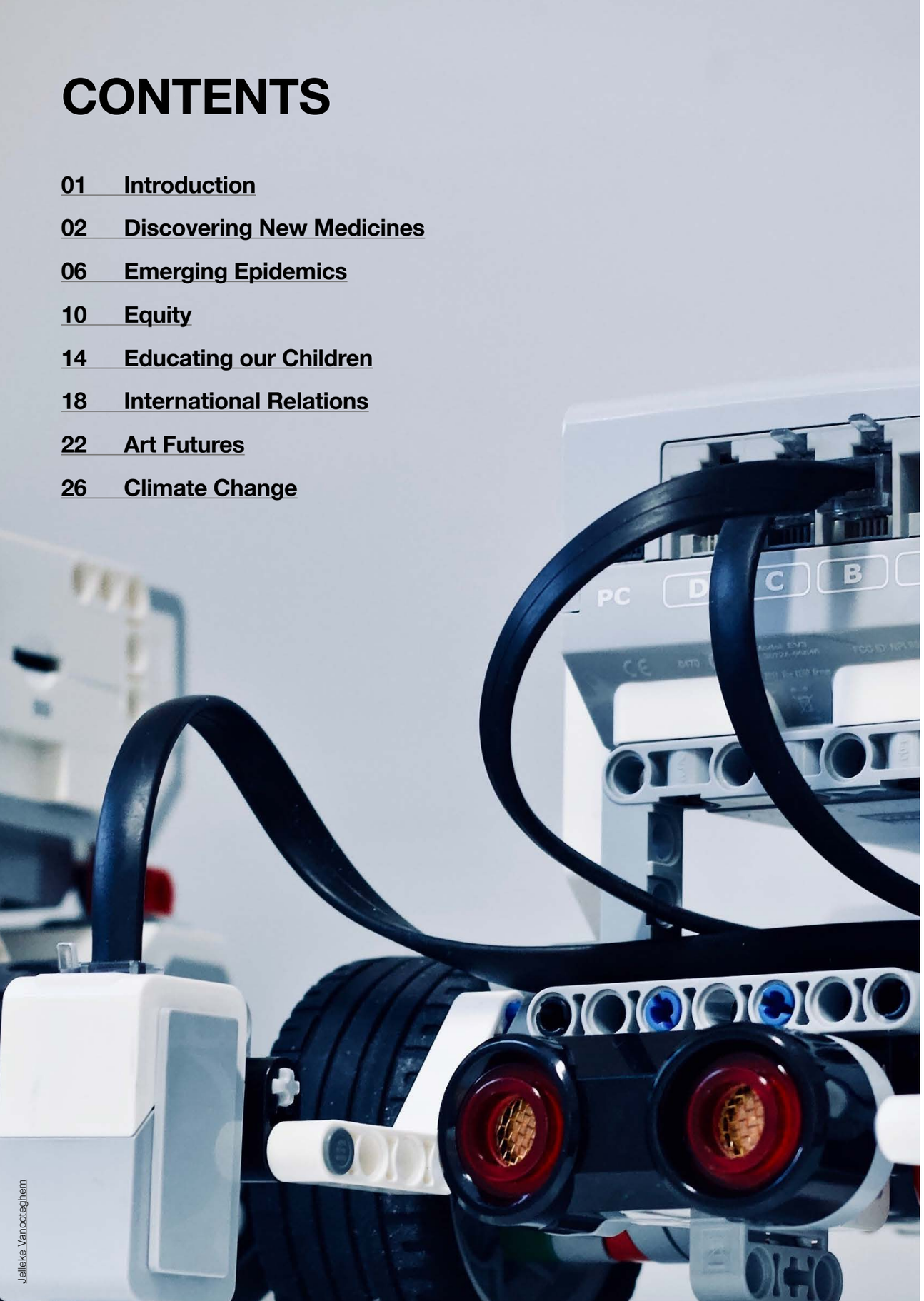
UCL

AI for **People** & **Planet**

Insights for Public Policy

CONTENTS



Jelleke Vanooteghem

Introduction

Artificial intelligence (AI) is transformational technology that is dramatically changing the world in which we work and live. UCL's vision, 'AI for People and Planet', encapsulates our belief that the purpose of research and innovation in AI and related technologies is ultimately to benefit people and societies around the world, and to make a positive impact on the planet.

Coordinated by UCL Public Policy and the Office of the Pro-Vice Provost (Artificial Intelligence) with support from UCL Grand Challenges and UCL Innovation & Enterprise, the AI for People and Planet roundtable series brought together leading voices from academia, policy and industry. The goal was to shine a light on the transformational effects of AI on society, and draw out the challenges and solutions to how we can live well with AI.

'AI-related technologies' are broadly defined as the constellation of new, data-driven technologies that generate intelligent outputs. This includes deep learning, which has had notable success in domains like image/speech perception, as well as other data-driven technologies including robotics, digital devices, Internet of Things (IoT) and other forms of big data analytics.

The series centred on seven topics, including: Equity; Art Futures; Educating our Children; International Relations; Discovering New Medicines; Emerging Epidemics and Climate Change.
The discussions focused on the following policy challenges in the context of each of the respective topics:

Using AI to inform policy-making

Understanding AI ethics and safety

Improving the provision of public services

Governance for projects using AI

This booklet consolidates the policy commentaries resulting from each roundtable. This booklet aims to inform policy professionals, academics, decision-makers and industry experts on where public policy efforts and research agendas should focus in the coming years. Each commentary was written under Chatham House rules by the respective roundtable chair in partnership with the UCL Public Policy team, in consultation with roundtable participants. If you would like to contribute to these discussions, please contact ai.vision@ucl.ac.uk.

Discovering New Medicines

The aim of the Discovering New Medicines roundtable was to identify how the UK can build on its strengths to meet the challenges and opportunities in integrating AI capabilities across research, drug discovery and clinical development and valuation.

Executive Summary

The key themes and recommendations highlighted during the discussion included:

- AI has the potential to systematically and correctly identify potential new drug candidates.
- A significant barrier to the uptake of AI modelling in drug discovery is the lack of publicly available data and integration of health data across the NHS.
- Connecting the specialist expertise in small and medium enterprises (SMEs) and universities with the resourcing and infrastructure of larger companies would support greater data sharing.
- Having a common set of principles that frames what 'public good' looks like in the context of AI and drug discovery and development may help to accelerate cross-sectoral collaboration.
- There is a skills gap in industry amongst new graduates who lack foundational skills in data science, AI and machine learning.
- Training programmes that partner with industry would provide students with real world experience. The upskilling of existing employees is also needed.
- Greater diversity (i.e. sex, ethnicity, age and comorbidities) is needed in clinical trials to ensure AI is fit for purpose. Precision and personalised medicine, including new methods such as decentralised clinical trials and smart synthetic data, could support this increase.

Opportunities and barriers to accelerated adoption of AI in drug discovery

Pre-clinical and clinical applications

AI has the potential to accelerate the development or repurposing of drugs. The predictive ability of AI is beginning to complement wet labs and reduce Design-Make-Test-Analyse (DMTA) times.

> ***The RECOVERY trial***
>
> The RECOVERY trial is the world's largest clinical trial assessing treatments for COVID-19 with more than 30,000 participants in the UK. Its success is partially due to its use of existing health datasets. The trial was able to identify that dexamethasone, an inexpensive and readily available drug, reduces death by up to one third in hospitalised patients with severe respiratory complications from COVID-19.

Concurrent to the RECOVERY trial, AI VIVO, a company that uses AI to accelerate drug discovery and development, also identified dexamethasone a a candidate treatment for COVID-19. This demonstrates the potential of AI for systematically and correctly identifying potential drug candidates.


Adam Niescioruk

Regulation

The drug discovery and development community need access to healthcare data sets, but the regulations around accessing these data often pose as barriers. While large companies might have the infrastructure and resources required, SMEs may struggle to comply with current governance requirements (for example, not having a qualified clinical epidemiologist on staff). There is an untapped collaborative market pairing smaller companies with larger ones to match specialist expertise from SMEs with the resourcing and infrastructure of larger companies. Health Data Research UK (HDRUK) is an example of an expert service platform that is not only responsible for elements of data curation, but also 'match-makes' companies who need access to data with people with clinical expertise.

While pharmaceutical companies are limited by the scope of current legislation, the COVID-19 pandemic has shown the possibilities when commercial opportunity combines with public utility. What is now needed is a collaborative effort so that all the regulators, authorities and research groups can work in concert. The UK needs to be attractive to pharmaceutical companies, for example, for the purposes of research and development (R&D) or to be a first-to-market launch area. Continuing to develop regulations and ethical protocols that protect people's data - but that are not overly prohibitive - will be important. This is especially the case given the recent proliferation of AI start-ups and SMEs in the UK, many of which have valuable offerings, but this growth could be suffocated if regulations do not account for their size and needs.

