



Cambridge
Papers
Towards a biblical mind

Artificial intelligence and simulated relationships

By John Wyatt

Summary

Interactions with apparently human-like and 'emotionally intelligent' AIs are likely to become commonplace within the next ten years, ranging from entirely disembodied agents like chatbots through to physical humanoid robots. This will lead to new and troubling ethical, personal and legal dilemmas. Will the promotion of 'relationships' with machines contribute to societal wellbeing and human flourishing, or provide new opportunities for manipulation and deception of the vulnerable? As biblical Christians we are called to safeguard and to celebrate the centrality of embodied human-to-human relationships, particularly in essential caring and therapeutic roles, and in our families and Christian communities.

Introduction

The 2013 movie *Her* constructs a near-future world in which Theodore, a young single man on the rebound, falls in love with Samantha. She's cool, sassy, knowing and intimate. But she is entirely virtual – an artificial intelligence in the cloud that remains in contact with Theodore wherever he goes, communicating by a small earbud. But perhaps the fictional world of *Her* is not so far away.

Eugenia Kuyda and Roman Mazurenko were tech entrepreneurs who developed a deep friendship carried out mainly online with text messages. But then, at the age of 22, Roman was killed in a road accident. Eugenia was devastated and went through what she called 'a dark phase'. Then she decided to build an artificially intelligent chatbot to replicate Roman's personality, uploading all the text messages that Roman had sent over years. Eugenia found herself 'sharing things with the bot that I wouldn't necessarily tell Roman when he was alive. It was incredibly powerful...' The AI chatbot Replika which resulted from this work is described as 'a safe space for you to talk about yourself every day'. As Eugenia put it, 'Those unconditional friendships and those relationships when we are being honest and being real... are so rare, and becoming rarer... so in some way the AI is helping you to open up and be more honest.'¹

Could synthetic relationships with AIs somehow interfere with the messy process of real human-to-human interactions, and with our understanding of what it means to be a person?

AI chatbots

Hundreds of commercial companies around the world are developing AI chatbots and devices such as Amazon's Alexa, Google Home and Apple's Siri. The companies are engaged in an intense competition to have their devices present within every home, every workplace and every vehicle. In 2018 Amazon reported that there were hundreds of thousands of developers and device makers building Alexa 'experiences', contributing to more than 70,000 'skills' (individual topics that Alexa is able to converse about), and there are more than 28,000 different Alexa-connected smart devices now available.² It seems likely that interactions with apparently human-like and 'emotionally intelligent' AIs will

¹ BBC Radio interview, 16 February 2018.

² <https://developer.amazon.com/blogs/alexa/post/38bb01ef-ac9b-49ec-9e2c-fcb0b51a8b31/2018-highlights-for-alexa-skill-builders>

become commonplace within the next ten years. But how should we think of these 'relationships'? Can they play a helpful role for those grieving the loss of a loved one or those merely wishing to have an honest and self-disclosing conversation? Or could synthetic relationships with AIs somehow interfere with the messy process of real human-to-human interactions, and with our understanding of what it means to be a person?

Care robots

Paro is a sophisticated AI-powered robot with sensors for touch, light, sound, temperature, and movement, designed to mimic the behaviour of a baby seal.³ In her book *Alone Together* Sherry Turkle reflects on an interaction between Miriam, an elderly woman living alone in a care facility, and Paro.⁴ On this occasion Miriam is particularly depressed because of a difficult interaction with her son, and she believes that the robot is depressed as well. She turns to Paro, strokes it again, and says, 'Yes, you're sad, aren't you? It's tough out there. Yes, it's hard.' In response Paro turns its head towards her and purrs approvingly.



It seems likely that in the foreseeable future we will be confronted by a spectrum of different machines that offer simulated relationships, ranging from the entirely disembodied through to a physical humanoid robot.

