EMOTIONAL AI: JAPAN & UK

FINAL REPORT ON A CONVERSATION BETWEEN CULTURES

"scores": {
  "anger": 1.0570484E-08
  "contempt": 1.52679547
  "disgust": 1.60232943E
  "fear": 6.00660363E-12
  "happiness": 0.9999998
  "neutral": 9.449728E-0
  "sadness": 1.23025981E
  "surprise": 9.91396E-1
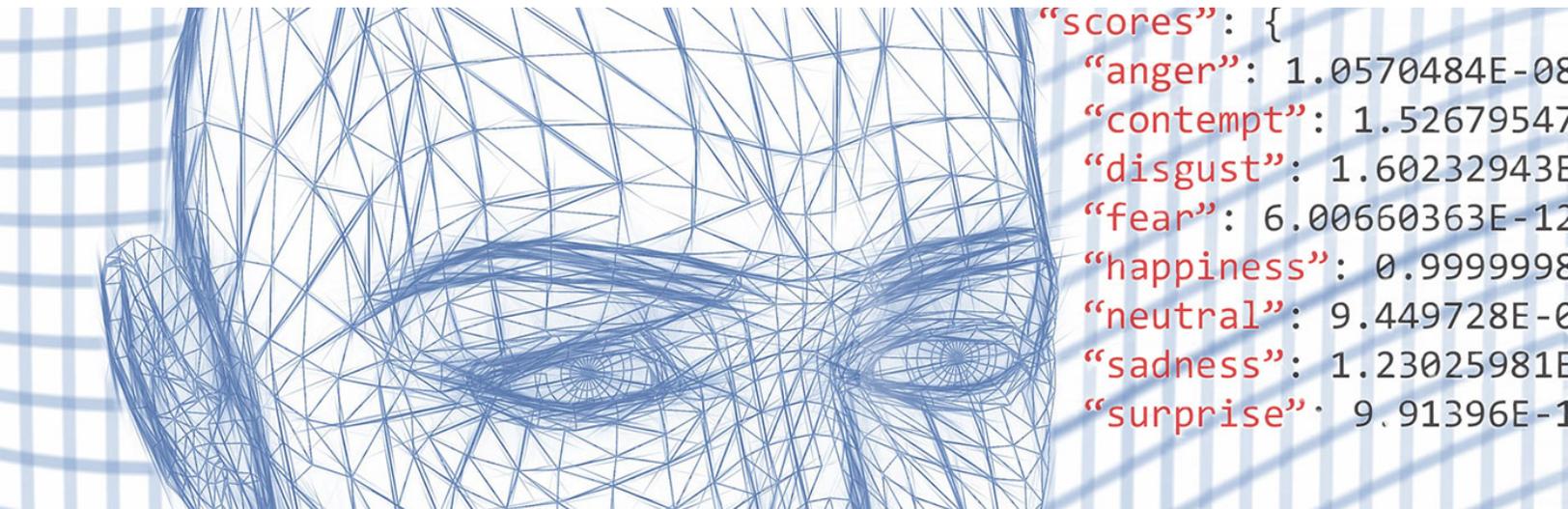
UK Research and Innovation · Arts & Humanities Research Council · PRIFYSGOL BANGOR UNIVERSITY · APU Ritsumeikan Asia Pacific University 立命館アジア太平洋大学 · E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL · THE UNIVERSITY of EDINBURGH

Project Investigators:
Prof Andrew McStay, Bangor University
Prof Vian Bakir, Bangor University
Prof Peter Mantello, Ritsumeikan Asia Pacific University
Dr Lachlan Urquhart, University of Edinburgh

With participants from:

New School for Social Research in New York,
Ritsumeikan Asia Pacific University (APU),
Bangor University,
University of Edinburgh,
Northumbria University,
University of Cambridge,
Keele University,
Kyoto University,
Freie Universität Berlin,
Meiji University,
University of Sheffield,
Rikko University,
University of British Columbia Chapel Hill University,
Chuo University,
Japanese Ministry of Defence,
Digital Catapult,
Sensum,
Nvidia,
Sensing Feeling,
Centre for Data Ethics and Innovation,
Privacy International,
Dyson School of Design Engineering,
Internet of Things Privacy Forum,
Coventry University,
ITD-GBS Tokyo,
Doshisha University
UK Cabinet Office,
Ernst and Young,
Independent artists

Introduction

Usage of emotional AI across all aspects of society in the UK and Japan, two of the most technologically advanced countries in the world, is fast increasing. It raises urgent questions about the body, intimacy, identity, relationships, intrusiveness, dignity, manipulation and reliability of data inferences.

The aim of this scoping project has been to understand ethical questions around emergence of these technologies in Japan and the UK, but also to find out what we can learn from each other.

To this end, we ran 3 workshops, 2 in Japan at the Ritsumeikan Asia Pacific University's Tokyo campus in July 2019 and 1 in the UK at the Digital Catapult in September 2019. The first workshop on July 9th considered the impact of emotional AI on commercial and civic life. The second workshop on 12th July examined security and policing uses of emotional AI. At the final workshop, we further reflected on lessons from Japan, and engaged with UK stakeholders for their perspectives on key emotional AI issues. This short report documents key themes that emerged from discussions at the various events. We had a wide diversity of participants from a range of disciplinary backgrounds, including: anthropology, new media studies, philosophy, computing, law, art, literature, journalism and criminology, to name but a few. This led to a range of opinions and insights.

The first workshop considered the potential of emotional AI, exploring social benefits and harms. We considered the technologies and their commercial applications in Japan and the UK. We also examined how citizens might feel about them, why they would feel this way and what laws and governance that guide these technologies are aiming to do in order to enable citizens to live well with emotional AI. With the second workshop, we explored the deployment of emotional AI in a range of security and law enforcement contexts, particularly with predictive policing and visual surveillance. We also examined recent trends around voice and facial recognition technologies, particularly at borders and in public spaces; the role of smart bots in manipulating and triggering user emotions in social media and their use in computational propaganda shaping civic discourse. With the third workshop, we examined our preliminary analysis from the Japanese workshops and considered future commercial and research agendas.