Skills training

Data science and AI-related courses are on the rise within academia. As the field becomes more established at both graduate and undergraduate levels, programmes need to have greater consistency across the fundamentals and minimum set of skills that are taught.

Greater emphasis needs to be placed on providing training programmes to equip students with the right skills for the work force. For example, AstraZeneca's newly launched Data Science and AI Graduate programme: R&D provides graduate students with the opportunity to build their technical skills and develop industry knowledge. Additionally, existing employees need to be upskilled so that people with domain expertise in the life sciences can not only understand the results of AI models and applications, but also create new ones.

Data and partnerships

Open data sets within the private and public sector would enable a closer understanding of what has worked and what has not. Current systems for accessing data housed within the NHS can be difficult to navigate and require specialist knowledge that SMEs will not necessarily have. Generally, large corporations see data as proprietary information that should not be freely shared. Work may need to be done to shift perceptions to enable healthcare data to be seen in the same way as information published in academic papers (i.e. available for people to access and use to inform their own hypotheses and decisions).

The potential of 'synthetic data'

Once a team manages to access a dataset, there is often great variation across the country in how the data have been collected and recorded. One way that universities, including UCL, have been trying to address this is through the development of 'synthetic data' (e.g. datasets that are generated by computer programs) on which to train machine learning models. Hazy is a UCL AI spin out company that focuses on AI-generated smart synthetic data. While anonymised data still pose a risk to re-identification, smart synthetic data are statistically equivalent to raw data, but pose minimal threat to privacy. However, synthetic datasets that are built off of or mimic real data still risk violating confidentiality. Particularly if health conditions are rare. This necessitates a careful balance between having synthetic data that is'real' enough to be useable, but preserves privacy.

Joshua Coleman

Cost effective, fair and equitable access to new medicines

Ethics

An acknowledged issue with AI is that it may inadvertently exacerbate inequalities that have been unintentionally built into datasets. For a further discussion, refer to the section on Equity. As the field of precision medicine and personalised medicine continues to develop, trials need to be diverse in terms of sex, ethnicity, age and comorbidities. While this might not be difficult in large cities, diversity may pose a challenge to other parts of the UK, particularly in rural areas. More work needs to be done to recruit participants so that cohorts are representative of the target population. Decentralised trials (e.g. where a trial is conducted remotely using telemedicine and mobile/local healthcare providers) have the potential to democratise the clinical trial environment as they are able to recruit people who live far away from research sites.

Social contract

At the core of any work with AI in drug discovery and development is the question of how patients in the NHS can benefit from NHS data. The NHS is faced with a dilemma where they are concerned about the ethics of selling patient data, but equally do not want to give away data for free. If patient data are used to develop a new algorithm, should the NHS then have access to the service it feeds at a reduced rate? Additionally, there is also the question of data ownership; who created the value, and for whom value should be maximised (i.e. the patient, the hospital, the university, the wider NHS etc.)?

The COVID-19 pandemic has seen a drive for open innovation, which has accelerated collaboration and development across universities, regulators and industry. Having a common set of principles that frames what 'public good' looks like in the context of AI and drug discovery and development may help to further accelerate cross-sectoral collaboration in the future.

Underlying principles should strive to place the value gained from better curating and connecting data where it would have the greatest impact. For example, this might be a local Trust where enhanced data and their availability would help to improve health outcomes in local populations.

UK ecosystem

Many businesses working with machine learning draw talent from, or are founded by, people with academic backgrounds. Newly developed technologies may originally be housed within a university, which acts as the bedrock for further development. However, a major barrier facing academics and students who wish to found a business based on their research is the time and resources needed to negotiate the university's financial terms for the right to use intellectual property (IP) and data. More permissive public-private technology transfer systems (i.e. systems that facilitate the transfer of technology from universities or government laboratories to companies) would offer a solution to these delays.

The COVID Symptom Study

Throughout the pandemic, the health science company, ZOE, has collaborated with King's College London (KCL) to develop the COVID Symptom Study app. Scientists from KCL are using algorithms and machine learning models to analyse self-reported data from four million people to understand how fast the virus is spreading in different areas, identify high risk areas and identify who is most at risk. The study has shown the successes that can be achieved when a private company, academia and citizen science join forces. For example, the study's data models can rapidly identify hotspots, having identified Leicester as hotspot ahead of the Government placing the city back in lockdown at the end of June 2020.

Matchmaking larger organisations with SMEs and universities could help. However, the process of matchmaking is not yet optimised to ensure that technology transfer (e.g. the flow of technical knowledge, data, designs, etc. from one organisation to another) can take place, not only between large and small enterprises, but also between public and private organisations.

Conclusion

The Discovering New Medicines roundtable discussed the opportunities and barriers to accelerated adoption of AI in drug discovery and development and explored whether AI can fulfill the promise of delivering cost effective, fair and equitable access to new medicines While AI has the potential to systematically and correctly identify successful new drugs, more publicly available data and better integration of health data across the NHS is needed. Connecting the specialist expertise in SMEs and universities with the resourcing and infrastructure of larger companies would support greater cross-sectoral data sharing and collaboration.

Additionally, having a common set of principles that frame 'public good' in the context of AI and drug discovery and development may help to accelerate cross-sectoral collaboration and democratise the drug discovery and development process. Lastly, addressing the data science skills gap in new graduates, as well as upskilling existing employees in understanding and developing AI models, would create a workforce that is well-equipped to meet the challenges and opportunities in integrating AI capabilities across research, drug discovery and development and clinical development and valuation.

Participants

Dr Jane Kinghorn (UCL Translational Research Office, Roundtable Chair)*

Professor Delmiro Fernandez-Reyes (UCL Department of Computer Science)

Dr Anne Lane (UCLB)

Professor James Thomas (UCL Institute of Education)

Professor Matthew Todd (UCL School of Pharmacy)

Dr Junaid Bajwa (UCLH; Microsoft Reach)

Dr Nathan Benaich (Air Street Capital)

Dr Paul Colville-Nash (Medical Research Council)

Dr Sarah Dickson (UKRI)

Dr Aldo Faisal (Imperial College London)

Ms Hilary Newiss (National Voices)

Professor John Overington (Catapult Medicines Discovery)

Dr James Weatherall (AstraZeneca)

Emerging Epidemics

The aim of the Emerging Epidemics roundtable was to convene a group of experts in AI, public health, epidemiology and infectious diseases to clarify thinking about the relationship between AI and emerging epidemics.

Executive Summary

The key themes and recommendations highlighted during the discussion included:

- The descriptive, predictive and prescriptive role of AI prior to the current pandemic has rapidly evolved since the outset of the outbreak, leading to innovations, capabilities and enthusiasm for data science that have not previously been seen.
- The current and past focus has been on human-to-human transmission, but there needs to be an equal emphasis on integrating research on the disease outbreak cycles in animals (i.e. epizootic cycles of pathogens) with research on disease outbreak cycles in humans (i.e. epidemic transmission cycles).
- A major challenge to the widespread adoption of AI is lack of standardisation, integration and availability of health data.
- Automation of the way in which datasets are brought together would assist in making informed predictions about future events.
- There are multiple streams of data that feed into decision-making. Determining which datasets are appropriate to inform decisions, and also understanding the aim of decisions, is crucial if AI is to be used effectively.
- Alignment between organisations and interdisciplinary approaches are key for ensuring that data are used ethically and democratically.

AI's role in epidemic preparedness and response before the COVID-19 pandemic

AI had three main roles before the COVID-19 pandemic, which are outlined in Table 1.

Table 1: AI's roles pre-COVID-19

Role	Description of role prior to COVID-19	Example of role during COVID-19
Descriptive – what is happening.	A certain amount of data was available to monitor people's health and the spread of diseases, but rich, context-specific data were missing.	Mobility data reports from Google, Apple and Facebook etc. have helped map the spread of the virus and understand patterns in transmission.
Predictive – what is about to happen.	Research was based on classic AI models developed within an environment where time was available to understand models.	The timeliness of results has been much more important.
	Previously, AI was not able to assist in mapping out the potential effects o policy decisions.	New AI tools and algorithms are now being used to help determine which kind of policy interventions are needed to achieve the best outcome.

A 'One Health' approach to data

The pre-COVID-19 conceptualisation of models and AI tools for predicting the jump from animals to humans was limited, as there were little data available. Whilst, previously, the focus has been on human-to-human transmission, there now needs to be an equal emphasis on understanding disease outbreak cycles in animals (i.e. epizootic cycles of pathogens), as well as disease outbreak cycles in humans (i.e. epidemic transmission cycles). Adopting a One Health perspective means looking at the interaction of human health, animal health and the environment. However, there is currently no direct or regular exchange of information and data between these areas, beyond high-burden diseases, such as malaria or dengue fever. While datasets are not currently adequately integrated, there is scope to automate the way we bring these together to start making informed predictions about future events. Greater public awareness and platforms for data scientists working in ecology to communicate their findings are also needed.

