LIBERTY

PROTECTING CIVIL LIBERTIES
PROMOTING HUMAN RIGHTS

Liberty’s written evidence to the Select Committee on Artificial Intelligence

September 2017

About Liberty

Liberty (The National Council for Civil Liberties) is one of the UK’s leading civil liberties and human rights organisations. Liberty works to promote human rights and protect civil liberties through a combination of test case litigation, lobbying, campaigning and research.

Liberty provides policy responses to Government consultations on all issues which have implications for human rights and civil liberties. We also submit evidence to Select Committees, Inquiries and other policy fora, and undertake independent, funded research.

Liberty’s policy papers are available at
http://www.liberty-human-rights.org.uk/policy/

Introduction

1. Liberty welcomes the establishment of the Select Committee on Artificial Intelligence and the opportunity to submit evidence to its inquiry.
2. The technological revolution is transforming society in numerable ways. In light of current and near-future seismic technological shifts, Liberty has expanded the scope of its work to include a new programme on Technology and Human Rights. We seek to provide the Committee with a brief human rights analysis of some of the great challenges and opportunities that artificial intelligence presents to the UK.
3. In this submission, Liberty wishes to focus on the following key issues:
    a. AI, privacy and data protection
    b. AI and equality rights
    c. AI, transparency and accountability
    d. AI and weaponry

The pace of change

4. The promise of scientific progress, clinical rationality and increased efficiency means AI is regarded as a highly desirable technology within relevant sectors. There is a clear rush to deploy AI in various areas of research and public life, making the pace of change rapid.
5. The rapid pace of change perhaps explains some of the most controversial current uses of AI in the UK. AI is already being used for predictive policing by Kent Police, for health research on NHS patient data by Google DeepMind, and for online advertising including political advertising.
6. Such a rapid pace of change often means that fundamental considerations such as the impact on human rights, data protection, transparency and public consent are neglected. This risks a silent deterioration of core values in the name of scientific 'progress'.

Techno-optimists and techno-pessimists

7. Whilst a tangible dichotomy in practice, the distinction between 'techno-optimists' and 'techno-pessimists' is not a helpful one. Both terms suggest a somewhat deterministic role of technology in society, with stakeholders predicting rather than determining what technologies are used, when they are used, and the outcomes.
8. In Liberty's view, it is paramount to actively uphold the rule of law and the human rights framework provided by the Human Rights Act 1998 throughout the ongoing

technological revolution. We firmly believe that human rights laws can play a leading role in safeguarding rights and liberties during this period of great change.

9. Human rights laws should not only guide the passage of new technologies into society – they should be hardwired into those technologies. Software should be designed with privacy, freedom of expression, accountability and civil liberties in mind as normative principles – ensuring fundamental rights underpin the digital sphere. ‘Rights by design’ may be more challenging where AI is concerned, as the software organically learns and adapts in response to its environmental input.

Summary

- **Privacy:** There is no tension between privacy or indeed any fundamental rights and socially beneficial scientific progression. Privacy is not a necessary sacrifice for positive applications of AI.
- **No discrimination:** Whilst data science clearly has the potential to illuminate and ameliorate discrimination, it is unclear why AI per se would be necessary for such progression. AI systems should not be reified as objective or decontextualised from the social context and ideologies within which they are constructed. Training datasets, and any social dataset used to train AI programmes, must be carefully assessed and controlled for patterns of historical and ongoing bias, or sampling deficiencies, to avoid the perpetuation or creation of discrimination and inequalities.
- **Diversity:** It is vital for the success of AI that workforces in the tech sector are representative of the diversity of experience and backgrounds within the society they seek to operate in.
- **Democracy:** The public should be informed of the areas in which AI is being applied and how it is being applied, whether in public policy or in relation to individual decisions. Liberty supports a growing public debate on the topic of AI.
- **Transparency:** AI systems should be transparent, open-source where possible, their functioning intelligible, their operation subject to democratic oversight, and both the systems’ and their developers’ decision-making accountable. AI-related decisions that engage the rights and liberties of individuals should always be challengeable.
- **Accountability:** AI and automatic processing must not be the sole basis for a decision which produces legal effects or engages the rights of any individual.
- **Stop Killer Robots:** Liberty joins the call for a ban, by way of international treaty, on lethal autonomous weapons systems that lack meaningful human control.

