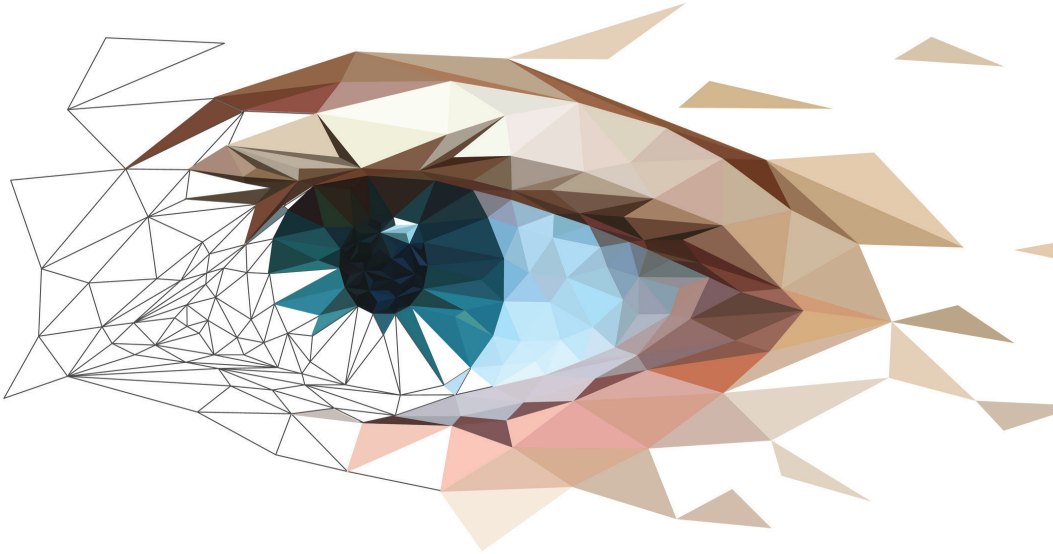
Long Distance
Christian

Artificially Intelligent?

Grappling with the myths, present realities and future trajectories of AI

Calum Samuelson



Artificially Intelligent?

This booklet dispels some of the sensationalism around Artificial Intelligence, asking instead how a fresh understanding of humanity can shape the trajectory of AI development. It draws on research interviews from ten leading AI practitioners and thinkers, and provides a distinctly biblical framework for understanding AI. Addressing all levels of expertise, the insights and guidelines provided will enable Christian leaders in church, business and public service to make informed responses to AI that are rooted in their faith.

"Informative, authoritative, insightful and practically orientated – I wholeheartedly endorse this paper which provides a helpful, theologically informed roadmap for the rapidly changing world of AI. Essential reading for church leaders and all those who wish to understand and engage with the challenges which lie ahead."
Professor John Wyatt, Professor Emeritus of Ethics & Perinatology, University College London

"Without doubt advances in AI look set to disrupt and transform the way in which we live our lives. This agile and evocative paper from the Jubilee Centre helps Christians to navigate this emerging landscape, identifying the risks and opportunities whilst grasping the significance of them within a biblical framework."
Nola Leach, Chief Executive at CARE (Christian Action, Research & Education)

"No field of technology is making faster progress than AI, and it may be than none has more potential to change how humans flourish. This thoughtful report will be accessible to a wide readership without requiring specialist knowledge of AI. Drawing on a Christian perspective on human relationships, it offers some challenging reflections and recommendations."
Professor Andrew Briggs, Professor of Nanomaterials, University of Oxford, author of *The Penultimate Curiosity* and *It Keeps Me Seeking*.

Calum Samuelson has earned two master's degrees in historical theology and is married with one son. Though deeply grateful for and dependent upon his AI-empowered tools, his favourite pastimes involve eschewing them for the simplicity and wonder of the great outdoors.

Jubilee Centre
BIBLICAL THINKING FOR PUBLIC LIFE

www.jubilee-centre.org

The Jubilee Centre's 'Long Distance Christian' series of booklets sets out a Christian perspective on contemporary issues by combining careful research with in-depth biblical analysis.

About the Jubilee Centre

Jubilee Centre's mission is to articulate a biblical framework for public life – especially the big issues around the economy, politics and society – and equip Christian leaders to be salt and light in a secular, pluralistic world.

How should Christians exercise public leadership? Is there a biblical alternative to capitalism or socialism? How does the gospel address individualism and consumerism? Grappling with these questions has forged the Jubilee Centre's mission for the past 30 years.

The way we do this is through research and publications, including the widely read Cambridge Papers, events and conferences. Often collaborating with like-minded organisations and churches in Britain and abroad, we seek to present radical, biblical responses to some of society's most pressing challenges, in a winsome and practical way.

We invite readers to explore the wide range of resources on our website (and those of our sister organisations), to sign up to receive Cambridge Papers quarterly, and to follow us on social media:

www.jubilee-centre.org

www.relationshipsfoundation.org

@JubileeCentre

Jubilee Centre,
St Andrew's House
59 St Andrew's Street
Cambridge CB2 3BZ

Email: info@jubilee-centre.org
Website: www.jubilee-centre.org

First published 2019

Printed in Great Britain

Design: Heather & Mark Knight

Proofing: Abigail Sykes

Contents

Preface

The Jubilee Centre has always been concerned with political, social and economic questions that may have an impact across multiple spheres of public life. Consequently, we have been keeping a watchful eye on the development and deployment of artificial intelligence (AI), with two particular reasons for this scrutiny. First, because it may lead to a 'fourth industrial revolution' causing significant disruption to patterns of work and substantial 'technological unemployment'. Secondly, reporting in the news media is often on the alarmist side, conflating AI with robotics, especially humanoids, with the implication that AI might end up running out of control.

Therefore, we thought it was an appropriate time to research carefully into this area, and analyse the assumptions, trends and prospects around artificial intelligence from a biblical perspective. This booklet is the result of that work; it is intended to provide church leaders and other Christians interested in AI with an introduction to the issue and offer a framework based on a biblical worldview to guide their responses.

The leadership at CARE invited us to contribute to a national conference on the Church, Robotics and AI in June 2018, and we were grateful for the opportunity to present some of the preliminary findings of our research at a workshop there.

Our prayer is that this report will bring clarity and understanding, as well as help Christians engage positively in the debate and responses to the opportunities and challenges of AI.

Jonathan Tame

Executive Director, Jubilee Centre

Introduction

It seems that more and more attention is being directed towards **Artificial Intelligence (AI)**. Amidst the various media stories and latest productions from Hollywood, it can be difficult to think clearly about AI developments and separate fact from fiction, much less be proactive about engagement with AI on a personal, organisational, or public policy level. Therefore, this paper aims to formulate a biblically-based framework for evaluating developments in Artificial Intelligence that will enable Christian leaders in church, business and public service to make informed responses that are inspired by their faith. This will involve several related—though separate—steps. Although not everyone reading this paper will be concerned about the so-called 'existential risks' associated with AI, this will certainly be true for some. Ergo, it is necessary to provide a measure of reassurance by dispelling some of the sensationalism and apocalyptic narratives surrounding AI ('What AI is *not*'). Dispelling more extreme narratives, however, does not permit us to become apathetic or dismissive of AI in general. Therefore, the second section of this paper will grapple with how AI operates and how to understand it more clearly ('What AI *is*'). A clear understanding of AI inevitably redirects focus to deeper and more foundational questions about humanity and society. Consequently, the third section will help inform some of the most pertinent issues in the AI conversation by drawing on three biblical themes ('Understanding humanity'). After gleaning some key insights from the wisdom of Scripture, section four will consider how the application or neglect of those insights can generate divergent trajectories as AI is applied in different sectors of society ('The impacts of AI'). Finally, seven general guidelines will be suggested for engagement with AI at both the expert and non-expert level ('What AI *can be*').

Amidst the media stories and Hollywood productions, it can be difficult to think clearly about AI.

As part of the research for this report, ten leading AI practitioners and thinkers were interviewed. Some of their general viewpoints are reflected throughout this paper, but because of diverging opinions and wishes to remain anonymous, individual interviewees have not been cited. Their names can be found in *Appendix A*. Finally, this report seeks to complement work already carried out by some groups and individuals, including Nigel Cameron[1] and the more recent report from the House of Lords.[2] In particular,

we agree with the sentiment that Britain seems to be especially well positioned to lead the way in thinking through AI for the rest of the world.[3] Due to the amount of potentially unfamiliar terms used in this report, the *Glossary* provides concise definitions for all terms in **bold type**.

1 What AI is *not*: exposing 'myths'

Many prominent figures are worried about AI.[4] Considerable thought has been invested in mapping the possible trajectories of different apocalyptic scenarios, which include both the intentional and indifferent elimination of humanity.[5] On the other hand, there are those who see AI as the key which will allow us to transcend the limitations of our human lives as we now know them.[6] Several of our interviewees spoke about the need to expose these types of myths. This paper uses the word 'myth' not to reject all plausibility of these scenarios occurring, but mainly to highlight the fact that they are rooted more in aspirations and fears (whether personal or cultural) than in science. Ultimately, we contend that AI on its own will neither spell humanity's ultimate doom nor ultimate salvation. Below, four interrelated myths are examined: unlimited exponential growth, superintelligence, computer consciousness and the singularity. Importantly, these myths operate in a type of succession, so that each builds upon the assumptions of the preceding myth(s).[7] Thus, like someone building a structure on an unstable foundation, each level becomes increasingly shaky. This is unfortunate, not least because these myths attract a great deal of attention, obscuring and inhibiting productive dialogue (especially—though not exclusively—at the popular level).

Unlimited exponential growth:[8] 'technology is improving at an ever-increasing rate and nothing will stop it'

The term 'exponential growth' is used frequently in discussions about AI and can generate fears of an increasingly powerful and unstoppable computer. It is linked with **Moore's Law**, which in layman's terms predicted

that computing power would double roughly every two years.[9] Perhaps the most serious flaw with 'exponential growth' as it is commonly used is that it presumes inevitable and unbounded progress. This presumption seems to disregard all the examples of exponential growth that occur only for limited periods of time—as in various chemical reactions or the growth of bacteria. In fact, outside of pure mathematics, there are no actual examples of unabated exponential growth in the real world. Whatever source is fuelling the growth eventually runs out. One example sometimes referenced is that of global population growth, but there are various reasons to believe that the global population will slow down and peak around 11 billion near the end of this century.[10]

Readers familiar with the concept of exponential computer growth may object that the record of computing power itself is an example of unabated exponential growth. In reality, although Moore's Law has proven reliable for the last several decades, most predict that it will break down in the next few years—even Gordon Moore himself.[11] This is because it will soon be impossible to make functional transistors on computer chips any smaller. When transistors become too small, they begin to experience *quantum tunnelling*, which disrupts the normal behaviour of the electrons they are conducting.[12] The absolute laws of physics triumph over Moore's 'Law' of computing power. Some defend the myth of exponential computer growth by objecting (fairly) that computing power will still continue to grow via other means, including increased **software** efficiency, hardware specialisation (e.g. three-dimensional silicon circuits), cloud computing and **quantum computing**. Even so, the point still stands that such progress is not inevitable, nor is it likely to be exponential. With this said, we should not be surprised if computer development continues at a surprisingly quick pace. Rather, we should recognise the real, physical limitations of our world (including the growing energy consumption of powerful computers)[13] and the fact that advancements are the result of admirable, persevering work from real people rather than inevitable laws of nature.

Technological advances are the result of work from real people, not inevitable laws of nature.