Limitations due to lack of data access

Methods for modelling infectious diseases using online user activity data, such as web searches, had been in development before the current pandemic. However, there have been criticisms of past effort due to perceived biases, as well as unreliability in models caused by barriers in access to datasets. For example, models created during the Swine Flu pandemic overestimated flu prevalence and led to the decommissioning of Google Flu Trends.

This issue has been partially solved by encouraging Google to offer access to aggregate search activity data as well as research groups developing more advanced machine learning models that can mitigate the issues of their predecessors. However, this body of work has focused on diseases for which there is a well-established evidence base. Limited focus has been placed on forecasting, which meant that existing modelling methods were unprepared to manage the onset of the COVID-19 pandemic. More rigorous modelling approaches for estimating influenza prevalence based on web searches and their incorporation into national health surveillance systems served as the foundation for building novel models.

Challenges and opportunities for the use of AI in mitigating effects of COVID-19

Challenge – validating models

Since the outset of the COVID-19 pandemic, there has been a rapid progress in AI's capacity to manage, monitor and forecast outbreaks. However, the issue of missing data, especially for people from minority ethnic groups in the Global North or people in the Global South, is still posing a major challenge to the validity of models. Lack of access to appropriate data is particularly problematic for resource-limited countries where routine, electronic health record systems for human health do not exist at the national level.

Initially, models were also being trained on highly skewed datasets, such as those with disproportionately high numbers of those without the disease. As clinical understanding of the virus has improved, models have been adapted locally, which has been key in the context of very little and often low-quality data. Additionally, having greater domain expertise in public health and epidemiology is also crucial for the relevance of models
and to contextualise results.

Challenge – value-based data

Value-based data (i.e. social and behavioural practices and personal beliefs) are not easily collected in datasets. Being able to systematically feed this kind of data into models still poses a challenge.

UCL COVID-19 Social Study

The UCL COVID-19 Social Study is the UK's largest study into the psychological and social impact of the pandemic. With over 70,000 participants, the study provides weekly data to decision-makers both nationally and internationally. Negative attitudes towards vaccines are a major public health concern and the Social Study was able to identify predictors of COVID-19 vaccine refusal.

Opportunity – leveraging enthusiasm for data science

There has been a newfound interest in data science amongst both the research and policymaking communities. The rapid genetic sequencing of the virus and subsequent sharing of this bioinformatics data was a remarkable feat that has continued into the second year of the pandemic. Genome sequencing has enabled researchers to explore different diagnostic, treatment and vaccine interventions and played a large role in the management of the outbreak. As the amount of data continues to grow, policymakers and decision-makers have called for dashboards to help make sense of the data in real time.

i-sense COVID RED

To address gaps in available data, UCL researchers have led the rapid adaptation of existing i-sense technologies, which aim to identify outbreaks of infectious disease, to create the i-sense COVID Response Evaluation Dashboard (COVID RED). COVID RED collates and presents data from the Office for National Statistics (ONS), Public Health England and NHS and is currently the only dashboard that explores the entire COVID-19 response system as a whole.

Opportunity – evidence synthesis

At the time of writing (March 2021), over 80,000 papers had been published on COVID-19. Using AI and natural language processing (the analysis of language in text and speech) to synthesise the relevant and high quality studies from this massive body of evidence could assist decision-makers by identifying effective interventions and key patterns and trends across studies.

Changes needed to enable AI to help build equitable and resilient health systems

Comprehensive data streams and frameworks

There are multiple streams of data that feed into decision-making, including health management information systems, routine disease-specific records, city information, latency operation data and satellite imagery. Determining what datasets are appropriate to inform decisions, as well as understanding the aim of these decisions, is crucial if AI is to be used effectively. Additionally, a greater understanding of which health and ecological data should be monitored is needed, especially data related to zoonotic origin diseases.

To improve standardisation in health data, machine learning could be used to read, decipher clean and standardise the data to facilitate its subsequent use in models. Robust policy frameworks are needed to create environments that include regulatory measures, incentive programs and research streams to address public health priorities.

Accounting for national readiness

An AI maturity model provides a framework for assessing an organisation or country's current AI readiness and capabilities. Assessing at which stage (exploring, experimenting, formalising, optimising and transforming) in the maturity model a country resides is crucial for determining what AI and digital health technologies can be implemented.

Democratising AI

Alignment between organisations in different policy areas and sectors and interdisciplinary approaches are key for ensuring that data are used ethically, with considerations for how the data are likely to be used and accessed in the future. Collaboration between academia and industry to increase the recruitment of people from diverse backgrounds into the field of AI would assist in democratising the field. Crucially, issues of equity and social justice must be considered at every point to ensure that AI models do not exacerbate existing inequities.

Conclusion

The ongoing COVID-19 pandemic has led to rapid advancements in the use of AI modelling and technologies for controlling and managing its impacts. In order for these advancements to continue to progress effectively, system-wide changes are needed to better integrate and standardise health datasets. While the past focus has been on human-to-human transmission, there now needs to be an equal emphasis on examining epizootic cycles of pathogens and epidemic transmission cycles. Beyond managing disease outbreaks, using AI-based algorithms for analysing health data will require policy and governance frameworks to ensure that AI is used ethically and democratically.

AI has the capability to play a significant role in predicting and managing emerging epidemics; what is now needed is alignment across sectors to realise its full potential. However, there remains significant variance in how different populations and communities either benefit from AI or are harmed by it. An interdisciplinary agenda of research and action is needed which acknowledges the importance of focusing on people, processes, and politics, going beyond the technical discussions of AI towards how it can be used to achieve impact in the real world.

Participants

Dr Lele Rangaka (Institute for Global Health, Roudndtable Chair)*

Dr Yipeng Hu (Faculty of Engineering Science)

Dr Vasileios Lampos (Faculty of Engineering Science)

Dr Audrey Prost (Institute for Global Health)

Professor Stephen Roberts (Institute for Global Health)

Mr Amane Dannouni (Boston Consulting Group)

Mr Nazir Halliru (Columbia University)

Dr Nick Jackson (Coalition for Epidemic Preparedness Innovations)

Dr Ramesh Krishnamurthy (World Health Organization)

Professor Trudie Lang (University of Oxford)

Dr Maimuna Majumder (Harvard Medical School)

Dr Mosa Moshabela (University of KwaZulu-Natal)

Dr Leah Mwai (UK FCDO East Africa Research and Innovation Hub)

Professor Barbara Prainsack (King's College London) Dr

David Redding (Zoological Society of London)

Dr Tara Vishwanath (The World Bank)

Equity

The purpose of the Equity roundtable was to understand AI's potential impact on Equity. To do this, discussion covered the range of threats, identified means to mitigate them and conceptualised how the considerations introduced could be integrated into existing ethical frameworks.

Executive Summary

The key themes and recommendations highlighted during the discussion included:

- When a policy is based on the result of an AI model, the public, in theory, should be able to assess the model and the principles on which the decisions were based as far as possible.
- Being able to understand what variables and assumptions a model is based on and compare similar models with each other is essential in helping to avoid biases. Conversely, it may not be necessary to know how a model works, but whether there is discrimination or bias.
- AI's ability to single out individuals is not just a threat to equity in making it possible to discriminate against them, but also a means of detecting and rectifying inequity.
- Full agency should not be given to the machine; accountability rests on the shoulders of humans.
- The principles of accountability and openness are particularly relevant to discussions on the interface between AI and equity. Moral accountability for the results of an algorithm should be ascribed to the person, or group of people (Government ministers, CEOs, etc.), who authorised the AI system.
- Having effective and advanced governance in the form of clear legislation, risk management structures, and monitoring and evaluation is crucial in accounting for equitable considerations that are adaptable to different contexts.
- To judge whether or not an AI model is 'good', we need clarity on the demonstrable difference to society its use could make.

Equity in the context of AI

It is important to distinguish between equality and equity. Equality refers to the uniformity of a treatment or an outcome; equity to its proportional sensitivity to circumstances of possibility and need. Equal treatment blindly applied across all groups, may unintentionally yield unequal outcomes. The moral consequences of each aspect need to be examined on a case by case basis.

The World Health Organization has published a report outlining equity considerations for the use of AI, which include:

- the digital divide (i.e. the gap between people at different socio-economic levels regarding their opportunities to access information and communication technologies);
- algorithmic bias and values;
- plurality of values across systems;
- and fair decision-making procedures.

The report also argues that AI technologies should be designed and implemented to actively redress or eliminate inequities and promote greater equity, while also not sustaining or exacerbating existing inequities.