> **Emotional AI: Derives from affective computing techniques and advances in machine learning and artificial intelligence (AI). It is a weak form of AI in that these technologies aim to read and react to emotions through text, voice, computer vision and biometric sensing, but they do not have sentience or emotional states themselves.**

HIGH-LEVEL INSIGHTS

- Emotional AI: it is not just about emotions, but affect, states, intention and empathy.
- Current methods and models of emotion are questionable.
- Among industry delegates, there was desire for regulation and ethical guidance to provide business certainty.
- Emotional AI will augment other technologies/practices.
- There are multiple ethical concerns in UK and Japan, but different focus points.
- Japanese personal and social experience of emotional AI differs from the UK.
- Japan-UK social, political and legal contexts are substantively different (although arrangements with EU General Data Protection Regulation exist).
- Ethical "toolkits" that include law need to go beyond just compliance.
- Profiling of bodies in policing has a long history, and emotional AI could be the next focal point.
- Understanding of nuances and the specificities of the cultural context is key.
- As in the UK, public responses in Japan to the impacts of new technology on civic life vary greatly.
- We need to consider respect as well as privacy (in both Japan and the UK).
- We need to consider "metaphysical/spiritual" dimensions of cities (to avoid the cybernetic, "cold" conceptions of efficiency dominating discourse).
- On emotional AI and disinformation: Consider low political participation among youth in Japan.

COMMERCIAL & CIVIL USES OF EMOTIONAL AI
WORKSHOP 1

In this workshop, we had a series of presentations on the nature of the technology and the emerging sector followed by exploration of cultural dimensions of emotional AI (EAI), culminating with an overview of the governance landscape. Talks were coupled with a series of commentators and subsequent breakout sessions. These summaries represent an assimilation of themes based on workshop discussions and our own observations, following these different activities.

Motivations for EAI and Sensing Practices: **EAI[1] is not only about measuring emotions** – it about gauging the **'state'** of a person (e.g. if they are stressed or uncomfortable; distracted, engaged or aroused; their level of valence and fatigue). Both human contact and contactless sensors are enabling third parties to read the body. **Wearables,** like heart rate sensors and galvanic skin conductors, typify the former. The latter typically involves **cameras and microphones**, but also **chips** with biometric and vibration sensing, mobile phones, and heat or lighting **sensors.** Contactless systems reduce the **'on body nexus'**, changing the nature of the interactions between subjects and sensors. This shapes the scope for meaningful control over data collection too. The goal of EAI is for machines and humans to have **empathic interactions**, with a focus on building human relationships with systems, as opposed to just sensing users as data points.

Emotional Modelling: **Cognitive science models of emotion need to become more nuanced** as they currently neglect the aspects of human emotion. Some companies are going beyond **Ekman's 6 Universal emotions[2]** (fear, anger, disgust, sadness, happiness and surprise) when modelling human emotion in their systems, but many are not. **The limitations of this model need deeper critical reflection from industry,** reflecting academic critique.[3] Valence and arousal[4] are increasingly important elements, as are contextual factors (e.g. location, ongoing activity). The latter requires greater sensing and data capture from the subject's environment. Edge computing may prove an important trend supporting EAI, due to more local, subject centric data processing.[5] This could move past universal models, towards tailoring experiences based on emotion, in situ. Similarly, considering why these imperfect but widely used emotional models are adopted cannot be ignored in critique. As engineers want to get a system to work, certain models are adopted[6], but historical debates in computing around use of formalist

---

[1] McStay, A. *Emotional AI: The Rise of Empathic Media* (SAGE: 2018)
[2] Ekman, P Basic Emotions in T. Dalgleish and M. Power (Eds.). *Handbook of Cognition and Emotio*n. (John Wiley & Sons, Ltd., 1999).
[3] Barrett, LF. Adolphs, R. Marsella,S. Martinez, A.M. Pollak. S (2019) Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements. *Psychological Science in the Public Interest* 20(1)
[4] Rubin, DC, Talarico, JM (2009) A Comparison of Dimensional Models of Emotion: Evidence from Emotions, Prototypical Events, Autobiographical Memories, and Words *Memory.* 17(8): 802–808.
[5] Crabtree et al (2018) Building Accountability into the Internet of Things: The IoT Databox Model", *Journal of Reliable Intelligent Environments*, DOI 10.1007/s40860-018-0054-5
[6] Aylett, R., & Paiva, A. (2012). Computational Modelling of Culture and Affect. *Emotion Review*, *4*(3), 253–263. https://doi.org/10.1177/1754073912439766

models of human behaviour in AI,[7] may come to replay for EAI too. Importantly, we may also see third party, 'objective' metricised, technologically mediated assessments of emotional state becoming more authoritative than subjective, introverted self-assessment of state. However, if EAI models are wrong, this disconnect could cause significant harm. Furthermore, who is training these algorithms, where and – particularly due to normative impacts of cultural differences – on what training data. We wonder if there is a role for scientists to test such systems before there is in-the-wild deployment (as is done with other technologies for addressing safety concerns e.g. electrical safety)?

EAI Applications **–** The EAI industry is diverse, with a number of start-ups (Sensum, Affectiva, Real Eyes) but also legacy firms are increasingly active too (Amazon Rekognition; Hitachi Public Safety; NEC NeoFaceWatch). Emotion sensing is likely to be a **layer** to complement other digital services, as opposed to a standalone product. This is particularly so in **smart cities**, but also in monitoring **in car** experience for safety, entertainment, playful and more empathic interactions. With safety, this could be a response to emerging legislation on telemetry and monitoring driver state/drunkenness in EU. In Japan, for example, EAI is being used to monitor fatigue of freight drivers, to monitor gaze (with surprisingly little pushback from drivers). This also has applications for an **aging Japanese population**, where preventing elderly car crashes is a public concern. At the other end of the spectrum, use of EAI in Japanese school classrooms to **monitor attention/engagement** requires careful management (particularly for school reputation) and teacher input, to understand any impacts on pedagogy.

Structural Concerns **-** Ethical fears stem from the intimacy of EAI data processing, particularly as it is **non-conscious physiological data. Even the subject is likely to be unaware of the implications of how they are being read**. Managing these risks could shape the field, where **legislation could present parameters** of what should be done with EAI (enabling creativity through constraints). Similarly, there are concerns it could **constrain innovation** in an emerging industry, reflecting longstanding debates in both countries about when best to regulate new technologies (i.e. the so-called Collingridge Dilemma[8]). Linked to this, is the need to consider what the **end vision of EAI** is in UK and Japan. In Japan, policy strategies such as Society 5.0[9] point to a human centred vision of human machine interaction.

New Routes for Dialogue **– There is pressing need for increased disciplinary collaboration and engagement** between social scientists, cognitive scientists, engineers, interaction specialists and roboticists. This is particularly so around how modelling emotions for EAI emerges, and the values being embedded in systems design.[10] There is also a need to appreciate

---

[7] Suchman, L. *Plans and Situated Actions* (Cambridge University Press, 1987)
[8] See Japanese Cabinet Office – Society 5.0 document
https://www8.cao.go.jp/cstp/english/society5_0/index.html
[9] Collinridge, D *The Social Control of Technology* (1980)
[10] See work of the *Model Emotion Project* at www.modelemotion.org.

cultural, historical, and social dimensions of emerging technologies in Japan and how citizens relate to these, with input from historians, anthropologists, architects and identity theorists.

