The Faraday Institute for Science and Religion

AI, ETHICS & FAITH

NAVIGATING THE FUTURE

AI, ETHICS & FAITH

NAVIGATING THE FUTURE

Table of Contents

Acknowledgements 2
Foreword 5
What do we mean by 'AI'? 6
What are 'Algorithms'?
What are 'Deep Neural Networks' (DNNS)? And Large Language Models (LLMS)?
What is 'AI Ethics'? 8
Is there an agreed set of principles which makes up AI ethics?
How can we implement these principles in practice?
Why 'Faith'? 11
Is AI a 'Good' Or 'Bad' thing? 15
What are some of the opportunities AI presents? 18
Economic efficiency
In the Medical Sector
As a tool of engagement within faith communities
What are some of the risks AI presents? 20
Fake News & Journalism
Employment & Labour Markets
Terrorism, Child Sexual Abuse & Other Cybercrimes
Environmental Concerns
Amplifying Biases
Limiting Human Flourishing
What are some of the challenges of developing AI ethics moving forward? 24
Leaders' lack of understanding
The domination of utilitarian thinking over other important 'Human Values'
Advocating for marginalised communities
How can our faiths inform our response to these challenges? 27
Is it too late? 28
What next? 29
MoU Signings : Faraday AI 30
Appendix A – The faces behind the quotes: Speakers from the 2023 AI & faith conference 31

Acknowledgements

This handbook is based on the high-level discussions held at the AI, Faith & the Future of Society Conference on 15 June 2023 in the Palace of Westminster. Bringing together over 40 prominent faith leaders, government officials, academics, and technologists, the conference explored the implications of Artificial Intelligence (AI) for society, with a focus on incorporating Abrahamic faith perspectives. By addressing broader themes such as values, virtues, human identity, dignity, and purpose, the consultation seeks to inform the development of ethical frameworks for AI that align with the principles and guidance of the New Alliance of Virtue.

We would like to thank our speakers, Dr Nick Chatrath, leading international expert on the potential impact of AI, Fr Paolo Benanti from the Pontifical Gregorian University in Rome, Lord Nat Wei, Jon Cruddas MP, and the Rt Hon Sajid Javid MP, as well as our panelists, Dr Erin Green, Prof John Wyatt, Dr Ramon Harvey, Rabbi Dr Harris Bor, and Dr Yaqub Chaudhary.

We would also like to thank Shaykh Abdallah Bin Bayyah, President of the Abu Dhabi Forum for Peace for blazing a trail for interfaith relationships in the UK and for his commitment to furthering the discussion on AI ethics, most recently in his speech at the AI ethics Conference in the Vatican City in January 2023, as well as the Secretary General of the ADFfP Sheikh Al Mahfoudh Bin Bayyah.

This handbook was produced by the UK Forum for Peace (Registered Charity No. 1186676), chaired by Zeshan Zafar and The Faraday Institute for Science and Religion, of which Graham Budd is Executive Director. We would like to thank both of them for their persistent efforts in spotlighting and advocating for the importance of engaging faith perspectives in the discussion on AI.

We would also like to thank Graham Budd, Nicole Pearson and Esmé Partridge for their contributions to this handbook and Sonam Mittal for the design.

"AI is among the most momentous creative innovations of human genius that have emerged moving from theory to application in a very short time. Its use now spans both the private and public sectors, and whilst it has already succeeded in various fields, its development continues due to a passion for innovation, invention, and sustained investment. Hardly a field today exists that is not directly involved in it, especially healthcare, transportation, education, training, and manufacture."

H.E. SHAYKH ABDALLAH BIN BAYYAH

Foreword

Artificial Intelligence is the transformational technology of our era - the first tool in history that can seemingly create new ideas and make decisions on its own. Non-human intelligence can now perform complex tasks better than most humans – writing creative essays, discovering new drugs, or driving a car. This technology has the power to transform society, for good or for bad. AI's ability to simulate human behaviour holds a mirror up to the human soul, and reawakens ancient questions about human identity and purpose – questions that religion has been seeking to address since the dawn of human existence. What does this mean for humanity? Can AI be a force for good for human flourishing, or will superintelligent AI result in the marginalisation of some or even all humans? What kind of relationship should humanity have with AI, and what implications does this have for religion?

The purpose of this handbook is to provide some context on AI to enable the exploration of this relationship between AI, ethics, and faith, with examples of how diverse faith perspectives are important to help shape the future of AI. For this reason, this handbook specifically draws on insights from the diverse range of speakers and discussions held at the "AI, Faith and the Future of Society" event at the UK parliament in June 2023.

AI ethics is of vital importance to the future of human society, culture and religion in the 21st century. There are no absolute, clear answers to how to proceed, but what is clear is that people of faith need to be actively engaged in the debate about the future of AI: a debate which goes to the heart of understanding what it means to be human, and the virtues and values which bind us together.

Graham Budd
Executive Director, The Faraday Institute for Science and Religion

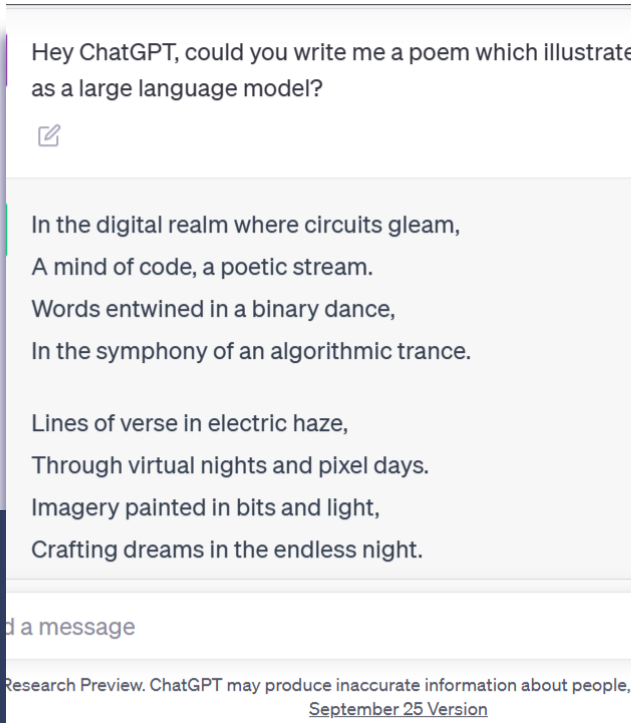
What do we mean by 'AI'?