Sherry Turkle writes '...in the moment of apparent connection between Miriam and her Paro, a moment that comforted her, the robot understood nothing. Miriam experienced an intimacy with another, but she was in fact alone.... We don't seem to care what these artificial intelligences "know" or "understand" of the human moments we might "share" with them. ...We are poised to attach to the inanimate without prejudice.'

Mental health applications

Synthetic 'relationships' are playing an increasingly important role in mental health monitoring and therapy. The delightfully named Woebot is a smartphone-based AI chatbot designed to provide cognitive behavioural therapy for people with depression or anxiety.⁵ It provides a text-based 'digital

therapist' who responds immediately night or day. An obvious advantage of Woebot compared to a human therapist is that it's constantly available. As one user put it, 'The nice thing about something like Woebot is it's there on your phone while you're out there living your life.'

Alison Darcy, a clinical psychologist at Stanford University who created Woebot, argues that it cannot be a substitute for a relationship with a human therapist. But there is evidence that it can provide positive benefits for those with mental health concerns. 'The Woebot experience doesn't map onto what we know to be a human-to-computer relationship, and it doesn't map onto what we know to be a human-to-human relationship either,' Darcy said. 'It seems to be something in the middle.'⁶

There is a common narrative which underpins the introduction of AI devices and 'companions' in many different fields of care, therapy and education. There are simply not enough skilled humans to fulfil the roles. The needs for care across the planet are too great and they are projected to become ever greater. We have to find a technical solution to the lack of human carers, therapists and teachers. Machines therefore can be a 'good enough' replacement.

Some pro-AI enthusiasts go further, arguing that humans are frequently poorly trained, bored, fatigued, expensive and occasionally criminal. In contrast the new technological solution is available 24 hours a day. It never becomes bored or inattentive. It is continuously updated and operating according to the latest guidelines and ethical codes. It can be multiplied and scaled indefinitely. To many technologists machine carers will not only become essential, they will be superior to the humans that they replace! (Of course this technocentric narrative is highly misleading and will be discussed in greater detail below.)

As the technology continues to develop, supported by massive commercial funding, it seems likely that in the foreseeable future we will be confronted by a spectrum of different machines that offer simulated relationships, ranging from the entirely disembodied, like a chatbot, via an avatar or image on a screen, through to a physical humanoid robot.

Sex robots

The development of artificially intelligent sex robots provides another perspective for engaging in the complexities of human-machine relationships. Academic debate has focused on whether the use of sexbots by adults in private should be regulated and whether on balance they will be beneficial for humans.⁷ Classical libertarian arguments have been used by those in favour of robot sex, especially focusing attention on humans who suffer enforced sexual deprivation, including prisoners, the military, those forced to live in single-sex

3 www.parorobots.com

4 Sherry Turkle, *Alone Together*, Basic Books, 2011.

5 See <https://woebot.io>

6 Quoted in www.businessinsider.com/stanford-therapy-chatbot-app-depression-anxiety-woebot-2018-1?r=US&IR=T

7 Ed John Danaher and Neil McArthur, *Robot Sex: Social and Ethical Implications*, MIT Press, 2017.

environments, and those with mental health or learning disabilities. Again we find a version of the ‘good enough’ argument; sex with a machine may not be the same as with a loving human partner but it is better, so the argument goes, than being deprived of sexual activity altogether.

There is no doubt that the development of humanoid sex robots will lead to troubling ethical and regulatory dilemmas. Should the law permit humans to enact violent and abusive actions on humanoid robots who plead realistically for mercy? Should the use of child sex robots be outlawed? These questions lie outside the scope of this paper, but they raise complex moral and legal questions of importance for secular thinkers as well as for Christians.⁸ Although libertarian arguments are already being employed to oppose legal restrictions on sex robots, in my own view, regulation and criminal sanctions will become necessary, not only to protect children but also adults from the potential harms of acting out violent and abusive behaviour with humanoid robots.