Principles of Public Life – accountability and openness

The Committee on Standards in Public Life offer guidance on The Seven Principles of Public Life, which apply to anyone who works as a public office-holders, for example, Government ministers, local government officials and civil servants. The principles include: selflessness, integrity, objectivity, accountability, openness and honesty. Accountability and openness are particularly relevant to discussions of AI and equity.

Accountability

While emphasis has been placed on ensuring that humans can override a result, the issue of responsibility is rarely discussed. Automated decision-making is the process whereby algorithms make decisions without any involvement, for example, in determining the decision of a loan application or an aptitude test for recruitment tests. Responsibility cannot simply be ascribed to an algorithm, so who is accountable for the outcome of an automated decision?

It may be argued that there are at least two forms of responsibility – causal and moral, differentiated in Table 2.

Table 2.

Form of responsibility	Description	Example
Causal	Purely causal responsibility arises when an agent causes an outcome, but without coherent purpose or intent. AI systems can have causal agency when, for example, their predictions guide policy decisions on welfare payments.	A child accidentally tripping and breaking his/her parents' vase is causally responsible for the damage.
Moral	Moral responsibility extends to causation whose agent can be meaningfully said to be responsive to reason and held to account by it. An AI system cannot be held morally responsible as it is not a rational agent.	If the tripping child in fact intended to break the vase and did so on purpose, she/he would then also be morally responsible.

If the argument in the table above is followed, on what grounds is responsibility assigned? Rather than placing it solely onto software developers, moral accountability would be passed to the person, or groups of people (Government ministers, CEOs, etc.), who authorised use of the AI system and approved its characteristics.

The A-level and GCSE results 'fiasco'

As the A-level and GCSE results 'fiasco unfolded, the Government's initial response was to ascribe responsibility to the algorithm itself. While Sally Collier, head of Ofqual, England's exam regulator, did resign from her post, many at the time called for Gavin Williamson, Secretary of State for Education, to resign, as the one with ultimate accountability. The decision for the algorithm to base students' results on those of peers from similar schools and backgrounds runs counter to the very principles of equity.

Openness

AI's ability to single out individuals (individuating model power) represents one of its greatest strengths, and conversely, one of its greatest risks. This ability to differentiate between individuals also makes it possible to discriminate between them. When a policy is based on the results of an AI model, the public, in theory, should be able to assess the model and the principles on which the decisions were based, as far as the necessary complexity of the model allows it. Equally, when a decision has led to instances of inequity, the openness and transparency surrounding how the decision was made is crucial.

There are technical solutions that can be used to reduce the risk of discrimination, such as limiting:

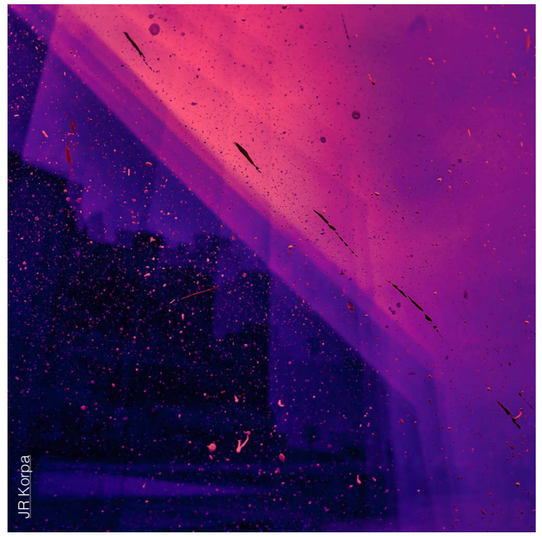
- individuating model power;
- the complexity of the algorithm;
- and autonomous applications (software that is able to make decisions without the authorisation of human operators).

The assumptions behind these solutions is that non-individuating models (those that cannot identify individuals from datasets) are inherently fair, simpler models are better, and the scope of complex models should be limited. However, simple models should not always be preferred merely because they are easier to understand, as reality does not always map onto a simple model. Any bias against complexity is misguided, as what matters is the assessment of whether the model works. There is a trade-off between having a system that is adequately complex to perform sophisticated functions and being able to explain the intricacies of the system to society if something goes wrong. While there are model fit diagnostics (i.e. tests used to evaluate a model's assumptions and investigate whether specific variable are having a large, undue influence on the analysis) the results are not always easily interpretable.

The role of governance

Decision-makers need to proactively consider the consequences of the use of models and should be able to evidence that a particular model being considered is the most appropriate. This requires people in leadership positions to understand the variables that go into an AI system and use their professional expertise and experience, in conjunction with the results from the model, to come to a decision. For decision-makers to understand the variables in the model, data scientists building the algorithms should produce documentation or metadata (i.e. file size, author, data was collected) that make finding and working with particular instances of data easier. Conversely, there is also the argument that it is not necessary to know how a model works; what is needed is knowing whether there is discrimination or bias.

A solution to ethical considerations is having effective and advanced governance and sufficient documentation – clear legislation, impact assessments, proper controls and risk management structures, setting responsibility for decision-making, monitoring and evaluation – that is adaptable to different contexts. It is the role of the person or group of people who commission the development of an algorithm to ensure that the appropriate guidance and checks are followed. Since ethical considerations are already entrenched in governance, the challenge now is how to enforce them.


JR Korpa

Equity in AI and public policy

Conveying the 'truth' within policy frameworks

Similar to issues raised in the other commentaries resulting from this roundtable series, the lack of data to test models on subgroups of people from different racial, ethnic and socioeconomic backgrounds has made it difficult to ensure that equity and diversity are being adequately accounted for. Being able to understand what variables and assumptions a model is based on and being able to compare similar models with each other is essential.

Achieving a model that is accurate in its individuation without being biased is a challenge, as fairness and equity cannot be analysed as dichotomous variables (i.e. 'equitable' or 'not equitable'). When people are exposed to risk, i.e. having their exam results predicted by an algorithm, it should be considered whether they would reasonably consent (or not) to this risk. When the model's outcomes negatively affect people, an equity lens must be used to examine whether it was wrong across all population groups (thus, in a sense, fair across all groups), or whether the model worked disproportionately against those from protected characteristics and other underrepresented groups.

Safeguards

Model transparency would help by revealing what variables were included in the model and how they were weighted. However, this is not always possible in a 'black box' AI system in which inputs and operations are not visible. AI models need to be assessed in a way that is workable for policy and, as far as possible, understandable by the public. Accountability should also be agreed on beforehand so that the appropriate structures are in place in the event that the model does make an error or the public disagrees with the results.

Conclusion

The Equity roundtable discussed AI's potential impacts and threats to equity and a conceptualisation for how the considerations that AI introduces can be integrated within existing ethical frameworks. AI's ability to single out individuals (individuating model power) represents one of its greatest strengths, and conversely, one of the greatest threats to equity. To mitigate these risks, effective and advanced governance in the form of clear legislation, risk management structures, and monitoring and evaluation is critical. Principles of accountability and openness must be considered alongside equity frameworks, with the moral responsibility of results of models being attributed to those who authorised the AI system as fit fo purpose. Being able to understand what variables and assumptions a model is based on and being able to compare similar models with each other is essential in helping to avoid biases. Conversely, it may not be necessary to know how a model works, but whether there is discrimination or bias. Ultimately, accountability rests on the shoulders of humans.

Participants

Professor Parashkev Nachev (UCL Queen Square Institute of Neurology, Roundtable Chair)*

Dr Zeynep Engin (UCL Computer Science) Ms

Amy Austin (Cabinet Office)

Mr Parker Mss (Genomics England)

Ms Zoë Porter (University of York)

Dr Roland Tegeder (SW7 Group)

Mr Paul Trenell (Government Equalities Office)

**Individuals marked with asterisks were co-authors of this particular commentary.*

Educating our Children

The aim of the Educating our Children roundtable was to assess how the UK education industry can leverage the benefits of AI effectively and to identify strategies for overcoming any obstacles on the way.

Executive Summary

The key themes and recommendations highlighted during the discussion included:

- Opportunities that AI, learning technologies/ tools and data science offer to education include assisting teachers in highlighting learning gaps, analysing large datasets to identify potential points for intervention and scaffolding content to facilitate students' progression.
- AI should be leveraged to not only personalise and tailor the education process for individual students, but to also include the students who would otherwise struggle to access basic education.
- Inequity in access to and data collected from AI technologies must be addressed in order to ensure systemic inequalities are not exacerbated.
- Education technology should support the development of agency and self-regulation in students to instil values of lifelong learning.
- Forming partnerships across sectors to create a more equitable, resilient and adaptive education system presents an opportunity for co-design as the UK enters the post-COVID-19 era.
- To put AI integration at the forefront of education policy in the UK, a future education group with stakeholders from different sectors, should be established.