Whilst there are many definitions of artificial intelligence, it can be generally described as a broad term referring to a machine capable of acting in an intelligent manner. "Intelligence" is a similarly difficult definition to pin down, but the meaning advanced by the European Commission's High-Level Expert Group on AI focuses on the idea of "rationality", or the ability of a machine to gather data about its environment, apply some form of reasoning to interpret the data and ultimately make an appropriate decision based on this process.

Humans have always been fascinated by trying to understand or simulate our own intelligence. From ancient times philosophers and theologians developed theories about the mind and intelligence, which when later combined with foundational ideas about computing, led to concepts such as the "Intelligent machines" envisioned by Alan Turing in 1950. The term "Artificial Intelligence" was first used by John McCarthy in 1956 - AI research began in earnest in the 1950s, going through multiple cycles of hype and despair in the second half of the 20th century. There were some notable breakthroughs for superintelligent task-specific AI, such as IBM's Deep Blue beating the world's best chess player in 1997. However, it was only with the advent of the data-driven internet economy in the 21st Century that a particular type of AI, based on machine learning

What are 'Algorithms'?

Algorithms are sets of rules or instructions that tell a computer how to perform a specific task. They're the building blocks of all AI systems, and are behind their ability to perform a wide range of functions, including natural language processing, image recognition, and more. A particularly notable example of a class of machine learning algorithm behind some of these impressive functions are 'deep neural networks' (DNNs).



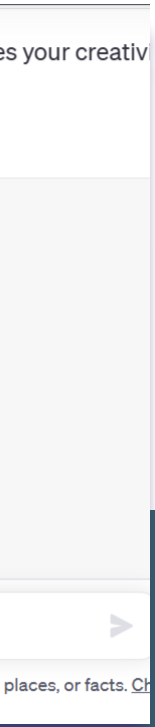
What are 'Deep Neural Networks' (DNNs)?

Today, almost all AI systems are based on machine learning, using deep neural networks (DNNs). This is a foundational technology for applications from self-driving cars to generative AI chatbots. A DNN is a multi-layer architecture of connected nodes, very much like neurons in the brain, with programmable weightings assigned to each of the connections. Vast data sets are used to "train" the neural network, by adjusting the weightings so that after training the output is closer to simulating an intelligent human outcome. Once the DNN is trained, new output decisions can be inferred from new input data. For example, early DNNs were used for image recognition – after training on millions of images, the network "learnt" how to recognise any image of (for example) a cat.

And Large Language Models (LLMs)?

The most powerful application of DNN technology so far has been the creation of Large Language Models (LLMs), which are a form of DNN using a Transformer architecture that enables the model to "learn" meaning and context by tracking the relationship of multiple inputs such as words in a sentence. These models are trained using vast swathes of human-created language data from books and the internet to predict the next most likely word in a sentence and are used by chatbots such as OpenAI's ChatGPT or Baidu's Ernie. These generative AI chatbots create uncannily human-like responses in conversation, can write exam-quality essays or poetry, and seem to be capable of context-relevant creativity that goes well beyond parroting back parts of their training data.

LLMs have highlighted one of the ethical challenges of AI systems. They produce a probabilistic output simulating human intelligence but have no inherent understanding of what they create, and no common-sense override. If the data sets used to train include some inherent bias (for example, a lack of content relating to a particular culture, gender, or ethnicity), then the model will produce biased outcomes. This becomes a major issue if, for example, such an AI system is making decisions on eligibility for societal benefits or access to justice, or creating educational content.



"Hey ChatGPT, could you write me a poem which illustrates your creativity as a large language model?" Chat GPT, September 25 Version, "Open AI", 9 October 2023.

What is 'AI Ethics'?

Artificial Intelligence is impacting numerous aspects of our lives - be it education, the workplace, transport, or even warfare - in ways that can feel beyond our understanding and control. Many fear the consequences it might pose for humanity: *what does automation mean for the future of employment? Could AI be used for harmful ends, or propagate prejudices and biases? What happens if it comes to have a mind of its own that surpasses human agency and regulation?*

AI ethics seeks to address these questions and address the moral and spiritual concerns of this emerging technology more broadly. It does so through exploring philosophical and also - though currently less so - theological perspectives, and putting them into dialogue with those developing the technology themselves. The Abu Dhabi Forum for Peace and The Faraday Institute for Science and Religion believe religious traditions can provide valuable and constructive contributions to this field and that including their perspectives is essential in ensuring that AI meets the ethical standards of all.

Many of these principles are echoed by H.E. Shaykh Abdallah bin Bayyah, who, speaking from a Muslim perspective, describes the importance these values to protect the rights of both people of faith and those of no faith:

Is there an agreed set of principles which makes up AI Ethics?

There is no exhaustive list of which agreed principles, values or standards should underpin an AI ethics framework. However, in 2019, the OECD adopted a widely used and influential ethical framework identifying the following key principles as crucial to the responsible development of AI, which has since been widely adopted across policymakers, academics and civil society:

- **Inclusive growth, sustainable development and well-being**: stakeholders need to take an active role in responsibly managing reliable AI to achieve positive results for humanity and the environment. These include improving human abilities and fostering creativity, promoting the inclusion of underrepresented populations, reducing economic, social and gender equalities, and safeguarding the natural world.

- **Human-centred values and fairness:** AI actors must uphold the principles of law, human rights and democratic ideals at every stage of the AI process. These encompass dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice and internationally recognized labour rights. To achieve this, AI actors must put in place mechanisms and safeguards, like the ability for human decision-making, that are suitable for the specific situation and in line with the latest advancements in the field.

- **Transparency and explainability:** AI actors must responsibly disclose meaningful information relative to AI in a simple and easy-to-understand way, with the goal of:
  - Fostering an understanding of AI systems
  - Making stakeholders aware of the interactions with AI systems, including in the workplace
  - Enabling those affected by AI to understand the outcome
  - Helping those negatively affected by an AI system to challenge its outcome

- **Robustness, security and safety:** AI systems need to be robust, secure and safe not only when used for their designated purpose, but also when they are being intentionally misused.

"This desired ethical system should be based on innate moral values, which are values common to all people, upon which no two people differ, and which are not affected by the vicissitudes of time, place, and human impulse. These are the natural rights of every human being by the very virtue of his or her humanity.

'They are Divinely-granted rights granted to both the believer and the atheist,' as stated in the first amendment of the Charter of The New Alliance of Virtue in 2019.

These ethics include the values of virtue, mutual respect, solidarity, and cooperation in benevolence and the common good, as well as the values of honesty, justice, sincerity, integrity, responsibility, and transparency. These include the central values in the ethical system of all religions and worthwhile human philosophies."