AI, privacy and data protection

10. Many AI systems are built on ‘big data’, whether ‘open data’ (publicly available data) or personal data. Some AI systems are built on smaller training datasets. Data protection, privacy and related rights must be closely regarded alongside the development of AI. In Liberty’s view, it would be beneficial to incorporate these topics into computer science and related academic programmes.

11. Personal data is often the fuel for AI – whether for research, commercial products, or personalised services. The unlawful data sharing of 1.6 million identifiable patient records by the Royal Free London NHS Trust to a Google AI start-up, Google DeepMind, is a prime example of the risks to basic rights in the rush for AI. There is little personal information that is more profoundly private than medical information. Liberty is deeply concerned by the effect that this reprehensible data sharing has had on patients. We are providing free legal advice to a number of Royal Free patients who have contacted us seeking help, having lost confidence in the confidentiality they are entitled to in the course of their healthcare.

12. Seeking to build AI tools by training software with personal data received in breach of the law is needlessly reckless. Privacy and consent are not only pillars of our democracy, the rule of law, public health, but they are also essential for technological innovation. There need be no conflict between privacy and innovation – innovators must simply respect the rule of law and human rights in the course of advanced software development.

13. The shrinking of the private sphere and the growth of a surveillance society are broad concerns amplified by the increasing AI applications that are fuelled by personal data – even when data is lawfully exchanged. For example, the increasing use of virtual personal assistants and growing personalisation functions in services requires software to learn from people by pervasively ingesting their data, effectively surveilling them. The BBC’s head of digital partnerships recently said:

> "*Just by listening to the voices in the room, your TV could automatically detect when there are multiple people in the living room, and serve up a personalised mix of content relevant to all of you in the room*.”[1]

[1] *Future BBC iPlayer could tell who is in the room and notice when the children have gone to bed* – Anita Singh, The Telegraph, 19 Aug 2017: http://www.telegraph.co.uk/news/2017/08/19/future-bbc-iplayer-could-track-familys-movements/ (accessed 21 Aug 2017).

This data exchange may be well-intended and deemed to be to the user's benefit. Accordingly, the data exchange could be constructed in a lawful way – for example, if the service were optional and on the basis of fully informed consent. Even so, it remains that the normalisation of pervasive monitoring, passive quantification and intelligent personalisation risks reshaping society in ways unconsidered. Never before have human societies been monitored and quantified in this way – pervasive monitoring leads to self-monitoring, whether conscious or unconscious, and inhibits the development of personalities and ideas. Truly private spaces risk being eradicated, even from the home. Constant personalisation creates risks too, both for a free press and freedom of thought. The benefits of exposure to diverse media sources risk being limited by the echo chambers artificially constructed around each person, fuelling radicalism and social divides.

14. This risk is compounded by the pervasive suspicionless surveillance people are subjected to by the State, as individuals can never be sure that the data they generate is *only* being used for personalised services, etc., and not also being aggregated by the authorities. Already, our phone call records, text message records, GPS location data, Automantic Number Plate Recognition (ANPR) data, TfL data, travel data, banking data, and internet browsing records – not to mention unnamed 'bulk personal datasets' – are hoarded and processed. The State's secretive approach to web logging and data gathering may cause people "*to feel that their private lives are the subject of constant surveillance*"[2] – even when services promise not to pass data on to a third party, State surveillance is always exempt and its shadow hangs over society, chilling free expression.

15. In addition, unique rights and ethics issues may still arise where data is interpreted and acted on using AI – again, even where data is exchanged lawfully. For example, Facebook has started using AI to identify users deemed at risk of suicide, in order to launch interventions.[3] The technology that exists to do this is increasingly advanced, and now predictive.[4] There is no evidence that suggests this type of 'intelligent' monitoring is in the best interests of vulnerable peoples' mental health and we are

[2] Joined Cases C-203/15 and C-698/15, *Tele2 Sverige AB v Post- och telestyrelsen, and Secretary of State for the Home Department v Tom Watson*, 21 December 2016, para. 100

[3] *Facebook Using Artificial Intelligence to Help Suicidal Users* – Aatif Sulleyman, The Independent, 2 March 2017.

