



Ollscoil Chathair  
Bhaile Átha Cliath  
Dublin City University

## Meeting on the Ethics of AI

Centre for Religion, Human Values,  
and International Relations

Mansion House, Dublin  
19th April 2024



**Meeting on the Ethics of AI  
April 2024**

**Centre for Religion, Human Values, and International Relations**





# Contents

SUMMARY	2
WORDS OF WELCOME	4
SCENE-SETTING ADDRESS	5
Response: Onesto Consulting	9
PANEL DISCUSSION ONE: THE PURPOSES OF THE AI ACT, AI AND SECURITY, AI AND THE INFORMATION ENVIRONMENT, AI AND STRUCTURAL BIAS	10
INTRODUCTION TO THE INSIGHT SCIENCE FOUNDATION IRELAND CENTRE FOR DATA ANALYTICS	17
MENTIMETER POLL	18
KEYNOTE PRESENTATIONS	19
PANEL DISCUSSION TWO: AI AND THE WORLD OF WORK, WITH REFERENCE TO EMPLOYMENT, PRODUCTIVITY, EDUCATION, AND EQUALITY	22
ROUND TABLES	25
Round Table 1: AI and education	25
Round Table 2: AI and the world of work	26
Round Table 3: AI and the information environment	28
Round tables 1, 2, and 3: AI and the role of faith communities	30
FINAL PLENARY	32
CONCLUSIONS	33
A: High-level conclusions (reflecting a broad consensus)	33
B: Examples of practical steps	34
ANNEX	35
Annex 1 - Programme	35
Annex 2 – List of participants	37
Annex 3 – A note on the politics of AI	39
However, late on in the article, the authors create an all-important space for dialogue when they recognise that the use of AI in warfare ‘opens a Pandora’s box’ of ethical and legal issues:	42
Annex 4: Case Studies	48



## SUMMARY

Under the auspices of the ‘European Future Talks,’ and as part of the follow-up to the Conference on the Future of Europe (CoFoE) in 2021/2022, officials of the European Union promoted consultations in 2023/2024 on major current issues with the participation of representative voices from the churches and faith communities. This exercise has been carried out in the spirit of Article 17, Treaty on the Functioning of the European Union. Reports from a number of such multi-stakeholder consultations will be shared at a joint meeting in Brussels on 20th November 2024, when the newly elected European Parliament and new Commission are in place. The Centre for Religion, Human Values, and International Relations at Dublin City University was asked to take responsibility for the topic ‘AI and its ethical implications,’ against the background of the coming into force of the European Union’s AI Act. Meetings on other topics were held in other EU member States and in Britain. The Centre was supported in preparing and guiding the meeting by Onesto Consulting.

The purposes of the meeting were as follows:

- to involve different sectors in dialogue on the significance of AI
- to conduct this dialogue in the light of high-level values
- to identify opportunities and challenges presented by AI
- to envision future frameworks for developing the governance of AI
- to support the European Union’s global leadership role in regulating AI

The underlying premise of the initiative is that it is useful for a diverse group of ‘social friends’ to arrive at some tentative conclusions, or to ‘disagree better,’ on a difficult topic such as AI. At this meeting on 19th April, the aim was to produce a consensus-based document as a contribution to public debate. This document (the present document) was to be ready well in advance of the Brussels meeting on 20th November.

The Lord Mayor of Dublin Daithí de Róiste gave us the use of the Oak Room at the Mansion House for our meeting, which took place over a full day on Friday 19th April 2024 with more than 60 in-person participants including distinguished international speakers; nominees of churches and faith communities; representatives of government departments and EU institutions; leading academic experts on AI in Ireland/Northern Ireland; civil society actors, including the European Movement, the National Economic and Social Council (NESC), and the Think-tank on Action for Social Change (TASC); and prominent figures from the cultural world, business, and law. The event was hybrid, allowing other attendees to participate remotely. The day consisted of a combination of keynote presentations, interactive panel discussions, breakout sessions, and real time polling activities using a mobile device (Mentimeter).

Seán Ó Fearghaíl, the Ceann Comhairle (Speaker of the Lower House of Parliament), made a keynote speech. A video message of support from the First Vice-President of the European Parliament, Dr. Othmar Karas, was played at the



beginning of the day. Axel Voss, MEP, rapporteur for the AI Act, contributed to the deliberations on-line.

The programme for the day and a full list of participants are provided in Annex 1 and Annex 2 to this report.

In addition to the ten high-level conclusions, the report offers eight examples of practical steps that can be taken here and now by the European Union and other public authorities, by individual businesses, or by the leading corporations engaged in the development of AI. A note on the politics of AI inspired by the day's discussions is at Annex 3. Annex 4 offers some practical examples of the use of AI and the difficult discernments that arise.

The present report finishes with ten high-level conclusions:

- i. The rapid development of AI has profound social and political implications at the global level. We cannot presume a priori that AI will make a beneficial contribution to the future of humanity and serve the cause of fraternity, freedom, and peace.
- ii. AI-based solutions make sense in a wide range of practical situations, including in science, medicine, and the workplace. Research should be encouraged by public authorities. Nevertheless, in deciding on the deployment of AI, a precautionary principle should apply.
- iii. A commitment by individuals to respect general principles in developing and deploying AI will not make the future secure. We need substantive regulation to provide innovators with the clarity they need. We should embed ethics in the design of systems, as well as in applications.
- iv. We also need to search for an overall vision that answers the question, 'What kind of reality do we want our children to live in?'
- v. AI should not reinforce inequality. The idea that the 'maximisation of shareholder value' is justified by collateral social benefits does not seem adequate as a guiding principle. 'Western' societies need to take specific steps to counteract polarisation and the loss of trust in institutions.
- vi. Proportionality in the allocation of resources in the light of a good that is common to all is a core democratic value that needs to receive greater attention in discussions around AI.
- vii. The military applications of AI have given rise to a dangerous inter-state competition in which previously accepted ethical parameters are set aside.
- viii. AI raises fundamental questions for the future of education and role of teachers. It is essential to maintain a balance between 'computational skills' and 'employability', and less quantifiable and more important human attributes such as the religious and historical imagination, ecological awareness, personal empathy and solidarity, creativity, and the ability to engage in dialogue and to persuade.
- ix. A 'holistic' framework of engagement at the global level to address AI can be achieved in the context of a well-designed post-2030 development agenda.
- x. Article 17, Treaty on the Functioning of the European Union, has the potential to serve as a 'space of shared projection' within which to enable a useful preparatory dialogue. If this works well, it can inspire a similar dialogue in other jurisdictions.

## WORDS OF WELCOME

The day began with words of welcome from **Lord Mayor Daithí De Róiste**; **Christian Gsodam**, (European External Action Service), Founder of the European Future Talks; **Daire Keogh**, the President of Dublin City University; and **Othmar Karas**, First Vice-President of the European Parliament (video message).

The opening speakers endorsed the inclusion of church and faith communities in the dialogue about the future. Another common element in their presentations was their emphasis on social cohesion and trust among citizens as key values to be upheld in a pluralist society. Dr. Gsodam focussed on the epochal nature of the social transformation that is currently underway. Dr. Karas noted the significance in global terms of the EU AI Act.

Dr. Gsodam briefed the gathering on the series of meetings that are taking place under the auspices of the European Future Talks. These include the following:

- Munich: social policy
- Vienna: migration
- Oxford: climate
- Cambridge: rule of law
- Dublin: ethics of AI
- Rome: peace and conflict





## SCENE–SETTING ADDRESS

**Dr. Jovan Kurbalija**, Executive Director of DiploFoundation (Geneva), brings an unparalleled depth of experience in the sphere of Internet governance and the impact of new technologies on politics and diplomacy. In his scene-setting address, Dr. Kurbalija anchored current AI developments within the broader history of ideas; discussed four different ‘layers’ where modern AI governance unfolds; and reflected on the temporal aspects of AI risks.

### **The Axial Age**

Beginning in the 8th century BCE in many different geographies, in the so-called ‘Axial Age’ (Karl Jaspers), there emerged social, political, and juridical spaces in which traditional ways of doing things could be examined critically, and new conventions could be established. The principle of verification produced a civilisational shift in terms of political transparency and accountability. New belief systems collectively forged a societal ‘operating system’ that continues to influence the present.

### **The European Enlightenment**

The Renaissance and Enlightenment serve as key knowledge bridges between the Axial Age and modernity. Key thinkers of the 18<sup>th</sup> century placed human rationality (as they understood rationality) at the centre of their reflections on society. Other factors in play included a scientific revolution, rapid technological and industrial change, the emergence of powerful nation States, a more unified global system increasingly controlled by Europe and the ‘West’, and a sense of the inevitability of progress. On the other hand, many thinkers of this period questioned a simplistic understanding of rationality and reintroduced the importance of emotions, faith, and human authenticity. Concerns about AI’s potential for technological dehumanisation can be traced back to the writings and debates of this period.

### **Vienna in the 20<sup>th</sup> century: no such thing as a private language**

Among the many thinkers of the first half of the 20<sup>th</sup> century, five stand out as particularly relevant to the AI era: Ludwig Von Mises (relevance of free choices), Joseph Schumpeter (creative destruction), Friedrich Hayek (power of knowledge), Sigmund Freud (psychological underpinnings), and Ludwig Wittgenstein (language, causation, and correlation). For Wittgenstein and others, truth and meaning are possible only on the basis of a shared language. This insight demands of us a conception of the human person that includes relationship (‘I-Thou’, ‘I am because you are’) and an understanding of psychology that undermines the picture of ‘economic man’ operating with perfect rationality on the basis of full information.

### **From the crisis of modernity to a new Axial Age**

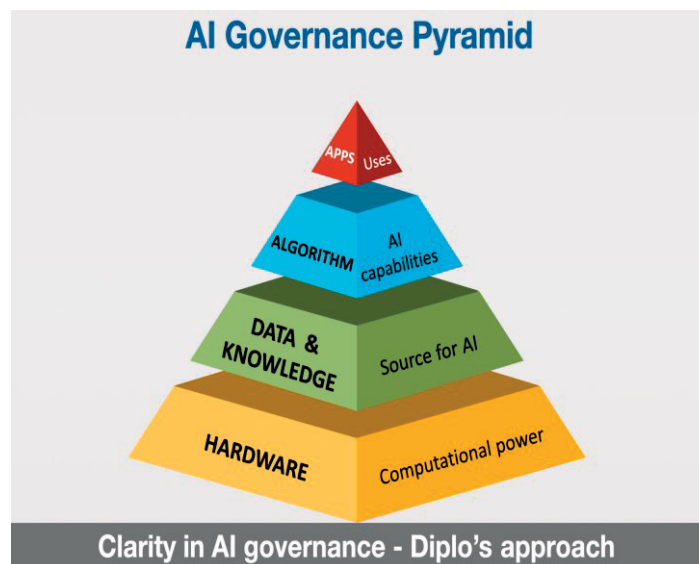
Dr. Kurbalija argued that without understanding the history of ideas, we cannot fully embrace the ethical and other dilemmas of the AI era. In the current ‘crisis of modernity,’ looking backwards can help us search for a new narrative – a

civilizational shift or change of mindset that could be described as an 'AI Axial Age.' In this perspective, AI's enormous potential is emerging at a watershed moment in human history. Can AI inspire a renewed sense of purpose and belonging among all citizens and a new narrative of community and sustainability? Should modern society introduce a new right to be 'humanly imperfect,' shielding us from AI-driven optimization and dehumanisation?

The ethical criteria currently applied to the development of AI do not seem to capture adequately the growing demand in many quarters for a unifying sense of purpose. A sense of purpose depends on prioritizing and proportionality. OpenAI is reportedly seeking to raise US\$7 trillion for chip production; in 2023, the entire GDP of Africa was US\$3 trillion. The principles applied to AI often appear to take it for granted that the pursuit of optimal or maximally efficient solutions within a profit-based economic model is acceptable, subject to transparency and other limiting factors. But the narrow context in which commercial decisions are taken, and the lack of contact with stated priorities at the global level, can be considered a legitimate concern in itself.