Nature of Human Machine Interactions – **In Japan, relationships with technology, both in terms of human attachment and temporality of these relations, differ to the UK.** With human attachment, there can be perceptions of **agency and spirit** within artificial systems. In Japan, there is a greater focus on designing and playing with human/non-human dynamics to create **characters** and personas for technology. This can **enable different types of relationships** where there is greater capacity for more intimate, emotional or companion-like bonds, particularly with robots (e.g. Lovot Kawaii style companion robot)[11]. In Japan, artists, use of screenwriters in addition to concepts like **Kansai engineering** [12], **moe anthropomorphism** [13] and 'techno animism' [14] principles, shape understanding of how to design technologies with personality (which in turn can foster greater emotional attachment).[15] In general, **anthropomorphism** is particularly salient in Japan due to its connection to Shintoism.These **personalities may not be gendered or even human-like**, which recalibrates what the nature of emotional interaction looks like, moving us beyond human-human orientated models. **Learning from Japan on how to move past gendered forms of interactions**, and utilising **technological affordances for emotive interactions[16]**, could address some of concerns in the UK about gender bias and female conversational agents (e.g. Amazon Alexa, Microsoft Cortana, Apple Siri). In some respects, EAI **simulating** emotion (and **stimulating** emotion by reading a subject's state and responding) could be seen as a **'consensual hallucination'**. The ontology of social reality is being played with. Japanese roboticists, for example, explore ideas of robots and 'hearts', where robots may affect a human heart, have their own or understand the human heart in order to better serve its needs.[17] Largely, however, the desire and ability for subjective relations with the machine changes how subjects relate to it.

Attached to this is how **temporality of devices interface with emotional attachment.** There is a sense of being-ness with robots that transcends their

[11] White, D (2019) The Future of LOVOT: Between Models of Emotion and Experiments in Affect in Japan *Platypus* http://blog.castac.org/2019/07/the-future-of-lovot-between-models-of-emotion-and-experiments-in-affect-in-japan/

[12] Nagamachi, M., *Kansei Engineering*. (1989, Kaibundo Publishing Co. Ltd, 1989)

[13] See in anime and manga where 'moe' qualities are given to non-human beings, objects, concepts, or phenomena.

[14] Jensen, C. B., & Blok, A. (2013). Techno-animism in Japan: Shinto Cosmograms, Actor-network Theory, and the Enabling Powers of Non-human Agencies. *Theory, Culture & Society*, *30*(2), 84–115. https://doi.org/10.1177/0263276412456564

[15] For a non-Japanese perspective on designing products with emotions in mind, see Norman, D *Emotional Design: Why we Love (or Hate) Everyday Things*. (Basic Books, 2004)

[16] See Robertson, J *Robo Sapien Japanicus: Robots, Gender, Family and the Japanese Nation*. (California University Press, 2017)

[17] Hirofumi Katsuno (2011) The Robot's Heart: Tinkering with Humanity and Intimacy in Robot-Building, *Japanese Studies,* 31:1, 93-109, DOI: 10.1080/10371397.2011.560259

4

physicality[18], as we see with Sony Aibo robot dog Buddhist funeral ceremonies at Kōfuku-ji temple,[19] where a physical machine dog was returned to the manufacturer, dismantled or used for 'organs'. With smarter, cloud based systems, will the robot 'soul' lives on beyond its material body, in remote servers? How might edge computing [20]change this? We may start to see interesting questions around when someone has emotional attachment to a device, but legal limitations in data portability[21] mean they cannot take the learned/inferred data an object holds (once support stops or the device breaks). This reintroduces the importance of the materiality to the 'personality' of the device.

Distribution of responsibility for actions of robots - **Where should responsibility lie for robot actions**, particularly where agency is not as strictly distinguished between subject and object? Should it be with designers, software, businesses, hardware? Whilst in Europe, legal personality for robots has been challenged by academics[22] and in the EU rules on Civil Liability for robots)[23], in Japan there is more interest in this prospect due to perspectives on non-human agency of technology.[24]

Importance of Respect and Etiquette - In Japan, the Society 5.0 strategy argues for human centricity in handing data but **what does this mean in practice?** We could return to Western centric principles such as dignity (*songen* in Japanese) and liberty (US), but in Japan **mutual respect (*sonkei*) is prioritized over these.[25]** Public displays of politeness are important, in Japan, and technology can challenge this. In Osaka, for example, a company was using facial recognition systems in a train station, and the system was discontinued, not only because of privacy fears but also because it was deemed to be creepy[26] and *disrespectful.[27]* **Privacy is a borrowed concept in Japan, whereas respect emerges as a focal point for how humans interact (particularly in terms of social hierarchy and forms of language).**

---

[18] Gal, D. Perspectives and Approaches in AI Ethics: East Asia, Forthcoming in Dubber, M Pasquale, F and Das. S *The Oxford Handbook of Ethics of Artificial Intelligence*. (OUP)

[19] Kubo, A (2010) Technology as Mediation: On the Process of Engineering and Living with the "AIBO" Robot. *Japanese Review of Cultural Anthropology*, vol, 11; White, D and Galbraith, P Japan's Emerging Emotional Tech. *Anthropology News*. http://www.anthropology-news.org/index.php/2019/01/25/japans-emerging-emotional-tech/

[20] Crabtree, A., Lodge, T., Colley, J. et al. J Reliable Intell Environ (2018) 4: 39. https://doi.org/10.1007/s40860-018-0054-5 - edge computing is where analytics is brought to the data on local devices, as opposed to centralising the data analytics in the cloud.