[4] *Artificial Intelligence Can Now Accurately Predict Suicide Attempts Two Years in Advance* – Paul Tamburro, Crave, 3 March 2017

concerned that a shrinking private sphere may indeed deter people from seeking social support and a safe space to freely express themselves.

16. This ‘intelligent’ processing of varied information should be subject to the data protection principle of fair and lawful processing. Liberty welcomes the clarity that will be provided by the EU General Data Protection Regulation (GDPR), which we anticipate will be passed into UK law via the forthcoming Data Protection Bill 2017, on the matter of specific, informed and unambiguous consent for data processing[5] and the right not to be subjected to automated profiling.[6]

AI and equality rights

17. AI systems are built from programmed rules, training datasets, and learned rules from past outcomes. Inherent in the generation of AI systems, therefore, is potential for in-built bias either from the developers, programmed rules, the training datasets, or past outcomes.

18. System developers may programme rules in AI systems that are influenced by unconscious biases. Constructing AI systems requires the selection of certain features and whilst this relies on the mathematical expertise of developers, the process will also inevitably be influenced by their intuition, underlying worldview, culture, and motivations.[7] Thus, transparency of the function of AI systems and close monitoring of how they operate in the real world is very important.

19. Training data, or in fact any data collected from society, may be reflective of patterns of discrimination and existing inequalities: *“to the extent that society contains inequality, exclusion or other traces of discrimination, so too will the data”*.[8] Social data come with complex histories, which may silently haunt the logic underpinning social policy if uncritically used. Patterns of social inequalities can be perpetuated through algorithmic processes – for example, Google advertises highly paid jobs to men more often than to women.[9] The institutionalisation of AI systems may have more serious effects still.

[5] GDPR Article 4(11) and Recital 32

[6] GDPR, Article 4(4); GDPR, Article 13 (2)(f), and in particular, GDPR, Article 22(1)

[7] *Algorithmic paranoia and the convivial alternative* – Dan McQuillan, Big Data & Society, July-December 2016, p.4

[8] *European Union regulations on algorithmic decision-making and a “right to explanation”* – B. Goodman, S. Flaxman, Aug 2016, p.3

[9] *Artificial Intelligence’s White Guy Problem* – Kate Crawford, The New York Times, 25th June 2016

20. Algorithmic profiling involves categorising and analysing people on the basis of a variety of categorisations and group memberships.[10] This is not only an issue when sensitive data, such as race, gender, health, religion, etc., are used for profiling. Even if race is prohibited as a category of profiling data (as may be the case under the GDPR), combinations of other categories of data can unintentionally serve as a proxy for race. For example, "*if a certain geographic region has a high number of low income or minority residents, an algorithm that employs geographic data to determine loan eligibility is likely to produce results that are, in effect, informed by race and income*."[11] This may be important to consider in uses of AI by law enforcement. For example, Kent Police use predictive policing software, PredPol, with little transparency – it is not publicly known what categories of data are processed. Our engagement with police forces about the potential for discriminatory biases in various algorithms they use has thus far revealed a concerning disregard for the issue.[12]

21. In addition, a predictive policing tool is likely to be based on large amounts of existing policing and crime data – but if this data reflects socio-economic, geographic or racially based discriminatory policing, those biases risk being entrenched in the tool the data seeks to produce, as early evidence suggests.[13]

22. Durham Police is using AI software, the Harm Assessment Risk Tool ('HART') in bail decisions, with little transparency. There is evidence that data science has perpetuated discrimination in criminal justice in the US. A recidivism algorithm called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions, by Northpointe, Inc.) was found to be twice as likely to incorrectly judge black defendants as at high risk of reoffending than white defendants.[14] This is despite race not being one of the categories of information ingested. Without open source algorithmic transparency, we cannot know exactly how or why the algorithm came to these conclusions. Such decision making should be challengeable, subject to an