### The AI governance pyramid: mapping of how and where AI should be governed

Dr. Kurbalija presented an AI governance pyramid, which consists of four main layers where most current initiatives, laws, and discussions can be located: hardware (computational power), data and knowledge (source of AI), algorithms (AI capabilities), and applications (uses).



### Hardware: computational power

In the ongoing 'AI race,' access to computing resources is considered a determining factor for the success of AI companies. These resources take the form of AI chips – specialised integrated circuits designed to handle the complex computational requirements of AI algorithms at high speed and efficiency. The AI chip ecosystem is complex yet concentrated, dominated by three key actors: US-based Nvidia (chip designer), Netherlands-based ASML (equipment manufacturer for chip production), and Taiwan-based TSMC (chip manufacturer). Geopolitical dynamics further influence this ecosystem's complexity. In the race for computational power, the USA and China are attempting to limit each other's access to AI chips (e.g., through export controls and sanctions), while other actors like the EU are trying to strengthen their own capabilities (the EU's Chips Act is an example).





### **Data and knowledge: sources for AI models**

Data is where AI derives its primary inputs. Generative AI models, for instance, rely on vast training datasets encompassing personal information, articles, papers, books, and more. However, the public knows little about what exactly goes into an AI model, leading to increasing calls for clarity on what – and whose – data and knowledge is used by developers. Privacy concerns, such as the extent to which personal information is used in AI training, and intellectual property protection are critical issues. Significant legal challenges have arisen as copyright holders contest the unauthorised use of their work in AI development. In response, some companies have begun negotiating licensing agreements, such as Apple’s discussions with publishers in late 2023.

### **Algorithms: controlling AI capabilities**

Regulating AI algorithms has become a prominent topic, particularly among those concerned about AI’s long-term risks to humanity. Governance at this level advocates for placing guardrails around developing advanced AI models to mitigate future ‘unknown’ AI risks. However, how such guardrails might look in practice remains a matter of debate, with concerns that stringent algorithm regulation could stifle AI innovation, including open-source solutions. On a more practical level, transparency and evaluation frequently feature in discussions about governing AI algorithms. There is currently very little transparency regarding AI models; for instance, we know little about the data fed into models or the weights assigned to parameters. Transparency is a prerequisite for evaluation and, ultimately, accountability in the digital realm.

### **Applications: regulating uses of AI**

AI governance on the application level follows the dominant practice of technological governance and focuses on the implications of system outputs regarding human rights, security, and consumer protection rather than regulating the algorithms themselves. As with traditional digital systems, responsibility and liability would be assigned to actors across the AI lifecycle (developers, deployers, and users of AI systems).

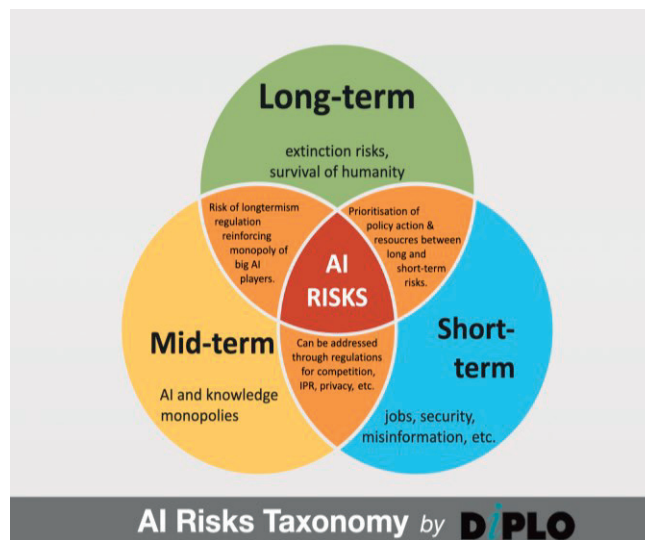
### **Where to govern AI?**

A shift toward AI regulation at the algorithm level (thus delving into the technology’s inner workings and ultimate capacity) would represent a significant departure from this established approach, with far-reaching consequences for technological progress. Deciding whether to govern AI at the level of computing power, data, algorithms, or uses will have profound implications for the future of AI and society.

## The temporal dimension of AI governance risk

Dr. Kurbalija distinguished between short-term, medium-term, and long-term risks.

**Short-term risks** include job losses, data and intellectual property-related issues, loss of human agency, mass generation of fake content, misuse of AI in education, and new cybersecurity threats. While familiar, these risks require more concerted efforts, often utilising existing regulatory tools.



**Medium-term risks** are those we can anticipate without being sure of their severity. Imagine a future where a few big companies control all AI knowledge, just as they currently control people's data. This concentration of power could lead to certain companies dominating business, life, and politics – a scenario reminiscent of George Orwell's dystopian visions. If nothing changes, we could face this reality in 5 to 10 years. Policy and regulatory tools like antitrust and competition regulation, as well as data and intellectual property protection, could help mitigate these risks.

**Long-term risks** are the 'unknown unknowns' – the existential threats where AI could evolve from servant to master, jeopardising humanity's survival. These threats dominate the global narrative with an intensity similar to nuclear armageddon, pandemics, or climate cataclysms. Addressing long-term risks is a major governance challenge due to the uncertainty of AI developments and their interplay with short- and medium-term risks.

## Response: Onesto Consulting

**Ashwini Mathur**, responding to Johan Kurbalija on behalf of the organisers, referred to the ‘ethics of not doing,’ a formulation that seemed to converge with the proposed ‘right to be humanly imperfect.’ Ashwini drew attention to a number of key questions in relation to AI:

- the importance of a holistic framework in measuring the impact of AI
- the importance of hope, a broad understanding of the scope of reason, and a sense of the sacred
- the need to understand the algorithms behind the analysis or generation of data
- the risks associated with using personalised data as the basis of the insurance industry
- the need for principles that will be acceptable across different geographies





## PANEL DISCUSSION ONE: THE PURPOSES OF THE AI ACT, AI AND SECURITY, AI AND THE INFORMATION ENVIRONMENT, AI AND STRUCTURAL BIAS

The first panel discussion was chaired by **Larry O'Connell**, Director, National Economic and Social Council, with the following participants:

- i. **Axel Voss**, MEP (online), rapporteur, AI Act
- ii. **Christian Gsodam**, EEAS Advisor for Strategic Communication/Foresight
- iii. **Abeba Birhane**, Adjunct Associate Professor, Trinity College Dublin, member of the AI Advisory Council; and member of the UN Secretary-General's High Level Advisory Body on AI
- iv. **Jane Suiter**, Professor in the School of Communications, Dublin City University
- v. **Catherine Prasifka**, Writer-in-Residence, Trinity College
- vi. **Archbishop Michael Jackson**, Chair, Dublin City Interfaith Forum

Comments from the floor were led by Professor William O'Connor, University of Limerick and member of the AI Advisory Council; Professor Stephen Williams, Queen's University Belfast; and David Donoghue. As Ireland's ambassador at the UN, David, with his Kenyan colleague Ambassador Machiara Kamau, co-facilitated the negotiation of the Sustainable Development Goals (SDGs) in 2014 – 2015. The discussion is summarised under a series of headings.

### EU AI Act


The EU AI Act is based on framework identifying four levels of risk:

1. Unacceptable Risk: AI systems posing a clear threat to people's safety, livelihoods, and rights will be banned.
2. High Risk: Critical infrastructures, employment, essential private and public services, law enforcement, management of migration, asylum, and border control, and administration of justice and democratic processes.
3. Limited Risk: AI systems will have to comply with specific transparency obligations.
4. Minimal or No Risk: The vast majority of AI systems fall into this category and can be developed and used freely.

The AI Act does not exist in isolation but is part of a broader EU digital strategy that includes for example, the General Data Protection Regulation (GDPR) and possible future legislation in the area of digital surveillance. The EU's approach to AI regulation fits in a broader context of attempting to balance innovation with ethical standards and human rights.

### Enforcement

The European Union will need to create a mechanism for enforcement of the Act effectively across all member states, taking into account the global nature of digital technologies and companies. Axel Voss emphasised the role of the



Commission in this regard. At the same time, the EU will need to foster international collaboration and work with other countries to develop global standards for AI. The recruitment of qualified experts by public authorities will be a key challenge. As of August 2024, the Commission is in the process of filling 140 full-time positions in the new AI Office in Brussels.

### **Harvesting of data**

Access to data poses a number of issues such as the need to allow for competition and questions of copyright. A more radical critique was that the harvesting of data on a mass scale, often without any form of consent, raises issues in itself. The assumption that everything is 'out there for us to take' is in tension with the historical concept of 'the commons' (Vinoth Ramachandra).

Is the privacy of data a matter of property or of dignity?

### **Definitions**

The EU will need to refine definitions so as to clearly define where the different levels of risk apply and to ensure that the criteria for these classifications are transparent and adaptable. Some issues are clear. For example, 'social scoring' such as practised in China is unacceptable. Less evident, in the category of 'unacceptable risk,' is what is meant by 'AI systems posing a clear threat to people's livelihoods.' Are there models, systems, applications or methodologies that cut across categories of impact in ways we are not accustomed to thinking about? The multiplication of financial transactions based on AI impacts on society in ways that should be measured and evaluated. (On the overall definition of AI, see Annex 3.)

### **Transparency obligations**

Algorithms designed to solve complicated problems are so sophisticated that it can become difficult for programmers themselves to understand exactly how they arrive at their results. This tendency is likely to accelerate considerably with the introduction of quantum computers that will operate not with binary circuits (semiconductors or microchips) but according to the highly complex laws of quantum physics.

### **AI and warfare**

The current legislation takes into consideration the security implications of AI, for example in connection with cyberattacks. Clearly, Government Departments, businesses, and other institutions need to introduce new security measures and to raise awareness about criminal activities on-line.

On the other hand, the EU AI Act appears to sidestep the increasing use of AI in warfare and the salience of dual-use technologies. There appears to be no agreed definition of 'Lethal Autonomous Weapons Systems' (LAWS) which can allow for pre-programmed machines to 'choose' to take the lives of human beings. Similarly, AI is used to 'generate targets,' using statistical information to identify suspected individuals and perhaps also to determine their 'value' as targets and to decide how many people it would be legitimate to kill along with the prime 'target'.



### **The 'national interest' versus ethics**


In 2021, the British and German Ministries of Defence published the paper 'On Human Augmentation – the Dawn of a New Paradigm' (Development, Concepts, and Doctrine Centre, MOD, and Bundeswehr Office for Defence Planning, 13 May 2021). This paper invites us to 'conceptualise the human as a platform.' This platform can be enhanced for command and combat purposes by genetic engineering, 'powered exoskeletons,' brain interfaces, 'cross reality,' and many other innovations. The authors argue that the information revolution is 'accelerating the speed and scale of moral change as different behaviours and attitudes become normalized through exposure.' In the light of 'moral change,' the paper concludes that the 'imperative' to use human augmentation in warfare is likely to lead to decisions by governments based on 'national interests' and not 'dictated by any explicit ethical argument.' 'The winners of future wars' will be those who use AI in association with other technologies to 'integrate the capabilities of people and machines.' Clearly, international rivalries are tempting some EU and neighbouring states to allow 'national security,' as traditionally defined, to prevail over initial ethical objections.

The British/German paper mentions the role of private sector corporations in driving military innovation. It is reported elsewhere that a German-based defence technology start-up has more than tripled its market valuation over the last year to an estimated \$US 5 billion and that the three leading European defence contractors will see their cash flow jump by more than 40% in 2026 as compared to 2021. In such a scenario, what are the respective responsibilities of governments, regulators, investors, employees, military personnel, and taxpayers in relation to Lethal Autonomous Weapons Systems, human augmentation, and the 'generation' and 'acquisition' of 'targets' through AI systems?

### **The need for a global vision**

It was noted that AI interfaces with 'grand challenges' in relation to the fracturing of global politics, climate tipping points, the loss of biodiversity, the spread of conflict and so-called 'grey zone warfare,' transnational organised crime, migration, the likelihood of another pandemic, the politics surrounding rare earth materials, and economic disparities that continue to intensify. A question put by one of the youngest participants in our meeting expresses very well what was also the central question posed by Dr. Kurbalija's introductory presentation: 'In what kind of reality do we want to live? Will AI reinforce dangerous trends?'