Leveraging the benefits of AI in education

Addressing existing social inequities

Opportunities that AI, learning technologies/tools and data science offer to education include assisting teachers in highlighting learning gaps, analysing large datasets to identify potential points for intervention and scaffolding content (i.e. breaking up the learning into chunks and providing a tool, or structure, with each chunk) to facilitate students' progression. However, these opportunities cannot be fully realised without first addressing issues of social inequity. The COVID-19 pandemic has deepened existing structural inequalities – a report from the Institute of Fiscal Studies found that children from poorer families spent 30% less time at home learning than their wealthier counterparts. Proponents of 'EdTech' (i.e. education technology) may overlook the safety net function of schools.

The integration of AI and other technology in education should be central to the Government's 'levelling up' agenda. Similarly, the pledge of extending broadband coverage across the country would facilitate the access to education nationwide.

> ***A-level and GCSE results***
>
> The A-level and GCSE results 'fiasco' in August 2020 demonstrated how the integration of AI into the education sector has to account for issues of equity and fundamental human rights. Current policies and regulations do not provide access to data from constituencies where students have the greatest support and learning needs. This raises questions as to how to ensure that data collection is inclusive and representative.

Infrastructure improvements

Any actions taken to make datasets more inclusive must be coupled with infrastructure improvements and capacity building amongst the education sector to manage increasingly large and complex datasets. Moreover, drawing conclusions from the data captured may be difficult simply because learning science as a discipline is still not well-understood by AI developers, and vice versa.

> ***Lexplore***
>
> Lexplore, a Swedish company, has developed a system that quickly scans for students at risk and detects dyslexia by tracking reader's eye movements. The company has expanded to the UK and the system is now supported by the British Dyslexia Association. Advancements that Lexplore has made have demonstrated the opportunities that can be realised when synergies between educators and AI developers takes place.

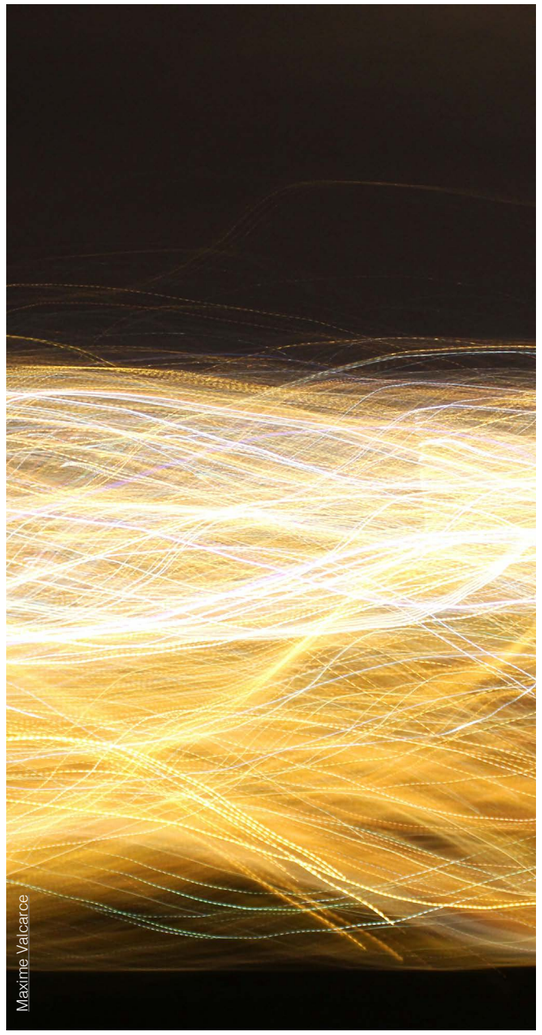
Supporting agency and self-regulation

The overall goal of education is to build a sense of agency and self-regulation in children and young people, allowing them to continue learning throughout their lives. Ideally, education technology should support this. However, in practice, software modelled on the gamification of learning imparts an extrinsic value for learning (i.e. if I learn this, I get a reward), when the emphasis should instead be placed on teaching an intrinsic value for education (i.e. if I learn this, I will learn skills I can apply to other aspects of life). AI could support the education process by assessing student's motivations (i.e. extrinsic or intrinsic) for those who are developmentally ready to engage with self-directed learning.

> ***Who are the 'stakeholders' in education?***
>
> In education, the term 'stakeholder' typically refers to anyone who is invested in the welfare and success of a school and its students. This could include administrators, teachers, staff members, students, parents, families, community members and local business leaders. Stakeholders may also be collective entities, such as local authorities or local businesses, as well as organisations that represent specific groups, such as unions, and professional associations.

UK education system

Despite the advances made during the pandemic, the UK's education system remains relatively 'low-tech', focusing on developing routine cognitive thinking in children. While it is not unusual for education systems to lag behind technological progress, a failure to remain abreast coupled with the automation of many jobs may lead to social crises and issues of unemployment in the future. This is accompanied (not only nationally, but globally) by the rapidly rising costs of education to the taxpayer and decreasing growth in productivity. Working with organisations such as the OECD's Directorate for Education and Skills, which has a cross-cultural and cross-sectoral reach, and private educational software companies, to introduce technological advances in pedagogy will enable the UK education system to be more resilient and adaptive as it enters the post-COVID-19 era.



Maxime Valcarce

Engaging across the education ecosystem

'Bottom-up' initiatives

Government leadership is important in progressing innovation in the education sector, but a diverse range of stakeholders, including teachers, parents and staff members, should be closely involved in the process, resulting in bottom up and market led innovation. The teaching community needs to be supported so that members can familiarise themselves with the new technology and then use it to address core problems. Involving every teacher and parent in the innovation process is likely to be too complicated and cumbersome. Therefore, the integration of AI into education could begin on a small, grassroots scale, while allowing sufficient time to address potential risks posed by new technology, before technologies are rolled out nationwide.

Cross-sectoral engagement

Progression relies on endorsement from the Government, as independent efforts from the market have not been successful to date. Policies are needed to determine the remit and structures in which EdTech companies should be operating. Education advocates also need to work with policymakers to facilitate step-changes in schools. Working across sectors and forming partnerships between stakeholders, including parents, teachers, software companies and public institutions to create more equitable systems presents an opportunity for co-design moving forward.

Funding, regulatory, ethical, policy and attitude changes needed

Demonstrating AI's potential

Changes in attitudes towards, and discourse around, AI are needed to modernise education in the UK and demonstrate how AI can facilitate progress in schools. Software companies and educators need to demonstrate how technology can quickly identify gaps in existing schooling methods. To put AI integration at the forefront of education policy in the UK, a future education group with stakeholders from different sectors should be established.

Central role of teachers

The lack of involvement of the teaching community in the innovation process has hindered the UK's uptake of technology in the education sector. Additionally, industry professionals have expressed concerns that the role of the teacher will be diminished to the point of obsolescence if advanced technology is integrated into education. Senior administrators and decisionmakers must recognise that teachers are crucial in spearheading changes in the classroom.

AI as a mechanism for achieving greater equity

Another crucial change would be to ensure that AI is leveraged, not only for personalising the education process, but also for including the students who would otherwise struggle to access basic education. This can be done through aggregating data to take account of each student's personal circumstances, in order to learn what their needs are. However, this in turn can only be achieved by ensuring that every child has the necessary hardware and software, including a stable broadband connection, to engage with the technology. Otherwise, the attainment gap may only become wider.

Conclusion

The Educating our Children roundtable discussed that, while AI is well-placed to improve access to education, there are challenges that need to be addressed before these technologies can be integrated into education systems on a wider scale. EdTech can highlight learning gaps, analyse large datasets to identify points for intervention and scaffold content to help students' progression. However, using AI technologies also risks exacerbating existing inequalities and care must be taken to ensure that datasets include as diverse and representative a population as possible. A centralised and coherent strategy for integrating AI into education is needed. Forming partnerships across sectors to create a more equitable, resilient and adaptive education system presents an opportunity for co-design as the UK enters the post-COVID-19 era. An adaptable and advanced education system that supplements the knowledge and expertise of teachers and parents is crucial for the next generation of young people in the UK to be equipped with the values of lifelong learning and to succeed in the workforce.

Participants

Professor Rose Luckin (Knowledge Lab, Roundtable Chair)*

Professor Allison Littlejohn (Knowledge Lab)

Dr Chris Tyler (STEaPP)

Mr Simon Allen (McGraw Hill)

Ms Andrea Carr (Rising Stars, Causeway Education, Sumdog, Langley Park Learning Trust)

Lord Tim Clement-Jones CBE (Queen Mary University of London, Institute for Ethical AI in Education, House of Lords)

Ms Tabitha Goldstaub (AI Council, CognitionX, CogX)

Mr Lee Hodgkinson (OSTC Ltd)

Ms Joysy John (Nesta)

Ms Priya Lakhani OBE (CENTURY)

Dr Dirk Van Damme (OECD)

**Individuals marked with asterisks were co-authors of this particular commentary.*

International Relations

The aim of the International Relations roundtable was to explore how AI-related technologies will shape global and international relations over the next 10 years.