[10] *European Union regulations on algorithmic decision-making and a "right to explanation"* – B. Goodman, S. Flaxman, Aug 2016, p.3

[11] *European Union regulations on algorithmic decision-making and a "right to explanation"* – B. Goodman, S. Flaxman, Aug 2016, p.4

[12] *Misidentification and improvised rules - we lift the lid on the Met's Notting Hill facial recognition operation* – Silkie Carlo, Liberty, Aug 2017

[13] "*Stuck in a Pattern: Early evidence on 'predictive policing' and civil rights*," - David Robinson and Logan Koepke, Upturn, August 2016

[14] *Machine Bias* - Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica, May 23, 2016

adversarial court proceeding - this opaque AI application is discriminatory and highly inappropriate.

23. Liberty has made requests for information to Durham Constabulary about HART under the Freedom of Information Act, which have been rejected – in our view, wrongly – under Section 21 (information reasonably accessible to the applicant by other means). We have recently resubmitted those requests. We were instead directed to an academic paper and Durham Constabulary's written submission to the Science and Technology Committee's inquiry into algorithms in decision making. Regarding the risk of discriminatory patterns in data being silently reproduced in algorithms, Durham Constabulary wrote in that submission,

> "*It would be wrong, and the error rates would increase, if the model failed to reflect reality. Before concluding that algorithms should therefore be viewed as biased, it is necessary to consider whether human judgement is more or less biased*."[15]

This attitude reflects a disregard of the duty police have to ensure equality and fairness in policing. 'Reflecting reality' is not where the bar for success should be set. The threshold for adopting AI in criminal justice and law enforcement should not simply be whether one function of the software exceeds human functioning on some measure – software must be considered holistically, accounting for new risks and long-term impact. Data analysts should, at the very least, attempt to discover and control for biases in existing datasets before using them to train AI tools for live deployment in the criminal justice system where they risk being embedded, reified and obscured from accountability. Seemingly progressive AI applications may produce socially regressive output and must not automatically be attributed as objective, or divorced from ideology.

24. Another reason that algorithms may not produce fair decisions for minority groups is that there may be too small a sample of data from which to generate predictions with confidence. For this reason, minorities are sometimes oversampled in public policy research.[16] An example of inadequate training data resulting in discrimination was seen in Google's photo app, which classified black people as gorillas. Similar examples of algorithmic discrimination include Nikon software reading photos of

[15] Written evidence submitted by Durham Constabulary (ALG0041; para. 22) in response to the Science and Technology Committee's inquiry into algorithms in decision making – April 2017

[16] *European Union regulations on algorithmic decision-making and a "right to explanation"* – B. Goodman, S. Flaxman, Aug 2016, p.4

Asian people as blinking and HP webcam software having difficulty recognising users with dark skin tones.[17] Low representation in training data could be a possible explanation for higher inaccuracy rates for the identification of women by facial recognition software.[18] Oversampling may be a solution in some circumstances, but if doing so requires increased surveillance or data collection on a particular group it could raise serious legal and ethical issues. We caution against subjecting a portion of society to increased surveillance for a technocratic 'greater good' – particularly for groups that are typically marginalised.

25. The performance of AI systems must be carefully tested before systems are operational, and continually monitored, as discriminatory flaws may not be easily discoverable and too often only come to attention once they already are having negative effect. Disturbingly, Liberty has witnessed a wilful ignorance from police towards potential discriminatory flaws, who seem confident in the professed objectivity of such software. For example, the Metropolitan Police told us they had no intention to monitor demographic accuracy (or even to ask the vendor if it had been tested for) in their use of facial recognition software. In addition, the Commissioner of the Metropolitan Police failed to respond to, or even acknowledge, a letter from Liberty and 12 other rights and race equality groups raising concerns about accuracy bias.[19]

26. Whilst the risks of embedding patterns of discrimination in AI are clear, it is possible to use data and design to control for and thus reduce discrimination. There is a view in the US that, if care is taken to minimise the possibility of biases or inaccuracies in the data, AI "*has the potential to improve aspects of the criminal justice system, including crime reporting, policing, bail, sentencing, and parole decisions*".[20] Whilst data science clearly has the potential to illuminate and ameliorate discrimination, it is unclear why AI per se would be necessary for such progression. Clearly, the COMPAS case study was not successful.