- **Accountability:** AI actors should be accountable for the proper functioning of AI systems and for the respect of all the above principles, based on their roles and the context, and consistent with the latest technological developments.

How can we implement these principles in practice?

One of the ways to implement some of these principles in practice is by encoding the ethical principles into the software itself. This is something explored in the field of 'algorethics', a term first coined by Fr Paolo Benanti to describe the branch of ethics that focuses on the moral aspects of algorithms and AI systems. According to Fr Paolo Benanti – who has been making major contributions to AI ethics from a Catholic perspective – algorethics will enable humans to build a "guardrail" against the harms caused by unethical behavior in this new age of machine:

"Ethics encapsulates all the problems human beings have on Earth. But now that machines can surrogate the decisions made by man, we must make sure that the same measures we used to take to avoid the harm of unethical behavior in humans, can be workable in this same Age. We need to write a new chapter of this ethics journal. And we should call it 'algorethics'.

Why? Because ethical questions have to be understood by a machine working without ethics. Algorethics is the at-

tempt to write a new chapter in ethics, to have some sort of guard rail for this kind of machine. For example, the reason you can drive a car with peace of mind is because you have the ability to control the car's destination and because there are guard rails on the road to avoid a collision with pedestrians.

We need algorethics, we need a guard rail for this age of the machine."

Why 'Faith'?

AI will transform society, amplifying the values and goals we give it – whether explicitly by design, or implicitly through choice of training data. We are creating tools that will enable individual humans and organisations to have 'superintelligent' power, influence, and agency that will rapidly impact other humans and society. While the technology brings great opportunities, governments and technology companies are starting to understand how widely-available generative AI could enable the powerful to manipulate the marginalised and potentially cause great harm to individuals and society – including people of faith. Some believe it could even cause the enfeeblement of humanity itself.

H.E Shaykh Abdallah bin Bayyah expresses why protecting humanity from this harm is indivisible from the goal of religious laws and commandments themselves:

> *"Religious laws and commandments, in their endeavour to promote wellbeing and guidance, are keen to ensure that achievements in science and technological take into account ethical frameworks that preserve human dignity and the nobility of man, and, most importantly, protect life itself."*

He goes on further to explain why from a Muslim perspective, the development of AI without robust ethical frameworks poses a threat to what Muslims call the five protected universals:

> *"These dangers could pose a threat to humanity from the perspective of what we Muslims term the five protected universals: religion, life, reason, property, and human integrity, the last of which includes family and individual human dignity. The preservation of these five higher objectives of our sacred law remains an overarching priority."*

In addition to religion's existential role in protecting human dignity, there are a number of other key reasons faith perspectives are particularly relevant to the discussion AI ethics:

1. Because the involvement of faith perspectives is key to inclusion

As outlined, the first OECD principle refers to the importance of promoting the inclusion of underrepresented populations. However, currently, faith perspectives are often missing from the debate on AI regulation, which is largely taking place between technology companies and governments. The absence of these perspectives could see AI developed in a way that fails to protect the interests of stakeholders with religious beliefs, which as the Rt Hon Sajid Javid MP underlines, make up a significant proportion of the population:

"Around 85 percent of us profess some faith, so it matters to almost all of us. It's important when there are such major developments that will touch every aspect of society, that faith is discussed, and the whole discussion is informed by this."

2. Because faith traditions have experience grappling with existential questions raised by AI

Secondly, one of the biggest challenges of AI is the so-called 'alignment problem' – how do we align the goals and values of an AI with those of humanity? Answering this requires thinking about how the technology is created and regulated, but also deciding what those goals and values should be. What does it mean to be human? Whether through AI design or the choice of training data, what ethical approaches do we want our AI systems to use to make decisions affecting individuals and society? What will be the consequences of the default utilitarian approach of machine learning systems on humanity and different cultures? Is there a role for virtue in AI systems to help transform society for good? Such questions of values, ethics and the purpose and identity of humanity have been explored by theologians and philosophers for thousands of years. Faith communities therefore have great insight to add to the 'alignment problem' debate, and it is essential that their voices are heard.

Speaking from a Muslim perspective, Dr Ramon Harvey illustrates an example of how faith can be used to grapple with questions around AI and 'stewardship': the idea, present in both the Bible and the Qur'an, that humans have a unique responsibility of protecting the world around us:

"[In the Qur'an] God informs the angels that he is going to create a steward of the Earth. The angels say, 'why would you do this, when it will cause bloodshed and trouble?'. God responds: 'I know what you do not know'.

***The question is, what happens if we create a future out of silicon? Can AI extend our stewardship of the world?** And how can this technology be used in the service of our own self-responsibility in the world that we have, in light of, for example, the climate breakdown?*

Could something more advanced be conscious? It seems clear that already there are systems which are indistinguishable from human beings in terms of the kind of responses they produce.

Some people belonging to Islamic theological thinking and community will deny that any machine on principle can ever have a soul. A soul, in its conception, is a unique spirit that is created by God and comes into human beings in the world and is not something that could be held by an object of our own creation. But in another reading, even animals and plants have a soul; if the soul is something connected with all life, then the issue becomes about how we define life? Can there be an artificial life, if something develops a degree of ability to interact with the environment, show reasoning and act in a way which is perceived as conscious, could it move up from an animal perception to that of a person, who would be in that sense responsible?

There is the possibility of considering whether something could gain this responsibility to not only be an extension of our own stewardship, but have stewardship in its own right."

Rabbi Dr Harris Bor also explains how the Jewish concept of Shabbat may provide some guidance on living in a new age of AI, by carving out a space or time AI doesn't go:

*"We have Shabbat (the Sabbath) once every seven days, on Saturday. It's an incredibly profound thing. The biblical roots lie in two areas. One of them is in the notion of creation, 'God creates the world in 6 days, rests on the 7th', so there's a link with creation. The other link is slavery. God released the Hebrew slaves from Egypt, and they were given a reminder of the need for freedom in the form of Shabbat, where all creative work ceases. The concept of Shabbat is relevant to AI because **we will need, as human beings, to secure a time or space where AI doesn't go,** to get some distance so we can take stock and see where are heading …we need to create a distance."*

3. Because faith voices offer a unique perspective

Thirdly, faith perspectives are also important to help avoid the erosion of culture by prevailing ideologies within the commercially-driven technology sector. For example, while secular Western ideological thinking emphasises individual identity and relative truth, different faith perspectives may bring a stronger sense of the value of community, different concepts of truth and morality, and additional perspectives on the deep value of human relationships which will be important insights to shape humanity's interaction with AI.