Executive Summary

The key themes and recommendations highlighted during the discussion included:

- The global development of AI standards, regulations, and ethical frameworks is uneven, but there is scope for influential actors to effect significant change.
- Although much emphasis is placed on the potential of AI to solve (or exacerbate) global problems, the core components of the technology – data, algorithmic source code, and engineering talent – lie with private companies.
- Novel relationships are needed between the state and private sector to ensure that AI-based technologies are developed and implemented in a way that advances societal values and progress.
- AI-based technologies provide opportunities for new actors and agents to engage in surveillance and defence activities, which previously only major powers would have been capable of.
- The COVID-19 pandemic has highlighted that the greatest global challenges can only be prevented and addressed with effective global governance.
- AI is disrupting relations and distributions of power and control.

Shifting dynamics of power and control

A global competition is now underway as major powers strive to position themselves at the cutting-edge of AI innovation and strive to benefit economically. This process of capacity building could alter the balance of power between states, and even increase the attractiveness of different political regimes, especially in emerging markets. China, for example, is building its domestic AI industry and hopes to become the 'leading AI power by 2030.'

AI-based technologies are also disrupting distributions of power between states, private companies, and individuals. Companies and states are using techniques such as AI-enabled surveillance, 'big nudging' (i.e. digital architecture that influence someone's behaviour) and online micro-targeting (using personal data to identify the interests of a specific audience) to manipulate and control the behaviour of individuals at scale. This appears to be continuing unabated, with wide ranging implications in almost every domain, most notably in elections, public discourse and access to information.

AI can also be a force for exclusion and discrimination, thereby reducing the autonomy and control of individuals or marginalised groups. The adoption of certain AI-enabled Internet of Things (IoT) technologies, for example, could exacerbate gender-based violence and domestic abuse.

Cooperation, coordination and competition

Novel relationships between the state and private sector

One of the most significant shifts in power is from the state to the private sector, especially with leading technology companies developing AI-based technologies. Although much emphasis is placed on the potential of AI to solve (or exacerbate) global problems, the core components of the technology – data, algorithmic source code, and engineering talent – lie mainly with private companies. Although many AI engineers are trained in public universities, the benefits of working in the private sector are often hard to ignore.

The lack of agility and AI expertise within governments is cause for concern, as the current system is essentially reliant upon the private sector to develop and deploy AI-based technologies for the good of humanity, which cannot be guaranteed. This evolving situation calls for novel relationships between the state and private sector – underpinned by ideological renewal – whereby government interventions foster the right conditions and increase the likelihood that AI-based technologies are developed, deployed, or even scaled back in a way which advances societal values and progress. For example, the European Union is attempting to build an AI 'ecosystem of excellence', predicated on new private and public sector partnerships, which creates incentives to 'accelerate the adoption of solutions based on AI'.

The potential for AI-based technologies to solve major problems depends upon the effective 'intelligence assembly' capabilities (i.e. combining expertise, technological systems and datasets from private and public sources for public benefit) of governments worldwide. Conversely, undesirable private-public sector relationships are emerging whereby authoritarian governments seek to access and synthesise private sector data sets, in order to advance agendas of surveillance and social control.

Opportunities for new actors and agents

AI-based technologies provide opportunities for new actors and agents, including 'amateurs' and groups with limited resources, to carry out activities that previously only major powers would have been capable of. This includes sophisticated surveillance and defence activities such as open source imagery, drone strikes, and even online investigations by groups such as Bellingcat. In this sense, there has been a democratisation of technological tools and capabilities, which governments may find difficult respond to or control, and whose long-term consequences are difficult to predict.

The uneven development of standards, regulations, and ethical frameworks

Existing AI governance strategies and ethics codes broadly agree on the importance of fairness, transparency, and accountability. However, how these principles are translated into concrete standards, regulations, and ethical frameworks is more varied.

The EU, for example, lacks large global technology companies but has aimed to exert influence through regulation. In the domain of ethical governance, human-centred AI, and democratically legitimate standards, the EU is currently forging a 'regulatory framework for trustworthy AI'. The European Commission's recent 'White Paper on AI' could be the first step towards new legislation, which in turn could lead to a 'Brussels effect', where global companies follow at least some EU AI rules in order to participate in its market. However, the ability of the EU to influence the development of regulations and ethical frameworks outside its borders is contested. Moreover, the EU's approach stands in contrast to other influential regimes, such as China. It is likely that highly divergent standards and ethical frameworks will emerge worldwide, and there could be pressure on states to follow specific models. For example, China could use its investments and economic partnerships in African nations in order to embed technological infrastructure underpinned by Chinese 'AI values'.

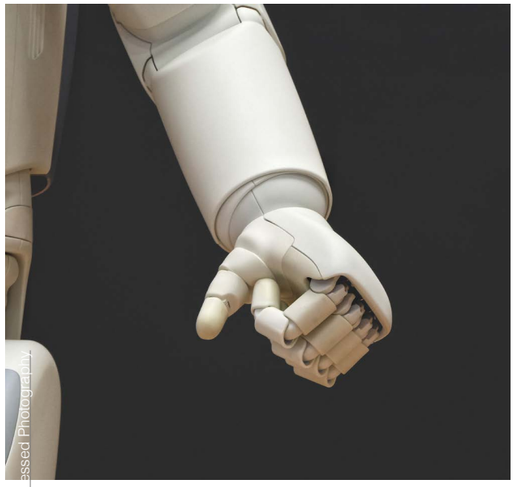
Various regulatory models are likely to emerge, and many nations will be following the model of one of the major powers. Global regulatory powers like China, the EU, and the U.S. will be considering how best to influence the states within and beyond their spheres of influence. An interesting question is the extent to which the UK will diverge from the EU's standards in this domain. Given the level of UK policy and research activity, UK divergence could be significant, with potential economic implications.

Priorities over the next 10 years

Look past the hype and focus on the 'boring'

The reality of AI does not yet warrant the dystopic 'hype' and fervour surrounding it. As such, decision-makers should focus on the more mundane applications of AI-based technologies, which may, unbeknown to most, already be widely deployed and driving significant social, political, and economic change. For example, in the armed forces sector, attention is placed on headline grabbing technologies, such as autonomous weapons or the anticipated use of AI in command and control functions. Focussing on such 'glamorous' or futuristic AI applications detracts from existing AI technologies, in domains such as supply chain management or predictive maintenance, which are transforming militaries worldwide.

Similarly, the rapid and widespread adoption of digital technologies driven by the COVID-19 pandemic is another example of existing AI-based technologies effecting significant change. Technologies that underpin social media, video conferencing, and other communication and collaboration tools are now being used at an unprecedented scale across virtually every domain, representing one of the largest social innovation experiments. For example, it is now standard practice for health care services and education to be delivered remotely through digital platforms. While the technologies that underpin these platforms are not 'glamorous,' there is a major opportunity to use them to reduce inequalities, advance social progress, and develop new norms. As such, consideration must be made as to whether digital technologies should be classified as a vital resource or whether access to these technologies should be a fundamental right.



International governance of AI

There are currently no global, multilateral bodies exclusively focussed on governing AI-based technologies that would enable states to deliberate, develop norms, and set agendas on issues ranging from algorithmic discrimination to AI in warfare. In an era of rising great power competition, the creation of new multilateral institutions or global AI treaties would be fraught by years of negotiations, and, moreover, is highly unlikely. As such, there is the risk of a 'governance vacuum', whereby global AI standards and innovation evolves in a disparate fashion, with a lack of coordination and cooperation among major powers.

Given the current global context, the focus should be on utilising, maximising, and strengthening the potential and scope of existing instruments and institutions, in order to advance the development of shared standards and solve global problems related to AI.

There are many existing forums in which (some) states, and other actors, such as civil society organisations and private companies, cooperate in this realm. For example, the Organisation of Economic Co-operation and Development members have agreed upon AI Principles, and organisations like the Institute of Electrical and Electronics Engineers play a key role in the development of global industry standards. Finally, the United Nations – and the implementation of the Sustainable Development Goals – may help, although this increasingly requires liberal democracies to compete with authoritarian states in UN bodies like the International Telecommunications Union.

Conclusion

This roundtable on AI and International Relations discussed how AI-related technologies will shape global and international relations along their political, security, economic, international development, social and business dimensions. Notably, there has been a shift in power from the state to the private sector, with the core components of the technology – data, algorithmic source code, and engineering talent – resting mainly with private companies. Novel relationships are therefore needed between the state and private sector to increase the likelihood that AI-based technologies are developed, deployed, or even scaled back in a way that advances societal values and progress.