The UN sustainable development goals (SDGs) are the closest thing we have to a 'holistic' global vision or common medium-term plan for humanity which avoids a North-South split. The SDGs imply an interdisciplinary approach to policy in the light of clear social objectives, including conserving and sustainably using our planet's marine and terrestrial resources, promoting sustainable lifestyles, and reversing the degradation of ecosystems. As we review the SDGs and consider the content of a post-2030 development agenda, there may be scope for a global negotiation, analogous to the SDG negotiations on 2014/2015, in which the impact of AI on all the above-mentioned issues is on the agenda, and civil society, faith communities, and private sector companies are involved. The draft concluding document of the UN Summit of the Future (September 2024) envisages that negotiations on the post-2030 development agenda will start in 2027.



In the context of the SDGs, attention was drawn to the potentially very significant environmental impacts of data-centres and chip production. These impacts relate to water and energy use, as well as the competition to acquire rare earth metals.

### **The information environment**

Much of the discussion focussed on the ‘systemic risk’ to the electoral process and ultimately to democracy posed by certain practices enabled by AI. The clearest example is the spread of ‘fake news’ or emotive messaging by actors intent on manipulating the political process, including external actors. This growing phenomenon is the subject of guidelines published by the EU in March 2024. Even apart from deliberate interference with the political debate, the sheer scale and spread of information and opinions, combined with a relative absence of occasions for reflection, can undermine the public sphere. So-called ‘generative artificial intelligence’ rearranges and recycles existing content, reinforcing a dominant narrative or orthodoxy. This said, it was acknowledged that the phenomena of ‘social bubbles’ and the loss of trust in institutions were observable well before the advent of AI.

### **Mental health, wellbeing, and recommender algorithms**


A major theme of the panel discussion was the impact of the AI-assisted social media on the psychological wellbeing of children, adolescents, and young people. New apps teaching ‘financial literacy’ to children enable companies to track relationships within families. Among teenagers, individual behaviour is ‘tracked, studied, and sold’ by companies. The companies have no financial incentive to stop. On the contrary, their usual business model is based on promoting ‘engagement’. This means measuring the time spent on-line by users, assessing the predictability of their attention, and identifying cohorts of regular users for the purposes of advertising.

This ‘monetising of attention’ is accompanied by the deployment of ‘recommender algorithms’ which push new content towards users on the basis of previous choices. In transmitting ‘social videos’ according to this pattern, it is standard practice to intensify the emotional impact along the way. Negative emotions such as anger, or a sense of helplessness, can strengthen the ‘engagement’ of users across time.

Three days before our meeting, RTE’s Prime Time documented the use of ‘recommender algorithms’ on the social media accounts of Irish teenage girls. The evidence is that young people were drawn into forms of ‘engagement’ in which serious self-harm was being normalised. Some commentators do not hesitate to use the word ‘addiction’ as a risk in this connection.

### **Editorial responsibility versus content moderation**

A series of eight podcasts on BBC Radio 4, ‘The Gatekeepers,’ was broadcast not long before our meeting. ‘The Gatekeepers’ underlines the foundational role of ‘Section 230’ in shaping our current on-line culture. In the 1990s, at a time of widespread deregulation in the US, ‘Section 230’ was included in the misleadingly entitled Communications Decency Act 1996 to provide immunity for online service-providers with respect to third-party content generated by its



users. Section 230 states that ‘no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.’ This formulation was developed in order to stop in its tracks an emerging argument in litigation that service providers should be treated as publishers and not just as ‘distributors of content.’ Section 230 was fundamental in enabling the commercial strategy of the major companies.

In recognition of the risks implicit in the legal framework created by Section 230, big name companies have accepted in the meantime a number of obligations in respect of ‘content moderation,’ mostly aimed at ‘taking down’ offensive material soon after it appears. However, the research undertaken by the programme-makers demonstrates that the measures in place are very often ineffective, for several reasons. The definition of what is unacceptable may fail to capture and restrain some of the major risk factors. The standard of ‘content moderation’ that has been achieved in respect of English language content is not maintained in respect of other languages. The sheer scale of messaging that is now possible (generative AI, chatbots) poses a challenge in itself, as it does in other sectors including law enforcement. Overall, it can safely be concluded on the basis of ‘The Gatekeepers’ that serious instances of self-harm, violence, and social conflict are attributable to the absence of accountability on social media platforms.

### **Structural bias**

It was forcefully argued by two panellists in particular that generative artificial intelligence is open to bias because it operates by searching big data for information and reassembling this material in the format required. Statistical factors are centrally important. The more a notion is repeated on the worldwide web, the more AI takes it into account, without having an overall capacity to sift for errors and preconceptions. In this sense AI is ‘reinforcing’; it runs the risk of legitimising ‘fake news’ and/or strengthening a prevailing narrative or dominant orthodoxy at the expense of communities at the margins of society – and to the detriment of original research and reflection.

This tendency towards introducing structural bias into the public sphere is exacerbated by other factors. First, and most obviously, the design of systems may reflect the unexamined assumptions of the designers, who for the time being are overwhelmingly in the ‘global north.’ For example, facial recognition products have failed to work for those with darker skin tones. The impact of the global north on the design of systems is magnified by the workings of generative AI, as mentioned above. AI may serve, even inadvertently, to prevent cross-cultural encounter and in particular the encounter with indigenous peoples which is increasingly recognised as an important course correction as a ‘globalised’ world seeks a new path.

Second, and also of great significance, is the impact of AI on our ways of thinking – the ‘optimisation’ of functionally useful knowledge – such as engineering skills – at the expense of emotional intelligence and our ability to explore the resonance of great fundamental words such as ‘love’ and ‘hope’.



Third, there is growing evidence that AI is contributing to a loss of quality in academic research as universities reward researchers for the volume and regularity of their published work, and time-strapped researchers turn to AI to help manage the sheer quantity of on-line material with which they are expected to be familiar (Katy Hayward).

Transparency around the purposing or tooling of algorithms, important as that is, may not be enough to contain the risks identified here.

**‘The limits of my language mean the limits of my world’**

One speaker suggested that in the presence of the inevitable uncertainties of adolescence, the social media are increasingly offering young people a range of premature, readymade explanations, thereby pre-empting the deeply personal process of exploration and discovery. In a second phase, these pre-packaged conclusions or diagnoses encourage a view of reality as a ‘network of data-points’ to which possible responses (‘sad’/‘happy’) are predefined. A sense of the complexity of reality is lost. One might contrast the ‘normal’ of the social media with the vicarious experience provided by lifelong engagement with a great work of literature. In a third phase, security services or employers may use



the facial recognition techniques of ‘affective AI’ to categorise individuals. Over time, a largely functional approach to defining one’s own disposition and that of others undermines the cultural conditions on which politics depends, including interpersonal communication and a capacity for shared discernment.

At stake throughout this process is what it means to be a person.



## Healthcare

There was no disagreement with the proposition that the adoption of AI in healthcare can streamline operational efficiency and contribute in practice to better decision-making and better patient outcomes. The challenge here is to identify the appropriate 'metrics', the competencies that are needed, and rules related to accountability and ethics. Society has a strong interest in improving clinical decision-making, screening, record-keeping, and research based on datasets.

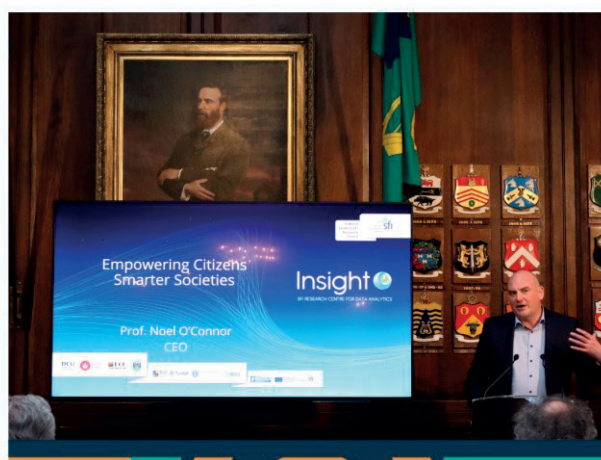
On the other hand, biases in decision-making by LLMs (Large Language Models) could, in principle, affect diagnostic accuracy. 'Human augmentation' – referring to many different interventions including prosthetics, genetic engineering, neurotechnology, and performance-altering drugs – is not to be endorsed without qualification. For example, in the sphere of gene-editing, human germline research is subject to careful scrutiny because of potential impacts across generations. At a meeting held in Beijing in March 2024, leading American, British, and Chinese experts issued a statement saying that a joint approach to AI safety is now needed to stop 'catastrophic or even existential risks to humanity within our lifetimes.'

# INTRODUCTION TO THE INSIGHT SCIENCE FOUNDATION IRELAND CENTRE FOR DATA ANALYTICS

**Professor Noel O'Connor** of DCU, CEO of the Insight SFI Research Centre for Data Analytics, Ireland's largest SFI-funded research centre, gave a brief but comprehensive presentation on the work of Insight. Insight is one of the largest data analytics centres in Europe. It seeks to derive value from Big Data and provide innovative technology solutions for industry and society. The Centre supports 450 researchers across areas such as the Fundamentals of Data Science, Sensing and Actuation, Scaling Algorithms, Model Building, Multi Modal Analysis, Data Engineering and Governance, Decision Making and Trustworthy AI.

The core science involved covers such areas as:

- Applied Mathematics
- Case-based Reasoning
- Computer Vision
- Distributed Ledger Technology
- Human and Societal Factors
- Lifelogging
- Enterprise Knowledge Graphs
- Natural Language Processing
- Open Data
- Privacy and Security
- Scheduling
- Signal Processing



Professor O'Connor instanced a number of areas in which the wide-ranging work of Insight helps to ensure that AI empowers citizens. These areas include Constraint Programming, Data and AI Ethics, Distributed Systems, Multimedia Analysis, Network Science, Optimisation, Recommender Systems, Trustworthy AI, and XAI (explainable artificial intelligence allowing human users to comprehend and trust the results and output created by machine learning algorithms).

Professor O'Connor's presentation also dwelt on the 'positioning' of Insight in relation to research and business partners at home and abroad.



## MENTIMETER POLL

With the assistance of convenor and facilitator Chris Chapman, we sought initial feedback from participants at the end of the morning, using the interactive software Mentimeter. Participants were invited to respond briefly to the question 'What are the most important things we have yet to address?' The overwhelming majority of the responses focussed on the global political implications of AI. Here are some examples of the responses we received:

1. Will the rest of the world follow the EU AI Act regulations?
2. What is it to be human? AI as a getting out of jail card for the mess we've made of the planet
3. Using AI to deliver zero hunger and a 1.5C temperature rise
4. The place of the Global South in the debate on the ethics of AI
5. AI and international relations (China, US, etc)
6. The effect on the human brain and our ability to think reason, and decide...
7. LLMs owned by big tech leading to knowledge slavery
8. The decoupling of AI from profit-driven corporations ... should AI be nationalised?
9. Practical usage and application... to advance social and economic challenges
10. Professional standards for computer engineers
11. Use of AI by malevolent actors/States for military purposes or population surveillance
12. The moral framework that is being encoded in algorithms is based on the subconscious ethics and motives of companies

## KEYNOTE PRESENTATIONS

By **Seán Ó Feargháil**, Ceann Comhairle (Speaker) Of Dáil Éireann;  
**Anja Kaspersen** (on-line), Carnegie Council for Ethics in International Affairs and Special Advisor at the Institute of Electrical and Electronics Engineers (IEEE); and  
**Vincent Depaigne** (on-line), DG Justice (Just), European Commission

A summary of this session is arranged thematically below.

### **Parliamentary diplomacy / the role of churches and faith communities**

Parliaments, including the European Parliament, can play an important role in intercultural dialogue and diplomacy. A parliament reflects the make-up of society. Parliamentarians have more freedom to engage in exploratory dialogue than the representatives of governments. There is no such thing as 'one-sided pluralism.' Long-term dialogue about the future must be inclusive.



