# Submission of Evidence to the House of Lords Select Committee on Artificial Intelligence

6th September, 2017

1.  ARTICLE 19: Global Campaign for Free Expression (ARTICLE 19) welcomes the opportunity to respond to this inquiry by the House of Lords Select Committee on Artificial Intelligence (AI).

ARTICLE 19 a is global human rights organisation that works around the world to protect and promote the right to freedom of expression and information ('freedom of expression'). Established in 1987, with its international head office in London, ARTICLE 19 monitors threats to freedom of expression in different regions of the world, and develops long-term strategies to address them. We advocate for the implementation of the highest standards of freedom of expression, nationally and globally.

2. Since 2014, ARTICLE 19 has pioneered efforts in technical communities to bridge existing knowledge gaps on human rights and their relevance to internet infrastructure. Our efforts have been geared towards integrating human rights into foundational documents at the Internet Corporation for Assigned Names and Numbers (ICANN),[1] the Internet Engineering Task Force (IETF)[2] and the Institute for Electrical and Electronics Engineers (IEEE).[3] At the IEEE specifically, ARTICLE 19 has taken an active part in the Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems.[4] In December 2016, we also published a policy brief on algorithms and automated decision-making in the context of crime prevention.[5]

3. The study and development of AI is over half a century old, with the term being coined in 1956.[6] The current momentum in this field is enabled by the availability of large amounts of data, computational power that is both affordable and widely accessible, the continued development of statistical methods and the mainstream adaptation of technology. Hence, ARTICLE 19 believes in the need to critically evaluate the impact of AI and automated decision making systems (AS) on human rights, and the various ways in which these technologies embed values and bias, thereby strengthening or sometimes hindering the exercise of these rights, particularly freedom of expression. The role of industry, governments, and individual developers must be grounded, at the very minimum, in existing standards of corporate responsibility and international standards on human rights. Given our mandate, this submission focuses on the issues most directly connected with defending freedom of expression and information.

4. **Terminology:** At the outset, ARTICLE 19 notes that the terminology around AI varies and can encompass different concepts, in particular:

---

[1] See, for example ARTICLE 19, ICANN's Corporate Responsibility to respect Human Rights, October 2015; available at http://bit.ly/1KgkV5n.

[2] See, for example ARTICLE 19, Internet Engineering Task Force discusses human rights in plenary meeting for the first time in its history, April 2017; available at http://bit.ly/2wwz037.

[3] See, for example ARTICLE 19, A New Frontier: Ethics, Artificial Intelligence and the Institute of Electrical and Electronics Engineers (IEEE), December 2016; available at http://bit.ly/2wwBEps.

[4] See, The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, Ethically Aligned Design, available at http://bit.ly/2plPsMc.

[5] ARTICLE 19, Algorithms and automated decision-making in the context of crime prevention, 2 December 2016; available at http://bit.ly/2gSnG9W.

[6] Stuart J Russel & Peter Norvig, Artificial Intelligence – A Modern Approach, Englewood Cliffs, NJ: Prentice Hall, 1995: 27.

- "Algorithm" can refer to any computer code that carries out some set of instructions, and is essential to the way computers process data.[7] They are "encoded procedures for transforming input data into desired output, based on specific calculations".[8]
- Automatic decision making execution "generally involves large scale collection of data by various sensors, data processing by algorithms and subsequently, automatic performance." [9] It is an efficient means to manage, organize and analyse large amounts of data and then to structure decision-making accordingly. [10]
- Artificial Narrow Intelligence is the ability of machines to approximate human intelligence in a *deliberate domain;*
- Artificial General Intelligence, also commonly referred to as "Singularity", is understood as the ability of machines to exhibit all aspects of human intelligence.

In this submission, we refer to AI only in terms of "Artificial Narrow Intelligence," as popular perceptions of "Artificial General Intelligence", is still, at the very least, decades away,[11] if not entirely implausible.[12]

### The pace of technological change

5.   Machine learning algorithms[13] increasingly influence the ways in which we interact with our environments, with applications in critical sectors. For example, AI currently determines the information we consume online through ranking and filtering online content, most visibly on social media platforms like Facebook, YouTube, and Twitter.[14]   AI is increasingly used for predictive policing,[15] countering violent extremism,[16] and removal of child sex abuse images or video.[17] Courts in the United States use AI to determine the risk assessments of defendants in criminal sentencing.[18] This is a trend that is also ready for deployment in the United Kingdom.[19] Machine learning algorithms also find application in the financial sector and are used to

---

[7] Centre for Internet and Human Rights, "The Ethics of Algorithms: from radical content to self-driving cars – final draft background paper" GCCS 2015, available at http://bit.ly/1D7IgTx.

