Institution: University of Cambridge

Unit of Assessment: 27 English Literature and Language

Title of case study: AI narratives: shaping public debate, policy and cultural responses to the future of Artificial Intelligence

Period when the underpinning research was undertaken: December 2016- ongoing

Details of staff conducting the underpinning research from the submitting unit:

Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Dr Sarah Dillon	Reader in Literature and the Public Humanities	April 2014-
Dr Kanta Dihal	Senior Research Associate	October 2017-

Period when the claimed impact occurred: 2017-2020

Is this case study continued from a case study submitted in 2014? No

1. Summary of the impact (indicative maximum 100 words)

'AI Narratives' investigates the representation of artificial intelligence (AI) in fictional and non-fictional narratives. Through media appearances, public-facing fora and popular publications, participation in cross-sector dialogue, and collaboration with a range of stakeholders, including Boeing and United Nations Educational, Scientific and Cultural Organization (UNESCO), Dr Dillon's and Dr Dihal's research has led to impact on policy, society, culture and media. The project alerted policymakers and industrial partners to the influence of narratives on perception of AI technologies, and to their importance for the development of public policy and regulation. It also engaged with other beneficiaries – including media professionals, creative practitioners, scientists, and citizens – in order to enhance critical reflection on AI narratives and AI itself.

2. Underpinning research (indicative maximum 500 words)

Dr Dillon and Dr Dihal both have long-standing research interests in literature and science, science fiction, and narrative forms. These interests inspired the AI Narratives project (2017-2018), based at Cambridge's Leverhulme Centre for the Future of Intelligence (CFI) in collaboration with the Royal Society. Dr Dillon was Programme Leader, seconded for a year from the Faculty of English, while Dr Dihal was the post-doctoral researcher.

The project's interdisciplinary research explored how AI has been and is currently portrayed in fictional and non-fictional narratives, incorporating methodologies and insights from, and contributing to, literature and science studies, science fiction studies, science communication studies, and history of science. Dillon, Dihal and others studied the limitations of prevalent fictional and non-fictional narratives to suggest how practitioners at the intersection of disciplines might move beyond those limitations. They also aimed to raise global awareness of the power of AI narratives, and to encourage public attention to the role they play in the research, regulation and reception of AI technologies.

The major publication arising from the project is *AI Narratives: A History of Imaginative Thinking About Intelligent Machines* (OUP 2020) **[R1]**. This edited collection contains a co-authored introduction and chapter contributions by Dillon and Dihal, and fourteen chapters by other contributors. The book is the first to map the history of imagining intelligent machines in this way. It makes the case that AI narratives are a social, ethical and political issue, shaping the technical field, and the acceptance and regulation of the resulting technology. It is based

on the premise that contemporary thinking about AI can usefully be informed by considering the role and influence of stories, old, new, and emerging. The book presents the history of imaginative thinking about intelligent machines in the Anglophone Western tradition in two parts. Part I covers antiquity to modernity, each chapter focusing on a specific historical period. Part II takes up the historical account in the modern period, focusing on the twentieth and twenty-first centuries, in which a greater density of narratives emerges alongside rapid developments in AI technology.

The edited collection is a centrepiece, but the project members have published an additional range of research findings in diverse forms and with diverse methodologies. These include new theories of the relationship between literature and science **[R2]**, a critique of the gendered nature of AI rhetoric **[R3]**, surveys of public perception of AI **[R4]**, a taxonomy of hopes and fears around AI **[R5]**.

Dillon has developed the AI narratives research further, with postdoctoral research fellow Dr Olivia Belton, in a case study investigating societal perception of AI-enabled flight, funded by Boeing (2019-20). The project produced a qualitative critical taxonomy elucidating existing thematic associations with autonomous flight as evidenced by science fiction literature and visual media. It also deployed narrative as research method, developing and trialling a collaborative storytelling game, played in focus groups, to understand and assess non-expert perceptions of autonomous flight **[R6]**. The findings revealed that technical failure, trustworthiness and climate anxieties are key to autonomous flight perception, and the research demonstrated the usefulness of narrative methods in assessing anticipatory assumptions and understanding technological futures with diverse stakeholders.