It may be that this myth retains vitality because of cultural notions about progress, which have gradually been revived in the unprecedented peace of the post-war period after they began unravelling in 1914. Advances in transportation, food security, life expectancy, literacy, leisure time and medicine represent considerable progress indeed.[14] But those who tout these

advances regularly ignore their dark counterparts, including the transience of society, sky-rocketing obesity,[15] high rates of elderly loneliness, fake news, an epidemic of apathy and boredom, and unprecedented levels of mental illness[16]—not to mention crippling levels of indebtedness and extensive loss of biodiversity. One could also argue that many of the assumptions made in AI draw on problematic, popular interpretations of Neo-Darwinism, which has itself come under serious scrutiny lately.[17] Ultimately, unlimited exponential growth (and the fear of 'runaway AI' that it gives rise to) is exposed as a myth because it mixes scientific observation with unfounded assumptions and predictions.

Superintelligence: 'computers will soon be far better at doing everything humans do'

Currently, AI can be classified as 'narrow' because it is only proficient in very specific tasks.[18] The concept of **Artificial General Intelligence (AGI)** extends AI *proficiency* to the entire range of tasks performed by humans, whilst the concept of **Superintelligence**[19] further denotes AI *superiority* in all human tasks. Nick Bostrom defines Superintelligence as 'an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills,' and this intellect may itself design and build even smarter machines than humans could.[20] Those concerned about Superintelligence insist that consciousness is not necessary for it to be detrimental to humanity, since simply misaligned goals could cause it to harm or eliminate humans in order to complete its objective(s).[21]

One flaw with the theory of Superintelligence is the way it reduces immensely complex human activity to mere 'tasks'. This notion is derived in part from the increasingly outdated view of humans as 'machines' as well as from a marked privileging of human intellect over soul and body.[22] The reality is that those things least like 'tasks' are what make us most human (loving, hoping, inspiring, striving, etc.).[23] Furthermore, the idea that there is some type of 'general' human intelligence greatly underestimates the diversity of human personality and activity. Could we someday have a single AI system that could defeat humans in every existing board game? Certainly. Might we someday have a single AI-powered robot that could defeat humans

Superintelligence is flawed because it reduces complex human activity to mere 'tasks'.

in every known sport? Probably not. Will we ever have a system that can do everything a human does as well as a human does it? Certainly not.

Perhaps the most important problem with the concept of Superintelligence, however, is how it underestimates and misunderstands *human* intelligence. There is often an implicit assumption that human knowledge is a unique tipping point that, once passed, will allow machines to become vastly more intelligent than humans. This assumption is not based in any actual evidence—it may be, in fact, that human intelligence is the absolute upper limit. Additionally, *even if* machines become as generally intelligent as humans, it is far from clear that this will enable them *also* to create new machines that are more—or even just equally—advanced.[24] One way to illustrate this is with reference to human knowledge. It may be possible for a brilliant professor to teach everything they know about a particular subject to a student, but this in no way guarantees that the student will *also* be able to teach that knowledge to someone else (which requires good communication skills, patience, a sustained relationship and more). As Benjamin Bloom and others have argued, teaching takes place at a higher cognitive level than merely replicating knowledge, and the process of *creativity* at an even higher level than teaching.[25]

In spite of these deficiencies in the concept of Superintelligence, proponents continue to argue for it and they rely upon the idea of unlimited exponential growth to explain how computers will *inevitably* become vastly more intelligent than humans. Thus, the weaknesses in the concept of Superintelligence and the fact that it relies upon another flawed concept, leads to the conclusion that this theory is also more myth than science.

Computer consciousness:[26] 'computer systems will eventually be self-aware'

Ever since John Searle proposed the '**Chinese Room**' in 1980, the hypothesis of computer consciousness has been hotly debated.[27] To be clear, no AI is even close to being conscious and there is no evidence to suggest that this will ever be true. Most of our interviewees were extremely sceptical about the possibility of machine consciousness and many researchers view it as a distraction from real AI developments.[28] The idea of computer consciousness builds upon the myths of Superintelligence and Unlimited Exponential Growth because proponents believe that once machines become much smarter than humans it is only a matter of time until they will eventually

become conscious.

Perhaps the central problem with the concept of computer consciousness is uncertainty and even confusion about what consciousness actually is. For some, consciousness necessarily includes sentience, such as the emotions exhibited by Hal 9000 in *2001: A Space Odyssey*. Others believe computers could be conscious without any feelings. Already, casual language in fields such as Computer Vision sometimes speak of an AI system being 'conscious' or 'aware' of its surrounding environment, but this is decidedly different from what we intuitively recognise as authentic human consciousness. AI is, of course, getting better and better at *mimicking* consciousness, but mimicking consciousness does not indicate a proximity to *actual* consciousness.[29]

AI is getting better at mimicking consciousness, but this does not indicate a proximity to actual consciousness.

Much of the discussion about computer consciousness relies upon the presumption that the human brain is the seat of consciousness and it is a system that can be replicated—either via '**wetware**' or an 'upload' (**Whole Brain Emulation**). Advocates of this endeavour draw comparisons between transistors in computers and neurons in the brain, but this analogy is becoming increasingly inadequate. Crucially, neural transmissions involve far more than the binary 'on and off' function of transistors, including several different types and degrees of signals as well as those that spread well beyond the synapse itself.[30] In fact, metaphorical and analogical language is probably the primary flaw in conceptions and narratives about computer consciousness. People have grown so accustomed to talking about 'smart' gadgets and 'intelligent' machines that they've forgotten that such language is fundamentally metaphorical. New AIs are getting very good at *measuring* blood flow in human faces and then matching that data with certain moods, but they do not actually *perceive* feelings or emotions.

Consciousness implies understanding, and understanding implies knowing how information relates to reality outside its own framework. As an example, we can consider one of the most impressive applications of AI: **Natural Language Processing**. Google recently amazed the world with a demonstration of its Duplex virtual chat assistant conversing in real time with unsuspecting humans via telephone.[31] Although this impressive AI can *detect* the word 'father', it does not associate this word with a particular, experiential understanding of what a father is (as humans naturally do). Language is an important case study because of the unique and even

profound ways it influences human experience and thinking.[32] The complexities of language as explored by those such as Ludwig Wittgenstein and Michael Polanyi reveal the significant gulf between experiential and non-experiential knowledge. One more example is instructive. When **AlphaGo** triumphed over Lee Sedol in the game of GO, the standard 19x19 board was used. If another round of games had immediately been played on a 20x20 board, AlphaGo would have failed because that's not the task it was programmed for. Its 'understanding' of the game is not just narrow—it's artificial. Some pundits have praised the fact that AlphaGo chose a 'new' move that had never been done in the history of the game. It would be more accurate, however, to recognise this as a *different* move that had not yet been discovered by human players.[33] If computers ever perfectly simulate certain capacities we equate with consciousness, they will be distinct from the actual phenomenon of human consciousness.

Singularity:[34] 'a point will come when humans are no longer dominant; we will be to computers what ants are to us now'

The Singularity is the most speculative concept of all because it compounds all the assumptions of the myths already discussed. It involves the idea of machines self-improving or 'bootstrapping' their own abilities to the point beyond which it becomes impossible for humans to comprehend. It is at this level that many of the runaway, apocalyptic, or snowball scenarios take root.[35]

The Singularity is the most speculative concept of all because it compounds all the assumptions of the myths already discussed.

There are numerous problems with belief in the Singularity, but two of the most important deal with notions of *speed* and *self-improvement*.[36] Increases in computing speed or power can be irrelevant for at least two reasons: 1) *optimal solutions may be impossible to improve upon* (in Noughts & Crosses, it becomes impossible to beat an opponent who defends well because the available options are limited); 2) *chance may render perfect solutions impossible* (vast computing power cannot guarantee the winning lottery number). In the apt words of Steven Pinker, computing power is not some 'pixie dust that magically solves all your problems.'[37]

Regarding self-improvement, advocates tend to assume that AI's ability to upgrade itself will increase either exponentially or at least linearly. However, there are two main reasons it is more accurate to understand the path of AI self-improvement as one of *diminishing returns* (or logarithmic growth)[38]: 1) actual historical progress in AI development has mostly been logarithmic; 2) the difficulty of overcoming obstacles in the way of advancement tends to *increase* with the sophistication of the technology being developed.[39]

Notwithstanding these major problems, many still believe strongly in the eventual onset of the Singularity. This is greatly affected by the way we *perceive* a potential cataclysmic event. Kevin Kelly has argued persuasively that this will never go away; the Singularity will always be near.[40] Others observe what has been called the 'AI Effect', encapsulated by the following quote: 'AI is whatever hasn't been done yet.'[41] Some have employed the familiar remark from Roy Amara in their speculations about the Singularity: 'We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run.'[42] But whilst this is certainly true to an extent, it is unhelpful to apply such thinking to the *distant* future (e.g. the year 3,000) because there are so many more pressing issues to address in the present and near future.

Now that we have shown how these concepts are more mythical than scientific, it is important to point out how such ideas associated with AI are motivated by agendas tangential or even contrary to pure technological development. Notable figures in AI often have ulterior motives, such as Bostrom, who is a passionate transhumanist and co-founded the World Transhumanist Association before authoring one of the most influential books on AI and acting as an advisor for government policy.[43]

Ideas associated with AI may be motivated by agendas tangential or even contrary to pure technological development.

2 What AI *is*: framing the conversation

The term 'AI' can imply a range of different things depending on who is using it. This is partly because its meaning has shifted and evolved since its inception.[44] It broadly refers to any system that can perform tasks in a way

that mimics humans, but has also been appropriated in recent years as a marketing tactic to sell everything from business management software to apps that help people follow a budget or sleep better. At its best, the term 'AI' is used as a digestible umbrella term by specialists when communicating with non-specialists; at its worst, 'AI' is simply a misnomer. This report views AI as a highly complex tool that helps humans perform repetitive tasks.[45] *This perspective was nearly unanimous among our interviewees.* Accordingly, we will use the term 'AI tools' to that effect.[46]

The building blocks of AI

Traditionally, computer systems have been given 'rules' in order to accomplish various repetitive tasks. These rules—like recipes—contain discrete instructions that must be followed according to a logical order. This is an **algorithm**. Programmers commonly refer to an algorithm that has been implemented within a specific computing language as 'code'.[47] Such code is found in everything from phones to cars, often with millions upon millions of individual instructions. Whereas these instructions have mostly been manually prescribed by humans in the past, **Machine Learning (ML)** enables a system to determine many of the 'rules' on its own without being explicitly programmed. Like AI, ML is not new (both have been around since at least the 1960s).[48] The main reason ML has been employed to such great advantage in recent years is because of the rapidly growing accessibility of large data sets or **Big Data**. In order to determine the right instructions, a system needs to scan for patterns, and the bigger the data set the more likely it will be that any patterns detected will produce accurate and effective rules. By way of illustration, although chess computer games have been around almost as long as computers themselves, IBM's **Deep Blue** was superior to older computers because it used ML to process hundreds of thousands of examples from grandmasters in order to determine good moves rather than being given a rigid list of rules by which to function.[49] This means that even brilliant new algorithms from a start-up stand little chance against giants like Google if the data sets they have access to are much smaller. Whilst some types of ML such as **Deep Learning** and **Generative Adversarial Networks (GAN)** involve higher degrees of complexity and autonomy, all ML must still be given directives to guide the

All AI will inevitably contain biases from its human programmers—there is no such thing as a purely unbiased AI.

patterns and correlations they scan for.[50] AI expert Rodney Brooks has even described ML as 'very brittle'.[51] Due to the human influence and oversight required, all instantiations of AI inevitably contain biases from their human programmers—*there is no such thing as a purely unbiased AI.*

Although some will inevitably quibble with this portrayal, one can think of AI as something which *performs* human-like tasks, ML as the *training* involved in preparing for those tasks, and data as the *resource* that determines the success of the training. Given the popularity of the Economist's suggestion that data is the 'new oil',[52] we could extend the analogy so that ML corresponds to refineries and AI to the final petrol or gasoline that is produced. One possible point of misunderstanding is that some will want to classify ML *itself* as something that performs human-like tasks. Using Deep Blue as an example, there are two main problems with such an understanding: 1) the intensive process of 'learning' how to play chess is not synonymous with the discrete human-like feat of defeating a grandmaster; 2) the methods used by ML in the intensive process of 'learning' how to play chess are themselves quite different from how a human learns how to play chess. Furthermore, even though many AI tools continue to make use of ML after the main 'training process', this can be seen more as a mode of customisation or fine-tuning rather than part of the intensive process of enabling the machine to perform a human-like task in the first place.