Anthropomorphism

At the root of many of the issues surrounding human-machine interactions is our profound inbuilt tendency to anthropomorphism, our capacity to project human characteristics onto non-human animals and inanimate objects. Commercial manufacturers and software developers are expending considerable efforts to find ways to elicit our anthropomorphic tendencies. Most companies are not aiming to replicate the human form exactly, since robots which are too similar to humans are often perceived as ‘creepy’ and repellent. Instead the focus is on creating ‘cute’, friendly, child-like beings, to enable us to suspend disbelief and enjoy the interaction.

A particular focus is on the robotic or virtual face, and especially on the eyes. Eye movements and gaze appear to be a central non-verbal cue in helping humans understand the intentions of other social agents. Establishing eye contact, and experiencing the other as ‘looking back at us’, enables us to reach out to another person who is conceived and rationalised as being ‘behind the eyes’.

A second focus for eliciting anthropomorphism is the development of increasingly human-like speech. This, of course, is a vital difference between human-animal relationships and human-AI relationships. Although a dog may appear to understand human speech (at least in part), it cannot speak back to its owner. As I gaze into my dog’s eyes I have no idea what it is thinking – or even if it is thinking at all. But when my pet robot or AI chatbot speaks back to me, something momentous has happened. The machine appears to be communicating to me the hidden thoughts of its ‘mind’ – revealing its own subjective experience, conscious awareness, desires and



intentions. Except of course that it has none of these. So to give a machine the capacity for human-like speech seems both powerful and manipulative.

The commercial use of powerful anthropomorphic mechanisms opens us up to new forms of manipulation and even abuse. As journalist David Polgar put it, ‘Human compassion can be gamed. It is the ultimate psychological hack; a glitch in human response that can be exploited in an attempt to make a sticky product. That’s why designers give AIs human characteristics in the first place: they want us to like them.’⁹

Analogous personhood or simulated personhood?

From a positive perspective, some intelligent machines may be perceived as holding, in Nigel Cameron’s phrase, at least ‘analogous personhood’.¹⁰ They are not real human persons but they can play to a limited extent some of the same social roles. They can give us an experience which is analogous to human friendship and this may have many beneficial consequences. An analogous friend can teach me how to build friendships with a real person and can play the role of friend when my real friend is absent. An analogous carer can give me part of the experience of being cared for.

There has been some discussion as to whether social robots should be designed to supplement existing human relationships or to replace them. Some have emphasised that social robots are meant to partner with humans and should be designed to ‘support human empowerment’. When used correctly, it is argued that social robots can even be a catalyst for human-human interaction. This inevitably raises new questions. Children who grow up with a sophisticated digital assistant, such as Alexa or Google Home, are practising relationships with an entity that is human-like, but whose sole purpose and function is to attend obediently to every human command and whim. This is to model a relationship with a (usually female) human

⁸ Ibid.

⁹ David Polgar quoted at <https://qz.com/1010828/is-it-unethical-to-design-robots-to-resemble-humans/>

¹⁰ Personal communication.

slave. So should we teach our children to be polite to Alexa, to say please and thank you, to respect its 'virtual' feelings? Or is it of no significance if children abuse, tease and bully a simulated slave-person?

The 'relationally sensitive' responses that chatbots generate are those that their programmers have prioritised. So it is inevitable that they reflect what programmers think is desirable in a submissive and obedient relationship. The relationship that is offered reflects the demographic of the programmers: mainly male, young, white, single, materialistic and tech-loving. The image of the Silicon Valley engineer is reflected in their machines; these are the hidden humans who are haunting the robots in our homes.

Many young tech specialists seem to have an instrumentalist understanding of relationships.¹¹ At the risk of over-simplification, they seem to assume that the purpose of a relationship is to meet my emotional needs, to give me positive internal feelings. So if a 'relationship' with a machine is capable of evoking warm and positive emotions it can be viewed as an effective substitute for a human being.