3. References to the research (indicative maximum of six references)

R1. Stephen Cave, **Kanta Dihal** and **Sarah Dillon**, ed., *AI Narratives: A History of Imaginative Thinking about Intelligent Machines* (Oxford: Oxford University Press, 2020); includes co-written introduction and essays by both Dillon and Dihal.
R2. Sarah Dillon, 'On the Influence of Literature on Science', *Configurations* 26:3 (2018), Special Joint Issue with *Journal of Literature and Science*: State of the Unions (Part 2): 311-16. doi:10.1353/con.2018.0027 URL: http://muse.jhu.edu/article/698847
R3. Sarah Dillon, 'The Eliza effect and its dangers: from demystification to gender critique', *Journal for Cultural Research*, 24:1, 1-15 (May 2020), DOI: 10.1080/14797585.2020.1754642
R4. Stephen Cave, Kate Coughlan, and **Kanta Dihal**. '"Scary Robots": Examining Public Responses to AI.' *Proceedings of AI Ethics & Society* (July 2019). https://dl.acm.org/doi/10.1145/3306618.3314232
R5. Stephen Cave and **Kanta Dihal**. 'Hopes and fears for intelligent machines in fiction and reality.' *Nature Machine Intelligence* 2 (February 2019): 74-78. DOI: 10.1038/s42256-019-0020-9
R6. Olivia Belton, **Sarah Dillon**, 'Futures of Autonomous Flight: Using a Collaborative Storytelling Game to Assess Anticipatory Assumptions', *Futures* (2020), https://doi.org/10.1016/j.futures.2020.102688.

All these publications arising from the project research were peer-reviewed at scholarly journals or a major University Press. R1 is a field-defining essay collection covering a comprehensive history of AI narratives (and a REF2021 output submission); R2 is a position piece, part of a special issue describing the state of the field; R3 is a specific case study on AI rhetoric, narrative and gender; R4 and R5 provide overviews of the field derived from surveys of public opinion (R4) and fictional representations of AI (R5); R6 is a peer reviewed publication on public perceptions of autonomous flight, presenting the results of an exploratory study of non-expert anticipatory assumptions through collaborative storytelling games.

Royal Society 2017, September 2017 – August 2018, GBP2013.10, Grant number 2017/138 RG91897, Dillon Principal Investigator

The Boeing Company, GBP92,294.97, Grant number 2018-STU-PA-343, Dillon Principal Investigator

AI and Gender, 2019, The Ada Lovelace Institute, GBP8,500. Dillon Co-Investigator

AI and Gender, 2019, PwC, GBP3,000. Dillon Co-Investigator.

Global AI Narratives – Sub-Saharan Africa Workshop, 2019, Cambridge-Africa ALBORADA research grant, GBP5,000. Dillon Principal Investigator until July 2019, from then Dihal Principal Investigator.

Global AI Narratives: Reframing current artificial intelligence narratives through fostering international perspectives and networks, 2019, Templeton World Charity Foundation, December 2018 – December 2022, GBP174,954.33, RG98757, Dihal Principal Investigator, Dillon Co-Investigator until May 2019.

Global AI Narratives Project, 2019, DeepMind Ethics & Society, December 2018 – December 2021, GBP175,000, G104648, Dihal Principal Investigator.

Diversifying Artificial Intelligence Narratives. 2018, ESRC Impact Acceleration Account, GBP20,000, Dihal Principal Investigator, Dillon Co-Investigator.