The effects of AI

Google CEO Sundar Pichai believes that AI will have a more profound impact than electricity or fire.[53] Although this may be an overestimation, it is clear that AI tools have the potential to accomplish both great good and great harm. On the one hand, AI is already helping restore abilities to the disabled,[54] combat crime, save crops, manage invasive species,[55] protect biodiversity, detect cancer,[56] perform surgery,[57] dispose of bombs and biohazards, create more efficient energy schemes,[58] and more. On the other hand, there are serious risks involved with AI. Some include glitches in programs, outliers and anomalies in data sets and simply unexpected consequences.[59] One of the most notorious cases involved Google's image-recognising AI identifying black

AI is already helping to combat crime, protect biodiversity, perform surgery, and more.

people as gorillas.[60] On their own, none of these risks would make the list of top ten problems in the world.[61] But when paired with environmental concerns, threats to democracy, plummeting birth rates, or the growing epidemic of loneliness, the risks of AI could easily factor in to any number of the world's most pressing problems. Some of the most urgent concerns identified by our interviewees and others are: large scale loss of jobs, autonomous weapons and data monopolies.

'Uncanny Valley' may soon become a thing of the past as AI becomes totally convincing.

In addition to these high-profile AI risks are others of a subtler nature. One such risk raised by several of our interviewees is the growing power of AI simulation. Google's somewhat deceptive Duplex technology will be followed by many more striking examples of a computer system simulating humans, and it may well be that the unsettling feeling known as the '**Uncanny Valley**' soon becomes a thing of the past, as AI becomes totally convincing. Additionally, the abilities of the **Deepfake** algorithms in fabricating realistic videos of actual human beings will likely mark a watershed regarding trust of online materials and interactions.[62] Paradoxically, many examples of AI which look the smartest (e.g. **Sophia**) are in reality quite simple, whilst others that look unimpressive (e.g. the **Echo** and **Jibo**) are actually capable of performing surprising tasks. Powerful simulative AI will continue to advance for the purpose of entertainment and experimentation and will raise several difficult moral questions. But whenever simulative AI transgresses the boundaries of entertainment and experimentation, it will tend to be ultimately deceptive and dishonest, producing *pseudo-relationships* that lack real authenticity. The probability of AI increasingly functioning in direct interaction with humans suggests a significant need for what has been called 'functional morality',[63] which among other things would help maintain a certain level of transparency about what the AI is and isn't. Because humans tend to anthropomorphise things naturally, the AI tools which are the least transparent present the greatest risks.

Communicating about AI

Ultimately, in order to make objective contributions and help advance the conversation surrounding AI, accurate language is vital. This has not been well-practised up until now and although journalists are frequently (and rightly) criticised for their faults,[64] this is also a serious problem among

specialists and academics as they tend to conflate speculation with clear explanation.[65] To some degree, this is understandable because it is helpful to use familiar language[66] and metaphors to communicate new discoveries and developments. But it is not right to confuse non-specialists who must rely on authoritative testimonies. Additionally, small companies and start-ups have sometimes used exaggerated or inflated language either to impress potential buyers or intimidate competitors. The fact that software is often embarrassingly inefficient[67] and that programmers often aren't sure how exactly they get the results they do has rightly prompted calls for increased honesty and the elimination of **black boxes** in many vital areas of AI use.[68] Great progress can also be made simply by avoiding anthropomorphic language, which can gradually slide from metaphor into an actual description. This task is not easy, but it is important because of the way our language influences our perception and understanding.[69] We can accurately say that AI tools 'detect', 'scan', 'process' and 'function' without resorting to equivalent anthropomorphic terms such as 'recognise', 'consider', 'feel', or 'think'. AI has the potential to help tackle some of the biggest problems in the world, but this will require clear communication so that governments, regulatory institutions and other organisations can be confident about real potential without digging through beguiling descriptions.

This section has argued that AI is fundamentally a tool—even if it is used for distraction, entertainment, deception, or violence. Consequently, we must conclude that AI is inherently neither positive nor negative, but neither is it ever neutral because of the inevitable human biases contained within it.[70] AI can potentially be utilised for nearly any *task* imaginable (remembering that not all human activity can be reduced to tasks) and leveraged towards almost any end. But this does not mean that AI can take on any *role* in society. Here we must make a crucial distinction between *objects* and *subjects*. Martin Buber famously wrote about this distinction by claiming that we can only engage dynamically with a subject because in it we encounter a free, authentic being which we address as 'thou' rather than 'it'.[71] We interact with AI as an object or an 'it', whilst AI often influences the way we interact with subjects or other objects. Although AI will increasingly *simulate* subjects, it is unlikely ever to be encountered as a truly free 'thou'. One way to understand this complicated set of interactions is by viewing AI as a type of intermediary or filter between human relationships with both other subjects and also various objects.[72] Interactions with chatbots like **Sophia** and game-playing AIs like **Deep Blue** are exceptions precisely because they do not fulfil any larger purpose (unless, of course, we view them through

the lens of *entertainment*). The vast majority of AI acts as an *intermediary between subjects* by helping them work, communicate, or understand more efficiently.[73]

If AI systems are essentially tools that extend or amplify the reach of humans, careful thought should be given to what it actually means to be human—and even to what society should look like as whole. Indeed, discussions about AI frequently end up as discussions about human beings, considering who or what we are and how technology can enhance or diminish human dignity.[74] Behind every example of AI are fundamental human dynamics that need to be addressed. A typical sentiment comes from MIT Professor Max Tegmark, who claims that 'we need to capture the meaning of life' in order to ensure safe AI for the future.[75] Although AI researchers acknowledge the importance of such human dynamics behind AI, there is currently little consensus about what exactly constitutes human flourishing—much less how it should be facilitated vis-à-vis AI. Some have talked about the dubious concept of **Coherent Extrapolated Volition (CEV)**, which takes moral progress for granted and considers what universally held morals might look like in the future when humanity is better than it is now.[76] Not much progress has been made so far, but it is becoming increasingly apparent that insight from fields outside of AI, computer science and neuroscience are required. It is for this reason that we now examine some of what the Bible says about humanity.

Discussions about AI frequently end up as discussions about human beings, considering who or what we are.

3 Biblical analysis: understanding humanity

The Bible provides invaluable insight regarding who we are as humans. Throughout history, human culture and knowledge have taken many forms and often progressed, but human nature itself has not changed. Technologies such as AI can help humans do many things, but unless we understand our weaknesses, purpose and trajectory, even the most advanced tools will simply make us more efficient in repeating the same mistakes we've always

made. Based upon the advice of our interviewees and other research, this section considers human nature by focusing on three biblical themes: the Imago Dei, the Fall and Eschatology. In particular, the Imago Dei helps identify which human qualities and characteristics AI should seek to facilitate or enhance. The doctrine of the Fall helps diagnose how human imperfections and malevolence influence the development and application of AI. Finally, biblical Eschatology helps us anticipate, imagine and yearn for our ultimate destination and think critically about different AI-powered futures.

Imago Dei

Exploring the best dimensions of humanity is essential to the task of using AI tools to amplify good and promote human flourishing. AI experts (including most of our interviewees) are eager to parse the distinction between humans and computers. In this effort, it is common to invoke illustrious human feats such as Michelangelo's paintings, Bach's symphonies, or Einstein's theory of relativity. This method of distinguishing between human and AI is unsatisfactory not least because it neglects most people who have ever lived. Most importantly, however, such cursory assessments of humanity's greatness fail because they measure accomplishments divorced from the role of purpose. The Imago Dei helps us better understand human purpose.

Given the perilous state of human identity in the postmodern world, it is hardly coincidental that humans are increasingly compared to computers.

The belief that humans are made in the image of God has rightly occupied a central position in Christian consideration of AI to date. Creativity, reason and morality have largely dominated as the primary dimensions of the Imago Dei in the last millennium.[77] Notwithstanding their importance, these dimensions may have had more popular currency in the Modern period (when there was more social consistency and structure) than they do today in the fragmented and pluralistic world where AI is making its mark. Indeed, considering the perilous state of human identity in the postmodern world, it is hardly coincidental that humans are increasingly being compared to computers. Consequently, this section considers the Imago Dei through the lens of *relationships, responsibility and self-giving love.*

According to the Bible, humans are explicitly created in the image of a

relational God, the implication being that we are only fully human when in meaningful *relationship* with others.[78] The metaphor of the Christian community as a body teaches that every member plays an integral role.[79] Also, the fruit of the Spirit is always manifested in relational contexts.[80] This raises important questions for the development of AI tools and leads one to conclude that there can be no single version of the 'ideal human' because each possesses different qualities and gifts in varying degrees and arrangements. Consequently, some are suggesting that it is better to design a range of AI tools to do different tasks rather than attempting to develop a single tool that mimics humans completely.[81] This also has important implications for current discussions about 'digital personhood' and 'digital subjects', since these terms suggest it is possible to know someone apart from a relational context.[82] Finally, the fact that humans are created for relationships can help explain the tendency to anthropomorphise, and can also shed light on human vulnerability to computers that simulate humans.

Another implication of humans being created in the image of God is that they have *responsibility*. God is the supreme, faithful sustainer of all Creation but has also entrusted humans with the unique responsibility of caring for and ruling over his creation. Much more than a mere task or goal, this responsibility requires the entire human being to act like an 'angled mirror' which simultaneously reflects the lordship of God to creation and the praise of all Creation back to God.[83] The importance of responsibility in the realm of AI may have been the most common exhortation among the interviews we conducted. Some of our interviewees understand their work with AI as a clear example of subduing the earth;[84] others of them think about their work with AI more as an aspect of serving people and society in love. Regardless, there is a clear difference between designing AI tools to aid in the responsibility of wisely ruling Creation and designing them to rule so that humans can shirk the weight of responsibility. Already one can perceive small ways in which humans are abdicating their responsibility of ruling, whether by using autonomous weapons, foetus screening, employee profiling, or criminal image scanning. Increasingly, if the AI says a decision is right, the human users will execute it. This is not only the definition of *irresponsibility,* it also dampens life's dynamism by assuming that difficult ethical decisions can be avoided or even eliminated.[85]

There is a clear difference between AI that helps us to wisely rule Creation and AI designed so that humans can shirk the weight of responsibility.