[21] Urquhart, L., Sailaja, N. & McAuley, Realising the Right To Data Portability for the IoT. *Pers Ubiquit Comput* (2018) 22: 317. https://doi.org/10.1007/s00779-017-1069-2

[22] Koops, BJ Hildebrandt M & Jaquet-Chiffelle, DO. (201) *Bridging the Accountability Gap: Rights for New Entities in the Information Society?* 11 Minn. J.L. Sci. & Tech. 497 Available at: https://scholarship.law.umn.edu/mjlst/vol11/iss2/4

[23] Urquhart, L, Reedman-Flint, D & Leesakul, N (2019) Responsible domestic robotics: Exploring ethical Dimensions of Robots in the Home. *Journal of Information, Communication and Ethics in Society.*DOI: 10.1108/JICES-12-2018-0096

[24] Inatani, T (2019) New Legal Framework for A.I. Devices *Forthcoming.*

[25] Miyashita, H (2011) The evolving concept of data privacy in Japanese law, *International Data Privacy Law*, 1(4) 229–238, https://doi.org/10.1093/idpl/ipr019

[26] Tene, O & Polonetsky, J (2014) A Theory of Creepy: Technology, Privacy, and Shifting Social Norms, 16 Yale J.L. & Tech https://digitalcommons.law.yale.edu/yjolt/vol16/iss1/2

[27] Japan Today (2014) Camera facial recognition program put on hold for Osaka train station building https://japantoday.com/category/national/camera-facial-recognition-program-put-on-hold-for-osaka-train-station-building

Respect for non-human agents in Japan can extend to seeing inanimate objects having 'spirit'.[28] **If something material lasts beyond 100 years, it has spirit.** Thus, relationships with AIs can be co-emergent and made of novel interactions beyond human or animal relations, to build relationships of value. Respect may require regulatory and policy ethics initiatives, but also consideration of ontological status of non-human things (spirit/quasi-life), particularly as there are fears companies could come to exploit this respect and emotional attachment. Furthermore, the mystique around how EAI actually functions may link into spiritual narratives.

Respect (and how to attain it) is a useful framing device when discussing EAI. It goes beyond the focus on harms in the UK, and implies a relationship which carries responsibilities (i.e. if you are not attending to those responsibilities to the other, you could be seen as disrespectful). As McStay has argued, there are **protocols** around privacy and codes of behaviour that guide relationships and interactions.[29] How might these be recalibrated for human/non-human relations mediated by EAI?

Regulatory Similarities (and Differences) – The difficulties we see in data protection governance of biometrics, CCTV and facial recognition may provide a precursor to the kinds of issues EAI poses. In terms of regulatory frameworks, Japan and the EU recently reached an **agreement on the adequacy of the Japanese Data Protection** regime, enabling the largest cross-border data transfer area in world. Despite this, there is **no Art 22 GDPR equivalent in Japan**, which enables human intervention and meaningful oversight of an automated decision. Art 22 and Art 15 GDPR create a right to an explanation, and what this might mean in practice has been a significant research focus in the UK.[30] The lack of Art 22 limits protection around profiling or need to provide logic about why decisions have been reached. Protection is normally provided through **heavy reliance on privacy policies and requirements to obtain consent** (but given well known issues with these models of governance, meaningful protection might be lacking).[31] With EAI, the lack of subject consciousness of collection in the first place and high potential for scope creep in what is done with data is concerning. Also, there are **no monetary sanctions** in Japanese DP law, instead **naming and shaming** of the organisation is the ordinary response. This speaks to the **importance of etiquette**, and losing face to others as a sanction. Similarly, unlike in the UK[32], there is **no notion of privacy by design** in Japanese law currently, although it is a popular concept in privacy policy circles.

---

[28] Jensen et al (2003)

[29] McStay, A. Privacy and Philosophy: New Media and Affective Protocol (Peter Lang, 2014)

[30] Edwards, L; Veale, M; (2017) Slave to the Algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Duke Law & Technology Review*, 16 (1) pp. 18-84

[31] Edwards, L. (2017) Privacy, Security and Data Protection in Smart Cities: A Critical EU Law Perspective, European Data Protection Law http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2711290

[32] Art 25 GDPR.

6

Towards an Ethical EAI Toolkit: What is needed?

To conclude we asked groups to consider where to focus energies in the future to create more ethical EAI.

Defining EAI red lines - Establishing the **red lines** that cannot be accepted around EAI, could **complement** the current trend towards aspirational codes of practice and ethics for AI. Red lines could be monitoring in private settings, such as the home or public toilets, but even those are increasingly made visible (particularly the home). **Testing technology before deploying** (e.g. learning from market research, audience feedback or user centred design) could establish a process to explore value tensions, and how to resolve these.

On this point, the notion of establishing global codes raises questions **about the values guiding these, and their appropriateness**. If they stem from Western liberal democracies, how appropriate are they to export globally, to ensure cultural sensitivity? UK-Japan could be a good place to learn lessons by finding commonalities and differences.

Enabling user control – There is a recognition that the same **personalisation systems** used to tailor services to our preferences could be **repurposed** to **protect citizens from harm too**. So, if we could computationally translate ethical preferences around EAI, could a P3P type or agent-based system act on our behalf in machine interactions in different contexts?[33] Similarly, **personal information management systems** and peer to peer data sharing (to bypass dominant platforms) might help increase user control over data. There is a need to balance new systems with usability and making them easy to use, to encourage wide adoption.

Remembering what we already know – We need to learn from existing regulatory frameworks and **guard against the narrative that new technologies mean we are in a rule vacuum**. For example, with GDPR, it may seem new, but it builds on the Data Protection Directive which has been around since 1995, hence its novelty should not be overstated as a justification to not act. It is also useful to consider what other frameworks we could turn to, such as bio-engineering, as it poses similar harms/benefits to society but remains emergent.

Finding the right metaphors – Speculative fiction can be a valuable educational and provocative tool for envisioning future relationships with EAI. In the UK, we often tend towards **dystopian** narratives of future relations e.g. Black Mirror. In Japan, technology is often portrayed in more **utopian** light e.g. Ghost in the Shell 2, hence exploring the media and cultural metaphors for EAI could help to situate the shift more clearly.

A Role for Design – designers can create artefacts to test responses from the public, enabling feedback and introducing a **playful dimension** to EAI interactions. Art can go beyond monetary or power considerations that other

---

[33] W3C Web page on P3P. https://www.w3.org/P3P/

stakeholders may prioritise, enabling exploration of different EAI mediated futures.

Future areas to study

To conclude, we are interested in what areas we need to explore in more depth to understand cultural differences around ethics and EAI in Japan and the UK.

- We need to adopt a **historical approach** to examining how EAI-like systems emerged and their route to becoming embedded in society (also how governance emerges alongside this). We could do this by **examining antecedents of AI**, the **social justifications** around their development and the **socio-cultural histories** surrounding these, to go beyond technology led narratives.

- We should unpack **current and future technological trajectories for EAI** through use of patent analysis, examining business strategies of companies, and ethnographic fieldwork with industry.

- We should examine **popular discourse/narratives around EAI** to explore if there are any patterns with other technologies. This could link to Sci Fi and future framings of AI-enabled society through cultural artefacts, to explore what is acceptable (and what is not).