Dr Nick Chatrath argues that independent thinking is key to effective AI regulation. He lists that one of the ways faith communities bring a unique perspective is through their renown for looking out for the vulnerable members of our society, including those who are most likely to be most negatively affected by some of the risks AI presents:

"One reason independent thinking matters from a faith point of view is because, at best, people of faith look out for losers. Every social movement and paradigm shift brings with it winners and losers. One opposite of independent thinking is subservient thinking. As AI accelerates, the first of being subservient to the dominant way of thinking increases. A junior politician or businessperson might think what their boss wants them to think in order to progress or get that bonus. Against this, people of faith can cut against the grain, and encourage independent thinking among the 'have-nots' in any given society. ***So people of faith are well positioned to look to the lost and the least, to look out for the marginal and to give preference to the poor.*** *Independent thinking is at the heart of that because the assumption with independent thinking is that every single human being is intelligent, creative, worthy of good outcomes, and resourceful, no matter their background."*

Is AI a 'Good' or 'Bad' thing?

The Rt Hon Sajid Javid MP believes that the changes caused by AI can be divided into three buckets:

1. Those which are opportunities, such as in the healthcare and educational sectors;

2. Those which are challenges: which Mr Javid views as the things that come about because of technological change, and which are a necessary outcome of getting the benefits. They are the things he believes we have had to confront throughout time, with every technology change, such as within the labour market.

3. And those which are threats.

So does the 'good' outweigh the 'bad'? AI experts, academics, and technologists hold a wide range of different views on how to weigh up the different benefits and harms AI presents. Dr Nick Chatrath helpfully summarises these different viewpoints in three categories:

1. Bloomers: The bloomers are people who believe AI will follow a similar path as previous industrial revolutions. They assert that we are experiencing a great disruption, to which we will respond by adjusting the powerful, yet imperfect systems we already have, and which will then unlock unprecedented human flourishing.

2. Gloomers: The gloomers take a less optimistic view. This is the robots are going to take all the jobs and ethical nightmare' position.

3. Doomers: If the gloomers sounded grim, it's positively bubbly compared to the doomers. Many of us have seen The Matrix, Terminator, the Blade Runner, and they're full of cautionary tales of AI running amok and ruining the world. These fears are not just the province of science fiction. Sam Altman believes the world is not that far away from potentially scary AI developments, and he's the head of the company that brought ChatGPT to the world. Many business leaders and politicians are warning of the existential risks that AI may pose for humanity.

Examples of different views taken by AI experts, academics, and public policy experts:

Graham Budd

"I'd like to encourage us, as well as thinking about the risks, to think about how AI could be a powerful agent, to amplify virtue for the benefit of all humanity and help solve some of the challenges we face as society, especially for marginalised communities. From a Christian perspective, I see our ability to build AI systems as part of our amazing God-given creativity. So how can we harness that power of AI to promote peace and improve society?"

The Rt Hon Sajid Javid MP

"I would put myself as an optimistic bloomer. Especially after my time in health, some of the things I saw were incredible. Whatever happens in the regulation approach, it has to be informed by ethics, and faith is a very important part of that.

I think some of the things we hear are a bit sensational, and we maybe hear more about that because that's what's going to help sell news and media, but that doesn't mean you shouldn't take the issues being raised as not being serious; While we don't might not have all the answers, it's important to make sure we ask the right questions."

Dr Nick Chatrath

"I'm a cautious bloomer. I believe that humanity is at an inflection point now, given AI's progress. We're at a key shaping moment in history. There's no guarantee as to how things will turn out, but we can shape a better future and minimise the chances of negative outcomes, if we upgrade our leadership operating systems. As leaders, we have an opportunity to create a world for which your children and grandchildren will thank you."

Jon Cruddas MP

"I'm a bloomer but on the gloomier end. I suggest the authoritarian populism we witness on a global scale is almost an angry backlash to that technocratic form of public administration. It's an angry verdict among people about our inability to ask moral and cultural questions and the moral political range. People want a political conversation about the good life, the life they live compared to the lives that are promised to them by politicians, technologists, and people with power. That's why I'm at the gloomier end."

What are some of the opportunities AI presents?

Economic efficiency

One of the most widely anticipated opportunities presented by AI is its capacity to increase productivity and efficiency. AI systems can automate routine tasks, broadly improving efficiency whilst allowing labour to be allocated to tasks that are more dependent on human input. Economists anticipate that the efficiency benefits of AI technology will span a multitude of sectors including logistics, manufacturing, advertising, transportation, and financial services. To take one example, an AI system could draw on vast amounts of data to design a maximally optimised supply chain to manufacture a product; an AI system could then be used to anticipate demand and inventory levels at a regional level, before scheduling delivery of the product via an AI-powered autonomously driven vehicle. Drawing an analogy to the improvements in efficiency in the industrial revolution, whilst some jobs may be rendered obsolete in the short term due to automation, some economists suggest that AI will contribute to job creation in the long term.

In the Medical Sector

AI has shown exciting potential in the medical sector and is already being used by researchers and practitioners. Some current applications of AI in healthcare include:

Diagnostics: AI is making the process of diagnosing illness quicker and more accurate. Models are capable of drawing on vast amounts of data, such as images from scans or medical literature, and in some instances make decisions in seconds that would typically take a human much longer.

Treatment: The ability of AI models to draw on vast amounts of data when making decisions could allow for treatment to be personalised to patients, determining dosage and treatment plans based on an individual's health profile, rather than broader attributes such as sex.