27. Since software, including AI, reflects the values of its creators it is important that workforces in this sector are representative of society. It is particularly urgent address

[17] *Artificial Intelligence's White Guy Problem* – Kate Crawford, The New York Times, 25th June 2016

[18] *Face Recognition Performance: Role of Demographic Information* - Brendan F. Klare ; Mark J. Burge ; Joshua C. Klontz ; Richard W. Vorder Bruegge ; Anil K. Jain, IEEE Transactions on Information Forensics and Security (Volume: 7, Issue: 6, Dec. 2012)

[19] *Misidentification and improvised rules - we lift the lid on the Met's Notting Hill facial recognition operation* – Silkie Carlo, Liberty, Aug 2017

[20] *Preparing for the Future of Artificial Intelligence* – Executive Office of the President, National Science and Technology Council Committee on Technology, Oct 2016, p.14

the under-representation of women - only 17% of those working in technology in the UK are female and just 7% of students taking computer science A-level courses are female.[21]

AI, transparency and accountability

28. It is important that decision making is accountable in a democracy – both of AI creators and their creations. However, accountability for their creations can be frustrated by several factors: the highly complex, multi-dimensional nature of many processing systems that are not easily interpretable if at all; the prevalence of commercial, proprietary systems and of secret systems (e.g. in state intelligence); the inability of probabilities to offer explanations for output beyond the assumption that past associations between data will be replicated in the future; and the inaccessibility of source code and complex algorithmic rules to many individuals, even where it is published.

29. The internal machinations of AI systems are often highly complex, opaque and sometimes near-impossible to 'reverse-engineer'. Although AI processing of big data can reveal previously invisible relationships, the processing itself is inherently opaque. AI systems are rarely designed to provide explanations for the output they produce or decisions they make. AI-based social decision-making "*renders individuals unable to observe, understand, participate in, or respond to information gathered or assumptions made about them"* and has been described as "*antithetical to privacy and due process values".*[22] It can even be argued that AI in social decision making can be "*authoritarian (…) in that it eludes democratic oversight and, so far, evades a social discourse capable of challenging its teleology".*[23] We must be alive to the potential that authoritatively 'objective', opaque calculations behind social policy in the future may evade challenges based on human reasoning, disempower the public and shift the relationship between the citizen and state.

30. Furthermore, AI systems are increasingly designed to learn, adapt and improve so their processes may change during deployment. Such systems pose *"perhaps the biggest challenge – what hope is there of explaining the weights learned in a*

[21] Women In Tech (womenintech.co.uk), accessed Sept 2017

[22] *Prediction, pre-emption, presumption: How big data threatens big picture privacy* – J. Earle & I. Kerr, Stanford Law Review Online 66:65, 2013

[23] *Algorithmic paranoia and the convivial alternative* – Dan McQuillan, Big Data & Society, July-December 2016, p.5

multilayer neural net with a complex architecture?"[24] This opacity, combined with perceptions of AI as producing perfect, rational, if unknowable calculations, creates a situation in which important decisions may be readily accepted but too reluctantly questioned or scrutinised. It is certainly true that *"algorithmic vision derives authority from its association with science (…) an aura of neutrality and objectivity, which can be used to defend against the critique that they carry any social prejudice."* [25]

31. An additional risk of opaque AI systems is that they may have unintended consequences or malfunction, without being readily noticed or understood. This idea is sometimes the source of dystopian science fiction, but could have some more everyday negative impacts. The recently agreed Alisomar Principles on AI state that developers must provide transparency to understand failures, plan for/mitigate catastrophic risks, and ask questions such as, *"how can we make future AI systems highly robust, so that they do what we want without malfunctioning or getting hacked?"*[26]

32. Researchers should design open-source AI systems where possible, and always maintain the ability to provide transparency as to their output and explanations for decisions made. Human interpretability is especially important where AI is used in public decision making and applications that could have human rights engagements or require ethical consideration. Both AI systems and their creators must remain accountable.

