of growth — even without hardware improvements."[9] But once again, exponential growth is hardly inevitable. **Ray Kurzweil**, sometimes touted as the ideal proponent of exponential tech growth[10] (including the advent of AGI), has himself admitted that only the *effort to produce AGI* is actually growing exponentially (i.e. more and more people are spending more and more time and money on creating AGI).[11]

An additional problem with the 'exponential growth' paradigm involves *how we measure intelligence*. The basic linear, single-dimension model proposed by thinkers such as **Nick Bostrom** has been

[5]Although some give important caveats about the occurrence of 'S-curves' in exponential growth, this still takes the overall narrative for granted.
[6] A nanometer is one billionth of a meter. The average human hair is roughly 90,000nm wide.
[7] https://en.wikipedia.org/wiki/Intelligence_explosion.
[8] It seems that some thinkers work more from the idea of exponential growth in global population or technologies in general, both of which are flawed in several ways. For an example, see the beginning of Sam Harris' TED Talk: https://www.youtube.com/watch?v=8nt3edWLgIg&t=4s.
[9] One of the most significant aspects that will continue to boost computing speed is *cloud computing*, which enables local systems to outsource the 'heavy lifting' to much larger ones via the cloud.
[10] Kurzweil's own phrase for this is the 'Law of Accelerating Returns.'
[11] See Kevin Kelly's article in Wired: https://www.wired.com/2017/04/the-myth-of-a-superhuman-ai/.

strongly criticised;[12] 'intelligence' is an extremely complex, contended subject, and there is plenty we still don't know about animal intelligence, let alone human intelligence. One way to shed some light in this area is to consider how both *quantity* and *quality* of information influences intelligence. In his TED Talk, **Sam Harris** defines intelligence as 'the product of information processing in a physical system'. Clearly a certain type of 'intelligence' or 'knowledge' will continue to grow as we collect and organise more information. Higher *quantities* of information don't necessarily correlate with higher *qualities* of information which are needed for intelligence and understanding.

However, problematic assessments of intelligence are most often casually circulated in casual and subtle ways. A fitting example is located in the applauded 2015 open letter from the **Future of Life Institute**. The unfortunate clause in the otherwise excellent statement assumes, 'everything that civilization has to offer is a product of human intelligence'. What about human qualities such as perseverance, loyalty, and love? Even things such as inspiration and optimism could be argued to be more important than intelligence. Such instances demonstrate how easy it is for us to get caught up in the various narratives surrounding AI.

Accordingly, several leaders in this field have been careful to emphasise that developments and progress have come about as programmers, statisticians, computer scientists, engineers, etc. have worked on **very specific solutions to very specific problems.** One does not simply set out with the task of, say, 'creating an algorithm to improve customer service.' Instead, incredibly complex facets of the interactions with customers must be analysed, and a very detailed solution must be articulated in order to even begin writing code that *might* work towards that end. Other misperceptions abound, including the huge gap between general 'nanotechnologies' (involving structuring at the molecular level—such as **'carbon tubes')** and the more vague, sci-fi inspired 'nanobots'.[13] In short, the exponential growth of AI—though very serious and significant—is neither automatic nor inevitable.

Exponential growth of AI is neither automatic nor inevitable.

Such understanding provides a critical basis from which to approach speculation about existential threats. Given the considerable range of disciplines and specialties that are now associated with AI, it should be no surprise that opinions regarding how humans will fare in the future are dispersed across a spectrum between more optimistic thinkers (**Ray Kurzweil**, **Grady Booch,** et al.) and more pessimistic

[12] As one example, see MIT article: 'Progress in AI isn't as impressive as you might think', https://www.technologyreview.com/s/609611/progress-in-ai-isnt-as-impressive-as-you-might-think/.
[13] For a helpful insight into this area see, 'The next step in nanotechnology': https://www.youtube.com/watch?v=Ds_rzoyyfF0.

thinkers (**Nick Bostrom**, **Elon Musk**, **Stephen Hawking**). This is 'the existential question', and it is the subject of this next section.

The Existential Question

Many of the most important developments are taking place not merely at the broad level of AI or even Machine Learning (which are nearly ubiquitous already in the lives of many urban dwellers), but rather at the level of **Deep Learning**. This involves complex neural networks, which 'learn' by mimicking the brain's network of organic cells ('deep' refers to the *number of layers* in the network, not some type of 'deep' or profound 'understanding'). According to the 2018 MIT Tech review, one of the top ten new technologies is **Generative Adversarial Networks (GAN)**, which pits two neural networks against one another in an attempt to make each other better. It is precisely the potential of such 'self-improving' systems that raises concerns about the future existence of humanity.

Interestingly, relatively few people seem to be afraid of AI development *in their own particular discipline*. Their own projects are always described calmly and sensibly as something that will greatly benefit humanity. It's the *other* applications of AI that are dangerous: sex robots, drones, medical implants, etc. One looks in vain for any signatories of the Future of Life open letter with a title similar to 'Weapons Developer' or 'Tactical Programmer', who presumably are precisely the type of people we would want signing such a statement.