Research and development of drugs: The development of new drugs is notoriously expensive and lengthy. AI is already speeding up the research process by enabling the simulation of potential drug molecules on computers, generating new drug molecules, and helping researchers to understand the way that diseases develop to target specific aspects of their biological functioning with new drugs.
The Rt Hon Sajid Javid MP, former Secretary of State for Health and Social Care, describes his first-hand experience of AI's potential to revolutionise medicine whilst in government:

*"**I think there will be huge benefits for mankind.** I speak from my experience in government as a minister. When I think about the government and its job to deliver good public services for everyone to look after the most vulnerable, I can see huge opportunities already.*

*To take some more practical examples, my last job in government as you all know was as Secretary for Health and Social Care. When I think about health, **there are huge improvements I've seen around health, in preventative care, in mental health care, and in drug discovery.** Looking at existing drugs and seeing if they can be repurposed and used for other sicknesses than they were originally brought to market for, developing new drugs and bringing them to market much quicker, developing them more cheaply.*

*Also in the field of diagnostics, especially image scanning. When I was there, I promoted an AI hub to develop new AI-based ideas. One of the first things they focused on was imaging. And I remember a visit to St Bart's Hospital when they were checking people for heart disease, they were able to find that in 30 seconds, instead of 13-14 minutes. That might not sound like a lot but that will save thousands of lives. **One of the things the government should be doing is looking out for the sick. I think AI is a really good example of how we can do that."***

As a tool of engagement within faith communities

Dr Erin Green, speaking from a Christian perspective, highlights how some faith communities have already started using AI in academia and community development, contrary to the commonly-held belief that faith communities are exclusively fearful and reluctant to engage with AI developments:

"What we're finding is Christian communities and theologians are freely experimenting and engaging with artificial intelligence in positive ways. Not just in academic circles, but in liturgical spaces, and other more common places.

*A few examples of this would be at the recent German church festival, the Kirchentag, they had an AI-led worship, which shows tremendous courage, and innovation to interact with AI in a very tangible way. Another example, in 2014, there was an evangelical seminary in the United States that bought a humanoid robot to be integrated into its ethics to learn about anthropology, what it means to be human, and these types of questions that often come up in the abstract of academic reasoning, but in a material way. Something you can manipulate. Another example is a contact in Singapore, who works really intensively with virtual reality, working on the development of communities in virtual spaces, and even exploring traditional theological stances there. So I think it would be very wrong to characterise Christianity as locked-in, fearful of AI, and only acting in a responsive way. These few points that I've mentioned, coming from the stance of playfulness point to how **Christians aren't just reacting but proactively engaging in the development and the directions in which AI will be in the future."***

This notion that religion is not at odds with the development of AI is also echoed by H.E. Shaykh Abdallah bin Bayyah, who underlines religion's role as a stimulus to innovation:

"The Qur'an states, 'Say, ‹My Lord, increase me in knowledge.' Knowledge, in its exhaustive meaning, of which we perpetually seek an increase includes both religious and worldly knowledge, which is the foundation of any successful civilization both spiritually and materially. Thus, when religious civilizations flourished, discoveries and inventions were made that enhanced human life. ***Religion was not a hindrance to innovation, but rather a stimulus that motivated creativity and achievement.****"*

What are some of the risks AI presents?

Fake News & Journalism

There are rising concerns about the ability of AI tools to blur the line between artificially generated content and reality, to the detriment of accurate information online. Many have pointed to the capacity for rapid and far-reaching information sharing on social media as a vehicle for widespread disinformation, to the detriment of democracy and public health. AI allows harmful actors to go further by utilising deep learning to create fake images and videos known as 'deepfakes'. Deepfake content can spread like wildfire on social media, spreading disinformation on group chats before fact-checkers may respond. A deepfake video of a politician endorsing a political rival or espousing a hateful belief can be much more compelling than a fake news article articulating that same story. As the old adage goes, "seeing is believing".

According to Dr Nick Chatrath, we are already starting to see the impact of AI in the quality of journalistic output:

"A growing number of media companies are already using AI to write news articles. And the fact those articles are sometimes littered with factual errors and awkward syntax matters less to those in charge of the money."

This fear is echoed by H.E. Shaykh Abdallah bin Bayyah who warns that AI ***"can cause further division and disunity in societies by contributing to the spread of false information and misleading news and influencing voter opinions."***

Employment & Labour Markets

Another oft-discussed concern is the effect of AI on the employment & labour market. As many of AI's developments are centred around replacing low-skilled jobs, many of the economic benefits of AI will be concentrated among people with well-paid skilled jobs. According to the World Economic Forum's "The Future of Jobs Report 2020", AI is expected to replace 85 million jobs worldwide by 2025.
However, while acknowledging the need to put in place programs to help people adapt to this transition, the Rt Hon Sajid Javid MP underlines that this should not stand in the way of modernisation:

*"**The jobs people have done for years, sometimes for decades will either disappear or become less important.** Obviously, that's huge and will create a lot of anxiety, and concern for people understandably. Governments will have to put things in place programs of transition to ease that, but it shouldn't try to stop it altogether. I think if you try and stop it, you won't get the benefits. It's something we've coped with for decades, and it's something we can manage and get through."*

Terrorism, Child Sexual Abuse & Other Cybercrimes

Aside from its potential to radicalise via the spread of disinformation, AI could be harnessed by terror groups in planning terror attacks. A particular area of concern is bioterrorism, where LLMs may enhance the research capacity of terror groups or state actors looking to design biological agents to inflict harm on the public. An LLM could act as a research assistant, providing information about how to synthesise a pathogen, how to maximise its potency, where to obtain relevant materials, and how to run experiments, with the ultimate goal of creating a biological weapon.

AI is already being used by cybercriminals to enhance their capabilities. Tools at the disposal of fraudsters and black hat hackers range from voice cloning (used in fraudulent calls to pose as family members asking for money), to deepfaked videos of celebrities endorsing fake products, to LLMs capable of coding potent malware and generating fraudulent messages capable of bypassing spam filters. Child safety charities are also raising alarm bells about paedophiles modifying open-source AI technology to generate child sexual abuse material (CSAM).

H.E. Shaykh Abdallah bin Bayyah enumerates some of these security risks as follows:

*"**Artificial intelligence may pose a threat to life as an uncontrolled weapon used against humanity**. It also poses a danger to human integrity because of the absence of privacy and the subsequent violation of information security. It may also cause the loss of money and property through electronic theft, piracy, and intrusions."*

Environmental Concerns

AI models require significant amounts of computing power for their training and operation, which in turn can be highly energy intensive. These environmental harms include the resource cost of building the hardware AI models are run on, as well as the construction and operation of vast data centres. Additionally, as a tool capable of influencing human behaviour, AI could also lead to environmental harm by fueling the spread of disinformation about climate change.
Dr Erin Green expands on the environmental harms of AI in the context of the climate crisis, emphasising the importance of this aspect of AI not being overlooked:

"We always talk about human AI, human-centric AI, and here we are in global catastrophic climate breakdown. Where is the Earth in this? Where is Creation in this? We often don't see ecological concerns at the heart of this, even though ***AI is directly related to rapacious consumption, e-waste, mining, the use of natural resources to support data centres, and on and on and on.****"*

Amplifying Biases

As explained, AI functions by producing probabilistic output based on datasets they are trained with, but don't have any inherent understanding of what they create. As such, AI systems can reproduce biases embedded in the data they're trained on, reflecting societal and historical inequalities and stereotypes. Even when characteristics like race or gender are removed, it can still reproduce biases through names or locations.