The *self-giving love* of God—which has always existed in the dynamic relations of the Trinity—flowed outward in the act of Creation and was eternally enacted in the sacrifice of Jesus on the cross. Because humans are created in this God's image, one characteristic of humans is the capacity to love in a manner that considers the needs of others above self. Culture at large often only praises this type of love if it is valiantly portrayed in Hollywood—even the deeply shameful Crucifixion has been turned into a grand act of heroism and fortitude. Others dismiss the value of sacrifice in favour of more empirical, scientific accomplishments. One influential statement about AI claims, 'everything that civilisation has to offer is a product of human intelligence'.[86]

Whatever the world says about love, Jesus claimed that there is no greater love than to lay down one's life for one's friends.[87] Within this logic, the widow who gave two mites was praised above the lavish tithers,[88] the quiet tax collector was the one made right with God,[89] and small children are singled out as possessors of God's Kingdom.[90] Human greatness is not found in mighty, memorialised achievements, but in seemingly 'small', self-giving acts of devotion, humility and sacrifice.

AI may help people feel happier or be more efficient, but if it does not improve human relationships it is ultimately misdirected.

As AI becomes increasingly common, it is important to remember that love always prioritises the other. AI may help people feel happier, be more efficient, obtain more knowledge and even feel more ethical, but if it does not improve human relationships it is ultimately misdirected. Accordingly, people should be very cautious about seeking to outsource or automate the most common and apparently mundane manners in which they give themselves in love to others. The simple gift of listening is rapidly being replaced by AI. The command to weep with those weep[91] is being threatened by AI tools that detect our mood and tell us how to fix it. The practice of hospitality in which one opens up one's home is being superseded by virtual interactions. If *agape* love were simply another task that required energy to perform, then it would make sense to continue designing AI tools that preserve energy. But *agape* love is not a separate task to be performed at the end of the day like other elements of leisure time. It is something that must be *practiced and developed*, and often the best way to do this is by washing the feet no one else wants to wash.[92]

Doctrine of the Fall

Whereas the section about the Imago Dei explored the goodness of humanity in Creation, this section highlights its shortcomings. A major concern among our interviewees was that secular thinking is not equipped to account adequately for or anticipate the realities of imperfection and malevolence in human nature and the world. Of course, programmers and developers acknowledge that glitches can plague a computer system and that sometimes criminals hijack a piece of good technology for a bad purpose. But for the most part, AI development buzzes with an optimism that believes sustained effort and education can eventually help humanity overcome all problems and perversions.

> Secular thinking cannot adequately account for the realities of imperfection and malevolence in human nature and the world.

The Bible sees things differently. God created a world that was 'very good', but it has fallen from that status because of sin—which is anything that obstructs relationship with God. Humankind can naturally recognise entropy, atrophy, disease, corruption and brokenness of all kinds as deviations from an ideal situation, but can also become tragically resigned to the idea that these things are simply woven into the fundamental fabric of the universe. Crucially, the doctrine of the Fall helps make sense of the tension between desired behaviour and actual behaviour, and helps Christians consider how this tension might influence the development and deployment of AI. In order to do this, it is necessary to examine both the *depth* and *breadth* of sin.

The *depth* of sin reaches to the very core of our being and cannot be encompassed within a binary system of 'rights' and 'wrongs'. Jesus taught that even perfect 'right' actions can be sinful if done with the wrong posture of heart. It may be possible to distance oneself from particular external sins, but no one is ever far away from the allure of pride and self-assurance. Applied to AI, this truth has two major implications. First, it means that attempts to transcend human faults and discover 'perfect morality' through AI are misguided. Indeed, perfection should never be attributed to machines because they have been created by imperfect humans. Second, it means that even the very best AI developments can have negative consequences. Some of these consequences are caused by glitches or programmer bias. More insidious, however, are AI tools that seem supremely good or helpful but end up turning hearts away from God (e.g. a financial tool that ends

up increasing greed or a voice replication tool that ends up enabling deception).[93] Some fear enslavement to AI through *oppression*,[94] but we are already becoming enslaved through the subtler route of *obsession*.[95] It may well be that efficiency and knowledge will be the predominant idols of the AI Age.

In addition to the *depth* of sin, the *pervasiveness* of sin touches every corner of the world. In a hyper-individualistic age, it's easy to interpret passages like Romans 3:23 as an indictment against particular, personal failures. But Paul's message carries a sweeping universality from which nothing can hide. *All* of creation groans for redemption as the effects of sin are felt. This means that sin is encountered both internally and externally, individually and structurally. Therefore, just as an excellent policy or strategy can be thwarted by external factors, so also can AI fail due to user error, corrupt data, or false information. It is conceivable that one party, nation, or culture could develop a genuinely productive framework for engagement with AI, only to have it disrupted or destroyed by a broken, sinful mindset or system. The pervasiveness of sin must also be considered in a diachronic sense. One of the great falsehoods connected with modern myths of inexorable progress (whether capitalistic, Neo-Darwinian, or even 'exponential') is the idea that human morality itself can continually improve. Whilst it is obvious that most humans in the West no longer pillage, rape, burn, imprison, or torture other people, one need not look far to uncover modern versions in the form of embezzlement, habitat destruction, child abuse, debt slavery and animal cruelty. Humanity doesn't get 'better' intrinsically, we simply get 'better' at devising ways to justify our crooked actions.[96]

If the world was not even called 'perfect' before the Fall, we should hardly expect that we can make it perfect through AI now.

At its most basic level, the pervasiveness of sin confronts the field of AI development in which progress, success, benevolence and good behaviour are simply taken for granted. One must not only consider the impact of individual sin, but also of sin within every other person and institution with which they interact. One direct implication for AI development could be insisting upon designing systems in such a way that *expects them as a rule* to break down, be misused and impact unexpected stakeholders.[97] If the world was not even called 'perfect' before the Fall, we should hardly expect that we can make it perfect through AI now.[98]

Eschatology

In addition to appreciating humanity's purpose and sinfulness, a holistic view requires comprehension about humanity's trajectory and ultimate destination. According to the Bible, this trajectory is anchored in the Redemption already inaugurated in the person of Jesus and headed towards a supremely good New Creation after the end of this age. One could even argue that the only type of true inexorable growth that is possible in the universe is growth in Christlikeness, which by the Spirit's power will continue for all eternity. Regardless of whether a linear or cyclical view of time is espoused, it is not uncommon for humans to yearn for an ultimate destination beyond time, and many generations have thought the world will end with them. It should be no surprise that much of AI dialogue also yearns for a different future and ultimate end for humanity. Several interviewees urged us to highlight the gravity of long-term effects and the need for goal-oriented trajectories of AI, and one of the best ways to do this is to consider what the Bible says about humanity's ultimate end.

The Bible points towards a a dynamic pan-ethnic community that challenges aspirations for seamless technological uniformity.

First, Paul clearly teaches that resurrected humans will not be spirits without bodies.[99] This has important implications for various agendas which view the human body as a disposable inconvenience and hope that AI will help humans to eventually discard it. Second, eschatological pictures in the Bible envision the flourishing of the rest of non-human creation. This point has vital implications for the care of animals and the environment, for it seems that ultimate symbiosis with the New Creation is meant to be an outflowing of human interaction with Creation in this life.[100] Third, the Bible portrays a dynamic pan-ethnic relational community existing in the New Creation. This challenges aspirations which posit seamless technological uniformity, compatibility, or even complete 'monism'.[101] Last, the Bible emphasises the importance of simplicity and purity[102] in the Kingdom of God, which belongs to the little children[103] whose play energises and characterises the perfect peace we will know there.[104] This is quite distinct from some secular narratives which aspire to vast knowledge, efficiency and complexity.

Whilst Christians can be confident in Christ's return, humility and vigilance are the best postures for discussing *how* this will come about.[105] Will AI help us save the environment and usher in a superior age free from fossil

fuels, or will it be the only recourse available after we have destroyed the biosphere?[106] Will Jesus return before or after the planet is hit by a super asteroid? The parable about the wheat and the tares can help Christians navigate seemingly conflicting reports about the world's trajectory, as it insists that both evil and good will continue to increase in the world until Jesus' return.[107] This means that neither fear, nor naïve optimism, nor apathy are appropriate mindsets, because Christians are called to be alert, joining in the work of the Spirit wherever it may be found. One practical way to live within this tension is by nurturing a theology of surprise.[108] Rooted in God's often unexpected works of redemption, this way of viewing the world *actively anticipates* God doing surprising things as Christians act as salt and light in the world. A theology of surprise protects against excessive commitment to narrow programs or agendas, as both God's warnings and blessings come in ways that cannot be predicted. Concerning AI, this may mean that Christians encounter real hope in the places they are least comfortable with and fear in the places they least expected to.

Eschatological topics can present a good way to engage transhumanists and technologists in meaningful conversation.

Dialoguing about what one ultimately yearns and hopes for can be powerfully inviting, and some may find that these eschatological topics present a good way to engage transhumanists and technologists in meaningful conversation. Public dialogue is increasingly turning to questions about what an ideal society *should* look like, and Christians should capitalise on this opportunity by looking forward towards what perfect eternity *will* look like. This practice is deeply demanding because it requires the active deployment of our imagination in tandem with the mysterious movements of the Spirit,[109] but for that very reason is also infinitely more valuable than anything Christians do without the help of God.[110]

4 Trajectories: the impacts of AI

Having considered some key features of human beings and also the advanced AI tools which help amplify those features, this section will argue that the human relationships which operate 'behind' the use of AI are the

most important factor to consider in this discussion. This section examines four broad areas where the effects of AI will influence relationships in significant ways. The four areas are *investment, employment, regulation* and *products/services*. As AI tools are increasingly implemented in societies, they will have *both* positive and negative impacts upon relationships. By default, many of the impacts are likely to be somewhat negative, further entrenching the current ideologies of capitalism, individualism and consumerism.[111] But it is also possible for AI tools to help bring reform, although this will require intentional and concerted efforts. This section aims to demonstrate how these opposing trajectories might play out in each area.

Investment: shareholders, start-ups and universities

The main element here—and a major concern amongst our interviewees—is increasing imbalance and disparity of wealth and risk in society. AI tools allow those with capital to leverage their resources to new degrees by increasing efficiency of production and eliminating many of the costs involved with labour. It may even be that AI tools could play a part in the collapse of the increasingly unfit-for-purpose capitalistic system as we know it.[112]

Negative trajectories of AI vis-à-vis investment (whether financial, intellectual, or other) are heavily connected with corporations. The stark reality is that most AI development in the West is being led by profit-driven companies.[113] Whilst it is certainly true that many new developments in AI originate in the academy or in start-ups, very few of these remain separate from the corporate world for very long. A case in point is the way that **Big Tech** companies have bought out nearly every competitor in order to secure their own growth.[114] Thus, Big Tech allows smaller companies to take most of the risks of innovation and then use their capital to acquire whatever innovations prove to be successful. In this scenario, shareholders and directors play a decisive role, both of whom are often primarily seeking short-term profits.[115] Even though shareholders are commonly viewed as the 'owners' of a company due to their financial investment, they do not share a proportionate amount of risk and most have very little involvement in the decisions that are made.[116]

Alternative trajectories, however, could actually help to *decrease* inequality as AI becomes cheaper and more accessible. Small companies[117] and organisations may be able more easily to customise AI tools for their

specific needs or even share resources with each other.[118] The difference between these trajectories will involve many factors, but can be significantly influenced by shareholders, directors and other managers seeking to promote human flourishing. Viewing investment more as *involvement* and reward more as *quality relationships* could help ensure that investment in AI research and training moves in the best direction. Considering the Facebook–Cambridge Analytica data scandal, people and organisations alike would do well to recognise the role that trust, honour and reputation play in the success of a company rather than focusing narrowly upon profit. Christians can help lead the way in making investment for social benefit more feasible and effective. They should also work to expand a cultural vision of such investment practices by articulating a holistic and integrated paradigm of human flourishing, rather than fixating on individual issues which may or may not be connected to a core conviction or belief.