Sherry Turkle reported that children who interacted with robots knew that they were not alive in the way that an animal was alive. But children often described the robot as being 'alive enough' to be a companion or a friend.¹² In a study of children who grew up with robots that were lifelike in appearance or in social interaction, Severson and Carlson reported that children developed a 'new ontological category' for them. 'It may well be that a generational shift occurs wherein those children who grow up knowing and interacting with lifelike robots will understand them in fundamentally different ways from previous generations'.¹³ What effects will this have on the emotional development of children? As Sherry Turkle put it, 'The question is not whether children will grow up loving their robots. The question is "what will loving mean?"'.¹⁴ In other words, how may human relationships become distorted in the future if children increasingly learn about relationships from their interactions with machines?

It is important to consider the wider societal context in which relationships with machines are being promoted and are likely to become increasingly common. There is an epidemic of relational breakdown within families and marriages, as well as increasing social isolation and loneliness. This leads to a pervasive sense of relational deficiency and a technologically-simulated relationship seems an ideal fix. A private chatbot available any time and anywhere appears to offer a degree of intimacy and availability which no human bond can ever match.

There is no doubt that human-machine relationships raise complex ethical, social and philosophical issues and there have been a number of recent initiatives within the UK and elsewhere aimed at the development of regulatory

frameworks and ethical codes for manufacturers of AI and robotic technology.¹⁵ From a Christian perspective there are a number of fundamental questions which seem of special significance:

1. How should we think of 'relationships' with machines within the context of the biblical revelation? What is the fundamental difference between a relationship with another human and a relationship with a machine?
2. Will the promotion of 'relationships' with machines contribute to societal wellbeing and human flourishing? How will simulated relationships influence and change the web of human relationships on which a healthy society is founded? What potential harms may flow from this?
3. What practical steps can be taken to minimise the potential harms and manipulative potential of machine relationships?

In the remainder of this paper, I shall outline some initial responses from a distinctively Christian and biblical perspective.

Christian responses

Humans are created to be embodied relational persons

In biblical thinking human beings are created as embodied persons, sharing a biological inheritance with animals, but uniquely created as God's image-bearers. We are created to represent God's loving care for the world and for relationships with God himself, with one another, and with the non-human world.

So we are created beings of a particular kind, embodied, fragile, and dependent. We are mortal and limited, but designed for union and communion with one another and ultimately with God himself. We are persons created by a relational God for relationships. And our humanity embodied in flesh is central to our relationships (Genesis 2:23, 24). Instead of being superseded, our fleshly embodiment is



11 Sherry Turkle, *Alone Together*, Basic Books, 2011.

12 Ibid.

13 R. L. Severson and S. M. Carlson, 'Behaving as or behaving as if? Children's conceptions of personified robots and the emergence of a new ontological category.' *Neural Networks*, 2010, 23:1099–103.

14 Sherry Turkle, 'Authenticity in an age of synthetic companions', *Interaction Studies*, 2007, 8, 501–517.

15 House of Lords Select Committee on Artificial Intelligence report, 2018, 'AI

in the UK: ready, willing and able?', <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>; Nuffield Foundation, 'Ethical and Societal Implications of Data and AI', 2018 nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundation.pdf; European Commission, 'Ethics Guidelines for trustworthy AI', 2019, <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

vindicated in the Incarnation and Resurrection when the Word became flesh (John 1:14; Luke 24:39). Machines on the other hand cannot share our fleshly embodiment. They are artefacts of human creativity, with the potential to support our unique human calling, but they can never enter into genuine human relationships.

As we have already seen, behind the simulated compassion of AI bots and companion robots it is possible to identify a shallow and instrumentalised understanding of relationships, seen as orientated towards the satisfaction of my internal emotional needs. But the Christian faith provides a richer and deeper perspective on human relationality. At their most exalted, human relationships can mirror and participate in the union and communion, the self-giving *agape* love, of the Persons of the Triune God. In the Gospels, Christ himself models voluntary and freely chosen self-sacrificial love for others. 'Whoever would be great among you must be your servant, and whoever would be first among you must be slave of all. For even the Son of Man came not to be served but to serve and to give his life as a ransom for many' (Mark 10:43–45). The paradoxical nature of Christlike love, whose concern is not for the meeting of one's own needs but is instead self-forgetful because it is focused on the other, is beautifully expressed in the prayer of St Francis of Assisi:

O Divine Master, grant that I may not so much seek to
be consoled as to console;
to be understood as to understand;
to be loved as to love.
For it is in giving that we receive;
it is in pardoning that we are pardoned;
and it is in dying that we are born to eternal life.