33. It is important to maintain the same standards of accountability and transparency when using AI – we believe AI, particularly where it interacts with individual's rights, should be rejected where transparency is not possible. It would be unacceptable to consider AI systems as immune from accountability.

34. It is important that the public is informed in which areas AI systems are being used, that changing behaviours of adaptive AI systems are closely monitored, and that AI decision-making is always transparent. The Science and Technology Committee described decision-making transparency as one of the *"key ethical issues requiring serious consideration* (…) [and] *ongoing monitoring".*[27] Similarly, a US Executive Committee on Technology recommended: *"As the technology of AI continues to*

[24] *European Union regulations on algorithmic decision-making and a "right to explanation"* – B. Goodman, S. Flaxman, Aug 2016, p.7

[25] *Algorithmic paranoia and the convivial alternative* – Dan McQuillan, Big Data & Society, July-December 2016, p.4

[26] See Appendix I

[27] *Robotics and artificial intelligence* – Science and Technology Committee, Sept 2016, p.36

develop, practitioners must ensure that AI-enabled systems are governable; that they are open, transparent, and understandable".[28]

35. Where it is argued that AI systems or their controllers cannot be transparent or provide explanation for decisions, for example where they are used in the intelligence community, we call for the maximum possible public transparency with full transparency allowed in a closed independent review or adversarial procedure. This is important to verify that the subject's rights and freedoms are safeguarded, particularly where a system has legal or other significant effects on a subject.

GDPR

36. In limited circumstances, EU citizens may soon have the right not to be subject to algorithmic decisions that would significantly or legally affect them. Article 22 of the EU's new General Data Protection Regulation, set to take effect from 2018, states:

> *"The data subject shall have the 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."*[29]

This principle does not apply if the decision is authorised by EU or state law so long as the data subject's rights, freedoms and legitimate interests are safeguarded.[30] It is also disapplied if it is necessary under a contract between the data controller and the subject,[31] or the subject has given explicit consent.[32] Nevertheless, this Article may prohibit private corporations using AI in various applications.

37. This Article leaves open the possibility for a citizen to be subject to a decision based *solely* on automated processing, that produces legal effects, without the right to human intervention or to express their view and contest the decision, so long as the state or EU has safeguards in place to protect the subject's rights, freedoms and legitimate interests. It is hard to envisage what safeguards could effectively protect against the risks inherent in this model.

[28] *Preparing for the Future of Artificial Intelligence* – Executive Office of the President, National Science and Technology Council Committee on Technology, Oct 2016, p.4

[29] *General Data Protection Regulation, Article 22(1)*

[30] *General Data Protection Regulation, Article 22(2)(b)*

[31] *General Data Protection Regulation, Article 22(2)(a)*

[32] *General Data Protection Regulation, Article 22(2)(c)*

38. In Liberty's view, individuals should have the right not to be subject to a decision by the state that is based solely on automatic processing and which produces legal or other significant effects for them.

39. Article 22 of the GDPR further offers EU citizens 'the right to explanation' regarding an algorithmic decision made about them with consent or under contracts. The right to explanation does not apply to the State's use of AI. Article 22(3) states that the subject maintains,

> "(…) *the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision."*[33]

Furthermore, Articles 13 and 14 state that subjects have the right to be given "*meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing".*[34] Article 12 states that communication with data subjects must be in "*concise, transparent, intelligible and easily accessible form".*[35] These regulations go some way to addressing concerns about transparency around proprietary AI systems, as well as concerns about the accessibility of transparency mechanisms such as open source code – but, given the exceptions, do not fully satisfy our concerns about the transparency of state AI systems. Transparency is arguably even more important where the decision originates from the State and has significant or legal effects for the subject.