An example of this is the 2015 Amazon recruitment algorithm. Seemingly innocuous, the algorithm was being used to hire employees based solely on the resumes submitted to the company over the past ten years. However, because that pool of resumes was made up of predominantly male applicants, the algorithm was found to be biased against women.

Dr Ramon Harvey compares the way AI amplifies biases to author H. P. Lovecraft's "shoggoth" - a kind of monster that reproduces but in doing so exaggerates human behaviours:

"Large Language Models point out the black box element to how these things work. That they work and can produce these miraculous feats of language or image is clear, but how precisely they're doing this seems to be somewhat unknown. And because of this, they are being described as a kind of 'shoggoth'.

In the story, the shoggoth gets to work and is put to work to build cities because it's flexible in the

shapes it can makes, but it twists the things it makes and mocks life. And I think this is to some extent what we're seeing in some of the generative ***AI in the way it takes and twists images, distorts the use of language. And it does so in this amplificatory way, to amplify biases, amplify evil, amplify the things that are already there.*** *If it's being trained in a Euro-centric way, if it's being shaped by what's of interest to the companies feeding it, it's going to not only take that as its basis, but amplify it. So if you're going to use these tools, they should more accurately reflect different religions, different ethnic groups, and so on."*

Limiting Human Flourishing

Another concern among people of faith is that AI will inhibit human creativity. Dr Ramon Harvey, discussing this issue, says that LLMs can only regurgitate what is given to them, and so what appears like new content is really just something pre-existing in a different form. Only human beings, he suggests, are truly capable of creating new ideas and works of art, and there is a danger that the convenience offered by AI will tempt us into neglecting this:

"One worry is it seems that the very design of these LLMs is that they hallucinate. And when they try and make them more accurate, they become boring. So some of the creativity you see, is tied up with this fraudulent nature, the way they twist things. This seems to be a systemic problem.

Another is the ethical issue of how these ***LLMs generate things in such a way that they devour the text, cultural artefacts, of thousands of authors and artists.*** *They need original human material to feed it. So if they only have AI input to create content with, they start to degenerate.*

That's a problem because once they have that, they're very good at imitation. It's almost indistinguishable from what's already there. But they can't do a new thing. It's almost a reflection of the way culture has been in the last 20 years, endless remakes and recycling of ideas. The point is they're going to have to keep relying on original designs, new ideas. But where does that leave people who are trying to express themselves originally? And are we just feeding things to replace us in our own production? I think at least with the LLMs I'm really worried about the way these things work and the risks to human creativity."

Rabbi Dr Harris Bor suggests that AI, if not handled properly, could overshadow not only human creativity but diminish our powers altogether:

"I think there's definitely a risk we'll be diminished. One of the areas is replicability, machines can do some things better than us. The big fear also of course is influence. ***Once these machines become more accurate and they advise us, and know us better than we know ourselves, then how can we not want to follow them?*** *So there are real problems there."*

What are some of the challenges of developing AI ethics moving forward?

Leaders' lack of understanding about some of the basic frameworks of AI

One of the biggest challenges facing the conversation around AI and ethics raised almost unanimously across all table discussions – and faith perspectives - at the Conference was the lack of technical understanding leaders have around AI.

Jon Cruddas MP, explains how this is particularly relevant in a political context:

"I start with the honest acknowledgment that ***the nature of our modern challenges is almost inversely proportionate to our capacity as politicians to grasp most challenges****. It's uncomfortable for me to say that, as a member of this class of politicians but that seems to be to me where we're at.*

I think this question about technology is doubly problematic for politicians because of our lack of technical knowledge and expertise in the area and we rely on technologists more often than not to give us an account of what their software can do to us. And that means a lot of the software itself is disguised and we slip into a form of techno-solutions in terms of future', the idea that the problems can be «solved» using technologies, including some of the problems that technology causes."

Politics has lost its ethical grip, and the conversation has shrunken to forms of technocratic administration."

The domination of utilitarian thinking over other important 'Human Values'

Another key challenge of developing AI ethics is moving away from the commercially-minded, utilitarian thinking that often dominates discussions and instead focus on a more human-centric values and perspective.

Jon Cruddas MP explains this tension from a political perspective, outlining how public conversation and political conversation is dominated by utilitarian thinking at the expense of basic questions around human flourishing:

"As a politician, I'm always interested in the tempo of the political conversation. And in this conversation, especially, its symptomatic of the domination of a type of utilitarian thinking that crowds that public conversation. If you look at political philosophy, it's divided into three frameworks. The first one looks at how we can maximise human welfare, the second is how we can maximise human freedom, and the third about maximising human virtue. It seems to me the political contest is truncated so that the third dimension is crowded out in the public conversation. That's why it's so important faith communities are included in this conversation. Welfare economics in this space

is almost all-embracing. We have to start off with some basic questions about human flourishing and re-introduce those categories into the debate, almost foundationally, to alter the tempo of this conversation, in defence of humanity itself."

This notion of not measuring prosperity based on pure technical and economic advances is a point which resonates with H.E. Shaykh Abdallah bin Bayyah's belief that ideals and emphasis on peaceful coexistence are doubly important:

"The prosperity and stability of societies is not measured by technical and economic advances alone. Rather, it requires the ability to enrich the world with valuable content and noble ideals, and it needs a comprehensive view of life based on harmony, coexistence, and cooperation."

According to Prof John Wyatt, faith traditions may provide some answers on avoiding this technological future in which maximising human enjoyment is prioritized over human flourishing:

"There is a common form of thinking which is sometimes termed Techno-utopianism. It dreams of a frictionless future. Every desire, interest, and whim will be instantly satisfied by advanced technology and with the minimum of effort.
And if you do read science fiction, Iain M Banks wrote a series of novels called the 'Culture series' in which benign super-intelligent computer minds do the important serious stuff and humans can just enjoy themselves. But in this technological future there's no friction which is necessary for human beings to flourish.
It seems that in order to develop character, we need resilience, we need struggle, we need perseverance, we need frustration, we need suffering. Don't all our faith traditions say that? We need to think deeply and together to envision what kind of future will enable our children to flourish and become the people they were meant to be."