Employment/work: companies, churches and communities

There is considerable disagreement regarding whether AI will ultimately create or eliminate jobs after society passes through a rocky 'transition' period.[119] Either way, there can be no question that the landscape of employment will be transformed by AI tools. It may even be that full-time work or employment becomes a thing of the past in more advanced economies, which has led to proposals for some type of universal basic income. Despite misleading headlines claiming that workers are being 'fired by a machine',[120] it is vital to remember that these changes are still the result of practical human actions and decisions.[121] Churches and communities should recognise the future opportunities and begin brainstorming about what fulfilling and meaningful work could be created without requiring any formal salary. Even if a universal basic income never becomes a reality, there will be an increasing need for re-training and re-skilling in the growing 'gig economy' that simply cannot be fully met by the state.

There will be an increasing need for re-training and re-skilling in the growing 'gig economy' that cannot be fully met by the state.

Negative trajectories see companies replacing human employees with AI tools in a race for the bottom line. Unsurprisingly, companies with the least relational capital and coherence will have the fewest qualms

about this. Although some employers may claim to implement AI without actually displacing any human employees, the effect will be virtually identical as they simply phase out existing posts as people move on or retire. This process is being led by big companies, and many of the small and middle-sized companies will feel forced to replace human employees with AI simply in order to compete. One irony that is not often recognised is that many middle-income jobs are actually more at risk because the pay-off for automating low-paying jobs (such as janitors and cleaners) is relatively low whereas the pay-off for automating middle-paying jobs is much higher. Another factor relates to the declining birth rates in most Western countries. The UK population doubled during the Industrial Revolution, which meant that many people struggled to find work due to automation in factories. Today, however, the UK birth rate is only about 1.76 children per woman, which is below what sociologists call the 'replacement rate' at 2.1.[122] Consequently, some employers are struggling to find qualified employees, especially in health & social care and education. Ultimately, it may be that the biggest threat is not so much the elimination of work across the board, but an even greater disparity between demeaning work and fulfilling work, leading to a growing underclass in society. If the thesis of Pickett and Wilkinson is correct, this disparity would actually be worse than if society at large found themselves universally 'unemployed' and in receipt of a basic income.[123]

Ultimately the biggest threat may be the disparity between demeaning work and fulfilling work, which leads to a growing underclass in society.

Alternative trajectories could involve AI replacing precisely the most onerous and demeaning work so that humans can do more fulfilling jobs. Some refer to this as **intelligence augmentation (IA)**, arguing that *enhancement* rather than replacement should be the ultimate goal of AI with regard to work.[124] However, the expenses involved will require employers to value more than the mere replacement of labour with capital. Just as the Cadburys and Rowntrees of the 19th century provided employment as a means of social integration and social good, society would do well to recover a holistic and enlightened view of business in the age of AI. To be most sustainable, this will require intentional involvement from *both* companies and consumers. Enlightened consumers are already placing higher priorities upon brand transparency and authenticity, and can continue to exert their influence by insisting that companies provide high-quality employment in addition

to high-quality goods. Enlightened companies can view employment as a blessing to society and imagine new, mutually beneficial forms of work. Similar to the factory jobs provided by Cadbury which did not require any special skills, directors and managers can help develop new, meaningful jobs by identifying and articulating nascent needs. It is possible that many of these will be in the 'emotional sector' since both manual and intellectual labour are being overtaken by machines and computers.[125] Due to the growth of remote work, longer commutes, zero-hour contracts and self-employment, work today suffers from fragmentation and loss of relational coherence. It is not difficult to understand why mental illness is becoming such a serious problem among workers. This reality does not provide a good context for the implementation of AI, but a keen focus on relationships can help encourage AI to restore meaningful and satisfying work. Rather than permitting implementation to increase the fragmentation of work, employers can focus on using it to help facilitate more human relationships and decrease levels of mental illness.[126] Perhaps just as physical illnesses and health risks in the workplace led to the introduction of HR roles, the epidemic of mental illness could lead to new professions which focus on emotional wellness in a proactive and dynamic manner (rather than primarily reacting to mental illness). Owing to the profound legacy surrounding work and vocation throughout Christian history,[127] there is tremendous opportunity for philanthropic Christian leadership within business of the unprecedented AI Age.

There is tremendous opportunity for philanthropic Christian leadership within business of the AI Age.

Regulation: Big Data, Big Tech, states and the environment

When the AT&T telecoms monopoly was finally broken up in 1982, the decision was made not because the US government feared their power or political influence, but because the monopoly was not good for competition in the economy in general. Today, some states do in fact have reason to fear the tremendous power of Big Tech, but also lack the means to exercise legal discipline because globalisation has put many aspects of the corporate world outside the effective control of current governance structures.

If AI development continues on its current trajectory, societies might increasingly be controlled by tech companies rather than political governments.[128] The most important aspect here is the use, control, protection and privacy of data. For all the good intentions behind desires to make data 'open' and available to all, there is a basic misunderstanding about the nature of data. Data is not useful to everyone, and much of the data collected by Big Tech would be meaningless to most people. The reason most individuals are happy to give away their data without qualms is because it is not inherently valuable to them. In the 19th century, native peoples had no use for the crude oil under their feet because they didn't possess any combustion engines, much less the refineries to process the oil. Put simply, not all data is equal. Thus, there is a crucial imbalance of motivations in the effort to make data public. The recent implementation of the **General Data Protection Regulation (GDPR)** may be a good step in managing the collection of personal data, but it disproportionately affected small companies, who didn't have the resources or staff to make all the necessary changes and acted more like a mere slap on the wrist to the companies that really matter.[129] As long as corporate profit and national GDP are the primary aims, AI regulation will be characterised by strained relationships between companies and authorities.

Following the current trajectory, societies might increasingly be controlled by tech companies rather than political governments.

Despite the extreme difficulty of regulating Big Tech currently, it is possible to envisage an outlook in which sophisticated data analysis significantly streamlines the relationship between companies and states by providing superior and timely information for policy planning. Two potential beneficiaries in this scenario are the natural environment and future generations. Because most electricity today is distributed via outdated grids, AI can vastly improve the efficiency of energy consumption.[130] Nonetheless, there must be a clear strategy in place in order for these gains to be translated to the environment rather than simply funnelled into the coffers of the state, energy companies, or even consumers themselves. Several encouraging efforts are being made specifically to harness AI for the good of the planet,[131] but success will be most likely if the public decides to support these endeavours. In order for future generations to reap the benefits AI is capable of producing, the tech industry will need to confront various inefficiencies and self-serving tendencies—especially with regard to quality and reliability.[132]

One way regulative bodies might help is to develop a system of 'employability permits' similar to those being implemented to control carbon emissions.[133] Such permits provide an economically efficient method to regulate the market by allowing jobs to be replaced by companies possessing a permit to do so. This would mean that companies making significant profits from AI tools that fulfil or replace human jobs would be required to pay money to facilitate human work elsewhere—thereby offsetting the overall loss of human jobs. In this scheme, large companies who can afford to purchase 'employability permits' (which exist in finite supply) will do so based upon how many people they could be employing relative to their profits and gross computing power/capacity.[134] Those who can't afford to purchase the permits will opt to provide more jobs for employees directly (who could work to make their AI more efficient in order to use less computing power) or indirectly through charitable causes and trusts. This scheme is a variation on proposals for a 'robot tax', but offers the advantages of providing some way to measure less-tangible AI productivity and allowing bargaining between companies to adjust the price of permits (rather than a governing agency simply imposing an arbitrary and fixed tax amount).[135] Although any form of 'robot tax' will be difficult to implement, it is worth pursuing whilst in the early stages of human labour displacement so that glitches can be corrected with minimal collateral damage.

Products/services: efficiency, entertainment and escape

Although AI tools are not limited to products or services provided to customers, these are the most frequent interfaces with AI for many people in society. Even if the primary form of interaction with AI for most people appears to be relatively low-level, such as through smart devices in the home or predictive algorithms on social media, these can still have a significant effect on how we view and understand ourselves. With specific regard to AI, there is an urgent need for the public to become aware of that fact that they are *simultaneously customers and products*. Most common applications of AI are provided for 'free', but in fact a company is making money from people's use of that tool. Just as aboriginal people groups failed to perceive that they were standing on 'black

The public must become aware of that fact that they are simultaneously customers and products.

gold', many users of AI fail to perceive the 'gold' in their daily technological consumption.

The current trajectory of AI involves many tools that operate with a type of inbuilt dependency upon the company—whether it is trusting Apple to store photos in iCloud or reliance upon current financial institutions for making payments. This, of course, can change the relationship between customers and companies to one of resigned dependency rather than one of freely-chosen loyalty. More importantly, however, this changes the way that many individuals perceive their use of AI tools. Rather than feeling fundamentally empowered by a tool that can help them achieve the tasks they want to achieve, people are increasingly expressing that they feel trapped in a cycle of pursuing tasks they never intended to do in the first place. As AI continues to boost efficiency, entertainment and escape will both become popular responses to the increased non-work time people experience, and it will not always be easy to distinguish between the two. One serious concern is the growing popularity of sex robots, which may be the most pertinent example of the power of simulation discussed above.[136] Without moral guidance, many people will use AI tools which eventually diminish their own humanity rather than increase it.

Without moral guidance, many people will use AI tools which eventually diminish their own humanity rather than increase it.

Accordingly, the best AI-powered products and services will be those that foster human flourishing by strengthening individuals' self-control and relationships with other people. In order for this to be done, people will need to re-evaluate the language of 'rights' and 'wants' in an individualistic, consumeristic society, which has not been able to deliver the type of society it has promised. Several groups have made important steps in this area,[137] and a core realisation is that AI tools work best in helping people achieve previously established goals as opposed to helping determine what those goals are. This is the difference between entering a website with the goal of purchasing a specific product and letting the shopping algorithms help find the best one, versus entering a website with an itch to buy anything that will satisfy and letting the algorithms determine what exactly that is. Customisation of advertisements and entertainment can be extremely useful, but unless they align with positive human traits like responsibility and self-giving love, people may ultimately find them distracting or even detrimental.

5 What AI *can be*: application

Much sound advice has already been offered by various groups regarding the development and deployment of AI.[138] This section both summarises and expands current thinking by offering several practical, general guidelines for engagement with AI in light of the biblical reflection earlier, which are applicable at both the expert and non-expert level.

Guidelines

a) Mastery

This implies using a tool in a way that is most effective, safe and beneficial. Hitting one's fingers is a painful and non-beneficial result of using a hammer without mastery. The more complex a tool, the more practice is required to master it. Although people tend to assume that they can quickly master their 'intelligent' tools such as smartphones, more often it is they who are 'mastered' by the tools.[139] A basic indicator is to consider how use of AI tools may or may not impinge upon the most important relationships in our lives. True mastery of AI tools will help channel and leverage people's brightest ideas rather than simply giving them more money, leisure time, or information.

b) Accountability

For developers, accountability can help mitigate errors in programs and eliminate wasted time from preventable mistakes—which is especially important when working with the massive amounts of data characteristic of AI. For the end-users of AI, accountability means that their engagement with AI tools should be in the context of relationships with others who have access to and familiarity with the same AI, and where honest conversations are held about the impact of the technology. In this context, accountability can act as a counterweight to isolating effects on individuals seeking things such as companionship, entertainment or escape via AI. Since AI is data-driven, it can easily generate reports on usage (e.g. reporting on the amount of 'screen time' spent with a chatbot).[140] However, it is especially important that accountability is more robust than mere surveillance; accountability requires two-way communication embedded in relationships. This is best

achieved when an individual wilfully and actively participates in a group which expressly strives for the common good of all its members. Practically, this type of accountability could shape positively the development of 'smart' public facilities, security systems and nursing homes; shared self-driving cars, open source software, ML platforms,[141] entertainment & recreation parks,[142] and allotments equipped with AI tools/sensors. Additionally, local churches might offer workshops, training, or common resources about AI tools which could be used by everyone in the congregation.