AI & lethal autonomous weapons

40. Lethal autonomous weapons systems (LAWS) could potentially identify and kill a target without human intervention. The Government has said it will not develop LAWS, but defines LAWS with a very vague and futuristic view of 'autonomy': "*An autonomous system is capable of understanding higher level intent and direction*".[36] Furthermore, the Government has opposed proposals for an international ban on the development of LAWS[37], or the development of any guidelines or additional

[33] *General Data Protection Regulation, Article 22(3)*

[34] *General Data Protection Regulation, Article 13(2)(f)* and *Article 14(2)(g)*

[35] *General Data Protection Regulation, Article 12*

[36] *Joint Doctrine Note 2/11: The UK Approach to Unmanned Aircraft Systems*, Ministry of Defence, 30 March 2011, para 205, p.14

[37] *UK opposes international ban on developing 'killer robots'* – Owen Bowcott, The Guardian, 13 April 2015, https://www.theguardian.com/politics/2015/apr/13/uk-opposes-international-ban-on-developing-killer-robots

legislation.[38] Both the UK and US cite international humanitarian law and ongoing international discussion as key to regulating the use of LAWS.[39],[40]

41. Precursors to fully autonomous weapons are being developed and deployed by the UK as well as the US, Russia, China, Israel and South Korea. A drone called Taranis is being developed by the UK's MoD and BAE Systems, and has reportedly been tested to autonomously locate and engage targets.[41] [42].

42. The US has more openly embraced autonomy in weapons systems. A US Executive Committee on Technology stated: *"The United States has incorporated autonomy in certain weapon systems for decades, allowing for greater precision in the use of weapons and safer, more humane military operations (…) Nonetheless, moving away from direct human control of weapon systems involves some risks and can raise legal and ethical questions."*[43]

43. UK Government has committed to keeping weapons under human control, but has not specified what it understands by human control. For example, 'human control' may mean authorising weapons release when prompted by a system to do so. A more appropriate goal may be 'meaningful human control' – a term that requires development but is orientated on the premise that humans must exert meaningful, cognitive control of combative actions and perform important decisive functions.

44. Even retaining meaningful human control, the site of accountability for such advanced weapons use may be problematic. Military personnel may not fully understand the internal machinations of these highly advanced systems – particularly of such closely guarded national security technologies. Should they be held accountable for using equipment that, whilst advertised as more safe, precise and rational than human intervention, they do not entirely understand?[44] Who would be held accountable, not just in law but on the political stage, for the malfunctioning of a

[38] *Statement on Lethal Autonomous Weapons Systems to the CCW Meeting of the High Contracting Parties*, United Kingdom of Great Britain and Northern Ireland, 12–13th November 2015

[39] *Preparing for the Future of Artificial Intelligence* – Executive Office of the President, National Science and Technology Council Committee on Technology, Oct 2016, p.3

[40] *Statement on Lethal Autonomous Weapons Systems to the CCW Meeting of the High Contracting Parties*, United Kingdom of Great Britain and Northern Ireland, 12–13th November 2015

[41] *Anglo-French UCAV Study Begins To Take Shape* – Toby Osborne, Aviation Week, 4th Feb 2016, http://aviationweek.com/defense/anglo-french-ucav-study-begins-take-shape

[42] *The United Kingdom and lethal autonomous weapons systems* – Article 36, April 2016, p.2

[43] *Preparing for the Future of Artificial Intelligence* – Executive Office of the President, National Science and Technology Council Committee on Technology, Oct 2016, p.37

[44] See also evidence to the Science and Technology Committee by Richard Moyes, Article 36, cited in *Robotics and artificial intelligence* (Sept 2016), para/ 56, p.21

LAWS? Clearly, maintaining transparency and meaningful human control of weapons systems' functioning will be vital to ensure proper accountability frameworks and the upholding of obligations under human rights law.

45. Liberty calls for a pre-emptive ban on the development and use of lethal autonomous weapons that do not involve meaningful human control. We concur with Human Rights Watch and industry leaders that such a pre-emptive ban is necessary now.[45] [46] We support the Government's commitment to maintaining human oversight and control over the use of force, but stress that such control must be *meaningful*.

Silkie Carlo

[45] *Killer Robots* – Human Rights Watch, accessed 25.1.16. See: https://www.hrw.org/topic/arms/killer-robots

[46] *Written evidence submitted to the Science and Technology Committee's inquiry on Robotics and Artificial Intelligence* (ROB0062) – Google DeepMind, May 2016