Politics and democracy depend on trust. The absence of trust, combined with other factors including the very rapid development of AI, poses a risk to politics. According to some experienced commentators (Geoffrey Hinton), this risk may be increasing exponentially. The challenge we face is to interpret and apply our high-level values in a world that is changing rapidly and faces many 'existential' questions. There is every reason to include churches and faith communities in dialogue, as happens in the European Union under Article 17, TFEU. Philosophical organisations, representing those whose worldview is not expressed in religious terms, are rightly a part of such a values-led dialogue. There is important work to be done, including at parliamentary level, on the concepts and organisational principles that can encourage a mutually beneficial engagement by political leaders and other stakeholders with relevant actors.

### **Who were the Luddites?**

The term 'Luddite' is derived from the legendary Robin Hood-type figure Ned Ludd, a name used as a pseudonym by early 19th campaigners against the replacement of skilled workers and traditional products in the British textile industry by new technologies (knitting frames, steam-powered looms) that enabled mass production based on new and simpler designs. In today's discourse the term 'Luddite' is used to impute small-mindedness and prejudice to those who criticise the introduction of new technologies. It was suggested that a closer look at the stated positions of Luddite leaders and at the economic and



social assumptions behind their suppression by military means could be a helpful factor in today's debates.

### **Is the tool a correct analogy for AI?**

AI is sometimes described as an extremely powerful tool. This in turn suggests a certain line of argument, namely that tools are neutral in themselves and that what matters is the use to which they are put. For example, knives are necessary; transmuted into weapons, they kill people. We use the fusion of atoms to produce energy; a similar technology allows the manufacture of weapons that in principle could destroy the world.

The 'tool' analogy does not take us as far as we might wish as a guide to the development of AI. The jury of humanity would say that some tools such as weapons of mass destruction (nuclear, biological, chemical) should not be produced in the first place. The same is true of many other tools, such as AI-manipulated 'deepfakes'. In the case of AI systems, the constant adaptation associated with AI can come between the control exercised by the user and the ultimate impact of the 'tool'. In such a world we need to embed ethics in the design of systems, as well as setting parameters for their subsequent use.

### **Ethical reasoning and our response to inhuman strangeness**

One speaker suggested restoring the word hubris to our ethical discourse. In everyday language, we speak of the 'enormity' of certain actions. There are times when the sheer scale or 'outlandishness' of what is done (such as murdering prisoners in wartime or polluting rivers to increase the dividends paid to shareholders) renders it humanly unworthy. A sense of proportionality, of spontaneous shame or anger in the face of gross facts (German: Heuristik der Furcht), is helpful, and can make the detailed work of defending human rights and protecting nature much easier.

### **Workers' unions**

In Kenya there is on-going litigation over the right of some categories of workers in the AI/social media sector to unionize. A number of major corporations have allowed workers' unions whose activities extend (sometimes controversially) to protesting the employer's trading policies. Related issues include, for example, the work of the EU Commission on the employment status of so-called 'platform workers' and litigation in the US on 'non-compete contracts.'

As AI develops, an important question is whether there is scope for a legally protected right of conscientious objection for employees. To give a hypothetical example, if there were a treaty banning Lethal Autonomous Weapons Systems and an AI-providing company or research facility failed to adhere to international obligations, would employees be entitled to go on strike or to act as whistle-blowers? Hundreds of thousands of engineers (half a million?) are affiliated to the Institute of Electrical and Electronics Engineers (IEEE). What would be the equivalent of a Hippocratic Oath for the AI professionals who play such a vital role in modern societies?



### **Procurement standards**

In assessing AI for public procurement purposes, public authorities are increasingly disposed to apply a list of criteria to any given AI system. These criteria should include transparency, reliability, accountability, and non-discrimination. Is there scope to add a broader criterion to this list such as the 'safety of society' or predictable social impact? The essential point is that the story of AI needs a broad narrative and cannot be reduced to bullet points. As long as the unfolding of the story remains unclear, there is a case for a 'safety first approach.' This argument echoed that of another speaker earlier in the day who referred to the 'boiling frog syndrome.' According to a standard definition, 'the essence of the boiling frog syndrome is that when our living conditions deteriorate gradually, we adapt to these conditions instead of getting rid of them, until we are no longer strong enough to escape.'



## PANEL DISCUSSION TWO: AI AND THE WORLD OF WORK, WITH REFERENCE TO EMPLOYMENT, PRODUCTIVITY, EDUCATION, AND EQUALITY

Noelle O’Connell, CEO of the European Movement, chaired the second panel.

The speakers were:

- i. **Molly Newell**, TASC (Think-tank for Action on Social Change)
- ii. **Kieran McCorry**, National Technology Officer, Microsoft
- iii. **Susan Leavy**, Assistant Professor, University College Dublin, and member of the AI Advisory Council
- iv. **Matthew O’Neill**, European Movement
- v. **John Gilliland**, Professor of Practice, Queen’s University Belfast
- vi. **Don Andrea Ciucci**, Holy See, Fondazione Renaissance

Comments from the floor were led by Barry Scannell, a specialist in AI Law and member of the AI Advisory Council. A summary of the discussion is given under a series of headings.

### **Incontrovertible benefits of AI in some sectors**

As in the discussion of healthcare during the morning session, it was common ground that there are some employment sectors in which the introduction of AI brings tangible and incontrovertible benefits. As one speaker said, there is a ‘wow’ factor that we should not overlook. A clear example of progress is the technology that is used in airplane cockpits. Greater safety has been achieved with fewer personnel. Another case in point is the use of AI-empowered investigations to reduce the need for antibiotics on farms.

PowerPoint, Outlook, and Teams have changed the way we work. No one is suggesting a reversal of these innovations.

It was argued that public authorities, who have an obligation to save money, would be remiss in not introducing AI where AI can reduce overheads in certain projects.


### **Citizenship and ‘algor-ethics’**

The question of ‘what it means to be a citizen’ came to prominence during the second panel discussion – much as the morning’s discussion highlighted the basic ‘structural question’ of politics, ‘What kind of reality do we want to live in?’

More than one speaker underlined the global dimension of our citizenship obligations: ‘will AI enable us to achieve the goal of zero hunger in a world that limits the increase in temperature to the 1.5C prescribed in international agreements?’

The representative of the Holy See referred to the Rome Call for AI Ethics (2020 – [www.romecall.org](http://www.romecall.org)) and circulated a brochure outlining its content. The Rome





Call aims to promote a sense of shared responsibility among all stakeholders based on a human-centred approach to innovation and technological progress. At the heart of the Call is the concept of ‘algor-ethics’, situated in an anthropological framework: ethics, education, rights, with three requirements that can be summarised as follows:

- i. non-discrimination
- ii. service of the common good of humankind
- iii. attention to the ecosystem and sustainable food systems.

Complementing the three anthropological chapters are six principles of ‘good innovation’: transparency, inclusion, responsibility, impartiality, reliability, and security/privacy. ‘Algor-ethics’ are intended to be applied to the entire process of technological innovation, from design through to distribution and use.

Many of the most influential companies have subscribed to the principles of algor-ethics. Some companies are spending ‘billions’ on awareness-raising and training/reskilling in the area of AI. Experimental projects are bringing AI to marginalised communities. Nevertheless, doubts were expressed as to whether the ‘ethics of innovation,’ as interpreted and implemented by the tech innovators themselves, plus a programme of reskilling for many of those affected by change, constitute an adequate response to the demands of citizenship in the face of current challenges.

### **Education for change**

At one level, the employment issue is a matter of numbers. It was reported that 50 per cent of diplomats’ jobs are a risk of being lost, a percentage that may be exceeded in other domains. This pattern of job losses raises ‘quality of life’ issues for everyone, as well as important challenges for employer/employee relations.

A further issue is the manner in which AI is introduced into the workplace. Will it involve liberation from drudgery or (as Amazon has acknowledged in relation to some of its own recent practices) the intrusive monitoring of employees with serious consequences for their wellbeing and physical health?

One speaker argued that what is most needed is not lifelong education in the skills associated with AI so much as a deeper understanding of the meaning of education and the meaning of citizenship. There comes a point at which ‘employability skills’ and functionality in the job market are pursued at the expense of the virtues of the citizen. The basis of a democratic society is the ability to discern the way forward in a range of practical scenarios, and to do this through discourse and persuasion in the light of consistent criteria of evaluation. It is difficult to see how this can happen without some sense of direction, valued for its own sake, which becomes a unifying focus. This is why the virtue of contemplation plays a large part in ethics and politics at the roots of the European tradition. For the same reason, active citizenship was seen at different historical moments as a means of education in itself; societies badly governed were seen as destructive of character and temperament.

### **Scale of malign activity enabled by AI/cultural disablement/content safety**

In a continuation of the morning’s discussion around algorithms, speakers pointed out that AI can enable malign activities at scale, for example phishing

attacks and computer-generated efforts to break into bank accounts. It is difficult for law-enforcement bodies to match the sheer scale of AI-enabled criminal activity.

In a further overlap with the morning's discussion, it was suggested that recommender algorithms and generative AI contribute to the creation of images of the self with which people identify, leading to forms of cultural disablement. On the other hand, it was argued that AI may soon enable new strategies to promote content safety.

A further consideration to which attention was drawn is the continuing deficit in the area of 'explainability'; the principle that AI must be understandable to all is far from being upheld in practice. By way of qualifying this assessment, it was argued that significant progress is already being made in the direction of greater 'explainability'.





## ROUND TABLES

The day concluded with round tables, each with a facilitator/rapporteur. Each round table was asked to consider in turn (i) an allocated area of focus, (ii) other challenges, and (iii) the potential contribution of churches, faith communities, and philosophical organisations. The areas of focus were:


1. AI and education (**Damian Jackson**, General Secretary, Irish Council of Churches)
2. AI and the world of work (**Fearghas O Béara**, Secretariat of the European Parliament)
3. AI and the information environment, political and social (**Patrick O'Donnell**, Onesto Consulting).

### Round Table 1: AI and education

In the sphere of education, what are the values and capabilities that we wish to impart to young people at the different stages of education? As young people engage more and more with AI and have an increasing need for computational skills, what are the implications for school teaching staff? What is the difference between, on the one hand, 'computational skills', and on the other hand, less quantifiable human attributes such as ecological awareness, personal empathy, and the ability to engage in dialogue and to persuade? In terms of the educational process, how do we ensure integrity in the use of AI in relation, for example, to the selection of students, equitable access to resources, the face-to-face participation of students in higher education, the preparation of assignments, and the grading of assignments? What are the implications of AI for university teaching as a profession? Are academics encouraged to use AI to become more prolific as researchers in a way that can be deleterious to serious academic work?

In the discussion on AI and education, a central observation was that contradictory values are sometimes at play in education. We need to negotiate different and changing expectations as to what education is for. On the one hand we should teach people to use the information that is at our fingertips, for which AI is very useful. Acquiring skills of this kind is often closely linked to the student's future prospects in the jobs market. On the other hand, we need 'non-utilitarian exchanges.' From a humanistic perspective, each person has gifts to offer. The role of education is to bring out this potential. Education should promote kindness and a capacity for awe and wonder. Non-utilitarian values are relevant from early childhood education through primary and secondary school to university-level education. The university should be a space for personal interaction, the exchange of ideas, and the critical examination of our assumptions; these objectives are broader than useful or functionally driven research and training. Some speakers spoke of a collective, collaborative model of education that would also bring out a positive relationship with nature. UNESCO proposes four pillars of education:

- learning to know

- 
- learning to do
  - learning to live together
  - learning to be.

These principles can underpin an overall assessment of the impact of AI on education.

A number of practical suggestions emerged. Universal education meant that rote learning systems were often favoured to meet the demands of the time. AI can deal with some educational processes so as to allow a more interpersonal approach. The principle would be 'to stretch all, but break none.' 'Embodied mediation' refers to an approach under which AI, on the basis of intentional personal mediation, provides lesson plans and ways of assessing students individually. This is different to a situation in which high stakes examinations incentivise the unethical use of AI to achieve good marks. At present, stress and pressure on students are excessive. We should even begin to think of a 'right to be lazy.'