Local churches might offer workshops, training, and resources about AI which could be used by everyone in the congregation.

c) Diversity

Biases are a major concern in AI, and it is no secret that many of the top AI developers are young, white males, often connected with a coterie of elite institutions.[143] Trained algorithms may never have mistaken black people for gorillas if there had been more reasonable ethnic diversity at Google.[144] Furthermore, the current atmosphere of AI development tends to foster a type of 'homogeneous thinking', which can be uncreative and stagnated even if it is not biased.[145] Intentional diversity can challenge this atmosphere by drawing from various perspectives and ideas to identify, articulate and solve problems in wise and creative ways.[146] Indeed, because wisdom entails more than technical knowledge or quantity of information, people without expertise of a given AI system may still be able to contribute invaluable insight to an overall project. Despite current obstacles to diversity in society, Christians in particular should be energised by the pan-ethnic vision of eternity in the New Creation and work to shape creative and wise applications of AI.

d) Transparency

Both the internal operations and ultimate purposes of AI tools are often somewhat opaque.[147] This can make such tools less effective in general and also allow some questionable functions to be smuggled in. A hammer has the clear purpose of hammering nails. But algorithms used by Amazon which purportedly help customers find products also have the goal of getting them to spend money. Largely due to the **Uncanny Valley**,

manufacturers have found success in producing AI-equipped robots in the likeness of animals instead of humans,[148] but have not always been clear about their intended purpose. Do they primarily monitor vital signs or entertain? Do they primarily *provide* information or *collect* information? Although perhaps less problematic than realistic humanoid robots, increasingly realistic animal robots still represent a serious breach of authenticity, not least because they are often intended for elderly people or those with learning difficulties who may be less able to discern what they are interacting with.[149] Beyond the purpose of experimentation or entertainment, there is no reason that AI-equipped robots need to sound or look like humans *or* animals.[150]

Except for experimentation or entertainment, there is no need for AI-equipped robots to sound or look like humans or animals.

e) Precision

Programming and developing AI often involves more trial and error than precise or direct routes towards an end goal—especially when working on large projects with many team members.[151] Consequently, there is an urgent need for clarity and simplicity[152] in the design process of AI tools, which will often necessitate agile, built-for-purpose programs constructed from the ground up (rather than recycling inefficient code from other projects or applications). Furthermore, there is need for precision in relation to the ways that AI tools are packaged for and used by consumers. Taking a wider example from smartphone development, today's models serve as alarms, cameras, barometers, music players, calendars and sometimes even telephones! Such conflation, if mirrored in AI development and use, could make it difficult to assess how effectively a tool is actually accomplishing tasks or serving its users (does **Alexa** help someone be more productive or simply more busy?). Additionally, the impressiveness of much AI can attract superfluous features which capitalise on novelty (e.g. **Siri** being programmed to tell jokes). Therefore, increasing precision with regard to purposes of AI tools will be crucial, and will also help developers better anticipate unforeseen consequences since they will be focusing on how one single function might err rather than how multiple functions might err.

f) Empowerment

AI tools should most often enable the *enhancement* of particular human tasks rather than their replacement.[153] As already mentioned above, there is exciting potential for how daily work can be made more meaningful with AI. Another major way AI tools can be empowering is by assisting people with disabilities.[154] This will likely require close communication between developers and disabled users to ensure AI tools are genuinely empowering rather than marginally helpful and/or frustrating. Some have highlighted the benefit of developing a range of very narrow AI-tools that each empower humans in distinct ways rather than expending resources on the goal of **AGI** that aims to mimic everything about humans.[155] Due to the time-intensive customisation involved, truly empowering AI will probably be less profitable for developers and will therefore require considerable lobbying and encouragement. Enlightened consumers can increasingly demand that the AI-powered tools they benefit from are also fully accessible for less abled members of society.

g) Efficiency

Society will continue to benefit from increases in efficiency brought about by AI tools, but most of the processes by which these tools function can actually be much more efficient than they usually are.[156] This is largely because the mindset in tech has been focused more on achieving functionality than on making sure it operates as efficiently as possible. Fortunately, this mindset is beginning to change (in part due to the imminent decay of **Moore's Law**) and although it will be labour intensive, the process of making algorithms and programs more efficient is relatively straightforward since it aims at the clear goal of increasing performance (maximum output with minimal input) and speed. Nevertheless, as society increasingly engages with AI tools that assist with human-like tasks, it is vital to recognise that a rigid technological understanding of efficiency is rarely the best way to think about improving the worth or value of more personal human activities, which often require great amounts of time or have no real measurable outcome at all.[157]

A rigid understanding of efficiency is rarely the best way to think about the worth or value of more personal human activities.

Conclusion

This report has sought to acquaint readers with the basics of AI and help them engage wisely as these new tools continue to impact our world. It has been argued that much of the popular dialogue about AI is based more upon assumptions and aspirations than upon actual facts. A sober view of AI recognises that it has the power for great good and great harm; this report has highlighted the importance of communicating clearly and realistically about both possibilities. Leaders of all types have the obligation to ensure that AI does not simply amplify the current trajectory of present realities such as individualistic capitalism, and it has been argued that a keen understanding of humanity is crucial for this endeavour. In particular, leaders must take seriously humankind's propensity towards malevolence whilst being rooted in its ultimate calling (Imago Dei) and directed towards its final end (New Creation). We have sketched out divergent trajectories of increasing AI in four different social and economic areas whilst suggesting the actions necessary to ensure that AI leads to the greatest good for society as a whole. Finally, seven general guidelines were offered for application in daily settings by both experts and non-experts alike.

Ultimately, AI tools should help people regain healthier notions about the purpose of life in general.

There can be no doubt that AI will transform the world as we know it. As ambassadors and servants of Christ, Christians especially should strive to direct the impacts of AI in ways that help people live life to the fullest and bless the communities, cities and countries where they live. Just as various benefits of globalisation have also accelerated the loss of indigenous languages and cultures, it is conceivable that mass, indiscriminate implementation of AI systems could make humans very good at doing things which are not in their best interest. Ultimately, AI tools should help people regain healthier notions about the purpose of life in general. Recapturing both the art of discipline and a sense of human purpose, people can learn to eschew effects of AI which produce burnout or laziness in favour of those that help them mature and thrive as stewards of Creation and citizens of heaven.

Appendix

The following are among the experts interviewed in the course of the research, and each of them responded in a personal capacity. None of them is mentioned by name or quoted directly, and the opinions expressed in this report do not necessarily reflect the position of any of the institutions represented.

Dr Andrew Basden, Professor of Human Factors and Philosophy of Information Systems, Salford Business School, University of Salford

Dr Andrew Briggs, Professor of Nanomaterials, University of Oxford

Rt Revd Dr Steven Croft, Bishop of Oxford

Professor Nigel Crook, Associate Dean, Faculty of Technology, Design and Environment, Oxford Brookes University

Dr Derek Roberts, Co-Founder, Solarflare

Dr Paul Roberts, Operations Director, Cambridge Medical Robotics

Dr Peter Robinson, Professor of Computer Technology, University of Cambridge

Dr Robert Song, Professor of Theological Ethics, Durham University

Dr John Wyatt, Professor Emeritus of Ethics & Perinatology, University College London

Glossary

Alexa: The name of Amazon's AI assistant, which makes use of NLP. It is the flagship feature of the Echo.

Algorithm: A set of steps or instructions to solve a problem.

AlphaGo: A computer system created by DeepMind that in 2016 defeated the reigning world champion in Go (a traditional Japanese board game).

Artificial General Intelligence (AGI): A computer system that can do virtually everything a human brain does, to the same standard that a human brain does it.

Artificial Intelligence (AI): A general term used to describe a range of computer systems which can accomplish certain repetitive tasks in ways that mimic humans.

Artificial Narrow Intelligence (ANI): A computer system which can perform a specific task with proficiency.

Artificial Superintelligence (ASI): A hypothetical computer system which far surpasses human intelligence in every area.

Big Data: Quantities or types of data that are unable to be stored and/or processed with traditional, analogue methods by humans.

Big Tech: Some of the largest companies in the world which make their money in tech and often AI. These include especially Amazon, Apple, Facebook, Google and Microsoft.

Black Box: In technology, this refers to any system which obscures the activity between input and output.

Chinese Room: A hypothetical scenario proposed by John Searle in which a person is locked in a room, unseen, with a comprehensive supply of Chinese language tools. Someone on the outside passes a note to them in Chinese and waits for a response. With the resources at their disposal, the person inside the room is able to interpret the message and give a response.

The person on the outside is unable to determine if the person on the inside actually knows Chinese or can only merely interpret it. Thus, computers merely simulate and do not actually understand.

Coherent Extrapolated Volition (CEV): The hypothetical ambitions or aims of a more mature and evolved humanity which eventually overcomes most divisions and disagreements.

Deep Learning: A type of ML that makes use of many layers of neural networks. 'Deep' refers to the number of layers, not a qualitatively different type of computation.

Deep Blue: A chess-playing computer first created by IBM in 1995. It defeated world champion Garry Kasparov 3 ½ matches to 2 ½ in 1997.

Deepfake: A type of super-realistic, AI-generated, fake video. It uses deep learning to superimpose existing video onto source video, notably for a realistic face-swapping effect.

Echo: A type of 'smart speaker' device first sold by Amazon in 2014 which employs NLP to act as a voice-controlled personal assistant.

General Data Protection Regulation (GDPR): Legislation enacted by the EU in 2018 which is designed to upgrade protection of personal data. It replaced the Data Protection Directive of 1995.

Generative Adversarial Network (GAN): A type of unsupervised deep learning pitting two neural networks against each other. One network acts as a generator whilst the other acts as a discriminator.

Hardware: Tangible, physical components of a computer system.

Intelligence Augmentation (IA): A concept that focuses more on computers enhancing rather than replacing human intelligence.

Internet of Things: The interconnection of various 'smart' devices via the internet. This could include home appliances, vehicles and agricultural sensors.

Jibo: An AI-equipped personal robot assistant. Developed by researchers at the MIT Media Lab, it stands 12 inches tall and features a large camera on what vaguely resembles a head.

Machine Learning (ML): A subset of AI that employs various techniques to help computers 'learn' without being explicitly programmed.

Moore's Law: A prediction based upon observations that the number of transistors in computer circuits doubles approximately every two years.

Moravec's Paradox: The recognition that many high-level human activities require little computational power whilst many basic human activities require vast computational power.

Natural Language Processing (NLP): An area of AI involving phonetics, grammar, syntax and semantics.

Neural Networks: Inspired by the function of neurons in the brain, these allow computers to sort and filter information in sophisticated, multi-step manners. These are a basic building block in ML.