4. Details of the impact (indicative maximum 750 words)

The AI Narratives research was disseminated across a broad spectrum: a range of broadcast media, including Dillon's contribution to Radio 4's 'Today' programme on 30th December 2017; Dihal's TedX talk in Thessaloniki, 'Is the Robot Rebellion Inevitable?' on 21 April 2018 (400 audience members plus livestream, now on YouTube) **[E1]**; involvement in the 2019 'AI: More Than Human' exhibition at the Barbican (88,811 visitors in London; it then moved to Groningen in 2019, but closed due to COVID, and was due to open in Liverpool in 2020) **[E2]**; and Dihal's essay 'Ancient Dreams of Intelligent Machines' in the Arts section of *Nature* (2018), which had a top 5% Altmetric impact score, and was recommended on 'In Our Time' (BBC Radio 4) **[E3]** (p. 3). The article influenced the work of technology company DeepMind (owned by Google / Alphabet Inc.): their Roundtable on stimulating effective public debate on the ethics of AI was shaped around the essay, with the Briefing Paper quoting it as an epigraph **[E3]**.

Impact on Cultural Life and Professional Practice

This dissemination programme created further opportunities to engage in artistic and commercial contexts. In 2018, Dihal advised *2030: Crowded Room*, an art installation by Crowded Room at the Junction Theatre in Cambridge as part of the Collusion 2019 Showcase. Made in collaboration with technologists and informed by conversations with young people in Cambridge, Kenya and Nepal, it challenged audiences with the risks of AI **[E4]**. The exhibition attracted over 1,750 visitors from diverse backgrounds (46% non-white British), including 29% of visitors who do not usually engage with the arts (attending once a year or less) **[E4]** (p. 24). One participant reflected: 'Does generate different thoughts and ideas regarding technology and its role and perception in today's society.' **[E4]** (p. 24). A user-experience designer at Cambridge Consultants, an innovation and technology-based consulting company with projects in over 20 countries, further noted: 'One of my biggest takeaways from this project was the necessity of creating technology that's readily available for everyone. I realised how important it is to consider the needs of the whole of society rather than for one demographic or group of people' **[E4]** (p. 29).

The project's research contributed to, and informed talks and panels at, CognitionX (CogX) 2017, 2018, 2019 and 2020. The participant count of CogX grew from 1,500 in 2017, to over 20,000 visitors in 2019, and 44,000 in the virtual meeting in 2020. CogX brings together attendees each year from research, business, industry, government and the public **[E5]**. One of the cofounders of CogX, and the UK government's AI business champion, notes that 'the work of the AI Narratives project … gave the community the language with which to talk about what the future of AI could look like. Their research has been central to framing the conversation for people, enabling fruitful discussions between industry, academia, and government' **[E5]**. She also notes, in her capacity as a popular science communicator, the

influence of the AI narratives research on her book *How To Talk to Robots* (2020) which 'is aimed at women and girls, who remain underrepresented in the artificial intelligence industry. *How to Talk to Robots* draws on the research of the AI Narratives project, and […] include[s] the project publication *AI Narratives: A History of Imaginative Thinking About Intelligent Machines* in its further reading list' **[E5]**.

Dillon and Belton contributed to the discourse on AI and Futures Literacy on 16 December 2019, when they contributed to a public-facing UNESCO Futures Literacy Forum in Paris. The event attracted over 400 participants, including Futures Literacy practitioners, designers, facilitators, teachers and researchers from around the world **[E6]**. Participants in the lab, based on the Boeing research, discussed the complex social issues and potential benefits of AI through storytelling. The Head of Futures Literacy at UNESCO noted: 'the lab designed and implemented by Dr Dillon, on the basis of her pioneering work in this area, generated a significant step forward for UNESCO and the global community interested in gaining a better understanding of the role of the future in what people see and do' **[E6]**.