At university level, students can be encouraged to use AI, but then they should have to explain what they have done and what this has achieved. It was noted that the more surveillance of the suspected misuse of AI, the more likely that the misuse will happen. There is scope for developing university degree courses on science, technology and society, focussing on the values we want to embody in our educational systems.

A large part of the discussion concerned the implications of AI for the teaching profession. At secondary school level, there is a challenge to the identity of the teacher and lecturer as 'the one who knows.' AI is taking agency away from the expert, certainly in relation to computational skills. Could this lead to a reawakening of the teaching profession? AI could do a lot to free teachers from administration and report writing, helping them to return to a vision of interpersonal engagement and the 'communication of wonder.' One of the risks at university level is that researchers will be encouraged to use AI to write and publish more, with less attention to the other purposes of a university.

## **Round Table 2: AI and the world of work**

The session began by each person presenting themselves and their background, organisation, and professional activity. This revealed a diversity in terms of exposure to and experience with AI in professional life. Backgrounds included farming, public administration, university researchers and teachers, faith communities. The facilitator read through the introductory questions provided by the Centre for Religion, Human Values, and International Relations at DCU and Onesto Consulting. These questions were as follows:

- I. How will AI impact on the world of work – bearing in mind the changes that have already occurred as a result of the automation of industrial processes and globalisation?

- 
- II. In which areas of public administration is AI most likely to make a positive difference?
  - III. How will this impact on the professional roles of public servants?
  - IV. Can we assure citizens that sensitive decisions will be taken by persons – and are therefore politically accountable?
  - V. How do the same considerations apply in business, healthcare, elderly care, agriculture and other sectors?
  - VI. Is there a connection to a wider debate based on the lessons learned during COVID?
  - VII. Can the principle of gradualness or the precautionary principle play a role?

Those around the table who had direct experience of using AI in the workplace shared their impressions with the group. Conclusions are best expressed in terms of the main points and questions raised in the discussion.

### **Overall approach**


The use of AI should be looked at in a balanced way, not just in terms of risks or dangers, but also in terms of its potential to bring positive change.

### **The need for a policy in each workplace**

Artificial Intelligence (AI) is all around us and people are often making use of AI applications in practice without being aware they are doing so. What has drawn a lot of attention over the past six months or so is specifically 'generative AI', which has the potential to disrupt many sectors, such as education. As teachers are finding out in school, so managers are finding out in offices and other workplaces, there is a 'shadow culture' of people using generative AI (for example on personal devices) outside of the official work tools or procedures, but with purposes linked to assigned tasks at work. It was felt that organisations should make efforts to bring such use out into the open so that employees know where they stand. Going forward, organisations should invest time and resources in developing a policy for the use of AI tools in the workplace so that employees are offered both technical support and ethical guidance. Writing a company AI policy presents challenges for organisations where the top management may not understand the technology themselves. Outside support could be needed. Once such a policy is in place, the question then arises of how to implement it. Do organisations have the resources and expertise required to monitor it adequately?

### **Security of employment – the need for more research**

Examples were given of how AI can revolutionise different sectors such as farming and teaching, bringing positive benefits in terms of the quality of the daily work, more free time, greater efficiency leading to greater profitability, and an 'upskilling' effect resulting in more highly paid work. At the same time, in many sectors, employees are concerned about negative impacts on the quality of the work they do or their overall job security. Are individual workers entitled to be assured that their work conditions and quality of life will not disimprove, and how can this be done? In this context, we need studies on (i) whether AI is likely to impact disproportionately, in a negative sense, on certain types of job or on certain sectors of society; (ii) whether AI is likely to exacerbate some existing disparities in terms of job security and quality of work; and (iii) whether the impact



will be greater in sectors where a higher proportion of women than men are employed.

### **Social cohesion and the ‘geography of discontent’**

Recent years have witnessed a ‘polycrisis’, from the financial crash to the pandemic shutdowns to the supply-chain disruptions following Russia’s invasion of Ukraine. Some sectors of the economy are more vulnerable than others to such disruptions. How does repeated job insecurity affect society as a whole? How does it impact on social cohesion? Some commentators speak of the so-called ‘geography of discontent,’ whereby regions or social groups with a sense they are being ‘left behind’ by changes such as the ‘Green Transition’ or the ‘Digital Transition’ can become disillusioned with the political system. AI is relevant in this context. What is the role of Government in facilitating greater job mobility in the course of such transitions? What are the safety nets that are needed, the skills programmes that should be put in place?

### **Outside the workplace**

Even outside the workplace, what do Governments need to do to bring AI to all people on an equal basis? Those on the margins of society, retired people or the elderly, those living in remote areas?

### **Personal wellbeing**

How will AI in the workplace impact on personal wellbeing? The discipline of ‘industrial psychology’ is growing in importance.


### **The social responsibility of AI providers**

Given the potential for AI to be put to positive or negative use, the producers have a clear responsibility to put appropriate safeguards in place. The speed of change is undeniable, accompanied by high levels of uncertainty and anxiety about the future. The industry should see it as part of its responsibility to build trust and to that end to improve general AI literacy.

For AI providers, a more challenging social responsibility is to orient their activities towards a well-understood vision of the common good for society as a whole. Given that AI providers stand to make trillions of dollars from the roll-out of their products and services, they should acknowledge their role in ensuring justice and equity.

### **Round Table 3: AI and the information environment**

In the public sphere, how should we analyse the impact of AI? Information is much more widely available than in the past, which brings benefits. On the other hand, manipulative techniques are now widespread including quantitative approaches to messaging, tailoring communications to specific cohorts of voters, and ‘fake news.’ The heightening of emotion at the expense of fact – a technique that has been described as ‘engage in rage and addict’ — contributes to polarisation. ‘Structural issues’ in the background include the progressive commercialisation of news ‘content’ and the different editorial values that are



upheld on-line and off-line. Some of the main points and questions raised in the discussion are set out below.

### **AI democratization and user perception**

The conversation opened with the democratization of AI, exemplified by tools like ChatGPT and Google's Gemini, designed to meet end-user needs. A significant challenge noted was that these AI models are programmed to always provide responses, often giving plausible but incomplete or incorrect answers, which may misleadingly lead users to attribute autonomous thought to artificial intelligence.

### **Consistency among AI models**

It was highlighted that AI models differ significantly in their approach. Google Gemini, for instance, adopts a conservative stance, particularly with potential intellectual property issues, such as refusing to cite the Bible without clear user assurances on handling possible IP implications.

### **A broader lack of regulatory evolution**

It was suggested that the urgency of the need for ethical standards in AI possibly stems from a broader lack of regulatory evolution in technology. Speakers pointed out that AI exploits gaps in data privacy and regulation, as seen with apps like TikTok. Although data is anonymized, AI's ability to profile individuals has intensified the potential for misuse. The AI Act was mentioned as a step towards addressing these concerns, though it too presents compromises that continue to fuel ethical debates.

### **AI and complex problem solving**

AI's role in addressing complex or 'wicked problems' was discussed. It was questioned whether such problems could be tackled effectively by AI, considering they often require solutions that evolve over time rather than definitive fixes.

### **Trust and Mistrust in AI**

The topic shifted to trust in AI, with proposals that fostering a 'positive mistrust' might be more practical than trying to legislate AI into trustworthiness. This approach encourages ongoing vigilance and adaptability in AI use.

### **Positive Impacts of AI**

The practical benefits of AI were discussed, particularly its time-saving aspects. AI's ability to summarize content and assist in language translation allows users to focus on other duties, demonstrating its utility in streamlining communication and administrative tasks.

### **Historical lessons from technological advances**

Finally, the discussion acknowledged the historical pattern where technological advances often lead to adverse outcomes, such as slavery arising from the agricultural revolution or child labour from the industrial revolution. The rapid pace of current technological change poses challenges in predicting and mitigating potential negative impacts.



## Round tables 1, 2, and 3: AI and the role of faith communities

### **The need for a pre-political culture**

Democracy depends on high-level values such as trust, loyalty, solidarity, and a capacity to reason together. We cannot create trust by decree. Legislative decisions on their own do not generate or guarantee the cultural conditions on which democracy depends. Society needs a 'pre-political' culture or standpoint from which to critique a status quo which is constantly changing.

### **Including churches and faith communities in dialogue**

Despite a history of secularization limiting their involvement in public discourse, churches and faith communities offer essential perspectives on ethics, shaped by centuries of guarding historical narratives and ethical wisdom. Several presentations earlier in the day endorsed the premise of the day's work, namely that there is good reason to include churches and faith communities in dialogue, as happens in the European Union under Article 17, Treaty on the Functioning of the EU. Faith-based actors, or other actors who come to the table with a deep cultural perspective, should be in a position to contribute positively to conversations about a more humane future in which we make creative use of AI. In addition, faith communities have 'social capital.' For example, their membership crosses boundaries of all kinds including national boundaries.

### **Duties of churches and faith communities**

For their part, churches, faith communities, and philosophical organizations ought to remain detached from narrow party interests and from commercial and geopolitical calculations. Arguably, new forms of leadership should be developed within the faith communities themselves. In the past, poor theology has led to the formation of overly obedient and passive subjects, and sometimes to a diminished awareness of humanity's long-term responsibility to promote the ecological and climatic conditions on which life depends. It was suggested that faith communities need to lament and repent, and to reframe and communicate a theology more in tune with the existential threat posed by environmental degradation and climate change.

### **Storytelling**

Faith communities can help us to preserve storytelling by serving as both custodians of tradition and sources of inspiration for new narratives. This approach, as one contributor observed, involves 'looking backwards in order to think forwards.' Thus, platforms such as TikTok can provide a medium for storytelling; a forum for inter-religious dialogue; and an archive for religious and cultural heritage.

### **Justice and connection**

Faith communities need to envision the implications of AI for human relationships and for justice and equality. It was suggested that spirituality is fundamentally about connection. This need of human connection is potentially undermined by AI. In the right circumstances, on the other hand, AI could help to strengthen connections.



### **Reimagining the dialogue with political authorities**

The usual perspective on churches and faith communities in the Irish political sphere is removed from the ambition implicit in Article 17, TFEU. The future conversation on this island cannot be about reclaiming some imagined past; faith communities should humbly contribute to a fruitful dialogue on education and other major issues. We need 'deft and humble' leadership. Faith communities can help to ensure that an emerging ethical consensus on AI is genuinely cross-cultural.

### **The role of parishes and congregations**

Faith communities can have a role in informing, reassuring, and even training people in congregations and parishes on the potential of AI.

### **The demonstration value of Article 17, Treaty on the Functioning of the European Union**

Article 17 can be understood as a 'space of shared projection' (Jonathan White), a readymade forum for advancing positive visions of the future without calling into question the day-to-day negotiations that take place elsewhere. Article 17 is also the world's leading example of a commitment by public authorities to engage in a structured dialogue on common challenges with churches, faith communities, and philosophical organisations. This dialogue, if it works well, can inspire a similar dialogue in individual jurisdictions inside and outside the European Union and encourage international organizations to create similar spaces.



## FINAL PLENARY

The final plenary was facilitated by **Ashwini Mathur** of Onesto Consulting. As there was little time left, Ashwini sought to focus on some of the major themes of the day:

- The complexity and potential unwieldiness of the topic: the debate on AI covers a galaxy of different issues
- The impact on the emerging generation and the plea from younger participants for basic values and a clear political vision
- The risk of hubris as the opposite of the precautionary principle (as mentioned by one of the keynote speakers)
- The importance of remembering AI as we consider the content of a post-2030 development agenda ('an 18th SDG')

### UN Summit of the Future/Global Digital Compact

The UN Summit of the Future, scheduled for 22-23 September 2024, brings together the member states of the UN under the theme 'multilateral solutions for a better tomorrow.' The draft concluding document of Summit, the 'Pact for the Future,' envisages that negotiations on the post-2030 development agenda will start in 2027. The 'zero draft' of a 'Global Digital Compact', a proposed annex to the 'Pact for the Future,' became available on-line in April 2024. The draft states that the UN should work 'in collaboration and partnership with all stakeholders, including governments, the private sector, civil society, international organizations and the technical and academic communities.' This list of stakeholders does not explicitly include the partners identified in Article 17, TFEU.