Paro: A small, fuzzy, baby seal robot first developed by Takanori Shibata in 2001. Described as a 'therapeutic robot', academic research has shown it to have positive effects on elderly people and especially those with late-life cognition disorders.

Polanyi's Paradox: Similar to Moravec's Paradox, it claims that much of human knowledge and behaviours are developed and held at subconscious levels of cognition. Thus, humans cannot explain many of the simplest things they do.

Quantum Computing: A new type of computing featuring quantum bits (qubits). Rather than traditional binary bits, these can store information in superpositions allowing far more complex computations to be carried out.

Singularity: A hypothetical point in the future at which humans are unable both to control and comprehend computer activity.

Siri: The name of Apple's AI assistant, which makes us of NLP. It is now

standard with most Apple devices.

Software: Intangible programs and applications that run or function on a physical computer.

Sophia: A life-like robot developed by Hanson Robotics in 2015 which can display more than 50 facial expressions. It is described as a 'social robot' since its primary purpose is to converse with humans.

Strong AI: A way of describing computer systems which either exceed human intelligence or possess some form of consciousness.

Turing Test: First purposed by Alan Turing in 1950, it traditionally employed conversational language to test whether humans could distinguish a computer from another human.

Transhumanism: A general term for movements which seek to help humanity move beyond its current biological and intellectual limitations.

Uncanny Valley: The perturbing response resulting from interaction with something that seems human but is not.

Weak AI: Any computer system that performs a specific human-like task. Essentially synonymous with Artificial Narrow Intelligence.

Wetware: Neither hardware not software, but an artificial material that is compatible with biological tissue.

Whole Brain Emulation (WBE): A hypothetical feat involving the complete copying of a biological brain into a digital form.

Endnotes

1 Nigel Cameron, *The Robots Are Coming: Us, Them and God* (CARE, 2017); Nigel Cameron, *Will Robots Take Your Job?: A Plea for Consensus* (Malden, MA: Polity Press, 2017).

2 Select Committee on Artificial Intelligence, *AI in the UK: Ready, willing and able?*, HL Paper 100 (House of Lords: April 16, 2018).

3 Ibid., p.5.

4 These include Elon Musk, Sam Harris and Nick Bostrom, to name a few.

5 Some posit, for instance, that a Superintelligence will simply have misaligned goals (differing from human priorities) which cause it to eliminate human beings as indifferently as stepping on a bug. For a helpful article and graphic of this and similar misunderstandings, see https://futureoflife.org/background/aimyths/.

6 As argued by the leading transhumanist association, Humanity+ (formerly the World Transhumanist Association).

7 Whilst there may be *multiple* paths that can be taken to arrive at superintelligence, the same assumptions are still involved in whichever route is taken. See Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: OUP, 2014).

8 It is worth noting that most often the term 'exponential growth' is used to refer to *hyperbolic* growth (where the slope of a function becomes infinite) rather than actual exponential growth (where the slope of a function only *appears* to become infinite, but actually always remains finite). See Toby Walsh, *Android Dreams: The Past, Present and Future of Artificial Intelligence* (London: C Hurst & Co Publishers Ltd, 2017), p.120.

9 Gordon Moore's original article suggested that the density of transistors on a computer chip would double every year. Some prefer to adjust the time increment to 18 months.

10 See https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html.

11 Gordon Moore estimated that it would break down around 2025 and IBM more recently put the mark at 2021.

12 The smallest transistors in commercial production are 7 nanometres and problems begin to arise with traditional silicon transistors that are smaller than this. A silicon atom itself is approximately 0.2 nanometres in diameter.

13 Energy is another serious factor, which has been raised in relation to BitCoin mining. Even with impressive advances in solar and wind energy, it is likely that some computer developments may be impeded by lack of power they could require. See https://www.theguardian.com/technology/2017/nov/27/bitcoin-mining-consumes-electricity-ireland. Ultimately, it seems that some kind of breakthrough in nuclear fusion will be necessary to sustain the immense power required by the computers of the future.

14 See https://www.ted.com/talks/steven_pinker_is_the_world_getting_better_or_worse_a_look_at_the_numbers#t-475968.

15 See https://www.theguardian.com/global-development/2017/nov/04/danger-to-future-generations-640m-pledged-third-of-world-malnourished-obesity-hunger-kofi-annan.

16 See https://www.telegraph.co.uk/news/2017/08/31/one-three-sick-notes-mental-health-problems-alarming-report/.

17 In November, 2016 the Royal Society met in London to discuss the validity of Neo-Darwinism.

18 Like the acronym ASI, some use ANI to designate **Artificial Narrow Intelligence**. Some also refer to this as '**weak AI**'.

19 Some use the acronym ASI to designate **Artificial Superintelligence**.

20 See https://nickbostrom.com/superintelligence.html.

21 Nick Bostrom outlines the oft-repeated scenario of how a Superintelligence whose goal is to make paperclips could potentially end all biological life in the process. See Nick Bostrom, *Superintelligence*, 123ff.

22 See the striking line in the influential open letter: 'everything that civilization has to offer is a product of human intelligence', available at https://futureoflife.org/ai-open-letter/.

23 This is similar to the concepts expressed by both **Moravec's Paradox** and **Polanyi's Paradox**. See Marvin Minsky, *The Society of the Mind* (New York: Simon & Schuster, 1986); Also Hans Moravec, *Mind Children* (Cambridge, MA: Harvard University Press, 1990).

24 Toby Walsh calls this the 'Meta-intelligence' Argument and explains it brilliantly. See Walsh, *Android Dreams*, pp.124–26.

25 This is best conveyed in Bloom's taxonomy.

26 Some might also use the term **Strong AI** when talking about computer consciousness, but this can sometimes be confused with the idea of Superintelligence *without* consciousness so it is avoided here for the sake of clarity.

27 John Searle, 'Minds, Brains and Programs', *Behavioral and Brain Sciences*, 3 (1980), 417–24.

28 Walsh, *Android Dreams*, p.91.

29 The 'Turing Test' is often referenced in this area. The test has various manifestations and continues to evolve, so is actually not a very helpful term any more. However, some discuss the idea of a Meta-Turing Test. See Walsh, *Android Dreams*, p.47.

30 Andrew Briggs and Dawid Potgieter, 'Machine Learning and the Questions It Raises', in *From Matter to Life: Information and Causality*, ed. Sara Imari Walker, Paul C. W. Davies, and George F. R. Ellis (Cambridge: Cambridge University Press, 2017), pp.470–76.

31 See https://www.theverge.com/2018/5/8/17332070/google-assistant-makes-phone-call-demo-duplex-io-2018.

32 This is known as linguistic relativity, or the Sapir–Whorf hypothesis.

33 This same argument applies to the 'new' poetry, music, etc. that AI systems have 'created'. Such productions are simply rearrangements of previous material and ideas, never a paradigm shifting or mould-breaking development.

34 Originally, the Singularity was a cosmological concept describing reality beyond a black hole. What we are discussing here could more properly be called the 'technological singularity'.

35 Some would associate the Singularity more with computers passing the Turing test, but being fooled by computers' activity is not the same as being completely unable to comprehend it.

36 Toby Walsh brilliantly considers these along with others. See Walsh, *Android Dreams*, pp.121–31.

37 See Pinker, S., 'Tech luminaries address singularity', *IEEE Spectrum*, June 2008.

38 The author owes this perspective almost entirely to the insight of Toby Walsh. See Walsh, *Android Dreams*, pp.126–27.

39 Paul Allen (co-founder of Microsoft) has called this the 'complexity brake'. See Allen, P. and Greaves, M., 'The Singularity isn't near', *MIT Technology Review*, October 2011, pp.7–65.

40 See http://kk.org/thetechnium/the-singularity/.

41 Douglas Hofstadter (mis)quoting Tesler's Theorem. See Hofstadter, *Gödel, Escher, Bach: an Eternal Golden Braid* (New York: Basic Books, 1979).

42 Susan Ratcliffe, 'Roy Amara 1925–2007: American futurologist', *Oxford Essential Quotations*, 4th edition (Oxford: OUP, 2016).

43 See https://www.fhi.ox.ac.uk/fhi-researchers-advise-uk-government-artificial-intelligence/.

44 See the helpful article by Rodney Brooks, https://rodneybrooks.com/forai-the-origins-of-artificial-intelligence/.

45 Although it can seem like AI operates on an integrative level, this is best explained by the tremendous speed at which it is able to sort through data and detect similarities, patterns and anomalies. This insistence on AI as a tool is shared by many working directly in the areas of AI and robotics; see, for example, the *RAS 2020* report (Special Interest Group - Robotics and Autonomous Systems: July 2014) which characterises Robotics and Autonomous Systems as tools that 'perform useful tasks for us in the real world, extending our capabilities, increasing our productivity and reducing our risks,' p.6.

46 Because robots are often mentioned in the same breath as AI, it is important to emphasise the basic difference between **hardware** and **software**. Roughly speaking, AI is in the same family as software, whereas robots are in the family of hardware.

47 In general, one can say that algorithms relate more to the design process of AI while code relates more to the development process.

48 Both of these were used in Apollo spaceships to optimise thrust, for instance.

49 It should be noted that the type of machine learning used in Deep Blue would now be considered quite primitive, but at the time was perceived as a legitimate type of AI; see https://www.aaai.org/Papers/Workshops/1997/WS-97-04/WS97-04-001.pdf.

50 Broadly speaking, types of ML include reinforcement, supervised and unsupervised learning, but even 'unsupervised' learning requires careful parameters to be given by humans.

51 See https://www.technologyreview.com/s/609048/the-seven-deadly-sins-of-ai-predictions/.

52 See https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

53 Alexandra Suich Bass, 'GrAIt expectations', *The Economist*, Special Report: AI in Business, 31 March 2018, p.4.

54 See https://www.projectrevoice.org.

55 See https://research.qut.edu.au/ras/research/rangerbot/ and https://medium.com/thelabs/robots-saving-the-reef-d0573cac1c21.

56 See https://software.intel.com/en-us/articles/using-artificial-intelligence-ai-to-detect-lung-cancer-nodules.

57 See https://www.roboticsbusinessreview.com/health-medical/ai-assisted-surgery-improves-patient-outcomes/.

58 See http://sitn.hms.harvard.edu/flash/2017/artificial-intelligence-will-revolutionize-energy-industry/.

59 For example, the increased use of AI in automated financial trading has the potential to exacerbate market volatility.

60 See https://eu.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/. At the time of writing, this problem had still not been resolved.

61 Walsh, *Android Dreams*, p.141.

62 See https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed.

63 Wendell Wallach and Colin Allen, *Moral Machines: Teaching Robots Right from Wrong* (Oxford: OUP, 2009).

64 See https://futureoflife.org/2016/04/26/ai-journalism-goes-bad/.

65 This superb recent paper strongly critiques these trends and others in ML research: Zachary C. Lipton and Jacob Steinhardt, 'Troubling Trends in Machine Learning Scholarship', July 11, 2018, available at https://arxiv.org/abs/1807.03341.

66 Marvin Minsky called these 'suitcase words'.

67 This article offers an incisive view of the software world from the 'inside': http://tonsky.me/blog/disenchantment/.

68 See https://www.wired.com/story/ai-experts-want-to-end-black-box-algorithms-in-government/.

69 See endnote #32 above.

70 This is a variation of Melvin Kranzburg's well-known first law of technology.

71 Martin Buber, *I and Thou* (Edinburgh: T & T Clark, 1937). Importantly, this can resolve some of the difficulties and confusions of Cartesian dualism that others such as Herman Dooyeweerd attempted to resolve.