- We need to systematically **analyse commonalities in the UK and Japan** around AI development (culturally, legally, socially, technologically) - what differs, what can we learn?

- **We could consider AI standardisation processes**, particularly the political powers shaping the adoption of some standards, in addition to their content.

- **We should explore UK and Japanese interaction design process** and underlying principles (exploring similarities and differences), to mutually learn.

- To unpack the **role of EAI in elder care**, particularly where empirical work in Japan has indicated preferences from robo-care to careworkers from another country.

- We need to assess the **notion of respect in data governance**, as it changes discourse from harm to a more positive question of how we live well with EAI.

- We have to conceptually map what we mean by **creepiness** in the context of EAI, considering application domains.

SECURITY & POLICING USES OF EMOTIONAL AI

WORKSHOP 2

***Policing Visions; Policing Reality* - Unpacking the motivations** behind why new technologies are used in policing is important (e.g. drones, predictive policing, body worn cameras). Similar reflection will be necessary for EAI. **Financial constraints** are a large motivation as forces around the world have finite financial and human resources**. Faith in technology and modernisation is another**, where technological solutionism can be seen to overcome organisational and operational limitations (which is particularly influential among senior, manager level officers). It is also important to be aware of the **disconnect between visions of policing** (and high-profile public trials) and the **reality**. Despite theatrical displays[34] and institutional aspirations around facial and emotion recognition, technology does not always work as intended. Indeed, the context that many police forces' officers face (particularly in UK) is lack of access to basic technologies, such as laptops.

***Reading the Body* -** The desire to **use the body to infer suspicion and criminality is not new**, and we need to situate new EAI trends in their historical context. [35] **Positivist criminological** notions [36] of inherent, biologically traceable criminality enabled by reading the body with phrenology and biometrics have longstanding traction in policing[37]. **EAI may introduce emotions into profiling approaches, through use of psycho-physiological metrics**. In criminal justice, we need to remain **skeptical of these systems** (as with DNA and fingerprinting), to reduce their **aura of objectivity and infallibility**. In particular, what happens if unsupervised machine learning (ML) creates new categories of suspicious behaviour through correlation? EAI may work 'correctly', insofar as it reflects training data (which in turn reflects societal aspects). However, **if** EAI should be used by police (this is a debate in itself and this report is simply reflecting the conversation in Workshop 2), we need to make the **normative choice to hold these systems to a higher level of scientific and social scrutiny**. Correcting these biases to avoid reinforcing discriminatory effects could be one approach. Human officers may not fully be able to account for their own decision-making process: as EAI makes emotions machine readable and auditable, there is scope for policing decisions based on emotions to be held to a higher level. Furthermore, sovereign functions may be delegated to private entities, posing further challenges for accountability. A key concern around EAI in policing is the **element of human choice.** If someone is algorithmically deemed to be likely to act in a deviant way (neurologically or biologically), how should we push back against this deterministic assessment? Again, objective and independent scientific scrutiny is required. This is joined by a need for institutional awareness among all officers (not just technologists and police data scientists) that these technologies are highly limited.

---

[34] Schneier B *Beyond Fear: Thinking Sensibly About Security in an Uncertain World* (Corr: 2003)

[35] Miranda D and Machado H. (2018) Photographing prisoners: the unworthy, unpleasant and unchanging criminal body. *Criminology and Criminal Justice*. 10(3).

[36] Ferrerio-Lombroso, G and Lombroso, C. *Criminal Man: According to the classification of Cesare Lombroso*. (Putnam Press: 1911)

[37] Sekula, A. (1986) The Body and the Archive. *October*, *39*, 3–64.

Resisting EAI – to resist being read by EAI, **citizens need to understand how it works in order to subvert it**. As EAI reads non-conscious metrics (heart rate, sweat), scope to control reading of these is limited. How might citizens be able to respond? Should there be a greater focus on 'internal emotion training'? There is both a need to educate users about EAI but also limitations to building literacy around how systems perceive them. [38] Motivations for disciplining the self may also be important, for example to conform to community norms or to avoid guilt and shame. A key concern is understanding the **harms that prompt a need to resist** in the first place. Currently it is primarily about fears of technical infrastructures enabling **human identification** (as we see with facial recognition), as opposed to the deeper level of trying to **understand human intent**. Furthermore, even if a form of ethical EAI can be created, people still may dislike it: hence, considering how/if people can stop it being deployed requires attention to power relations between stakeholders driving deployment of this technology.

In Japan, the existence of the same governmental party in power for a long time; a greater focus on benefits of technology in public discourse; less critical press (in contrast to the UK e.g. on facial recognition) and (contested) perceptions of a culture of consensus and less fear of authority in Japan, are all elements which could **shape acceptability or resistance to EAI.** The value placed on risk aversity in Japan may legitimise more securitisation if it stops crime and terrorism, in contrast to the UK where there may be more emotive responses to public space surveillance. Also, certain populations may find themselves targeted by EAI more than others, where white collar criminals may be less of a focus than shoplifters (although sentiment analysis of emails for suspicious tone, is used by some investment banks).

Technological Advance in EAI - For the emergent EAI sector, holding large datasets (and having trained their systems on these) remains a metric of success[39]. Current exploration of **less invasive monitoring** through periodic sampling could also be useful. This corresponds with the importance of edge computing for EAI, keeping this intimate data closer to the user (and their control)[40]. The fear of scope creep when technology is used in policing is considerable, particularly driven by security rhetoric and terrorist events.[41] **How archives might change, for example,** in terms of identifying and monitoring populations could further shift with EAI. Traditionally large state-run fingerprint and DNA databases are now the norm globally, but more decentralized archives could emerge, beyond the state. With EAI, criminality can be understood in a more **psycho-physiologically granular and ephemeral manner**. For policing by consent and adhering to the rule of law, knowing how police are reading citizens, when (if it is retrospective, real time

---

[38] Relatedly, see Murakami, Y., Tatsumi, T., Otani, T., & Harada, Y. (2017). Ethics of Information Education for Living with Robots. *ORBIT Journal*, *1*(1). https://doi.org/10.29297/orbit.v1i1.21
[39] E.g. Splash page of leading EAI firm Affectiva https://www.affectiva.com/ states 'Affectiva Human Perception AI Analyzes Complex Human States 8,333,646 Faces Analyzed'
[40] Crabtree A et al (2018) Building Accountability into the Internet of Things: The IoT Databox Model. *Journal of Reliable Intelligent Environments*, DOI 10.1007/s40860-018-0054-5.
[41] Mantello, P. (2016) The Machine that Ate Bad People: The ontopolitics of the pre-crime assemblage". *Big Data and Society,* 3(2). 1-11.

or both) and what they do with their assessments means archiving, cataloging and predictive practices around emotional information will be important.[42] **Cultural context** is key here, as what publics deem to be the appropriate balance between liberty and security (and what desire there is for scrutiny over AI systems for trustworthiness) will differ between Japan and UK. The validity of the **social contract** can turn on this perceived legitimacy, and EAI prompts a renewal of this.