### Other major themes in the final discussion included:

- The impact of AI on the teaching profession. The teacher will no longer be the 'sage on the stage'
- In a benign scenario, this may lead to a re-imagining of the guiding and supporting role of the teacher in our search for truth – as a source of wisdom on the residue that will always evade quantification
- The relevance in this connection of UNESCO's four pillars of education, as cited above: learning to know, learning to do, learning to live together, learning to be





# CONCLUSIONS

## A: High-level conclusions (reflecting a broad consensus)

- i. The rapid development of AI has profound social and political implications at the global level. We cannot presume a priori that AI will make a beneficial contribution to the future of humanity and serve the cause of fraternity, freedom, and peace.
- ii. AI-based solutions make sense in a wide range of practical situations, including in science, medicine, and the workplace. Research should be encouraged by public authorities. Nevertheless, in deciding on the deployment of AI, a precautionary principle should apply.
- iii. A commitment by individuals to respect general principles in developing and deploying AI will not make the future secure. We need substantive regulation to provide innovators with the clarity they need. We should embed ethics in the design of systems, as well as in applications.
- iv. We also need to search for an overall vision that answers the question, 'What kind of reality do we want our children to live in?'
- v. AI should not reinforce inequality. The idea that the 'maximisation of shareholder value' is justified by collateral social benefits does not seem adequate as a guiding principle. 'Western' societies need to take specific steps to counteract polarisation and the loss of trust in institutions.
- vi. Proportionality in the allocation of resources in the light of a good that is common to all is a core democratic value that needs to receive greater attention in discussions around AI.
- vii. The military applications of AI have given rise to a dangerous inter-state competition in which previously accepted ethical parameters are set aside.
- viii. AI raises fundamental questions for the future of education and role of teachers. It is essential to maintain a balance between 'computational skills' and 'employability', and less quantifiable and more important human attributes such as the religious and historical imagination, ecological awareness, personal empathy and solidarity, creativity, and the ability to engage in dialogue and to persuade.
- ix. A 'holistic' framework of engagement at the global level to address AI can be achieved in the context of a well-designed post-2030 development agenda.
- x. Article 17, Treaty on the Functioning of the European Union, has the potential to serve as a 'space of shared projection' within which to enable a useful preparatory dialogue. If this works well, it can inspire a similar dialogue in other jurisdictions.



## B: Examples of practical steps

- i. The European Parliament, working with the Commission, should pay close attention to the implementation of the AI Act, which will require considerable resources. The Parliament should urgently initiate a dialogue on the overall social purpose that the applications of AI are intended to serve. The Parliament should promote a dialogue with other jurisdictions on the future of AI, perhaps through the agency of parliamentary ‘delegations’.
- ii. The European institutions (Parliament, Commission, Council) should make creative use of Article 17 in developing a holistic vision (or visions) of the future in which AI will have its proper place.
- iii. Governments should facilitate a positive, multi-faceted reflection on the future involving the creative use of AI. This will mean developing new frameworks for multi-stakeholder engagement, giving keen attention to social indicators and social safety nets, and envisaging new forms of public investment. Relevant experience has been acquired through the use of wellbeing indicators in many OECD countries and through the implementation of the European Union’s Green Deal.
- iv. Individual organisations should invest time and resources in developing a policy for the use of AI tools in the workplace. This is essential to avoid a ‘shadow culture’ where people use generative AI (for example on personal devices) outside of the official work tools or procedures, but with purposes linked to assigned tasks at work. Writing a company AI policy presents challenges for organisations. Outside support could be needed.
- v. In the educational sector, the concept of ‘embodied mediation’ should be explored further. This refers to an approach under which AI, on the basis of intentional personal mediation, provides lesson plans and ways of assessing students individually.
- vi. Current business models in the ‘virtual’ world should not be taken for granted.
- vii. The major AI companies are spending billions on education and socially oriented pilot projects. Building on these good practices, corporations should fund ‘champions’ for the delivery of values-led research projects focusing on the overarching social and anthropological questions raised by this report.
- viii. One such project could be the development of options for a ‘Hippocratic Oath’ or Charter for the AI engineers and other AI professionals who will play such a vital role in the coming years.



# ANNEX

## Annex 1 - Programme

### **Artificial Intelligence (AI) - An Ethical Challenge for Humanity** *European Future Talks 2024*

19<sup>th</sup> April 2024

The Oak Room, Mansion House, Dublin

**10:00– 10:30 Arrival, registration, tea/coffee**

**10:30 – 10:50 Words of welcome**

**Lord Mayor Daithí De Róiste**

**Christian Gsodam**, European External Action Service, Founder of European Future Talks

**Daire Keogh**, President, Dublin City University

**Othmar Karas**, First Vice-President of the European Parliament (video message)

**10:50 – 11:15 Keynote speech, presentation of questions**

**Jovan Kurbalija**, Executive Director of DiploFoundation, Geneva (keynote)

**Philip McDonagh**, Centre for Religion, Human Values, and International Relations, and **Ashwini Mathur**, Onesto Consulting

**11:15 – 12:15: Panel Discussion 1: The purposes of the AI Act, AI and security, AI and the information environment, AI and structural bias**

Chair: **Larry O’Connell**, Director, National Economic and Social Council

- i. **Axel Voss**, MEP (online), rapporteur, AI Act
- ii. **Christian Gsodam**, EEAS Advisor for Strategic Communication/Foresight
- iii. **Abeba Birhane**, Trinity College Dublin
- iv. **Jane Suiter, Professor**, Dublin City University
- v. **Catherine Prasifka**, Writer-in-Residence, Trinity College
- vi. **Archbishop Michael Jackson**, Chair, Dublin City Interfaith Forum

Comments from the floor, led by **Professor William O’Connor**, University of Limerick, and **Professor Stephen Williams**, Queen’s University Belfast

**12:15 Introduction to the Insight Centre for Data Analytics**

**Noel O’Connor**, Professor DCU, Insight SFI Centre for Data Analytics

12:25 Initial real-time feedback, using an app (facilitated by **Chris Chapman**, Facilitator, Burren College of Art)



### 13:30 – 14:15 Words of welcome in relation to the dialogue with churches and faith communities and overview of the state of the debate on AI

**Seán Ó Fearghail**, Ceann Comhairle (Speaker) of Dáil Éireann

**Anja Kaspersen** (on-line), Carnegie Council for Ethics in International Affairs and Special Advisor at the Institute of Electrical and Electronics Engineers (IEEE)

**Vincent Depaigne** (on-line), DG Justice (JUST), European Commission

### 14:15 – 15:15 Panel Discussion 2: AI and the world of work, with reference to employment, productivity, education, and equality

Chair: **Noelle O’Connell**, CEO, European Movement, Ireland

- i. **Molly Newell**, TASC (Think-tank for Action on Social Change)
- ii. **Kieran McCorry**, National Technology Officer, Microsoft
- iii. **Susan Leavy**, Assistant Professor, University College Dublin
- iv. **Matthew O’Neill**, European Movement
- v. **John Gilliland**, Professor of Practice, Queen’s University Belfast
- vi. **Don Andrea Ciucci**, Holy See, Fondazione Renaissance

Comments from the floor, led by **Barry Scannell**, AI Law Specialist

### 15:15 – 16:45 Round tables

There will be six round tables, each with a facilitator. Round tables will consider in turn (i) their allocated area of focus, (ii) other challenges, and (iii) the potential contribution of churches, faith communities, and philosophical organisations. It is suggested that they divide their time into segments of 40, 30, and 20 minutes.

Round Table 1: area of focus, AI and education (facilitator: Martin Hawkes, Burren College of Art)

Round Table 2: area of focus, AI and education (facilitator: Damian Jackson, ICC)

Round Table 3: area of focus, AI and the world of work (facilitator: Damian Thomas, NESC)

Round Table 4: area of focus, AI and the world of work (facilitator: Fearghas O Béara, Secretariat of the European Parliament)

Round Table 5: area of focus, AI and the information environment, political and social (facilitator: Gary Carville, Commission for Social Issues & International Affairs)

Round Table 6: area of focus, AI and the information environment, political and social (facilitator: Patrick O’Donnell, Onesto Consulting)

### 16: 17:45 Final plenary

Co-Chairs: Philip McDonagh, Centre for Religion, Human Values, and International Relations, and Ashwini Mathur, Onesto Consulting

Brief reports from the round tables

Further real-time feedback from participants (Chris Chapman, using an app)



## Annex 2 – List of participants

### **Oireachtas (Parliament), European Union**

1. Seán Ó Fearghail, Ceann Comhairle, Speaker
2. Jill Gray, Oireachtas
3. Christian Gsodam, European Future Talks
4. Michael Jansen, European Future Talks
5. Axel Voss, MEP (Online)
6. Vincent Depaigne, EU Commission Online
7. Fearghas Ó Béara, European Parliament Secretariat
8. Othmar Karas, MEP, Vice-President, European Parliament (video message)
9. Meabh De Burca, Commission Office

### **Dublin City University (DCU)/Centre for Religion, Human Values, and International Relations**

10. Daire Keogh, President, DCU
11. Philip McDonagh, Adjunct Professor
12. Timmayo Thumra, Centre
13. Josh Treacy, Centre
14. Jane Suiter, Professor
15. Fiona Regan, Professor
16. Noel O'Connor, Professor
17. Alan Smeaton, Professor, AI Advisory Council

### **Churches and faith communities**

18. Archbishop Michael Jackson
19. Fr. Andrea Ciucci, RenAIssance Foundation (Holy See)
20. Damian Jackson, General Secretary, Irish Council of Churches
21. Msgr Joe McGuinness Secretary, Catholic Bishops Conference
22. Gary Carville, Catholic Bishops Conference
23. Adrian Cristea, Executive Officer, Dublin City Interfaith Forum
24. Edwin Graham, Northern Ireland Interfaith Forum
25. Norman Richardson, Northern Ireland Interfaith Forum
26. Kevin Hargaden, Jesuit Centre for Faith and Justice
27. Fr. Seán Ford, Carmelites
28. Rev Andrew Irwin, Church of Ireland
29. Rev William Hayes, Presbyterian Church
30. Michael Briggs, Methodist Church
31. Dáire Campbell, Methodist Church
32. Ms Alison Wortley, Baha'i Community
33. Revd Myozan Kodo, Zen Buddhism Ireland
34. Ms Hilary Abrahamson, Dublin Jewish Progressive Congregation
35. Swami Purnananda Puri, Hindu Community
36. Mr Imran Haider, Shia Community
37. Mr Shaheen Ahmed, Islamic Cultural Centre of Ireland
38. Fr Anish John, Indian Orthodox Church
39. Dr Jasbir Singh Puri, Sikh Community Ireland
40. Pr Dare Adetuberu, Redeemed Christian Church of God

- 
41. Fr Paul Glynn, Columban Missions
  42. Lynda Gould, Northern Ireland Council for Voluntary Action
  43. Ed Petersen, Clonard Reconciliation Mission

#### **Keynote presenters**

44. Jovan Kurbalija, DiploFoundation, Geneva
45. Anja Kaspersen, Carnegie Council /Special Advisor IEEE (Online)

#### **Government**

46. Joy Hadden, Office of the First Minister/Deputy First Minister
47. Eugene Farrelly, Department of the Taoiseach
48. Mary Keenan, Department of the Taoiseach
49. Tomas O Ruairc, Department of Education
50. Sarah Glavey, Department of Public Expenditure and Reform
51. Karen Gillanders, Department of Foreign Affairs

#### **Business**

52. Kieran McCorry, National Technology Officer, Microsoft
53. Ashwini Mathur, Onesto Consulting
54. Patrick O'Donnell, Onesto Consulting
55. Stephanie Anderson, Google

#### **Academic (other than DCU)**

56. Stephen Williams, Queen's University Belfast
57. Niall Robb, Queen's University Belfast
58. Katy Hayward, Queen's University Belfast
59. Siobhan O'Sullivan, Royal Irish Academy
60. John Gilliland, Professor, Queen's University Belfast
61. Dr. Abeba Birhane, Trinity College Dublin, AI Advisory Council
62. Susan Leavy, Professor, UCD, AI Advisory Council
63. William T. O'Connor, Professor, UL AI Advisory Council

#### **Civil Society**

64. Damian Thomas, National Economic and Social Council
65. Larry O'Connell, Director, NESC
66. Martin Hawkes, Burren College of Art
67. Chris Chapman, Facilitator
68. Barry Scannell, AI Law Expert, AI Advisory Council
69. Catherine Prasifka, Author
70. Molly Newell, Think Tank Action on Social Change
71. Sunniva McDonagh, Irish Human Rights Commission
72. Noelle O'Connell, CEO, European Movement
73. Matthew O'Neill, European Movement
74. David Donoghue, Former Diplomat





## Annex 3 – A note on the politics of AI

### Definition of AI

Participants in our meeting mostly relied on the OECD definition of AI:

AI refers to machine-based systems that can, given a set of human defined objectives, make predictions, recommendations, or decisions that influence real or virtual environments. AI systems interact with us and act on our environment, either directly or indirectly. Often, they appear to operate autonomously, and can adapt their behaviour by learning about the context. (OECD, 2021)

There is a case for revisiting this definition, taking into account the most recent developments, especially in relation to generative AI.