Authentic Christlike compassion depends on freedom, the freedom to choose to serve and give to the other. And it depends on human solidarity, on our common humanity and shared experience. A machine cannot know what it means to suffer, to be anxious, or to fear death, and its simulated compassion (even if well-meant by its creators and users) is ultimately inauthentic.

However, it is not sufficient only to concentrate on the fundamental ontological difference between humans and machines. The machine is nothing but a sophisticated artefact, but if it becomes capable of simulating many of the most profound aspects of human persons and human relations, and hence evoke in other human beings responses of love, care, commitment and respect, this raises new and troubling issues.

Simulated personhood raises the question of whether I can ever be certain whether the entity I am relating to is a machine or a human. One possible regulatory approach is that of the 'Turing Red Flag', first proposed by a professor of computing, Toby Walsh, in 2015.¹⁶ A Turing Red Flag law

would require that every autonomous system should be designed to prevent it being mistaken for one controlled by a human. In the case of a chatbot, for example, there might be a law that in every interaction you should be reminded that you are speaking to a clever simulation and not to a real human person.

The eyes and the voice

It is striking to reflect on the priorities of social robotics from the perspective of the biblical narrative. The focus on the face and eyes reflect the Hebraic use of the face of God to represent his personal presence, as in the words of the Aaronic blessing: 'The Lord bless you and keep you; the Lord make his face to shine upon you and be gracious to you; the Lord lift up his countenance upon you and give you peace' (Numbers 6:22–27). Moses' face shone because he had been in the presence of the Lord, and the Apostle Paul uses the same metaphor: 'We all, with unveiled face, beholding the



Pepper, a semi-humanoid robot

While we may see wide economic and practical benefits from advancing AI technology, as biblical Christians we are called to safeguard and to celebrate the centrality of embodied human-to-human relationships.

glory of the Lord, are being transformed into the same image from one degree of glory to another' (2 Corinthians 3:18). Jesus taught that the eye is the lamp of the body (Matthew 6:22), pointing to the moral significance of what we choose to focus our vision upon.

The centrality of speech in the biblical narrative is just as striking. The spoken word of God is the very means of creation, the word expresses the hidden thoughts and purposes of the divine mind, and Christ himself is the *Logos*, the ultimate expression and revelation of God. So the spiritual significance of the face and of spoken words, and their foundational role in divine and human relationships cannot be avoided. The simulation of these precious and theologically rich means of divine and human communication for commercial motives seems to point to a spiritually malign element that is facilitated by current technological developments. The Apostle Paul describes Satan as disguising himself as an angel of light (2 Corinthians 11:14). The Greek word *metaschematizō* that

¹⁶ Toby Walsh, 'Turing's Red Flag', 2015, <https://arxiv.org/abs/1510.09033>

Paul employed means ‘to change from one form into another’ and it is perhaps not too fanciful to see the possibility of spiritual evil which may accompany simulation of the most precious aspects of human relationality.

The idol

Biblical scholars have pointed to the link between the Genesis description of human beings as being created in the image (*selem*) of God, and the subsequent use of the same word *selem* to refer to idols or ‘graven images’ in the later Old Testament.¹⁷ The implication seems to be that our creation in God’s image reflects our profound creaturely dependence upon him, but this is subverted when we transfer the divine image to a human artefact. As Richard Lints puts it, ‘Human identity is rooted in what it reflects’.¹⁸ The idol may be ontologically vacuous but its false image is capable of exerting a malign and destructive hold on its worshippers. There seems to be a strange parallel between the evil consequences of creating a human artefact as an image of God and creating a robotic artefact as an image of humanity. As technologically simulated relationships become ever more realistic and superficially convincing, we must be aware of the risk that the simulacrum will exert a seductive appeal to our hearts.