Cross Cultural Differences - Events like the 2020 **Olympics in Japan** are prompting new surveillance mechanisms, such as use of sentiment analysis. As the UK recently had a similar experience with the London Olympics 2012, finding lessons on policing of large-scale events could be a useful exercise. Similarly, we need to be aware of **projecting societal homogeneity** in how we frame Japan and UK. In Japan, notions of being a more monocultural society are controversial, but it can be a narrative that is used to imply cohesion, despite presence of subcultures: **how does this interface with notions of threats and of 'othering' that security systems need to address**? In the UK, in Northern Ireland, attitudes to security, policing and technology differ to other parts of the country due to 'the Troubles'. Indeed, many surveillance technologies, like CCTV, emerged in part due to terror threats surrounding the Troubles. We need to consider how companies are exploring **cultural nuances in how individuals live with technology**: how do they design for different socio-economic backgrounds and political contexts (e.g. London, Tokyo, Belfast, Beijing and Dubai)? An interesting dimension of this is how EAI may emerge in smart cities, and how this concept and vision is being **'sold'** in Japan. In the UK, it is often framed in cold, 'cybernetic' language of efficiency, performance and feedback loops. But in Japan, given the importance of human/non-human relationships, is it going to be different? Is there going to be more focus on companionship, spirituality and 'friendlier' machine relations? This is important for how companies place themselves in Japan (particularly UK firms seeking business there). How policing uses EAI in smart cities could also shift perception of the city.

Emotion and False Information in UK and Japan – understanding citizen feeling is **critical for political campaigns to further the reach and impact of their message**.[43] We have seen this already with Cambridge Analytica/SCL Elections, and their role in granular profiling of target audiences in order to develop engaging, emotive, and sometimes deliberately deceptive microtargeted messages within election campaigns around the world.[44] Their

---

[42] Sekula, A. (1989) The Body and the Archive. In *Routledge Handbook of Visual Criminology*; see Edwards, L., & Urquhart, L. (2016). Privacy in public spaces: what expectations of privacy do we have in social media intelligence? *International Journal of Law and Information Technology*, *24*(3), 279-310. https://doi.org/10.1093/ijlit/eaw007, https://doi.org/10.2139/ssrn.2702426 - on distributed state dossiers with social media policing

[43] Bakir, V & McStay, A 2018, 'Fake News and the Economy of Emotions: Problems, Causes, Solutions', *Digital Journalism*, vol. 6, no. 2, pp. 154-175. https://doi.org/10.1080/21670811.2017.1345645

[44] DCMS (2019) *Disinformation and 'fake news': Final Report*. Digital, Culture, Media and Sport Committee, House of Commons 1791. https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf
DCMS (2018) *Disinformation and 'fake news': Interim Report*. House of Commons 363.

influence is compounded by human nature, where **false news often travels faster than the truth**, as citizens spread it due to the fear, disgust and surprise it prompts.[45] Thus finding strategies to handle emotive disinformation and misinformation turns on the **nature of the press, and fundamental questions of how democracy functions**. In the UK, encouraging a pluralistic media economy, educating the population on digital literacy (how to avoid confirmation bias) or trying to get digital intermediaries to downgrade false information in online platforms are among many solutions proposed by various stakeholders.[46] **Cross cultural differences in the UK and Japan mean UK strategies may not translate well between the two nations:** deeper comparative analysis of the digital media ecosystem in UK and Japan would be of value when working out how to prevent the spread of emotive, false information online.

As Japan has had the same political party in power for decades, the spread of false political information online is **seen as less of an issue due to this stability.** Furthermore, **lack of political engagement by the younger population** means it is hard to prompt public discussions on the topic. Historically, universities were very political in Japan, but nowadays political discussion by young people is largely **stigmatised**, where talking about politics in everyday discourse is seen as '**not being fun'.** Furthermore, young people's identities are often more tied to commercial and sub cultural interests, than political ones. [47] Perhaps, **preparation** for online emotive, false information (such as 'fake news') and the role of EAI therein, is needed in Japan, as the issue has had less prominence than in the UK. Finding strategies for **resilience** is important, particularly **beyond broadcast media**, where online platforms have significant power (both during and beyond political campaigns). Given this novelty, there are a range of questions to explore in the future around emotion and false information. We also need to consider **the location, shape and style of the contemporary public sphere in Japan**. Is protection of different types of speech legally valued (such as commercial speech in the US), and how might censorship type approaches be received in Japan?

***Handling the Future: Deep Fakes and automated journalism**–* there is scope to **destroy public trust in audio-visual content**, as '**deep fakes' created by AIs** recalibrate notions of what is real: seeing may no longer be believing. Similarly, **mundane news could be weaponised too,** where it could be subtler, and less evocative than fakes of presidents' speeches (but similarly may get less scrutiny accordingly). [48] Using affordances of

https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/363.pdf
[45] Vosoughi, S et al (2018) The spread of true and false news online *Science* 1146-1151
[46] Bakir V and McStay A. (2017) *Combatting fake news: analysis of submissions to the fake news inquiry*. MECCSA. http://www.meccsa.org.uk/news/three-d-issue-28-combatting-fake-news-analysis-of-submissions-to-the-fake-news-inquiry/
[47] Brasor, P (2019) Japanese Politics Finds a Way to Target Youth. *Japan Times* https://www.japantimes.co.jp/news/2019/06/29/national/media-national/japanese-politics-finds-ways-target-youth/#.XTOv4C2Q3Uo
[48] BBC (2019) 'Dangerous' AI offers to write fake news, 27 Aug. Https://www.bbc.co.uk/news/technology-49446729

technologies, such as **Internet of Things**[49], for disseminating fake content is another vector to shape perceptions in subtle ways. The scale of spreading fake news can be escalated through automated journalism that computationally generates content. How EAI further complicates this picture requires attention, to guard against using this route to further manipulate citizens.