### AI and humanity

Positioning the emergence of AI on a global historical timeline was not a named point on the agenda for the meeting in April. However, many speakers offered an informed judgement that the advent of AI is of the profoundest historical and civilizational significance. The industrial revolution, and perhaps also the agricultural revolution, provide points of comparison. These revolutions led to step-changes in productivity, but also to dramatic new forms of inequality. We now face changes potentially far more significant than previous transformations in the means of production. The competitive, commercial pressure to analyse vast quantities of existing data using algorithms and to use algorithms to generate new data, including images, is impacting on the shape of society, the sources of wealth, and our self-understanding as workers and as human persons.

The defence-related study quoted in the body of our report identifies the ‘speed and scale of moral change as different behaviours and attitudes become normalized through exposure.’ This perception of a changing ethical reality is linked to the idea that we should conceptualize the human being as a ‘platform’ capable of being ‘enhanced’. In some university settings, official staff guidance within the university acknowledges the growing integration of AI into other software, such as plug-ins for word processing. That these practices are ‘blurring the notion of authorship and pushing the boundaries of collaborative intelligence’ is apparently not called into question (paper by Katy Hayward, forthcoming). The meshing of persons and machines can happen in other spaces as well, as we note in Annex 4 below.

These AI-related developments are occurring globally without the ‘slack in the system’ provided in past eras by the separation of continents. The pace of change is increasing. In the meantime, the conditions on which human life depends are being eroded by climate change, environmental degradation, and many other factors.

There is an evident deficit in our collective engagement with shared challenges. Many citizens feel powerless. The sense that time is not on our side makes it



harder to structure dialogue and build bridges across cultural divides. The urgency imposed by economic and military competition indirectly feeds populism and polarisation. It would be foolish to presume a priori that the development of AI will make a beneficial contribution to the future of humanity and serve the cause of fraternity, freedom, and peace.

### **Spaces of shared projection**

Many of the day's contributors in April looked forward to the development of 'spaces of shared projection' (Jonathan White) or even a new 'holistic' framework of engagement to address the overarching question (as formulated by one speaker), 'What kind of reality do we want to live in?' Another contributor suggested, in the Mentimeter poll, 'AI can become a getting-out-of-jail card for the mess we've made of the planet.' The most obvious way of developing a 'holistic' framework is to address AI in the context of a well-designed post-2030 development agenda ('an 18th SDG,' as Ashwini Mathur said at the final plenary). Eight parameters or criteria for such an approach are suggested here in the light of comments made in the course of the Mansion House meeting.

#### **1. Asymmetrical partnership in support of democratic decision-making**

The status of the different stakeholders in deliberation is a central question. Politics serves a good that is common to all. In this perspective, legitimacy and authority flow from the people as a whole and their representatives. Therefore, the status of actors other than public authorities needs careful attention.

A situation where religious authorities claim the last word *on practical decisions* in the public sphere is widely recognised as a danger. However, it is equally dangerous for a political authority to claim the last word on questions relating to *ultimate values and a future-oriented sense of right and wrong*. Democratic majorities make mistakes. Even the most powerful individual leader does not claim to determine the meaning of human experience. Nurturing an effective public truth is a multi-stakeholder task in the course of which political actors and citizens with a faith or worldview should enter into dialogue. In this process, *dialogue and communication* constitute an ineluctable first principle – ineluctable because to challenge this starting point is to find oneself back again in dialogue and communication. A decision made top-down or by a majority does not cancel the experience of the citizen, the meaning of that experience, or the co-presence of citizens to one another. Aristotle and Confucius have in common that politics is a communication system aligned with human nature.

Equally dangerous for deliberation is the suggestion that corporations and governments are 'interlocking institutions' that engage as equal participants in political decision-making. It is difficult to construct a reading of politics in which de facto economic power confers legitimacy or a popular mandate (though this has been attempted in the past). On the contrary, entrenched narrow interests are a threat to democracy and legitimacy unless overseen by an inclusive law-making process that is credible and transparent. Partnership in policymaking by governments, parliaments, transnational institutions, companies, civil society actors, and faith communities is by definition an asymmetrical partnership.



## 2. Planetary reconciliation


In 2022, the Archbishop of Canterbury, Justin Welby, published *The Power of Reconciliation* (London: Bloomsbury). The vision of Archbishop Welby is that reconciliation in today's world flows in part from acknowledging the dangers threatening the planet. Under the heading of racial differences, Archbishop Welby goes further and addresses the legacy of slavery and empire (pp. 252 – 257). This discussion is worth quoting in part as it provides a valuable 'macro' perspective on the development of frameworks of engagement in response to AI and its challenges:

*The question of reparations will have to be faced, and an answer found that is a sufficient sign and symbol of genuine relief of the needs caused by past actions ...*

The 'architecture of modern empire' (Arundhati Roy) takes many different forms. Given the salience of 'western' colonial history in a world of widening inequality and deteriorating climatic conditions, the European Union and other 'western' societies should consider investing in a broadly-conceived post-2030 agenda for social and economic transition as an essential and achievable form of 'transitional justice.' An adequately funded global project of this kind can provide a helpful context for exploring the potential of AI.

## 3. Perspective and proportionality as ethical criteria

The ethical criteria currently applied to the development of AI fail to capture the issue of proportionality. As pointed out by our scene-setting speaker in April, OpenAI is reportedly (as of February 2024) seeking to raise up to US\$7 trillion for chip production. OpenAI is not the only corporation whose planning is on a vast scale. Another statistic in play is that a small number of leading AI companies may soon have a collective valuation of \$15 trillion (cited in the Guardian podcast Black Box). In the meantime, in 2023, the entire GDP of Africa was US\$3 trillion. In 2023, according to FAO figures, 20 per cent of the population of Africa was facing acute food insecurity, a figure that is higher in some countries. Almost one billion live in countries where interest payments on debt exceed spending on health and education. Globally, financial assets are four times (or more) the size of the real economy. In 2021, assets held by financial corporations were estimated at \$510 trillion. According to the OECD, the size of the world economy today is more than \$105 trillion. Current western ODA (Official Development Assistance) is approximately \$224 billion. In the climate change negotiations, developed countries committed to mobilize collectively \$100 billion per year to support developing countries throughout the world in reducing emissions and adapting to climate change. The US Administration's proposed military budget for 2024 is of the order of \$840 billion. Global military spending amounts to more than \$2300 billion and is increasing. As attention turns to a post-2030 development agenda and a renewed commitment to climate targets, the figure of \$1 trillion for external financial aid to the global south is sometimes mentioned as an ambitious target, as part of a considerably greater sum to be raised by other means.



Any discussion of the international financial architecture needs to address the UN Secretary General's wide-ranging policy brief published in 2023. This document states the following:

*The international financial architecture, crafted in 1945 after the Second World War, is undergoing a stress test of historic proportions – and it is failing the test ... [it] already had structural deficiencies at the time of its conception ... [it] is entirely unfit for purpose in a world characterized by unrelenting climate change, increasing systemic risks, extreme inequality, entrenched gender bias, highly integrated financial markets vulnerable to cross-border contagion, and dramatic demographic, technological, economic and geopolitical changes ... The existing architecture has been unable to support the mobilization of stable and long-term financing at scale for investments needed to combat the climate crisis and achieve the Sustainable Development Goals ...*

The progress that is now needed at the global level largely depends on perspective and proportionality in relation to the scale and allocation of resources. The commercial ambitions and social obligations of the major AI-focussed corporations cannot be excluded from this discussion.

#### **4. Restraining the arms industry and avoiding geopolitical competition**

The former Chairman of the US Joint Chiefs of Staff and the former CEO of Google published a joint article in *Foreign Affairs* on 5<sup>th</sup> August 2024 under the heading 'America Isn't Ready for the Wars of the Future/And They're Already Here.' The authors' perspective is that 'the nature of war is, arguably, immutable. In almost any armed conflict, one side seeks to impose its political will on another through organized violence.' Without distinguishing one historical era from another, the authors list examples of technology-enabled organized violence ranging from the introduction of cavalry in the ancient world to the use of the atomic bomb in World War Two. With particular reference to Ukraine in 2024, they describe in positive terms the ways in which AI is increasingly used to manage autonomous weapons/drones and identify targets. The authors' conclusion is straightforward. They call for a major overhaul of the US military in favour of AI-enabled weapons; the US as the leader in developing AI for military purposes should not be outpaced by its rivals.

However, late on in the article, the authors create an all-important space for dialogue when they recognise that the use of AI in warfare 'opens a Pandora's box' of ethical and legal issues:

The Israeli military has used an AI program called Lavender to identify potential militants and target their homes with airstrikes in densely populated Gaza. The program has little human oversight. According to *+972 Magazine*, people spend just 20 seconds authorizing each attack.

Finally, the authors conclude:

In the worst-case scenario, AI warfare could even endanger humanity.



Anduril, the California-based defence tech start-up (which has its European equivalents), is a seven-year-old company valued at \$12.5bn in mid-2024. In a recent interview (Financial Times, March 2024), Anduril's 31-year-old co-founder explains his company's purpose as follows:

The way to frame it would be that I want to give ourselves a technology that turns the world stage, as it pertains to warfare, into the United States being an adult in a room full of toddler-sized dictators.

In many parts of the world, the race to produce AI-enabled weapons is embedded in a vision based on three dubious assumptions: (i) geopolitical competition without end; (ii) machine-assisted warfare as an 'immutable' aspect of the human condition; and (iii) the assertion that to protect our 'values' we may need to moderate our 'ethics'. These elements generate in turn a binary framing of the differences between societies ('clash of civilisations') and a weakening grasp of the interests that governments and states have in common. In the sphere of AI, a bleak understanding of the future is being allowed to prevail over ethical objections that governments, company executives, employees, and private investors took seriously until only a few years ago.


The UN Charter points towards a broad understanding of security as arising from cooperative relationships and the habits and assumptions that flow from this. 'General disarmament under international control' is a core objective under the Charter. Arguably, the emerging discourse around AI-enabled weapons systems carries with it an epoch-shaping loss of trust in the purposes and principles of the UN Charter.

There are conflicts of interest in the space where gigantic private investments are made in new weapons. Studies on the role of firearms in the colonial period, of machine guns in the years before World War One, and of carpet-bombing techniques and weapons of mass destruction in World War Two suggest that once military technologies are developed at great cost there is a perceived political necessity to ensure that they matter in practice. Another obvious factor is that the volume of drones or missiles used on the ground, and the duration of a particular conflict, impact on profits and on the likely scale of future demand.

The progressive loss of inhibition in relation to AI-enabled weapons involves cultural impoverishment – the disappearance of historical memory (notably in some major European countries), a desensitisation towards the devastation caused by military activity (as illustrated by the use of euphemistic formulations such as 'outsmarting the enemy'), and a cold and detached approach that masks one's own share of moral responsibility for the destruction over time of entire societies.