A fascinating case in point comes from **Stuart Russell**, a well-known figure in the world of AI, who helped create a recent, evocative video about autonomous, weaponised drones called Slaughterbots. Let it be said that the overall thrust concerning 'autonomous weapons' conveyed in this video is certainly legitimate, and as MIT Prof. **Max Tegmark** observes, we should seek to replicate the successful approach previously taken with bioweapons to prevent their dissemination. But the valid concern that certain parties may utilise autonomous weapons to achieve their own *human ends* is simply not synonymous with the fear that such weapons—by means of their 'autonomy'—may overthrow human control in order to accomplish their own *robot ends*. Fundamentally, we would do well to separate the ways in which intelligent technologies will augment *humanity's potential to level existential crises* from the ways in which some believe super intelligent systems may themselves level such existential crises.

There is really no limit to the ways AI can be used. BUT, it requires considerable care, creativity and resources. A precisely fine-tuned and time-tested algorithm used by Amazon cannot simply be transferred to your furniture company to help you sell sofas. The flipside of this is that with a clear objective and vision, competent programmers can devise

genuinely new uses for the opportunities afforded by AI. Optimists love to dream about the wonderful applications that 'we can't even imagine right now'.

In this vein, *it may be worth focusing energies on* ***IA (Intelligence Augmentation)*** *which works with humans, rather than AI which is popularly believed to replace tasks previously performed by humans.* To echo the plea of several visionaries, we need to think clearly about *living with* IA rather than wasting time worrying about a dramatic 'takeover' by 'AI'. In fact, one way to summarise a crucial concern of some 'experts' is to say that they are worried about the public *misunderstanding and overreacting to technological developments*.

When considering the nature of challenges and threats that lie ahead, it seems probable that they will be less like an Orwellian-type of 'control through *oppression*' and more like a Huxleyan-type of 'control through *obsession*'.[14] We don't need to worry so much about intelligent machines oppressing us by watching our every move (although it is nearly certain that such systems will become more invasive and oppressive in some particular industries, companies, and countries), but rather about autonomous systems satiating our wants and desires so completely that we become inextricably dependent on them. The fact that such satiation is already largely realised among many communities in the West (e.g. social media, online shopping, instantaneous entertainment, and encyclopedic knowledge in our pockets) seems to strengthen the proposition that our future will involve ever-increasing degrees of 'satiation saturation.' Additionally, the mounting pressure placed upon programmers to move beyond their now obsolete maxim of 'move fast and break things'[15] has arisen not because people have been forcefully coerced to do things against their will, but because they have been subtly nudged to allow platforms like Facebook to increasingly provide them with dopamine.

The future may be less Orwellian, and more Huxleyan in nature.

However, it should be maintained that our impending Huxleyan addiction to various technologies will be *qualitatively distinct* from the types of strong, organic bonds that develop between family, friends, and spouses. In this sense, it is appropriate to acknowledge that social media provide us primarily with dopamine, not relational meaning, security, or fulfilment. Despite the fact that some claim AI systems will (or should) be endowed with *emotional intelligence*,[16] a sober assessment suggests that human emotions are even less understood than intelligence and are therefore even farther removed on the timeline of humanity's future.

[14] The themes of *1984* and *Brave New World*, respectively.
[15] Mark Zuckerberg.
[16] For one example see https://sloanreview.mit.edu/article/planning-for-the-future-of-work/.

In a similar vein, **Rodney Brooks** talks about 'technology gone wrong', which *wrongfully displaces* workers and—crucially—cannot be *understood* by normal workers. He wants to give workers *tools* (and already has) that they actually understand and can 'program' *by themselves*. He does not envision robots as 'meaningful companions' so much as 'servants' which accomplish onerous tasks on our behalf.

On this point, **Kevin Kelly** of *Wired* helpfully describes the chief asset of AI as a way of solving problems that is *different* from our way as opposed to strictly *better* than our way. For this reason, he alludes that it may be better to understand 'AI' as 'alien intelligence'. This line of thought can be illuminating. Humans think in a unique manner and have unique purposes, just as AI will 'think' in a unique manner and have unique purposes. For instance, the way AlphaZero 'plays' Go is very different from how a human 'plays' Go (it doesn't try to trick or intimidate), and the purpose for which it plays Go is also different (it plays to win, not to have fun or pass the time). In short, AI will never be able to encompass all conceivable purposes simultaneously. **This means the concept of 'general intelligence' (AGI) is itself fundamentally flawed** (or at least ill-defined)**.** There is no human who possesses a 'general intelligence' and—more importantly—no such thing as a general intelligence that can be associated with homo sapiens. We do many things very well, but we are not even as good as a squirrel at remembering the location of acorns. Thus, perhaps the most penetrating question is: what is the *specific purpose* of humankind? What are we 'programmed' to do? It would sound harsh to say that most 'AI' experts haven't a clue, but they certainly don't talk about it much. Admirably, a growing number of voices within the AI community are encouraging fresh pursuit of ethics, philosophy, and theology for this very reason.

None of this, of course, discounts the fact that our landscapes of work, war, sex, play, etc. are all likely to change radically in the coming decades; it just reiterates the classic belief that the true purpose or *vocation* of humanity cannot be reduced to these categories. In his recent book, Max Tegmark concludes that in order to ensure we have safe AI for the future, 'we need to capture the meaning of life.'[17] Though this may seem daunting to some, it may serve as a timely corrective to some of the assumptions of a post-Enlightenment world and hopefully be received as a wake-up call for people of faith all around the globe.

[17] Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (London: Allen Lane, 2017), p.279.