[8] Gillespie T. "The relevance of algorithms" in Gillespie T, Boczkowski P., and Foot K., "Media technologies: essays on communication, materiality and society", 2014, Cambridge MA:MIT Press (p.167).

[9] M Perel et al,"Accountability in algorithmic copyright enforcement" 2016, Stanford Technology Law Review, Forthcoming.

[10] *Ibid.*

[11] Rupert Goodwins, "Debunking the biggest myths about AI", Ars Technica, 21 December 2012; available at http://bit.ly/2f2eqhi.

[12] Luciano Floridi, "Should we be afraid of AI?", Aeon, 9 May 2016, http://bit.ly/1q8UXOz.

[13] Machine learning is most successful subset of AI techniques, which enables an algorithm to learn from a provided dataset using statistical methods.

[14] See, for example, Casey Newton, "How Youtube perfected the feed", The Verge, 30 August 2017, http://bit.ly/2vOBbee.

[15] Haley Dunning, "Predictive policing gets a boost from 3m grant", Imperial College London, 21 March 2017, http://bit.ly/2ncWHZh.

[16] Matt Burgess, "Google's using a combination of AI and humans to remove extremist videos from YouTube", WIRED UK, 19 June 2017, http://bit.ly/2gFBC54.

[17] "Toddler hand inspired AI sex abuse tool", BBC, 1 December 2016, http://bbc.in/2eDXWLO.

[18] Julia Angwin et al, "Machine Bias", ProPublica, 23 May 2016, http://bit.ly/2f2eP3w.

[19] Nick Statt, "AI driven policing has arrived", The Verge, 10 May 2017, http://bit.ly/2pnXN6V.

determine the eligibility of individuals for loans and mortgages based on credit scoring,[20] and corporate bond trading.[21] Algorithms are also increasingly used for network management of critical infrastructure, from the electrical grid[22] to Internet routing.[23] While the use of AI may increase efficiency of these various processes in the future, its success so far has been limited and its use often controversial.

6.    ARTICLE 19 believes that at present, there is limited understanding of the ethical and legal implications of the training, development, and control of AI systems. Machine learning currently trains algorithms on datasets with a definition of "success", i.e. a definition of what the machine must look for, what features of the data it must train on. The choice of dataset, and the definition of success ultimately shape our interaction with these technologies. For example, some applications of AI have shown that they can exacerbate the problem of discrimination by excluding minority groups from services,[24] products,[25] or embedding biases against marginalised populations.[26] AI can hamper freedom of expression by unduly flagging legitimate content for takedown.[27] Facial recognition AI can increasingly identify people in a crowd,[28] undermining the right to privacy and anonymity in the public sphere. This generally indicates failures in making sound legal and ethical choices at the point of training data and defining success for the algorithm. However, this only partially explains questionable outputs. There is quite simply little understanding– and therefore accountability – at present of how machines produce outputs.

7.    Although the development of AI is not new, the digital environment will make it more enabling in the future, with greater volumes of data, computational power, and advances in statistical methods. Looking ahead, there is a strong tendency to implement AI across the board, making its potential even more pronounced. However, the need to think carefully through where, how, and whether AI should be implemented is an important one.[29] For example, AI may not be appropriate for tasks that require an understanding of context and judicial determination. A worrying trend is that this increased capability is not accompanied by an increase in scrutability, i.e the ability to not only see, but also to understand and investigate decisions made by, or on the basis of AI.

---

[20]Nanette Byrnes, "An AI-fueled credit formula might help you get a loan", MIT Technology Review, 14 February, 2017, http://bit.ly/2lLJBzt.

[21] "Goldman expands algorithmic bond trading", Financial Times, 16 August 2016, http://on.ft.com/2x2tt1S.

[22] "How Artificial Intelligence is shaping the future of energy", Open Energi, 9 February 2017, http://www.openenergi.com/artificial-intelligence-future-energy/.

[23] For example, see Hao Bai, "A Survey of AI for Network Routing Problems", http://bit.ly/2eK3WTP.

[24] Jonathan Vanian, "Amazon bows to pressure to bring same-day delivery to poor areas", Fortune, 6 May 2016, http://fortune.com/2016/05/06/amazon-ignore-poor-neighborhoods-deliveries/.

[25] Will Knight, "AI Programs are learning to exclude some African-American voices", MIT Technology Review, 16 August 2017, http://bit.ly/2wS2W7n.

[26] See footnote 18, above.