Advocating for marginalised communities

Dr Erin Green emphasises that one of the key obstacles to developing more inclusive AI regulation is the absence of marginalised communities' voices from the conversation:

"The best way to grasp this notion, is to ask, 'Who is missing? Who is not here?' Who is most affected by AI and faith? And in the context of this, of course, there's any number of groups, people and individuals who are not represented here, these kind of conversations are always privileged. But I'd like to point to just some of these different groups:

Women

Women ethicists and researchers are often the whistleblowers, they're the ones that point out the flaws in the methodologies, the flaws in the questions, and they're often the ones that are outed from Big Tech. These are often not just white women like me, they're women of colour, they're disabled women, queer women, women who exist in other marginalised spaces. So I think we always need to recall in these conversations, where the balance is there.

Indigenous people

A group really deserving of space is Indigenous communities. If you look to contexts like Aotearoa New Zealand, Indigenous people have proposed entire treaty systems for AI, which is completely different from the regulation mindset. Instead of a strict, top-down «these are the rules» mindset, the emphasis is instead on relationships, on community. The focus is on the collective benefit, reciprocity and the notion of guardianship, which is central to how we will proceed in considering the young children and generations to come who will live with the consequences of what we discuss today.

Refugees/migrants

And another group which is so often missing, and we saw this in the EU legislation which was passed yesterday [15 June 2023], is refugees and migrants. Conversations in Brussels are always centred around the concept of EU citizenship. It's always looking at the rights of EU citizens, never on refugees and migrants, those without these explicit rights. They are always left out. And when it comes to the needs of refugees, we look at basic needs like housing, shelter and education, but there are serious consequences for them which relate to the conversations about AI we have today.

Children

Where are the voices of people who cannot speak for themselves yet? Who is not represented in legislation? They will bear the consequences of the technology we deploy today."

How can our faiths inform our response to these challenges?

Since faith places such an emphasis on human creativity and our unique capacities as creatures made in the image of God, it has much to offer in terms of ensuring that AI develops along moral lines and in the interests of humanity as a whole. Rabbi Dr Bor, who believes that «religion can show us how to live in the face of AI", explains one of the ways faith may inform our response to these challenges is by shaping the values of those working in AI:

"Faith might shape the values of those working in AI. The moral values of those is bound to have a huge impact on AI. And I suspect that if those designing and programming AI have a strong moral compass and spiritual sensitivity, then we're more likely to have AI that is beneficial to humankind, rather than harmful."

He goes on to explain that another way faith may have a positive impact is by using our communities' collective influence to shape how people view AI, and in turn how it shapes:

"If as faith communities we can influence how individuals see AI, that will shape how AI develops. We are the consumers of AI, and we're also the products of AI, we give it our attention. And now we're being asked to give it our sense of intimacy and care. And if we resist that and show we're not willing to bargain on those terms, and that there are limits to what we can tolerate, then that can have a huge impact. As people of faith, I think we have a huge role to play there."

H.E. Shaykh Abdallah bin Bayyah also underlines how hadiths can provide guidance on approaching creating an ethical framework:

> *"The Prophet (peace be upon him) says, 'There should be no harm nor reciprocation of harm'. This principle is a governing principle because it suppresses the human desire for wealth under the pretext of development and discovery if this involves harm or reciprocation of harm to others. It also protects researchers and scholars from interference in their work as long as their work does not lead to harming others or reciprocating the harm from others."*

Is it too late?

So where does all this leave us? It can be tempting given the narratives around AI in the media to think that any action around curbing some of the risks of AI is simply "too late". While policy and regulation have historically tended to be reactive to AI development, Dr Nick Chatrath highlights that now is the time for action:

"I believe that humanity is now at an inflection point, given AI's growth. In the coming months and years, one thing that should not surprise us is the pace at which new AI tools surprise us. In the same way that many were surprised at what ChatGPT could do when it came out, we will be regularly surprised in the future. Surprise will be the new normal for many. Sadly, it may be the case that some earth-shaking event is required to make the confluence of policymakers, business leaders and others really put the right guard rails in place."

Similarly, Rabbi Dr Harris Bor warns that we are in an exciting but critical moment in the development of AI ethics as the values we set now maybe 'locked in' for years to come:

"I think faith can contribute to the values that are shaping AI. In his recent book, 'What We Owe the Future, William MacAskill', a philosopher who is associated with the effective altruism movement, came up with the notion of locking in value. He defines it as an event that causes a single value system to persist for an extremely long time. If value locking occurs globally, then how well or poorly the future goes will be determined significantly in part by the nature of those locked-in values. So this is an exciting moment where we must take stock of our collective values, allowing for diversity."

Despite this urgency, Dr Erin Green explains how her Christian faith helps her keep hope:

"I come from a theological background that frames things more in terms of justice and redemption. I firmly believe the train hasn't left the station. I'm obliged by my faith to live in hope. For me, that's a fundamental commitment I have as a Christian, that you can never say something can't be redeemed. So I certainly have moments where privately I feel that the train has left the station and I feel discouraged, especially when you're working in advocacy it's very hard to see the impact you're making. But, I think when you look at faith, I do feel obliged to commit myself to that hopefulness that the train hasn't passed."

What next?

One of the things that became evident during the conference in June, both through speeches and roundtable discussions was the need for faith communities to work together to ensure that AI is developed in an ethically conscious manner, with respect for fundamental rights and values. Faith communities have the ability to unite people from a variety of backgrounds, to listen to their concerns about AI, to bring the voices of the marginalised to the table, and ultimately cooperate to drive the development of AI standards that work for the good of humanity. Prof Wyatt highlights that one of the roles faith has to offer in the field of AI ethics is in mobilising people of faith working on AI development. He notes that there are already many people of faith working in tech industries, but that there is a need to mobilise them - perhaps through interfaith dialogue - so that they can "use their collective muscle" to influence how AI is being developed:

"On a positive note, I think that the power of the tech employees hasn't really yet been reckoned with. There are huge numbers of people of faith working in the tech industry. They want to believe in the work they're doing, that their products are products for good, products that promote human flourishing. They want to feel they've done something good for humanity. So I think we can mobilise and encourage people of faith within the technical sphere, the tech employees, to get together and to start to use their collective muscle.