---

[49] See BBC Box available at https://www.bbc.co.uk/rd/projects/living-room-of-the-future what would negative look like -how might when people watch nightly news, the lighting, the temperature, the volume? Also. Sailaja et al (2019) Living Room of the Future. *ACM TVX '19.*
http://www.cs.nott.ac.uk/~pszaxc/work/TVX19.pdf

UK PERSPECTIVES ON EMOTIONAL AI

WORKSHOP 3

Our final workshop at the Digital Catapult in London delved back into themes and topics from the previous Japanese workshops. We began by considering the technological emergence of emotional AI, with a particular focus on seeking responses to this from NGOs and industry. We then considered the use of emotional AI in Japan, reflecting on the cultural and interactional implications and to what extent these have corollaries in the UK. We presented our high-level findings from the first 2 workshops back to the group as provocations. Many of the conversations mirrored our earlier findings, but we note some of the new directions below.

Regulatory and Ethical (Un)certainty

Certain industry actors expressed a **growing desire for regulation[50] and clear ethical guidance**, in response to the growing tide of well intentioned, but (arguably) imprecise, discursive, dialogue driven ethical AI frameworks.[51] Whilst ethics is clearly important for building trustworthy EAI, some felt **stricter regulatory frameworks provide a higher degree of business certainty** of what is needed for compliance. Strict rules make it easier to understand what EAI technologies can (and cannot) be built. The interpretation led nature of law means it can also lead to uncertainty and gaps in knowledge on how to apply it. There was a recognition of the value of ethics in complementing legal requirements, but there was a concern that in both cases, **ambiguity costs money**. Companies, particularly start-ups need support, particularly as they may seek to 'contain' risks, as opposed to shutting down novel applications altogether. They lack the resources to be able to get sufficient legal advice to give certainty of safety of their product, in contrast to larger firms which can dedicate time, money and staff to engaging with ethics. As such, this may lead to a more cautious industry, which if we consider the precautionary principle, could be desirable, but does feed into debates about innovation being stifled.

In response, **checklists and certification to verify** their approaches are appealing for businesses[52] (as seen in other domains e.g. environmental protection certification[53]). Emerging standards on EAI[54] could also be valuable in prescribing how to build in appropriate safeguards **at scale**. However, we advise caution and recall that **ethics is a process**.[55] A desire for further precision from industry is understandable, but we suggest there is a reflective, complementary richness that technology ethics brings to the table, beyond law.

---

[50] Snaith, E (2019) Mark Zuckerberg calls for more government regulation of internet: 'We have too much power' *The Independent.* https://www.independent.co.uk/news/world/americas/mark-zuckerberg-facebook-regulation-internet-government-washington-post-a8847701.html

[51] Fjeld, J et al (2019) Principled Artificial Intelligence: A Map of Ethical and Rights Based Approaches https://ai-hr.cyber.harvard.edu/primp-viz.html

[52] See Art 40-41 GDPR and European Data Protection Board Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies under Regulation 2016/679 (EDPB, 2019)

[53] See BRE Environmental Certification Schemes - https://www.bre.co.uk/page.jsp?id=1763

[54] IEEE P7014 Standard Working Group on Emulated Empathy in Autonomous and Intelligent Systems Working Grouphttps://sagroups.ieee.org/7014/

[55] Verbeek PP. Technology Design as Experimental Ethics. In: van der Burg S., Swierstra T. (eds) *Ethics on the Laboratory Floor*. (Palgrave Macmillan, London, 2013)

Ethical codes of conduct and guidance may imply that ethics is something that can be 'done' if these principles are adhered to, when in reality it is an ongoing negotiation among stakeholders, that adapts as human values do. Responsible innovators need to build in processes to reflect shifting concerns over time, not just at design stage[56].

Ethics as Branding:

Contradictorily, there were discussions about the **value of being perceived as acting ethically**. There is concern of notions of **'ethics washing'** around AI currently. Yet ethics is still, to a large extent, becoming a branding exercise, much like **corporate social responsibility.** Companies positioning themselves in the market need to differentiate themselves from competitors and signal their virtues. Crafting their key 'ethical principles' can be a way of doing this. This, in turn, moves away from more accountable norms (like law), towards controlling the terms on which they are judged publicly. Firms may claim to go **beyond regulation** to build in resilience and a sustainable business that stays ahead of regulatory shifts. Others may use the same rhetoric but not adhere to this in practice. For an emerging EAI sector, there is concern about such differentiated approaches on the sector, as we see with public backlash to facial recognition.[57]

One challenge for EAI stems from **sources of funding and reputational impacts** of **associations with particular industries**, for example the security and military industrial sector. Navigating whether or not money should be accepted is difficult for cash strapped start-ups, but can have significant impacts on how they are perceived. In general, there are concerns that ethics remains within the purview of wealthy firms who can afford to control the narrative and be more selective of funding.

There are also interesting questions about the **source of values in EAI ethics[58]**. For example, in Japan, relationships with technology are in part influenced by religious values (from Shintoism), particularly with respect for inanimate objects and acceptance of blurred boundaries within human/non-human interactions.[59] It is valuable to consider how the UK, as a multicultural nation, might have differentiated responses to EAI, shaped by religious,

[56] Stahl, B., Eden, G., & Jirotka, M. (2013). Responsible research and innovation in information and communication technology—identifying and engaging with the ethical implications of ICTs. *Responsible Innovation*, 199–218; Urquhart L (2018) Ethical Dimensions of User Centric Regulation, *ACM SIGCAS* 47(4) 81-89 https://dl.acm.org/citation.cfm?doid=3243141.3243151

[57] Doffman, Z (2019) Hong Kong Exposes Both Sides Of China's Relentless Facial Recognition Machine *Forbes.* https://www.forbes.com/sites/zakdoffman/2019/08/26/hong-kong-exposes-both-sides-of-chinas-relentless-facial-recognition-machine/#62e881de42b7

[58] Friedman, B., Kahn, P. H. J., & Borning, A. (2008). Value Sensitive Design and Information Systems. In K. Himma & H. Tavani, The Handbook of Information and Computer Ethics. Wiley and Sons. But also see Le Dantec, C (2009) Values as Lived Experience: Evolving Value Sensitive Design in Support of Value Discovery *CHI 09* https://dl.acm.org/citation.cfm?id=1518875

[59] Discussed above.

cultural and social values of users. Some communities might have different levels of acceptance of AI mediated emotional interactions.

Responding to EAI Harms

EAI can pose privacy harms from both **accuracy and inaccuracy**.[60] If it is too accurate, intimate features of the subject are visible, but similarly, inaccurate sensing may create false positives (e.g. in an airport bodily metrics may indicate an aggressive state as opposed to a traveller being flustered). In general, we need greater debate about contextual implications of EAI, and acceptable uses in different domains.

