AI changing world: generating opportunities and challenges

Professor Gavin Shaddick, Royal Holloway University of London

The rapid development of Artificial Intelligence (AI) models has recently been the subject of a great deal of debate and is high on the political agenda. Much of the focus has been on the use of large language models (LLMs) that utilise the explosion in the availability of vast amounts of text data, including books, websites, and social media posts to train models that can generate natural language responses in a variety of contexts. LLMs have multiple applications, including virtual assistants, chatbots, or text generators, such as ChatGPT.

However, the potential uses of AI go far beyond these applications and the ability to generate new insights through identifying patterns within multiple, diverse, data sources means that AI has huge potential to contribute positively across all aspects of society, including health, safety, security, education, environmental sustainability and future prosperity.

Examples of potential uses of AI

Sustainable Development: The United Nations' 2030 Sustainable Development Goals (SDGs) call for a plan of action for people, planet and prosperity. Quality, accessible, timely and reliable disaggregated data is essential for understanding the scale of the challenges and in measuring progress to ensure that no one is left behind. AI tools and models can play a pivotable role in assessing the sheer volume of data required to produce SDG indicators and, given the substantial differences in the quality and availability of data in different regions, that they are consistent over time and location.

Healthcare: applications of AI in the healthcare field include predictive models for diagnosis and treatment based on the interrogation of vast amounts of data, including images (e.g. x-rays, scans), at speeds that would be impossible if performed manually, and would be impractical in terms of the resources that would be required. As AI tools develop, they will increasingly incorporate expert knowledge and physical constraints of the system under investigation, rather than being entirely data-driven.

Climate change and Net Zero: AI has crucial role to play in facilitating pathways to Net Zero through the integration of AI technologies in energy, water, transport, agricultural and other environmentally related systems and by empowering individuals, organisations and businesses through the provision of personalised information that will support behavioural change.

Data from environmental monitoring and modelling could be integrated with socio-economic and other data sources to predict the differential impacts and resilience of climate change across different populations and marginalised groups. This would allow the effects of different policies and interventions to be assessed 'virtually' across a range of outcomes (e.g. human health, biodiversity and infrastructure).

Industry: AI could become an integral part of product design, supply chain optimisation, quality control and predictive maintenance. For example, models based on historical data of a manufacturing process could be integrated with real-time measurements from sensors in a Digital Twin to optimise production based on environmental conditions.

Creative industries: there are a wealth of opportunities across a range of applications, including: (i) content creation, (ii) information analysis, (iii) content enhancement and postproduction workflows, (iv) information extraction and enhancement, and (v) data compression[1].

Education: the use of generative AI tools and LLMs could revolutionise teaching, learning and learning outcomes. To date, much of the debate in HE related to generative AI models has focused on risks to the integrity of assessments, but there is also an opportunity for universities to lead on frameworks for the ethical and fair use of AI more widely and to provide/equip society with the skills to fully utilise AI (for example, the skills required in setting AI prompts and fully understanding potential biases and limitations.

The need for increased understanding

One of the key features of AI tools is their ability to use data from multiple sources to provide information and insights that are predicated on a

[1] Anantrasirichai, N., Bull, D. Artificial intelligence in the creative industries: a review. doi.org/10.1007/s10462-021-10039-7

much wider evidence base than has traditionally been possible. However, this does mean that data are increasingly being used to answer questions and form the basis of decision making outside of the sphere of its intended use. The differences between using routinely available data and data collected as part of a carefully designed experiment requires understanding the effects of all aspects of the 'data journey', including collection (including potential biases), governance, analysis, interpretation, communication and the future use of resulting composite datasets in other fields. However, methods for analyses, testing, and performing inference have not kept up with the enormous growth in the variety of data sources and the quality of the data that they produce. It is one of the most important aspects of AI, but one that is not well understood and includes a multitude of challenges. These include the ability to identify where errors matter the most. In tasks where individual where mistakes are cheap, such as web search and presenting advertisements, decisions are based upon global error rates, but in 'high stakes' applications, such as medical diagnosis, very prediction matters and more sophisticated, measures of success need to be developed and adopted.

As we move towards the large-scale use of AI in decision-making, there is an urgent need for a deeper understanding of the potential effects of these, and many other, issues.

AI and applications at Royal Holloway

Royal Holloway, University of London (RHUL) has a proud history of AI research: our Department of Computer Science at RHUL was established in 1968 and is one of the oldest computer science departments in the world. It includes world leading researchers in algorithms and complexity, artificial intelligence, bioinformatics, distributed and global computing, machine learning, software language engineering and type theory. The Information Security Group was formed in 1990 and is recognised as a UK Academic Centre of Excellence in Cyber Security Research and collaborates extensively with government and industry in the area of cyber-security.

RHUL is home to pioneering work in machine learning, from early work on support vector machines to the more recent invention and development of conformal prediction.

The University's recent strategy has been to expand our portfolio of challenge-led research and educational programmes to ensure that we respond to the challenges of our modern times. This builds upon our strengths in discovery and methodological research and was supported by the establishment of four Catalysts: (i) Transformative Digital Technologies, Security and Society; (ii) Living Sustainably; (iii) Advanced Quantum Science and Technologies; and (iv) StoryFutures (Creative Arts). Each of these has a strong link to data science and AI. The Catalysts bring together researchers across a wide variety of disciplines with experts in computer science, data science, artificial intelligence, together with external partners and stakeholders from government, business and industry to promote inter-disciplinary collaborations. RHUL has recently become a member of the Alan Turing Institute's University Network, which offers an opportunity for our researchers to expand partners and engage more widely across the Turing network.

Partnership with Royal Botanic Gardens Kew

RHUL has a historically strong partnership with Kew and we are exploring the ways in which we can expand our collaboration to fully utilise the opportunities associated with AI. Kew curates one of the world's great complex datasets: with millions of images of collected specimens of all the species in the world, with metadata, species labels, free text anatomical descriptions, and line drawings. We are embarking on a collaborative programme of research that will develop neural representations of specimens, for intelligent search and comparison within the vast collections. The key aims are to *accelerate taxonomy* to rapidly identify new species; and to enable new specimens collected in the field to be *matched directly* with both labelled *and unlabelled* specimens, meaning that millions of specimens can be fully used for the first time in global biodiversity monitoring.

Alongside technical developments and implementation, we are also responding to the crucial need for research into the ethical use of AI, ensuring that the decisions made by AI do not discriminate, are transparent, and preserve privacy. This is one of the core challenges to be addressed in the Transformative Digital Technologies, Security and Society Catalyst: *How will transformative digital technologies, such as Artificial Intelligence, impact a rapidly changing world?* For example: How can we ensuring that the benefits of AI are enjoyed by society as a whole, and that the risks are mitigated so as not to disproportionately burden certain people or groups?; What does it mean to ensure automated decision-making systems act fairly?; How can we ensure that 'black box' algorithms perform their functions transparently?; and How can personal data be protected securely? (Examples from the Alan

Turing Interest Group on Fairness, Transparency and Privacy[2]).

[2] https://www.turing.ac.uk/research/interest-groups/fairness-transparency-privacy