[27]Julia Carrie Wong, "Mark Zuckerberg accused of abusing power after Facebook deletes Napalm girl post", 9 September 2016, The Guardian, http://bit.ly/2c2e0GI.

[28] Shaun Walker, "Face recognition app taking Russia by storm may bring end to public anonymity", 17 May 2016, The Guardian, http://bit.ly/23VMZpb.

[29] Ryan Calo, 'Artificial Intelligence Policy: A Roadmap', https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3015350.

8. The excitement currently surrounding AI lacks clarity. There is a tendency to conflate machines that use powerful statistical and probabilistic methods to solve problems, with machines that exhibit human intelligence across domains. At present, AI can surpass human understanding within *narrow, deliberate* domains that these technologies are trained in, as was demonstrated by AlphaGo in May 2017.[30] However, we are still far away from building machines that can truly 'think'. Even the most complex AI currently cannot begin learning without direction, a human must still guide the machine and train it on what to look for, which also involves AI amplifying the preferences and values of the trainer. While the potential of AI is indeed exciting, the current hype is largely *misinformed* by popular coverage of developments in the area. Excitement surrounding the 'Singularity' focuses on fictional threats while ignoring the more urgent and immediate considerations for AI.[31]

**Ethics**

9. As AI is increasingly deployed in various sectors, ARTICLE 19 considers that there is a need for a shared ethical framework within which these algorithms can function. **The development and use of AI must be subject to the minimum requirement of respecting, promoting, and protecting international human rights standards.** This would at the very least, ensure a minimum level of fairness and accountability in these processes. It is only through this minimum standard that legal-ethical considerations like fairness, accountability can be reached.

10. At the development stage, AI is embedded with values, i.e the model is made to optimise for some specific attributes, for a specific outcome, determined by the developers of these systems. The ethical implications here are the choice of the dataset used, the prioritisation of various attributes, and the safeguards put in place to ensure the promotion of fairness, scrutability, inclusivity, and accountability. For instance, the implications of a non-inclusive training dataset were exposed in 2015, when Google's photo app was found to be tagging black people as gorillas.[32] Similarly, the issue of identifying attributes and gender discrimination in machine learning was brought to test by a 2015 study at Carnegie Mellon University which found that Google advertisements for high paying jobs were more likely to be shown to men, as opposed to women.[33]

11. At the stage of implementation, ARTICLE 19 finds that the *manner* in which AI is used gives rise to both ethical concerns and concerns for the protection of human rights, **particularly freedom of expression.** AI enables censorship in the form of content removal, prioritization, filtering, and blocking algorithms. The detection and removal of content relating to online extremism and child sex abuse images through 'hashes' in the UK and the US relies on AI, but at the same time risks over blocking and operates without judicial oversight, thus setting dangerous precedent.

---

[30] "AlphaGo beats planet's best human Go Player", Tech Crunch, 25 May 2017, http://tcrn.ch/2rk3Mue.

[31] Caroline Sinders, "Dear Elon - Forget killer robots, here's what you should really worry about", Fast Code Design, http://bit.ly/2wbnisu.

[32] "Google photos identified black people as gorillas but racist software isn't new", Splinter News, 1 July 2015, http://bit.ly/2gFs767.

[33] "Questioning the Fairness of Targeting ads online", CMU, 7 July 2015, http://bit.ly/2w5Mwa3. Also, see https://www.theladders.com/p/26101/ai-screen-candidates-hirevue as critique on the objectivity of AI.

Similarly, YouTube's content removal algorithm, ContentID, asymmetrically privileges content owners over content creators, even in the case of legitimate speech.[34] Presently, AI is very poor at understanding context,[35] but is made to carry out tasks that require it to do so, which means that sometimes these technologies block or enable the removal of legitimate content, evidenced most recently by Facebook's takedown of a Pulitzer prize-winning Vietnam war photograph of a naked girl.[36]

12. Potential ethical implications of AI are difficult to determine because usually, the use, underlying values and problems within these systems become apparent only when a harm arises.[37] Resolving negative implications needs to start with ensuring that the ethical framework within which AI functions has a strong grounding in international human rights standards as a minimum level of protection.

13. A relative lack of transparency, i.e black boxing, or making the logic or data being used by an AI system selectively available, is acceptable **only** where absolute transparency would involve the violation of fundamental human rights, particularly the disclosure of personal or sensitive data of individuals.[38]

14. While transparency in AI systems is desirable, it is not in and of itself sufficient to hold algorithms accountable.[39] It is important to stress here that the requirement for transparency in AI systems is only meaningful when it leads to the end goals of fairness, accountability, or intelligibility.[40] It is far more effective to embed values of fairness, accountability, and non-discrimination at the time of building AI systems.