Public Policy and Civil Society

The AI Narratives research informed UK government policy, through the House of Lords AI Select Committee, the AI Council (an expert committee of independent members from industry, the public sector, and academia, which provides high-level expertise and priorities to the UK government), and learned societies. The Chair of the Council remarks: 'I... can attest to the fact that Dillon and Dihal's research has been a contributing factor to the UK Government's policy on AI. The AI Council has a working group on "narratives", the creation of which was directly influenced by my engagement with the AI Narratives project' **[E5]**. In 2017, Dillon, Dihal and other members of the AI Narratives team gave an oral presentation to the House of Lords AI Select Committee. The Chair of the Committee calls the input of the project team 'revelatory' **[E7]** adding 'We had not considered the impact of fictional narratives on the public perception in artificial intelligence in such a manner before' **[E7]** (p. 188). Following receipt of written evidence from the AI Narratives team, the Select Committee report (Chapter 2, 'Engaging with Artificial Intelligence') focuses on 'the public's understanding of, and engagement with, AI and its implications, and how it can be improved' **[E7]** (p. 25).

The project's research further informed the programme and content of the Global Governance of AI Roundtable at the 2019 World Government Summit (Dubai). The GGAR sub-committee 'AI Narratives: Underrepresented Narratives' was included as a direct consequence of the work of the team. Dr Dillon served as the Expert Group Chair for the sub-committee **[E8]**. A November 2018 Royal Society report was co-authored by the project team. *Portrayals and Perceptions of AI and Why They Matter* was aimed primarily at policymakers **[E9]** (pp.1-28). The report has informed civic debate around AI narratives in the field of Knowledge Management, with RealKM Magazine, for example, using it to inform discussion of the question 'Is the popular narrative harming development of AI?', explaining the importance of the report in 'understanding why distorted narratives emerge', as well as making key 'recommendations for practitioners to adopt to improve matters' **[E9]** (p. 30).

Impact on industry

In the Boeing-funded project, Dillon and Belton investigated public perceptions and stories of autonomous flight. Autonomous flight, and particularly AI-controlled aircraft, has been depicted in science fiction for decades, but the technology is still far from widely available. As Boeing notes: 'Dr. Dillon's research effort on Social Perceptions of Autonomy will help us in future research and development efforts around autonomous systems and strengthen our focus on the human aspects' **[E10]**. Through collaborative storytelling workshops the project critically reflected on avenues for understanding emergent technology, essential 'in the perception and acceptance of autonomous flight' **[E10]**. To this end 'Dr. Dillon's report was received very well and helped multiple groups within Boeing understand the perception considerations when

building and operating autonomous aircraft. We look forward to leveraging findings from the work' **[E10]**.

5. Sources to corroborate the impact (indicative maximum of 10 references)

[E1] BBC Today Radio 4 programme screenshot 30.12.2017 (Dillon appears at 1.53.13 link);
Screenshot of Youtube video of Dihal's presentation at Tedx Thessaloniki link
[E2] Report from 'AI: More than Human' exhibition.
[E3] Evidence relating to Stephen Cave and Kanta Dihal, 'Ancient dreams of intelligent machines: 3,000 years of robots', Books & Arts, *Nature* (July 2018): Altmetric page 14.09.2020; BBC Radio 4 'In Our Time, Automata' 20.09.2018; Deepmind briefing 'Stimulating effective public debate on the ethics of artificial intelligence', September 2018.
[E4] *2030: Crowded Room* impact reports and feedback_Collusion.
[E5] Testimonial from the co-founder of AI advice platform CogX, 03.11.2020 and email confirming 2020 CogX participants 10.12.2020
[E6] Testimonial from Head of Futures Literacy, UNESCO, 24.09.2020
[E7] House of Lords evidence: House of Lords, Select Committee on Artificial Intelligence, Report of Session 2017-19, 'AI in the UK: ready, willing and able?' March-April 2018; Leverhulme Centre for the Future of Intelligence – Supplementary written evidence (AIC0238) 2018 link: Letter from the Chairman of the Select Committee on Artificial Intelligence, 21.11.2017
[E8] Dillon evidence of participation Global Governance of AI Roundtable
[E9] *Portrayals and perceptions of AI and why they matter.* The Royal Society, November 2018 link; Gaskell, *Real KM magazine* 'Is the popular narrative harming development of AI?', 23.01.2019;
[E10] Boeing testimonial regarding workshops and report, 30.11.2020