The global development of AI standards, regulations, and ethical frameworks is uneven, but there is scope for influential actors to effect significant change. Decision-makers can ensure they remain at the forefront of developing trends and subtle changes in global affairs by placing greater emphasis on the routine use of AI-based technologies. Doing so will enable governments to build public acceptance and engender trust in the technologies that are increasingly shaping the world.

Participants

Dr Nick Wright (Roundtable Chair)*

Professor Joanna Chataway (STEaPP)

Dr Zeynep Engin (Department of Computer Science)

Professor Sir Geoff Mulgan (STEaPP)

Mr Oliver Patel (European Institute)*

Maj Gen (Retd) James Chiswell

Dr Rogier Creemers (Leiden University)

Dr Al Fisher (Buro Happold Engineering)

Sir Lawrence Freedman (King's College London)

Ms Claire Hancock (HM Government)

Dr Bryn Hughes (Defence, Science and Technology Laboratory)

Mr Shashank Joshi (The Economist) Professor

Helen Margetts (University of Oxford)

Mr Paul Nemitz (European Commission, DG JUST)

Sir David Omand (King's College London)

Ms Agnieszka Wierzbicka (European External Action Service)

**Individuals marked with asterisks were co-authors of this particular commentary.*

Art Futures

The aim of the Art Futures roundtable was to identify ways to leverage the transformational effects of AI to enhance the UK s creative strengths.

Executive Summary

The key themes and recommendations highlighted during the discussion included:

- To support the adoption of AI, Arts Council England (ACE) aims to provide programmes and platforms for artists to experiment with the AI technology, but this will require more comprehensive, long-term sectoral funding.
- Universities can bridge the gap between an arts education and an arts career. This could be done through creating forums across the country for people from different sectors and building capacity amongst emerging artists to understand and use AI technologies.
- Building cross-industry partnerships between private corporations and the arts sector would benefit the Creative Industries by giving artists the resources, visibility and expertise to expand and explore their practice.
- AI and the creative industries would benefit from long-term innovation funding and investment as well as Research Council support for PhDs and Doctoral Training Centres.
- Ethics and equity are major concerns for the application of AI, as individual biases, preconceptions, and assumptions will be amplified by algorithms.
- Artistic platforms could be used as environments in which to identify and address issues in new technology and introduce these to wider audiences.

AI-enabled ambitions for the creative and cultural sector

The need for a long-term strategy

Arts Council England (ACE) recently published their new 10-year strategy, 'Let's Create'. Developmental goals include:

- broadening access to creative experiences, specifically to those in culturally deprived areas;
- fostering resilient and cohesive cultural communities;
- and making sure the UK can maintain its reputation for high quality, innovative arts and culture.

AI can facilitate the achievement of these goals. To support the adoption of AI, ACE aims to provide programmes and platforms for artists to experiment with and understand the impact of AI and other technologies on existing intellectual property laws and advocate for changes where needed. This will require wider sectoral funding. The UK government's £1.57 billion investment to protect Britain's cultural, arts and heritage institutions provided a vital lifeline to the sector, but many jobs remain unprotected.

Commercial sector

As the UK begins to reopen its arts and cultural sectors post-COVID-19, instead of innovating, museums and other arts institutions may feel compelled to engage in proven and profitable strategies to recover from the crisis rather than pursuing more innovative models

Academic sector

Universities are well placed to bridge the gap between an arts education and an arts career, as there is often no clear pathway for emerging artists. This could be done through creating forums across the country for people from different sectors (e.g. software programmers, artists, corporations, etc.) so that new artists can develop research questions and projects that would then elevate them into the professional artists' community. Capacity building is also needed to ensure artists have a sufficient level of understanding of AI and technology to enable them to navigate the industry. While some industry giants, such as Adobe and Facebook, have artist-in-residency programmes, such opportunities are limited in number. More sustainable and long-term funding and programmes are needed so that new creators can understand AI and produce art based on it.

Leveraging the uptake of technology to progress effectively and ensure the benefits of AI are available to the creative sector

Cross-industry partnerships

Creative Industries (i.e. businesses that centre on creativity, for example, design, music, film and video, visual arts, TV and radio, and the performing arts) in the UK would benefit immensely from collaboration with domestic technology companies so as to have the resources, visibility and expertise to expand their practice. Currently, many artists, such as Karen Palmer, a digital filmmaker and storyteller who attended the roundtable, are left to establish these partnerships on their own, which requires considerable time and resources. UK Research and Innovation (UKRI) can help the arts sector broker such relationships and act as an intermediary to strengthen connections. UCL Innovation & Enterprise offers one mode for such assistance through providing partnership support at the interface of academia and industry, facilitating introductions and developing long-term, multi-disciplinary strategic partnerships. Conversely, AI developers, especially those still at university, could benefit immensely from partnerships between creators and academic institutions.

Living Archive

Living Archive is a collaborative experiment between Studio Wayne McGregor and Google Arts and Culture in which machine learning technology learns and recreates the style of any dancer. The collaboration arose organically - a Programme Manager from Google Arts and Culture who had previously worked with Wayne McGregor invited the studio to meet with the Director of Google Arts and Culture and the Lab in Paris. The crucial link was having someone in the tech institution with experience in the arts who could curate artists, as well as having a project of mutual interest that pushed the boundaries of both partner organisations. Living Archive offers a valuable case study of how effective partnerships can be built between a large private corporation and the arts sector.

Funding

Nesta's report on 'The Art in the Artificial' recommended that the UK should fund PhD studentships as a way to reduce the cost of exploring potential routes for collaboration between AI research and the creative industries. Nesta notes that 'this would allow what is technically, artistically and commercially possible to be discovered more quickly and easily, enabling the UK to reach its full potential in this area.' AI and the creative industries would benefit from long-term innovation funding and investment as well as Research Council support for PhDs and Doctoral Training Centres.

Ethics, business and policy implications for equitable AI and Art Futures

Ethics and equity are major concerns for the application of AI. Even the most advanced machine learning algorithms are designed by humans. Individual biases, preconceptions and assumptions will be amplified by algorithms.

The arts, CreaTech (the emerging field in which technology enables the creative sector to produce new products, services or experiences) and the creative industries need to be aware of the threats to equity associated with AI. Artists can use the platforms available to them to introduce such issues to a wider audience and artistic platforms could be used as environments in which to identify and address issues in new technology. The large audiences visiting museums and galleries provide perfect conditions for verifying theories related to human interaction with machine learning or other software.

Immersive technology

Karen Palmer's 'Perception.io' immersive experience places participants 'in the shoes' of a police officer during a confrontation with a potential suspect. The actions taken by the participants, as well as involuntary reactions, such as eye movement, are recorded and reproduced to form a comprehensive picture of one's intuitive preconceptions.

Conclusion

The AI for Art Futures roundtable explored ways to leverage the transformational effects of AI to enhance the UK's creative strengths. Strategies to increase the adoption of AI in the creative industries will also require long-term funding. Universities can play a key role in supporting these strategies, through providing funding, capacity building and fostering relationships between emerging artists and industry partners. However, ethics and equity are major concerns for the application of AI, as individual biases, preconceptions and assumptions will be amplified by algorithms. Artistic platforms could be used as environments in which to identify and address issues in new technology and introduce these to wider audiences.

The discussion raised some important questions, including:

- How do we establish lasting partnerships between artists, technology companies, government entities and higher education institutions to ensure continued development of AI in the creative sector?
- Is there a way to maintain a sustainable pool of funding, beyond emergency funding, for AI and art initiatives, even in the face of the global recession?
- How can an artist's path from education to a career involving AI be facilitated?

The creative industries play a crucial role as a communicator, demystifying AI and portraying the challenges and opportunities it creates to a wider audience. Challenges remain regarding ways to incorporate AI in creative work and avenues for artists to engage with technology companies. However, provided that open discussions such as this roundtable can continue to take place and generate ideas, there is potential for innovative solutions to be developed.

Participants

Ms Jo Townshend (Innovation and Enterprise, Roundtable Chair)*

Dr Anna Donovan (UCL Faculty of Laws)*

Mr Kieren Reed (Slade School of Fine Art)*

Professor Yvonne Rogers (UCL Interaction Centre)

Mr Jake Elwes (Artist)

Mr Graham Hitchen (UK Research and Innovation)

Mr Wayne McGregor CBE (Studio Wayne McGregor)

Dr Chris Michaels (National Gallery)

Ms Tonya Nelson (Arts Council England)

Ms Caroline Norbury MBE (Creative Industries Federation)

Ms Karen Palmer (Artist)*

Mr Ben Vickers (Serpentine)

**Individuals marked with asterisks were co-authors of this particular commentary.*

Climate Change

The aim of the Climate Change roundtable was to explore how AI can assist in the transformational changes needed to decarbonise systems, infrastructure and societies.