72 One of the most common examples is probably the internet itself.

73 This is complicated by the fact that AI becomes one of several layers of intermediaries. The internet, for instance, is already a complex intermediary between subjects (who are often anonymised).

74 Much of this debate is happening through science fiction; see recent popular television series such as *Black Mirror* and *Westworld*.

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

76 Eliezer Yudkowsky, *Coherent Extrapolated Volition* (Berkeley, CA: Machine Intelligence Research Institute, 2004). This concept features prominently in Bostrom, *Superintelligence*.

77 Roughly speaking, one can recognise that creativity was explored in the Renaissance, reason in the Scientific Revolution and morality in the Enlightenment.

78 Genesis 1:26.

79 1 Corinthians 12:12–31.

80 Galatians 5:22–23.

81 Briggs and Potgieter, 'Machine Learning and the Questions It Raises,' p.478.

82 Olga Goriunova, 'The Digital Subject: People as Data as Persons' in *Theory, Culture and Society*, Special Issue: Transversal Posthumanities (forthcoming 2018).

83 Tom Wright, *The Day the Revolution Began* (London: SPCK, 2016), p.100.

84 See Genesis 1:28.

85 Whilst advocating for a robust understanding of (and emphasis on) *human* responsibility, this paper does not address the particulars of *how* ethical decision-making algorithms should be constructed because this will vary across cultures, contexts and companies. There is already a large body of interdisciplinary work being done in this area, particularly around guidelines for autonomous weapons and self-driving cars.

86 See https://futureoflife.org/ai-open-letter/.

87 John 15:13.

88 Luke 21:1–4.

89 Luke 18:9–14.

90 Matthew 19:13–14.

91 Romans 12:15.

92 John 13:1–17.

93 Crucially, idolatry or obsessions facilitated by AI tools can grow much faster than other types since they increase efficiency by definition.

94 There are valid concerns around AI's potential use for oppression, especially in authoritarian states. The AI-assisted Social Credit Register introduced by the Chinese government is perhaps the best

example of an effective surveillance and social control mechanism, see https://www.wired.co.uk/article/china-social-credit.

95 The plots of George Orwell's *Nineteen Eighty-Four* and Aldous Huxley's *Brave New World*, respectively, chart these trajectories vividly.

96 C.S. Lewis expressed a similar sentiment in *The Abolition of Man*: 'Education without values, as useful as it is, seems rather to make man a more clever devil.'

97 See https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/36302.htm.

98 Orthodox Christians have long understood Adam and Eve more as innocent children than perfect humans. Additionally, some would point out that the Garden could not have been perfect if it contained a deceptive serpent and a tree containing the knowledge of evil.

99 1 Corinthians 15.

100 The traditional reading of 2 Peter 3:10 has tended to emphasise the destructive nature of the fire, but several scholars are trying to recover the true reading as a 'refining fire'. See Richard Middleton, *New Heaven and New Earth: Reclaiming Biblical Eschatology* (Grand Rapids: Baker Academic, 2014), pp.160–63.

101 See Jan H. Naude, 'Technological Singularity and Transcendental Monism: Co-producers of Sustainable Alternative Futures', *Journal of Futures Studies*, 13.3 (2009), 49–58.

102 Ephesians 5:26–27.

103 Matthew 19:13–14.

104 Zechariah 8:3–5.

105 Similar to the way that transhumanists and technologists have overestimated the pace at which 'exponential' AI development will usher in the New Age, so also have Christians over anticipated the full arrival of the Kingdom God. It is interesting that a (minority) stream of fatalism regarding environmental destruction also runs through both groups. Some groups of fundamental Christians in the USA see destruction of the environment as a step in ushering in Christ's return and some futurists believe that biological life will be superseded since machines can run on solar-generated electricity. The ultimate demise of carbon-based life forms is what drives the urgent search for life beyond the need of biological resources.

106 See endnote #100 above.

107 Matthew 13:24-30.

108 This is associated especially with Lesslie Newbigin.

109 Tom Wright's modern classic, *Surprised by Hope*, offers several inspiring ideas about how this may unfold practically in the lives of believers.

110 See Psalm 127:1.

111 See Calum Samuelson, *The Steering Wheel: Confronting the ideologies driving western culture and society* (Cambridge: The Jubilee Centre, 2018).

112 One interesting alternative suggested by some is a major switch to platform cooperatives. See https://www.yesmagazine.org/peace-justice/when-robots-take-our-jobs-platform-cooperatives-are-a-solution-20180420. One major problem, however, is who provides the initial capital to get things running.

113 The reality is obviously different in places like China, Russia, North Korea and other authoritarian regimes. See https://www.technologyreview.com/s/608324/china-plans-to-use-artificial-intelligence-to-gain-global-economic-dominance-by-2030/.

114 See https://techcrunch.com/2018/06/10/the-largest-buys-of-techs-big-five-a-look-at-ma-deals/?guccounter=1.

115 It is a common misunderstanding that directors are legally required to produce a *financial return* for shareholders. Rather, they are required to act in the *best interest* of the shareholders, which ultimately boils down to doing what is best for the company. See https://www.nytimes.com/

roomfordebate/2015/04/16/what-are-corporations-obligations-to-shareholders/corporations-dont-have-to-maximize-profits.

116 See https://www.hbs.edu/faculty/Pages/item.aspx?num=52623.

117 Most companies in the UK are either small or medium sized (<250).

118 One difficulty with this possibility is the quality and size of datasets available to these smaller companies.

119 The well-known report from the Oxford Martin School contends that 47% of US jobs are at risk. See Carl Benedikt Frey and Michael Osborne, 'The Future of Employment: How susceptible are jobs to computerisation?' (Oxford: Oxford Martin School, 2013). Far more optimistic, however, is the recent report from Capgemini, which claims that a majority of companies have already created new jobs because of AI; see https://www.capgemini.com/gb-en/service/artificial-intelligence-where-and-how-to-invest/.

120 See https://www.bbc.co.uk/news/technology-44561838.

121 In the case of the BBC story above, it is apparent that the computer system only fired the employee because his manager had recently been laid off and had failed to renew the employee's contract.

122 See https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsummarytablesenglandandwales/2017. South Korea is worst with a rate of 1.05, see https://www.economist.com/asia/2018/06/30/south-koreas-fertility-rate-is-the-lowest-in-the-world.

123 Kate Pickett and Richard Wilkinson, *The Spirit Level: Why Equality Is Better for Everyone* (London: Penguin, 2010).

124 See https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7.

125 See https://sloanreview.mit.edu/article/planning-for-the-future-of-work/.

126 See https://www.psychologytoday.com/gb/blog/feeling-it/201208/connect-thrive.

127 See Darrell Cosden, *A Theology of Work: Work and the New Creation* (Eugene, OR: Wipf & Stock, 2006); see also Calum Samuelson, *The Enduring Power of Vocation* (Cambridge: Jubilee Centre, 2017).

128 There are, however, some optimistic predictions that Big Tech monopolies will come under scrutiny from various regulatory bodies and eventually be broken up by government antitrust legislation; see https://www.economist.com/business/2018/04/26/americas-antitrust-apparatus-prepares-to-act-against-big-tech.

129 See http://www.business-money.com/announcements/forum-fears-new-data-protection-bill-could-damage-small-businesses.

130 See http://sitn.hms.harvard.edu/flash/2017/artificial-intelligence-will-revolutionize-energy-industry/.

131 One example is the recent establishment of Microsoft's AI for Earth; see https://www.microsoft.com/en-us/aiforearth.

132 See endnote #67 above.

133 See https://academic.oup.com/reep/article-abstract/1/1/66/1548600.

134 It is worth noting that a company's data centres are often located in different jurisdictions from their headquarters or places of human employment. This will need to be taken into account by regulatory bodies, especially if international borders are crossed.

135 Economists view these types of permit systems as desirable alternatives to taxes because they allow the market to function more efficiently. Permits determine supply and allow the market to adjust the price. Another possibility involves some type of marginal cost pricing.

136 See https://www.forbes.com/sites/susannahbreslin/2018/01/30/paris-sex-doll-brothel-france/#6a82b64d3946.

137 See http://humanetech.com/.

138 For just one example, see the report from AI Now, https://assets.ctfassets.net/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf.

139 See Guy Brandon, *Digitally Remastered: A Biblical Guide to Reclaiming Your Virtual Self* (Edinburgh: Muddy Pearl, 2016).

140 One new chatbot is extremely popular with teenagers: https://replika.ai.

141 See https://www.tensorflow.org/.

142 See https://www.technologyreview.com/s/611580/forget-about-vr-in-the-living-room-this-summer-its-on-waterslides-and-in-arcades/.

143 This was brought into the limelight recently when Google fired James Damore for 'perpetuating gender stereotypes'. See https://www.vox.com/identities/2017/8/8/16106728/google-fired-engineer-anti-diversity-memo. See also https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html.

144 At the time of the incident, 60% of Google's employees were white, 31% were Asian and only 2% were black. See endnote #60 above.

145 Some of the most influential institutions in AI are MIT, Stanford and Cambridge; see https://www.forbes.com/sites/mariyayao/2017/05/01/dangers-algorithmic-bias-homogenous-thinking-ai/#68ddc46f70b3.

146 Although they can greatly aid a group in this process, machines are currently very poor at articulating *why* something is actually a problem. AI is increasingly being used in some fields of academia with PhD students, but heavily relies on the context of what has already been studied and determined to be valid. In other words, an AI system wouldn't challenge well-established theories like Einstein and others have done.

147 See **black box** in the glossary.

148 Whether to classify animals more as subjects or objects is a difficult and contested dilemma, but the Bible seems to overcome this dichotomy by viewing animals as a type of subordinate and non-responsible subject; even donkeys are given a Sabbath rest and the ox must not be muzzled while threshing the grain (Deuteronomy 5:14 and 25:4, respectively). This makes good sense if we understand our relationship with animals as more similar than dissimilar. After all, both animals and humans have 'living souls'. See Genesis 2:7 and 2:19 where 'nephesh' is used to describe both animals and man.

149 The fact that robots like **Paro** yield positive improvements in 'comfort levels' in old people's homes is somewhat misleading because it is not being compared with human or animal companionship; it is literally better than nothing. See https://www.thelancet.com/journals/laneur/article/PIIS1474-4422%2813%2970206-0/fulltext.

150 Accordingly, AI-powered robots such as **Jibo** (which resembles neither a human nor an animal) seem like the least problematic option.

151 See https://www.theatlantic.com/technology/archive/2017/09/saving-the-world-from-code/540393/.

152 Programmers refer to this type of code as 'elegant' since it fully accomplishes its purpose in a minimal fashion.

153 Once again, this raises the idea of **Intelligence Augmentation**, which emphasises the need for various tasks to retain a measure of human oversight rather than being completely outsourced to AI.

154 See https://www.abilitynet.org.uk/news-blogs/how-ai-could-transform-lives-disabled-people and https://blogs.microsoft.com/on-the-issues/2018/05/07/using-ai-to-empower-people-with-disabilities/.

155 Briggs and Potgieter, 'Machine Learning and the Questions It Raises', p. 478.

156 See http://tonsky.me/blog/disenchantment/.

157 As Jacques Ellul argued several decades ago, there is a great risk that the drive for efficiency can begin to shape the way people do things in general. See Jacques Ellul, *The Technological Society* (Toronto: Vintage Books, 1964). Cf. Calum Samuelson, *Redeeming sport?*, Cambridge Papers vol. 27, 3 (Cambridge: Jubilee Centre, 2018).