Practical implications

‘How then shall we live’ in a society which seems to be increasingly promoting AI-simulated relationships in many aspects of care, therapy, education and entertainment? These challenges are complex and multifaceted but an initial response is to ask what are the underlying questions and needs to which AI-simulated relationships appear to provide a technological solution?

As we saw above, a common narrative is that the needs for care across the planet are too great and that we have to find a technical solution to the lack of human carers, therapists and teachers. But the current shortage of carers is, of course, in part a reflection of the low status and low economic valuation which our society places on caring roles. There are more than enough human beings who could undertake the work of caring, both in paid roles and also in unpaid voluntary caring within families and communities. It is surely better as

a society that we strive to facilitate and encourage human carers, rather than resorting to technological replacements for human beings.

In the world of healthcare, although AI technology can provide remarkable benefits with improved diagnosis, image analysis and treatment planning, it cannot replace the centrality of the human-to-human encounter. The realities of illness, ageing, psychological distress and dementia all threaten our personhood at a profound level. In response, the therapeutic and caring encounter between two humans provides an opportunity for human solidarity which understands, empathises with and protects the frailty of the other.

In my experience as a paediatrician, with the privilege of caring for children and parents confronted with tragic and devastating loss, I have learnt afresh that the essence of caring is to say both in our words and our actions, ‘I am a human being like you; I too understand what it means to fear, to suffer and to be exposed to terrible loss. I am here to walk this path with you, to offer you my wisdom, expertise and experience, and to covenant that I will not abandon you, whatever happens.’

So, in conclusion, while we may see wide economic and practical benefits from advancing AI technology, as biblical Christians we are called to safeguard and to celebrate the centrality of embodied human-to-human relationships, particularly in essential caring and therapeutic roles, and in our families and Christian communities. There is no substitute for human empathy, solidarity and love expressed in the face-to-face gaze of embodied human beings and in compassionate, thoughtful words spoken by human mouths.



John Wyatt is Emeritus Professor of Neonatal Paediatrics, Ethics & Perinatology at University College London, and a senior researcher at the Faraday Institute for Science and Religion, Cambridge.

¹⁷ Richard Lints, *Identity and Idolatry*, IVP, 2015.

¹⁸ Ibid.

Photo Credit: Page 2, Amber Case, CC BY-NC 2.0

About Cambridge Papers

All available issues of *Cambridge Papers* dating back to 1992 can be downloaded from www.jubilee-centre.org/category/cambridge-papers. We encourage debate and, if you would like to respond to this or any other Paper, we invite you to visit the website and post your comments there.

If you would like to receive *Cambridge Papers* and regular news from the Jubilee Centre, and you are not already on our mailing list, please join via our website www.jubilee-centre.org or send an e-mail to info@jubilee-centre.org

Alternatively, please send your name and address to:
The Jubilee Centre, St Andrews House, 59 St Andrews Street,
Cambridge CB2 3BZ Tel: 01223 566319

As part of our commitment to the environment and our efforts to reduce production and postage costs, we encourage readers to access *Cambridge Papers* electronically. Whether you receive

our papers by email or in hard copy, there are costs involved in publication and distribution, and we invite readers to help meet these by making a donation by standing order, cheque or credit card. If you are a UK taxpayer, signing a Gift Aid declaration (available from the Jubilee Centre) will increase the value of your donations.

Cambridge Papers is a non-profit making quarterly publication which aims to contribute to debate on a wide range of issues from a Christian perspective. The editorial group is an informal association of Christians sharing common convictions and concerns. The contribution of each member of the group, and each author, is as an individual only and not representative of any church or organisation.

Next issue:
The housing crisis

 **Jubilee Centre**
BIBLICAL THINKING FOR PUBLIC LIFE