There is a spectrum of EAI uses and thus we cannot say it is bad inherently e.g. what about health led or playful applications in contrast to policing streets and borders, in persuasive advertising or in court lie detection. It isn't necessarily appropriate to advocate outright bans of EAI, and instead, demarcating what contexts are not appropriate is important[61]. As a starting point, this may turn on the purpose of the technology. Is the application seeking to **support individuals in reaching their goals, or to profile and target them as a subject of suspicion? [62]**

Across the spectrum of EAI uses, this may also shape what explainability looks like. The wide variety of uses of EAI may also shape the need for **contextualized explanations**. Despite a regulatory and technical desire for explainability[63], it is difficult to find the balance between high level abstract explanations and overly granular, technical ones, both of which are largely useless to users[64]. **Designing for specific EAI application domains** and actual user interactions could help create legible interfaces that support users to be more critical of how they are being read by EAI.[65]

Legal boundaries used to justify not explaining AI often focus on risks of **sacrificing intellectual property, trade secrets or confidentiality**.[66] However, participants suggested that often the business novelty is less in the algorithms themselves, and instead in the business service model. This is especially so where publicly accessible **patents associated with the work exist**.

---

[60] Bayamlioğlu, E., Baraliuc, I., Janssens, L., & Hildebrandt, M. (Eds.). *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling The European Citizen*. (Amsterdam: Amsterdam University Press 2018) DOI:10.2307/J.Ctvhrd092
[61] See McStay (2018) on EAI in public survey results on public acceptability.
[62] This ties in with longstanding discussion in HCI about the need to build systems that support users e.g. Rogers, Y (2004) New theoretical approaches for human-computer interaction *Annual Review of Information Science and Technology* 38 (1) 87-143
[63] Edwards, L and Veale M (2017) Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for *Journal Duke L. & Tech. Rev.* 16. 18
[64] Crabtree, A and Urquhart, L and Chen, J (2019) Right to an Explanation Considered Harmful. *SSRN* https://ssrn.com/abstract=3384790
[65] HDI Principles foreground importance of legibility see Mortier, R. Haddadi, H. Henderson, T. McAuley, D. Crowcroft, J. and Crabtree A. *Human Data Interaction*. in Institute of Interaction Design, *Encyclopaedia of Human Computer Interaction* (IxD, 2019)
[66] For general discussion of interplay of IP and AI see WIPO Resource page on AI https://www.wipo.int/tech_trends/en/artificial_intelligence/

Motivations for EAI

There were concerns about how to justify the use of EAI particularly, as it may be to complement or replace what humans are already doing e.g. monitoring public space like train stations using human security guards. With EAI the issue is often one of scalability and how this impacts accountability. Whilst there is an awareness of the **limitations of the basic emotional models** in industry, they are challenged by the **desire of clients** who seek these simplistic models (e.g. faces labelled with an assessment of emotional state).

Building on discussions of respect and etiquette, there **remains uncertainty around acceptable forms of behavior with forms of AI** (which may have emotional aspects). For example, how appropriate is it for adults and children to be rude to conversational agents? The system could embed manners by design, requiring cue words of please and thank you, or building more human attitude into agents when they are treated improperly. This also turns to questions of what **kind of personalities objects should have in different contexts** e.g. an extroverted character for a home entertainment speaker to an introverted character in car navigation systems.

Impact of cultural representations in shaping our expectations and acceptability.

As in the Japan workshops, the role of speculative and **science fiction was queried**, particularly to what extent sci-fi creates blueprints for design and research (e.g. telecoms in Star Trek; gesture control in Minority Report).[67] The expectations raised by sci-fi can **shape our expectations of technology**, with both aspirations and disappointments. Contrasting images of **how we emotionally live with artificial agents** can also be informed by considering sci-fi: e.g. in Star Wars the humanoid C3PO is more concerned with human feeling, while the more utilitarian R2D2 is more callous. In the UK, sci-fi is also educating and raising awareness of ethical challenges around technology (with Black Mirror) which play an important role in public **acceptability.**

Further research topics

In conclusion, we also considered topics for further research from the UK perspective:

- How are EAI interactions impacted by gendering of AI interfaces? Related, what are the social, cultural and legal implications of EAI systems predicting and using gender of users (e.g. impact on trans users who may dislike being categorized in binary ways)?[68]

---

[67] Haraway, D The Cyborg Manifesto (1985)
[68] Keyes, O. (2018). "The Misgendering Machines: Trans/HCI Implications of Automatic

- Where is **training data for EAI sourced** from? Concurrently, given the hunger for more data in the industry, what mechanisms are needed to ensure sufficient oversight of the legality of its acquisition and use (particularly, guarding against surreptitious sources).

- To what extent will wider digital economy market dynamics and **lack of competition impact access to training data**?[69] Will major firms in a dominant market position control access to, or creation of, datasets? Currently, pretrained models owned by big firms are provided as a service to smaller entities. How does this align with government led and incubator investment to support start-ups with resources for funding, training and testing their products?[70]

- **What new challenges does EAI as a layer in existing technology pose for governance**? This requires focus beyond the system that EAI is augmenting e.g. smart speakers with additional emotion sensing, to reflect on emergent issues, what has already been addressed and what new strategies are proposed.

- What **opportunities are there for use of data sharing trusts** to mediate access to data?[71] How can these be run in the best interests of users, as opposed to being a purely commercial service?[72]

- How can we critically engage with the **quality of datasets** and query if they reflect 'actual' emotions vs more performative interactions with devices?[73]

- How can we extend **user testing with EAI** in different contexts (e.g. public, domestic, workplace) to obtain feedback and insights into their likes and dislikes, which could inform governance and design?[74]

19th September, 2019

Gender Recognition".2018 Conference on Computer-Supported Co-operative Work. Jersey City, NJ.

[69] Competition & Marketing Authority *Unlocking Digital Competition – Furman Review* (CMA, 2019) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/7855 47/unlocking_digital_competition_furman_review_web.pdf

[70] See Digital Catapult Machine Garage - https://www.digicatapult.org.uk/projects/machine-intelligence-garage/

[71] For early work on this, see L Edwards (2004) the Problem with Privacy, IRLCT 18(3) 263-294 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1857536

[72] ODI (2019) *Data Trusts: Lessons from Three Pilots* - https://theodi.org/article/odi-data-trusts-report/

[73] Goffman, E *Presentation of Self in Everyday Life* (Anchor Books, 1959)

[74] Ada Lovelace Institute *Beyond Face Value: Public Attitudes to Facial Recognition Technology - Survey Results* (ALI, 2019) https://www.adalovelaceinstitute.org/beyond-face-value-public-attitudes-to-facial-recognition-technology/