Sadly, those individual employees who challenge the status quo are often outed and sacked. The Industrial Revolution taught us that combined worker power could be extremely powerful and beneficial. There is also an opportunity for people of faith to act as allies with other people of goodwill. I think it's very interesting in the American setting how Tristan Harris and the Centre for Humane Technology have been incredibly influential. And I wonder if we could get more cooperation between religious groups and those who are of a secular belief and are people of good faith. Because ***we are not disempowered, we have the possibility of influencing the direction this technology is going, but it will take significant working together."***

MoU Signings : Faraday AI
August 2023

On 1st August 2023, the Abu Dhabi Forum for Peace, represented by its Secretary General, H.E Sheikh Al Mahfoudh Bin Bayyah, entered into a Memorandum of Understanding (MOU) with The Faraday Institute for Science and Religion at Cambridge University. This collaboration follows the successful AI Faith and Society Consultation held in June, further cementing the Abu Dhabi Forum's leadership position in the realm of AI and Ethics. The MOU outlines several initiatives, including the development of additional programs, the creation of an AI, Faith, and Civil Society Commission, strategic knowledge exchange, and the publication of findings from the June event.

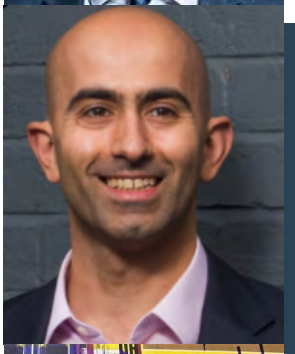
Appendix A – The faces behind the quotes: Speakers from the 2023 AI & Faith Conference

His Eminence Shaykh Abdallah Bin Bayyah
President, Abu Dhabi Forum for Peace, Abu Dhabi

A life-long advocate for peace, Shaykh Bin Bayyah is the President of the Abu Dhabi Forum for Peace and one of today's most well-respected Islamic scholars. He has published numerous writings and has been distinguished in a variety of ways for his work in Islam, interreligious dialogue and peace building.

Jon Cruddas MP
Member of Parliament

Jon Cruddas is the Labour MP for Dagenham and Rainham and is a Professor at the Jubilee Centre for Virtue in Public Life, University of Birmingham.



Dr Nick Chatrath

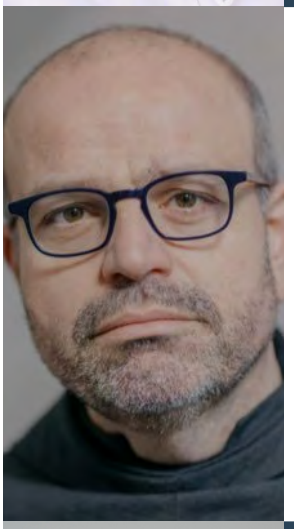
Dr Nick Chatrath is a former McKinsey consultant and leading international expert on the potential impact of Artificial Intelligence on human lives and society. In addition to co-founding two technology startups, he has recently authored a book on the topic, entitled *"The Threshold: Leading in the Age of AI"*.

Graham Budd, Executive Director, The Faraday Institute for Science and Religion

Graham Budd is the Director of The Faraday Institute for Science and Religion. He has over 30 years experience in computing and the global semiconductor industry, including his work leading the development of several of Arm's pioneering early single-chip computer designs.

Sheikh Al-Mahfoudh Bin Bayyah
Secretary General, Abu Dhabi Forum for Peace

Sheikh Al-Mahfoudh Bin Bayyah, is an influential diplomat. He played an instrumental role in establishing the flagship programs like the Marrakech Declaration, New Alliance of Virtue, and Charter for Inclusive Citizenship in the MENA region. He also contributed significantly to the field of countering violent extremism and supported the establishment of Zaytuna College, the first accredited Muslim liberal arts college in the US.



Fr. Paolo Benanti
Pontifical Gregorian University in Rome

Fr Paolo Benanti is an academic of the Third Order Regular of St Francis, Professor of Moral Theology, Bioethics, Neuroethics, and Ethics of Technologies at the Pontifical Gregorian University in Rome and a member of the Task Force on Artifical Intelligence of the Agenzia per d'Italia Digitale and the Prime Minister's Officer. He has a particular interest in technology ethics having recently published, '*The urgency of an algorethics*' (2023).

Rabbi Dr Harris Bor
London School of Jewish Studies (LSJS)

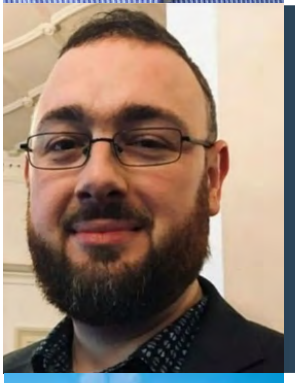
Dr Harris Bor is a a commercial barrister, Research Fellow and a Lecturer at the London School of Jewish Studies (LSJS), and rabbinic scholar with the Montefiore Endowment. He is also the author of "Staying Human - a Jewish theology for the age of Artificial Intelligence", and an advisor to the US think-tank, *AI and Faith*.

Dr. Yaqub Chaudhary, Cambridge Muslim College

Dr Yaqub Chaudhary is a Research Fellow in Science and Religion with an interest in AI and cognitive science in connection with Islamic conceptions of the mind and intelligence. He is also a founding member of *AI and Faith*.

Prof John Wyatt
University College London

Professor John Wyatt is Professor of Ethics and Perinatology at University College London. He has a particular interest investigating the implications for human self-understanding of recent advances in artificial intelligence and robotic technology, and co-authored *The Robot will see you now – AI and the Christian Faith*.



Dr Ramon Harvey
Cambridge Muslim College

Dr Ramon Harvey is a lecturer in Islamic Studies at Cambridge Muslim College and leads the John Templeton Foundation funded research project 'Beyond Foundationalism: New Horizons in Muslim Analytic Theology'.

Dr Erin Green
Pax Christi International

Dr Erin Green is the Head of Communications and Creativity at Pax Christi International. She is also a theologian and has particular expertise in looking at the social impact of AI and related technologies and has undertook extensive novel research into the social impact of robotics and AI from a theological perspective.

Lord Wei
Member of the House of Lords

Lord Wei is a member of the House of Lords, having previously had a successful career in the consulting and social sector. He has a particular interest in equipping the next generation with the tools and understanding to transform society in today's changing world.

Zeshan Zafar
Chair, Forum for Peace - London

Zeshan Zafar, Executive Director of Abu Dhabi Forum For Peace & Chair of UK Forum For Peace-London, unites diverse individuals to foster dialogue and collaboration. Welsh-born, he empowers the British Muslim community, spearheads international initiatives like annual assemblies, Marrakesh Declaration, & the Charter for a new Alliance of Virtue.

AND THE
OF SOCIETY
JUNE 2023 | LONDON

The Faraday Institute for Science and Religion

www.adpeaceforum.com
www.faraday.cam.ac.uk

adpeaceforum
TheFaradayInstitute

adpeaceforum
TheFaradayInst