Executive Summary

The key points and recommendations highlighted during the discussion include:

- Many technology solutions to achieve decarbonisation in energy systems already exist. AI can play a role as the 'software' solution that manages this complexity and makes it feasible to adopt the hardware into the system.
- Decarbonising the energy system and achieving net zero by 2050 will require behaviour change in people's everyday lives. AI has the potential to identify synergies & trade-offs of solutions as well as interventions that have the greatest likelihood of public acceptance.
- Digital twins are virtual models of physical energy infrastructure, energy markets and market actors (i.e., people and businesses). They have the potential to explore different policy and weather scenarios, manage resources and monitor the health of equipment.
- The predictive ability of AI can play a role in forecasting, modelling and managing supply and demand for renewable energy sources.
- Addressing the talent and skills gap is key, and it is vital that employment pathways are provided at the intersection of data science and climate change.
- Creating small community-based demonstrators could offer one way to introduce new concepts and technologies to the public, while also enabling decision-makers to explore ways to roll out the technologies.
- Greater collaboration and sharing of data and metadata between energy companies is needed, as well as improved access to nationally held data.

What are the applications of AI that could have the greatest impact? What do we need to know to increase such impact?

Forecasting, modelling and managing energy supply and demand

As the energy grid grows in complexity due to an increase in the number of fuels, actors, and components, AI becomes an important means to monitor, optimise and plan energy supply and demand. For example, as weather-dependent renewable energy sources such as wind and solar increase, the prediction of energy generation becomes a challenge. Using AI technologies to predict how weather scenarios might affect renewable energy generation would help to show where there might be significant drops or surplus in energy supply across the energy network.

AI can help to manage energy supply (i.e. forecasting the number of sunny days or windy periods) and demand (i.e. temperature changes that would affect heat usage)

While the potential of AI in forecasting applications has long been understood, a less developed application area of AI is in supporting faster roll-out of sustainable technologies or data-driven scenario development. Moreover, solutions depend on their acceptance by citizens. AI can play a role here by intelligent optimisation of limited resources or displaying costs of alternative options - for example, a targeted installation programme can be developed that makes best use of the limited workforce of installation technicians in the UK by identifying homes that would benefit most from a heat pump.

Innovation

Digital twins are a virtual representation of a system that spans its lifecycle, is updated from real-time data, and uses simulation, machine learning and reasoning to help decision-making. These models have enormous potential in smart cities and the energy sector. For example, Carson City, Nevada, has used a digital twin to manage the city's water supply, and has subsequently reduced the number of human hours needed to operate the system by 15 percent while maintaining the same level of supply. Additionally, GE Renewable Energy is developing a software platform to create a virtual version of GE's gas, steam and wind turbines to test different scenarios and monitor a turbine's performance. The energy sector is highly fragmented, geographically dispersed and greatly influenced by human behaviour. Therefore, being able to develop simulations that attend to these challenges could support scenario planning and big picture thinking.

Having innovative information management frameworks are also needed to optimise the potential of AI technologies and promote wider sharing of data between companies and prosumers. Open banking models (i.e. the financial ecosystem where transaction data can be anonymously shared with financial institutions and Fintechs to access better financial products and services) have opened up the banking sector and could offer an example for the climate sector.

Access to data

Data science and AI technologies, such as digital twins, could be used to help identify opportunities for decarbonising. However, efforts to incorporate AI into the energy sector are fragmented. While organisations such as the Open Data Institute are developing standards for data sharing, and the Modernising Energy Data programme at the Department of Business, Energy and Industrial Strategy (BEIS) is implementing the recommendations from the Data Task Force, a central AI and climate change position statement is needed to create a cohesive policy framework.

With increased competition for energy providers to offer green energy solutions, companies are less likely to be willing to share their data and metadata with their competitors. A mandate, possibly paired with financial incentives, is needed for companies to build-up capacities for data stewardship and to share data and models to foster greater collaboration and coordination across the industry.

Federated learning (sometimes referred to as collaborative learning) is a machine learning technique that trains models across multiple decentralised servers that hold local datasets, without exchanging them with the other servers. While industries such as defence and pharmaceutics use these techniques, federated learning is not currently widely used in the energy industry. However, given that federated learning enables multiple stakeholders to build a common, robust machine learning model without sharing data, these techniques present a solution to critical issues such as data privacy, data security, data access rights and access to heterogeneous data.

Domestic decarbonisation

The key to reducing emissions is to decarbonise the energy and material systems, including the built environment (i.e. the human environment including homes, buildings, zoning, streets, etc.), transportation system and electricity system. AI has the potential to reduce the costs involved with these changes and identify interventions that have the greatest likelihood of public acceptance.

While behavioural aspects are beginning to shift and more households are installing smart meters, many of these still only take monthly data captures. Having more frequent readings, for example, would provide valuable, granular data to help gain an accurate understanding of energy usage patterns and ways to maximise existing technologies. AI technology in homes could help to identify how behavioural aspects and technologies interplay and help to prevent potential loading problems that cause interruptions to energy supply across the network.

Nasa

Barriers to the successful deployment of AI technologies

Employment pathways

While data stewardship (i.e. implementing data governance policies and procedures in an organisation) and data science are becoming more recognised fields, the pathways to careers at the intersection of data stewardship, data science and climate change are not clearly established. There has been an increase in hackathons and data science events with climate focused problems being put to the AI community, such as the Ofgem climate emergency hackathon or the 'Learning to Run A Power Network' challenge at the Neural Information Processing Systems (NeurIPS) meeting in December 2020. However, more long-term solutions are needed to enable AI researchers to build careers that positively impact the climate change emergency. The pathways from academia into graduate training programmes and placements in industry that are beginning to be established in the AI and drug discovery sector (explored in more detail in the Discovering New Medicines section) offer a complementary model on which the energy sector could base its own career pathways.

Public acceptance for AI and climate change

A lack of public trust in AI technologies and acceptance for the use of personalised data, regardless of whether it is anonymised or unidentifiable, is a major barrier to the uptake of new AI technologies in the energy sector. Regardless of the technological innovations made, lack of public acceptance will result in new technologies not being adopted. The use of AI and large-scale data analysis to develop climate services (i.e. products that enhance users' knowledge and understanding about the impacts of climate on their decisions and actions) could leverage the enthusiasm of citizen scientists (members of the public who collect and analyse data), facilitate sustainable development in low- and middle-income countries, and increase transparency of decision-making. Climate services represent a way to bring greater individual involvement and engagement to the global decarbonisation movement.

Mobile phone-based climate services

In Ghana, the Climate Change, Agriculture and Food Security research program delivered tailored seasonal forecast information to farmers via their mobile phones. This information enabled farmers to adapt their farm management decision-making to seasonal variabilities due to climate change.

Gaining a critical mass of people who are willing to switch to sustainable energy sources is essential; efforts are needed to increase the uptake of, and acceptance for, green energy technologies. Creating small community-based demonstrators could offer one way to introduce new concepts and technologies to wider populations, while also enabling decision-makers to explore how such technologies could be rolled out across the country.

SGN's H100 Fife pioneering project

In November 2020, Ofgem announced funding for a scheme to demonstrate hydrogen distribution and in-home heating performance in Levenmouth, Fife, Scotland. By 2022, approximately 300 homes will be part of the 100% hydrogen demonstration network, with the hydrogen being produced by a local offshore wind turbine. This project will contribute toward understanding decarbonisation options for heating and provide evidence of hydrogen's performance in a real-world domestic setting as a carbon-free energy source.

Conclusion

The Climate Change roundtable explored how AI can assist in the transformational changes needed to decarbonise systems, infrastructure and societies. Innovations such as digital twins, community-based demonstrators, scenario explorers and innovative information management structures could facilitate the deployment of sustainable technologies and creation of smart cities. Additionally, if open and sharable data agendas are progressed rapidly, AI can play a role in transitioning to net zero through identifying the most cost-effective and interventions with the greatest potential for return. Underscoring all of these innovations and agendas is the need for strong leadership to drive the successful uptake and implementation.

Participants

Dr Aidan O'Sullivan (UCL Energy Institute, Roundtable Chair)*

Dr David Barber (UCL Centre for Artificial Intelligence)

Professor Tadj Oreszczn (UCL Energy Institute)

Mr Jay Ashton-Butler (Office for AI)

Mr Rafi Cohen (Quadrature Climate Foundation)

Professor Olga Fink (ETH Zürich)

Mr Alex Howard (Centre for Net Zero)

Dr Lynn Kaack (ETH Zürich)

Ms Catherine Newman (Limejump)

Mr Simon Pearson (Energy Systems Catapult)

Ms Laura Sandys (Energy Systems Catapult)

Professor Valeria Jana Schwanitz (Western University of Applied Sciences)

Dr Emily Shuckburgh (University of Cambridge)

**Individuals marked with asterisks were co-authors of this particular commentary.*

The AI for People & Planet Roundtable Series is supported by: