

Institution: University of Sheffield		
Unit of Assessment: B-11 Computer Science and Informatics		
Title of case study: Shaping international policy and public debate on lethal autonomous weapons systems (LAWS)		
Period when the underpinning research was undertaken: 2001–2014		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Noel E. Sharkey	Professor of AI and Robotics and Professor of Public Engagement (now Emeritus)	1994–2013
A.J.C. Sharkey	Senior Lecturer	1994–2019
Period when the claimed impact occurred: August 2013–2020		
Is this case study continued from a case study submitted in 2014? Y		
<p>1. Summary of the impact (indicative maximum 100 words)</p> <p>Research into the capabilities and ethics of artificial intelligence (AI) conducted at the University of Sheffield has fuelled debate at national and international levels on the development of lethal autonomous weapons systems (LAWS). Evidence from the research prioritised the issue on the UN agenda and ensured that technological issues were fully understood and considered by member states. In addition, it underpinned a global civil society campaign against LAWS, accomplished in large part through a coalition of NGOs co-founded by Professor Noel Sharkey, whose ranks swelled from 7 to 172 NGOs in the submission period. The arguments of the campaign persuaded national governments and political unions worldwide to support a ban on the development of LAWS.</p>		
<p>2. Underpinning research (indicative maximum 500 words)</p> <p>Two strands of Professor Noel Sharkey's research underpin the impact: one on robotics and another on the ethics and legality of the use of robots in LAWS.</p> <p>Throughout the 2000s, Sharkey extended upon his earlier novel research on the close relationship between the physical embodiment of robots and their control systems to understand the potential and limitations of future robots. In a 2001 study comparing human and robot embodiment, he argued that the two are fundamentally different and that strong embodiment, either mechanistic or phenomenal, is not possible for present day robots [R1]. This conclusion both provided an argument against strong AI and underpinned Sharkey's developing views on robot ethics.</p> <p>Sharkey investigated robot programming, learning, and construction, particularly with regards to the use of neural networks for learning robot behaviours. Examples include (1) work on robot localisation – the problem of determining a robot's location quickly, reliably and accurately – to which he and A.J.C. Sharkey proposed a novel solution exploiting self-organising maps and ensemble techniques [R2]; and (2) work on robotic arm control – the problem of how to rapidly adapt a robotic arm controller given the new geometric space that arises when sensor position is changed or sensors are replaced – to which he proposed a new solution combining genetic algorithms and neural networks [R3]. This body of work yielded in-depth insights into the</p>		

capabilities and limitations of autonomous robots and established Sharkey's technical credibility as a robotics expert.

In 2005, Sharkey began to investigate the ethical issues surrounding the use of robots in various applications, particularly military applications, leading him to explore whether the use of robots in LAWS could be deemed either morally defensible or legal according to international laws and conventions governing weapons and warfare. With reference to specific robot technologies and military robots, Sharkey analysed the perceptual and cognitive capabilities required for robots to distinguish between combatants and non-combatants, showing that current robot technologies fall far short of possessing these capabilities, despite the claims made by arms developers and military organisations. He therefore argued that autonomous robot weapons fail to meet two key principles of international humanitarian laws governing warfare (such as the Geneva and Hague Conventions): discrimination (it must be possible for an attacker to distinguish combatants from non-combatants) and proportionality (the anticipated loss of life and damage to property incidental to attacks must not be excessive in relation to the concrete and direct military advantage expected to be gained).

Sharkey also explored the concept of autonomy in robots, determining the degrees of autonomy that military robots can exhibit and finding that the failures of current non-autonomous, "human-in-the-loop" remote military weapons, such as drones, are even more likely to occur in LAWS. Supported by a Leverhulme Trust Senior Research Fellowship, this research has led to 11 publications in military ethics, law and technology journals (e.g. **[R4]** on the use of automated robots in wars and the new type of battle stress they introduce), as well as high-quality science and engineering journals (e.g., **[R5]** on the application of AI to discriminate between innocents and combatants in modern warfare), with over 185 total academic citations.