## **5. Antitrust/competition policy**

In the US Congress, interest has been growing in 2024 in the relevance of antitrust policy to the future of AI. Much of this attention is focussed on TikTok, with an obvious read across to trade relationships and US national security interests. However, there is a wider point as well. The current trajectory of AI



risks benefiting the few and creating insecurity among the many. Traditionally, antitrust policy in the US was oriented towards the distribution of power in the economy and the welfare of citizens broadly understood. Since the 1980s, partly because of globalisation and its perceived imperatives, there has been a shift towards an antitrust policy based on the single idea of lowering prices for consumers. This narrower understanding of 'antitrust' is losing credibility in some sectors.


In many parts of the US, traditional middle-sized farms are unable to compete with a handful of major corporations which increasingly control inputs, decisions on production, distribution, marketing, and storage, as well as owning farmland. In addition, processed foods marketed in the US or exported from the US by these corporations creates a major 'externality'. Diet-related diseases according to one study cost \$3.7 trillion per year to treat in the US alone. Disproportionate energy consumption and water use by data-centres can perhaps be regarded as an externality comparable to the downstream impact on health of ultra-processed foods.

One way or another, it seems likely that in a not-distant future, antitrust and competition policies will play a greater role in the sphere of AI.

## 6. The anthropological question

The 'anthropological question' posed by AI revolves around the meaning of human experience and history:

- If every action aims at some good, is there a higher good, such as happiness, which is valued for its own sake and becomes a unifying focus?
- Is there a common life or collective well-being that is more than the sum of our private interests?
- Is exploration and discovery the essence of human identity or is there a point at which stability or sharing should come first?
- Are we prepared to suffer for the sake of others?
- Are there human activities where knowledge for its own sake, practical discernment, and technical skill are co-present – 'intrinsically worthwhile activities' such as the work of craftspeople in certain domains or the exercise of democracy?
- Are such intrinsically worthwhile activities inherently part of a wider pattern?
- Are human beings entitled to a moral relationship with things, and how is such a right to be reconciled with wholesale private ownership of the means of production?
- Is there a law of progress or perfection in human history?
- Is emotion self-verifying and if not, what is the standard of truth?
- Is there always an element of 'givenness' in human creativity and happiness?
- How does a political dispensation based on coercion become a dispensation based on freely given consent?

- 
- Is it rational to hope for a ‘future not visible in the alternatives of the present’?


Over time, we can expect to discover a clear ‘dialogical’ relationship between our answers to questions such as these and our understanding of AI and its place in society. A central issue is whether we can avoid conflating what is characteristically human – such as our conscience, our intuition, our relationality, and our understanding of the resonance of words – with the processes and products of machine-learning and deep learning. It is a logical fallacy to argue from ‘X is Y’ (‘the human organism/brain performs in many respects like a computer’) to ‘X is nothing but Y’ (‘human empathy is not distinctive’).

## 7. Politics as a journey

Politics is ‘complex’, a journey into the future in which each step we take connects with steps taken by others. Therefore, as a necessary preamble to practical decision-making we should apply our reason to the understanding of political processes. Prejudice is often a collective failure, as when 19th century thinkers took for granted the supposedly different capacities of different racial groups. A ‘positivist’ reading of history (Oswyn Murray, *The Muse of History*, 2024) is one in which our theory of change is placed beyond the reach of rational argument.

As democracy was developed in the 5th century BCE, the leading thinkers saw clearly that the granular provisions of established law are an inadequate foundation for life in society, for several inescapable reasons. First, the law is incomplete: many of our responsibilities are not enforced by our codes of law. Second, lawmakers will not have reckoned with the precise circumstances of every case (an insight connected with the jurisprudence of equity). Third, circumstances are different from one society to another or may change. In times of political upheaval or social disintegration, a citizen’s obligations under the law can become unclear. Do we serve a revolutionary government or an occupying power? How do we define our moral obligations under rapidly changing international circumstances?

Of these three ideas, the most decisive for present purposes is the first, namely the possibility that the law as declared will in fact represent a distortion of justice. The defence paper quoted above notes that ‘different behaviours and attitudes become normalized through exposure.’ For Thucydides, the ‘normalization of certain behaviours through exposure’ is symptomatic of social breakdown: ‘war is a savage teacher.’ ‘The speed and scale of moral change’ under the circumstances identified by the respective Defence Ministries is for Thucydides and other thinkers of the 5th century BCE analogous to the spread of a plague. These thinkers contrast the forward-oriented politics of detachment, dialogue and deliberation with stasis, understood as the angry, vengeful and ‘unexamined’ pursuit of a perceived self-interest by separated clusters of citizens who are no longer in communication with one another. Two generations later, Aristotle pictures a political leader capable of connecting his intimation of ‘noble and divine things’ with particular choices. In other words, the criterion of evaluation for any specific political choice is that it should be forward-looking and fit within a



worldview that is itself independent of day-to-day politics—or perhaps we should say, a worldview that is always in a dialogical relationship with day-to-day politics.

Aristotle's *Politics* opens with the famous statement that 'the polis exists by nature' and that 'man is a political animal' (*politikon zōon*). Aristotle connects these two assertions with our human ability to reason together on questions of right and wrong: 'among living creatures only human beings possess *logos*.' From the Greeks we can learn that 'politics' is a forward-looking enterprise based on interpersonal communication in a shared space. Through politics we give expression of what it means to be human in the decisions of everyday life.

## 8. Rational hope and a 21st century Axial Age

The search for a global dimension to citizenship was already underway in Aristotle's time, as we see, for example, in the 'cosmopolitan' vision of the Stoics. Today, our political vision must include a long-term responsibility to promote the social, ecological and climatic conditions on which life depends. It is essential to ask how AI interfaces with 'grand challenges' in relation to the fracturing of global politics, climate tipping points, the loss of biodiversity, food insecurity, the spread of conflict and so-called 'grey zone warfare,' transnational organised crime, migration, population growth, the likelihood of another pandemic, the politics surrounding rare earth materials, and economic disparities that continue to intensify.

It is therefore for consideration whether rational hope is the primordial political value, an inner resource implying a readiness to engage with our circumstances and act where possible, even in the face of steep odds. When our societies look to the future together, we are by definition co-workers in a project whose detailed design is not personal to us. Changes bringing out what is best and most essential in our culture may make themselves felt only following decades of shared deliberation. Nevertheless, appraising the truth of the here-and-now, and acting in consequence of this, can sow the seeds of a wider change, because actions that conform with rational hope will be in harmony with the similar actions of other people elsewhere.

Our common criterion of evaluation cannot be the standard of mere self-interest, which by definition pushes individual actors in different directions. Any common criterion of evaluation at the local or global level will of necessity link one situation to another and one generation to another and enable a variety of actors to pull together in giving the future a distinct shape or character, even before the overall picture becomes clear. To paraphrase Voltaire, if rational hope did not exist it would have needed to be invented.

The standard of 'rational hope' will have greater impact if our societies can coalesce around a readily understood narrative. Our keynote presenter in April spoke of the so-called Axial Age, beginning in the 8th century BCE in many different geographies. At that time, in different places, there emerged a social, political, and juridical space in which traditional ways of doing things could be examined critically, and new conventions could be established. The principle of





verification produced a civilisational shift in terms of political transparency and accountability. The socio-ecological transition to which we look forward can usefully be described as the transition to a global 'Axial Age.' In this perspective, the enormous potential of AI is emerging at a watershed moment. If the 'political economy' of the present moment is well managed, AI's potential can perhaps inspire a renewed sense of purpose and belonging among all citizens. Our rational hope in the possibility of a 21st century Axial Age can bring together all those who face the future determined to be part of the solution. We are fortunate that Article 17, Treaty on the Functioning of the European Union, serves as a readymade 'space of shared projection' within which to advance a vision (or visions) of a better future.



## Annex 4: Case Studies

This annex lists some practical examples of the difficult issues that arise in regulating specific applications of AI.

### **The GRADE algorithm**

Consider the GRADE algorithm, developed by the University of Texas at Austin. It was implemented to streamline PhD admissions in the computer science department from 2013 to 2019. This AI-based tool was trained on data from previously successful applicants to identify potential candidates aligned with the department's preferences. According to the developers' research, GRADE significantly optimized the admissions process, reducing the necessity for comprehensive application reviews by 71% and decreasing the overall review time by at least 74%. Consequently, applications scored lower by the algorithm received less consideration from the admissions staff. The Computer Science department stopped using it from 2020 and said "... the code had the potential to pick up unfair biases ... it was difficult to maintain..."

This experience prompts some potential questions: Since this was an AI application, it needed to be assessed for risk: could it be in the high-risk area, because university education is an important 'private / public service.' Was there a human being involved in the decisions? If not, it is high risk. Did a decision for an individual have an impact on that individual's access to education? Was there transparency in giving details of the AI algorithm used for admission decisions to the candidates? What kind of data was used to train the algorithms? Is there bias in the data which could get propagated further by continual usage of the algorithm? How would an algorithm like this be addressed by EU AI act?

After EU AI act is implemented, two other well-known algorithms which could be questioned even before they could be used more deeply are the Amazon's AI Recruitment Tool and IBM's Facial Recognition Technology. Both have already been removed from usage.

### **Amazon's AI recruitment tool**

The AI recruitment tool was designed to screen job applicants' resumes. However, it was reported in 2018 that the algorithm exhibited bias against women. The AI was trained on resumes submitted to the company over a 10-year period, most of which came from men, reflecting male dominance in the tech industry. This led the system to unfavour women candidates for technical roles. After EU AI Act, this algorithm would be very difficult to roll out today.

### **IBM's Facial Recognition Technology**

IBM's Facial Recognition Technology was stopped from being used. IBM's CEO cited concerns over the use of such technology for mass surveillance and racial profiling as reasons for the company's decision. This sequence of events highlighted ethical concerns about AI and privacy and the potential for misuse in law enforcement and surveillance.



## AI and the judicial process

(This example is cited from the address of Pope Francis to the G7 on 14 June 2024)

Artificial intelligence is designed in order to solve specific problems. Yet, for those who use it, there is often an irresistible temptation to draw general, or even anthropological, deductions from the specific solutions it offers.

s

An important example of this is the use of programs designed to help judges in deciding whether to grant home-confinement to inmates serving a prison sentence. In this case, artificial intelligence is asked to predict the likelihood of a prisoner committing the same crime(s) again. It does so based on predetermined categories (type of offence, behaviour in prison, psychological assessment, and others), thus allowing artificial intelligence to have access to categories of data relating to the prisoner's private life (ethnic origin, educational attainment, credit rating, and others). The use of such a methodology – which sometimes risks de facto delegating to a machine the last word concerning a person's future – may implicitly incorporate prejudices inherent in the categories of data used by artificial intelligence.

Being classified as part of a certain ethnic group, or simply having committed a minor offence years earlier (for example, not having paid a parking fine) will actually influence the decision as to whether or not to grant home confinement. In reality, however, human beings are always developing, and are capable of surprising us with their actions. This is something that a machine cannot take into account.

### Governance issues

Besides scientific and ethical concerns, governance for AI usage in industry is challenging. A recent example of this challenge was faced by Google. Google's Advanced Technology External Advisory Council (ATEAC) was set up as a means of driving strong AI governance. However, it was dissolved just over a week after its formation. Google had established ATEAC to draw on the expertise of philosophers, engineers, and policy experts. The selection of certain individuals sparked a significant backlash. The appointment of Kay Coles James, the President of the Heritage Foundation, known for her controversial views on LGBT rights, climate change denial, and anti-immigrant sentiments, was met with strong opposition from Google employees. They were concerned that AI technologies' flaws disproportionately affect marginalized communities, thus arguing that James was not an appropriate choice for an advisory council focused on ethical AI development. Another controversy involved Dyan Gibbens, the founder of Trumbull Unmanned, which further intensified debates about Google's involvement with military projects.

This incident, as well as multiple other examples of controversial AI algorithms, underscore the complex interplay between technology development, corporate ethics, and societal values, highlighting the importance of transparent and inclusive approaches to governance in the AI era.



**For more information:**

Philip McDonagh  
Adjunct Professor  
Director  
DCU Centre for Religion, Human Values,  
and International Relations

E: [philip.mcdonagh@dcu.ie](mailto:philip.mcdonagh@dcu.ie)  
W: [dcu.ie/religionandhumanvalues](http://dcu.ie/religionandhumanvalues)