**The role of the Government**

15. ARTICLE 19 believe that the government has a role to play when it comes to AI. In particular, the government should:

   (i) **Ensure respect for international human rights standards:** A one-size-fits-all approach cannot work in context of the regulation of AI because of the sheer variety of AI systems and capabilities, varying degrees and instances of application, the stakeholders involved, and the nature of decisions being made. However, the minimum requirement for all AI and applications arising from AI should be compliance with international human rights standards.

---

[34] https://www.digitalmusicnews.com/2016/02/29/youtube-alters-response-to-takedown-complaints/.

[35] https://www.eff.org/files/AI-progress-metrics.html#Reading-Comprehension

[36] https://www.theguardian.com/technology/2016/sep/09/facebook-reinstates-napalm-girl-photo.

[37] Brent Mittelstadt et al, 'The Ethics of Algorithms: Mapping the Debate', Big Data & Society, 3(2).

[38] Also, see recommendation from the Council of Europe on protection of human rights with regard to search engines here: https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805caa87.

[39] Mike Ananny, 'Toward an Ethics for Algorithms: Convening, Observation, Probability and Timeliness', Science, Technology and Human Values, 1-25, 2015.

[40] Frank Pasquale, The Black Box Society, Harvard University Press, 2015.

(ii) **Ensure accountability of self regulatory mechanisms**: As ARTICLE 19 has previously stated in its briefing paper[41] on algorithms and automated decision making in the context of crime prevention, AI applications are generally used by online intermediaries to block, filter, takedown and remove content. This usually takes place on a self-regulatory basis, though often as a result of government pressure. In practice, this means that wrongful restrictions on access to content are placed beyond judicial scrutiny. At a minimum, the role of the government here is to ensure that individuals have a remedy to challenge decisions based on AI that interfere with their human rights.

(iii) **Promote a multi-stakeholder approach**: As AI is developed by various actors, and impacts a wide range of actors, policy-making in the area of AI should happen in a multi-stakeholder fashion.

### Learning from others

16.    ARTICLE 19 notes that the following lessons can be learned from other initiatives:

(i) **The need for meaningful mechanisms to challenge decisions based on AI:** The European Union's General Data Protection Regulation 2016/679 ('GDPR') provides a right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.[42] The recitals clarify that "in any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision".[43] While this is a step in the right direction, the scope of what has been dubbed the 'right to an explanation' is limited.[44] In particular, it does not apply if the decision: (i) is necessary for entering into, or performance of, a contract between the data subject and a data controller, (ii) is authorised by law; or (iii) based on explicit consent.[45] It also does not apply if the decision does not produce a legal or similarly significant effect on the data subject.

(ii) **The importance of aligning policy with technical capabilities**: Policy requirements must be aligned with solutions that are technically feasible, meaningful, and practical. This has been a challenge so far. For example, transparency standards that have traditionally been considered to be a prerequisite to accountability, may not be meaningful or desirable in context of AI systems due to their sheer complexity. Policy requirements must also be practical. For example, the GDPR stresses the importance of meaningful explanations of decisions made by

---

[41]Article19's brief on Algorithms and Automated Decision Making, http://bit.ly/2f219pd.

[42] Article 22 GDPR.

[43] See in particular Recital 77.

[44] Sandra Wachter et al, "Why a right to explanation of automated decision making does not exist in the General Data Protection Regulation", International Data Privacy Law (forthcoming), December 28, 2016.

[45] Article 22 (2) GDPR.

autonomous systems. However, this is in inherent tension with the explanation that machine learning algorithms are developed to offer.[46]

(iii) **The need to think outside the box**: Considering the new questions that the development of AI raises, for instances in the case of automated vehicles and embodied AI (robots), it is vital to engage in a broad scoping exercise of regulatory efforts around the world. For example, the Committee should take note of the Japanese Robot Strategy,[47] the forthcoming legal developments in South Korea,[48] and the calls for a European Agency for robotics. [49]

---

[46] Lilian Edwards and Michael Veale, "Slave to the algorithm? Why a right to an explanation is probably not the remedy you are looking for", Duke Law & Technology Review, Forthcoming, 3 July 2017, available at http://bit.ly/2j10atE.

[47] "Japan's Robot Strategy was compiled", January 2015, http://bit.ly/2eF9ZbY. Also see http://bit.ly/2vL4frU.

[48] "Legal preparation for AI era", Business Korea, 17 February 2017, http://bit.ly/2eEETkx.

[49] "Robots: Legal Affairs Committee calls for EU-wide Rules", 12 Jan 2017, http://bit.ly/2x9Wg8m.