Drawing on his background in psychology, Professor Sharkey conducted further research to explore what "human control" means in relation to LAWS. By referencing theories of automatic and deliberative human behaviour and examining 60 years of data on human "supervisory control" of machines, he introduced a new framework that reframes autonomy in terms of supervisory control and allows for greater transparency and allocation of responsibility **[R6]**.

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

Sheffield staff and students in **bold**.

- R1. Sharkey, N. E., & Ziemke, T.** (2001). Mechanistic versus phenomenal embodiment: Can robot embodiment lead to strong AI? *Cognitive Systems Research*, 2(4), 251–262. [https://doi.org/10.1016/s1389-0417\(01\)00036-5](https://doi.org/10.1016/s1389-0417(01)00036-5). Cited by 62.
- R2. Gerecke, U., Sharkey, N. E., & Sharkey, A. J. C.** (2003). Common evidence vectors for self-organized ensemble localization. *Neurocomputing*, 55(3–4), 499–519. [https://doi.org/10.1016/s0925-2312\(03\)00391-6](https://doi.org/10.1016/s0925-2312(03)00391-6). Cited by 13.
- R3. Rathbone, K., & Sharkey, N. E.** (2002). Evolving lifelong learners for a visually guided arm. *Integrated Computer-Aided Engineering*, 9(1), 1–23. <https://doi.org/10.3233/ica-2002-9101>. Cited by 2.
- R4. Sharkey, N.** (2011). The Automation and Proliferation of Military Drones and the Protection of Civilians. *Law, Innovation and Technology*, 3(2), 229–240. <https://doi.org/10.5235/175799611798204914>. Cited by 46.
- R5. Sharkey, N.** (2008). Cassandra or False Prophet of Doom: AI Robots and War. *IEEE Intelligent Systems*, 23(4), 14–17. <https://doi.org/10.1109/mis.2008.60>. Cited by 78.

R6. Sharkey, N. E. (2014). Towards a principle for the human supervisory control of robot weapons. *Politica & Società*, 2, 305-324. <http://doi.org/10.4476/77105>. Cited by 17.

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

Impact on a global campaign against the development of LAWS

From 2007 to 2013, as part of an engagement campaign, Sharkey presented his research-based technological and ethical case against the development of LAWS, which select targets for lethal force without human intervention [R2-R5]. He became the leading voice of expertise on this subject in the media, giving evidence to national and international military and government bodies and calling urgently for formal international discussion. This formed the basis of a REF2014 impact case study.

With formal international discussions on LAWS beginning in November 2013, Sharkey continued to draw on his research [R1, R6] to drive international debate through both the **International Committee for Robot Arms Control** (ICRAC - an NGO that he co-founded in 2009 to bring together interdisciplinary academic experts who shared his concern about the use of robots in LAWS) and the **Campaign to Stop Killer Robots** (CSKR - a coalition of 7 humanitarian NGOs, including ICRAC, that he co-founded in 2012 whose mission is “to ban fully autonomous weapons and thereby retain meaningful human control over the use of force”) [S1a].

Sharkey supported the mission of the CSKR by providing expert knowledge and research-based evidence to CSKR stakeholders in the debate throughout the submission period. The CSKR campaign leader, a **Nobel Peace Laureate**, confirmed, “Professor Noel Sharkey’s research into the ethics and capabilities of AI has provided the inspiration and the evidence base to support the Campaign’s goals since its inception and continues to be crucial to our work in driving this debate forward today” [S2]. Sharkey presented the evidence base [R1, R4-R6] in support of a ban to many technologically savvy potential donors, which was instrumental in winning their support. The CSKR campaign leader noted, “These activities have generated a significant amount of funding for CSKR, which we have used to enable NGOs and members of the campaign from less wealthy, often underrepresented countries to work with their own governments and attend UN events so their voices could be heard in the debate” [S2]. CSKR reports virtually all campaign funding has been provided by a Geneva-based group of anonymous private donors (\$1.3m in financial year 2019-2020) [S1b].

Since 2013, the CSKR grew from 7 NGOs to a global movement of 172 national and international organisations (including Nobel Peace Prize-winning **Human Rights Watch (HRW)**, **Amnesty International** and the **Nobel Women’s Initiative**) in 65 countries [S1a], with new regional subgroups continuously being established – the latest being CSKR South East Asia created in 2019 [S1c].

Impact on discussions of the UN Convention on Certain Conventional Weapons (UNCCW)

In November 2013, the **UNCCW** decided to convene a meeting of experts to formally discuss LAWS. The decision was in large part a response to two reports (in 2010 and 2013) by the Special Rapporteur on Extrajudicial Executions for which Sharkey provided key evidence [S3], as well as Sharkey’s engagement campaign conducted since 2007. Formal discussions of LAWS were held at annual international meetings of experts (2014-2016) and, as the issue increased in importance, at international Group of Governmental Experts (GGE, 2017-2019) meetings. Sharkey represented the ICRAC at all of these events [S4], contributing as an invited expert in 2014 and producing a report on meaningful human control in 2018 [S5] [R1, R6]. The report demonstrated to the delegates the crucial difference between *automatic* and *directive*

control and led to the incorporation of meaningful human control into the draft guiding principles for LAWS. One meeting chair described the value of Sharkey's input as follows: *"His in-depth understanding and research-based knowledge of potential capabilities of Lethal Autonomous Weapons and Artificial Intelligence in the military domain provided a factual and very much appreciated contribution that informed the group's discussions, in particular on the subject of meaningful human control"* [S6].

In addition to his direct contributions, Sharkey has indirectly influenced debate surrounding LAWS through the contributions of the CSKR at these expert meetings, organising side events and producing briefings and statements [S1c, S4]. At the vast majority of the expert meetings, CSKR members accounted for all (or all but one) of the NGO members present [S4].

Impact on national governments and political unions

Sharkey and the CSKR have engaged directly with **national governments** (Helsinki, Brussels, Berlin, Paris, Buenos Aires, and Rio de Janeiro) and **groups of nations** (the European Parliament, African Union, Non-aligned Movement, and Nordic Group) to generate support for a ban on LAWS by the UN [S1c, S4]. Some of the funding Sharkey helped raise for the CSKR has been given to members to hold local events [S1b, S2], and the group has actively lobbied and co-hosted side events at the **UN General Assembly** [S1c, S4].

A total of 99 countries have now raised LAWS in their remarks at the UN General Assembly, with dozens more aligning themselves with statements by political unions [S1d]. The number of countries mentioning LAWS during their statements rose from only 16 states in 2013 to 37 in 2020. Previously, 42 did so in 2019, 49 in 2018, 34 in 2017, 36 in 2016, 32 in 2015, and 23 in 2014 [S1d]. During 2020, Brazil, Japan, and Germany have hosted their own international meetings on LAWS, with Austria scheduling its meeting for early 2021 [S1d].

The UN Secretary General (2018), as well as the Pope (2020), the Dalai Lama (2014) and other faith leaders (2014), have joined 30 nations from Europe, Africa, South America, Asia and the Middle East in calling for a ban on LAWS [S1c, S1d, S1e]. However, since progress towards a UN accord has reached a stalemate, with Russia, Israel, the USA, South Korea, and Australia opposing any kind of regulation, the CSKR has refocused its efforts on finding a national sponsor for an international treaty [S2], which would not need unanimous UN backing to become law. The CSKR proposed the key elements of an international treaty [S7] with an emphasis on human control [R1, R4-6] and HRW has identified this as a key interest among nation states [S8, pages 1-7].

Guided by Sharkey's research, the CSKR targeted Germany and France as potential sponsors through public events and media campaigns, resulting in **French and German foreign ministers** identifying the threat from LAWS as one of six issues requiring urgent and priority multilateral action at the 2019 UN General Assembly. They also led their counterparts from 16 other countries to co-sign a **political declaration** endorsing the objective of "developing a normative framework" that would address autonomous weapons [S8, page 1 footnote 2].

Impact on UK political debate

In 2017, Sharkey was invited to give evidence on LAWS [R1, R6] to the **House of Lords Committee on Artificial Intelligence** as part of their inquiry 'AI in the UK: Ready, Willing and Able?' The former chair of the committee commented, *"His expert evidence made our members aware of the issues caused by the lack of clear definitions for 'autonomous' when applied to weapons, and also of the fact that the United Kingdom's definition differs significantly from the global consensus"* [S9]. At the committee's request, Professor Sharkey provided a report [S10]

on the various definitions of autonomous weapons, which directly underpinned their recommendation in a wide-reaching 2018 report that the Ministry of Defence change their language on autonomous weapons and meaningful human control to align more closely with the international consensus [S9].

Sharkey also provided evidence to the **All-Party Parliamentary Group (APPG) on AI**; the co-chair of the APPG attested that his contributions, “*have greatly contributed to our discussions on this and other areas of ethical AI development and deployment as well*” [S9].

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

S1. Combined: CSKR website information (All accessed 20th Jan 2021).

- a) About CSKR. <https://www.stopkillerrobots.org/about/>
- b) CSKR annual report 2019 reports income and expenditure activities (<https://bit.ly/3tP8491>) pp.7-8 & 20-22.
- c) History and achievements. <https://www.stopkillerrobots.org/action-and-achievements/>
- d) 75th UN Assembly Meeting. <https://www.stopkillerrobots.org/2020/10/un-diplomacy/>
- e) Positions of countries on the call to ban fully autonomous weapons (July 2020). https://www.stopkillerrobots.org/wp-content/uploads/2020/05/KRC_CountryViews_7July2020.pdf

S2. Confidential testimonial statement from the 1997 Nobel Peace Prize Laureate and CSKR campaign leader (2020). Corroborates the importance of Professor Sharkey’s research in the CSKR and ICRAC efforts.

S3. Confidential testimonial and report from the UN Special Rapporteur on extrajudicial, summary or arbitrary executions (2013). <https://www.ohchr.org/en/issues/executions/pages/srexecutionsindex.aspx>. Corroborates Professor Sharkey’s contribution. (Accessed 20th Jan 2021).

S4. List of UN events where Professor Sharkey has contributed (2014 - 2019).

S5. ICRAC report authored by Professor Sharkey “Guidelines for the human control of weapons systems” (2018). (Accessed 28th Jan 2021). <https://bit.ly/3tQHvXv>

S6. Confidential statements from a Chair of the Informal Meeting of Experts of the UNCCW (2015 & 2016) and a Chair of the Group of Governmental Experts of the UNCCW (2019) confirming Professor Sharkey’s contribution to the debate.

S7. Elements for a treaty on fully autonomous weapons proposed by the CSKR (2019). (Accessed 20th Jan 2021). <https://bit.ly/3f4Heps>

S8. Human Rights Watch report “Stopping Killer Robots - Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control” (August 2020). (Accessed 15th Dec 2020). Corroborates signing of political declaration (page 1 footnote 2) and calls for an International Treaty (pages 1-7) https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and#_ftn7

S9. Confidential testimonial letter from the former Chair of the House of Lords Select Committee on Artificial Intelligence and Co-Chair of the All Party Parliamentary Group (APPG) (2020). Corroborates Professor Sharkey’s role and contribution to the committee discussions.

S10. Professor Sharkey’s report submitted, on request, to the House of Lords Select Committee on Artificial Intelligence (2018). (Accessed 15th Dec 2020). <http://bit.ly/3tiIzDr>