#### **Collaborative Learning Discussion 1**

#### **by Vasilisa Lukashevich**

#### **(Unit 2 Research Methods and Professional Practice May 2023)**

**Summary Post:**

We had an engaging discussion with fellow students on the topic of the ethics code, based on an illustrative case from the Association of Computing Machinery website: “Automated Active Response Weaponry”. In this particular case, a company violated several rules of the ACM Code of Ethics and Professional Conduct. For instance, developers ignored Principle 1.2 (*avoid harm*), Principle 1.1 (*contribute to society and human well-being*), Principle 1.6 (*respect privacy*), and Principle 1.4 (*do not discriminate*). Furthermore, the engineers of this company acted correctly by resigning and publicly expressing their awareness, despite breaking their confidentiality agreements. This action aligns with Principle 1.7 (*honour confidentiality*), Principle 2.5 (p*rovide comprehensive evaluations of potential risks*), and Principle 3.1 (*ensure that the public good is of central concern*).  
  
It is evident that there are also clear violations of the British Computer Society Code of Conduct. The company failed to uphold *public interest with regards to public health, privacy, security, and well-being*, as well as *the promotion of equal access to the benefits of IT*.

Nils Linhoff (Linhoff, 2023) emphasised that this post raises significant questions about the role of military AI startups in today's business environment.

Constantinos Kyriacou (Kyriacou, 2023) pointed out that humanity must take full responsibility for our behaviour and always ensure that our actions do not exceed predetermined limits.

In conclusion, I have not come across recent information from scholars regarding ethical issues in the use of AI in the military. The most recent studies on this sensitive topic were published two to three years ago. For example, "How viable is international arms control for military artificial intelligence? Three lessons from nuclear weapons" (Maas, 2019). However, I believe this theme remains relevant, and scholars should raise more questions about it in the context of ethics.

**Initial Post: Weapons and ethics:**

This post has opened the floor for discussion on applying the ethics code to an imaginary situation taken from the Association of Computing Machinery website.

According to legend, an international defense contractor named Q employed facial recognition algorithms (using Machine Learning) to potentially apply automated non-lethal responses, such as tear gas, pepper spray etc. to some threatening individuals in protests. Then a few engineers from the company resigned, expressing concern that the non-lethal weapon had inadequate level of protection, for instance, against a potential replacement tear gas with a lethal poison. After that Q sued them for violating these workers in terms of confidentiality employment agreement.

It turns out that Q failed many rules of the ACM Code of Ethics and Professional Conduct. According to Principle 1.2, the designed systems must avoid harm, and according to Principle 1.1, they must contribute to society and to human well-being, because all people are stakeholders in computing. Also, the systems ignoring Principle 1.6 (respect privacy) and Principle 1.4 (not to discriminate). Especially when we know about racial discrimination in face recognition technology (Najibi, 2020).

Moreover, the engineers were right by expressing their awareness publicly, breaking their confidentiality agreements, based on Principle 1.7 (honor confidentiality), Principle 2.5 (give comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks), and Principle 3.1 (ensure that the public good is the central concern during all professional computing work).

There are, of course, clear violations of the British Computer Society Code of Conduct too. The Q’s systems failed public interest, regarding public health, privacy, security and well-being, as well as promotion of equal access to the benefits of IT. Q violates Professional Competence and Integrity which is said to avoid injuring others, their property, reputation etc. And it breaches the BCS principle of Duty to the Profession, which includes accepting personal duty to uphold the reputation of the profession and not take any action which could bring the profession into disrepute.

On the other hand, what can we do about the fact that nowadays business is booming for military AI startups? (Heikkilä, 2020) For example, Pentagon requesting $1.8 billion for Artificial Intelligence (Harper, 2023). Where do the ethics principles end?

**Peer responses from me 1:**

It is very interesting to observe, in retrospect, how scholars discussed the topic of implantation and ethics. For example, about 20 years ago, Hansson (2005) wrote about “a few promising applications for neural interface implants”, such as brain implants for bladder control. However, most of these ideas are still under development.

One reason is the complexity of the brain, which is not yet fully understood, as evidenced by the recent advancements of Neuralink. Another contributing factor is the lack of a unified approach to the ethics of implantation, which overlaps with several subdisciplines of bioethics. For instance, in the same 2005, researchers expressed concerns about the development of an "artificial hippocampus" and other technologies such as mood control and memory enhancement through ICT implants, which were considered as “a threat to human dignity and democratic society” (Erden, 2005).

Regarding the case of the fictitious "Corazón" implant mentioned by my respected fellow student (Kyriacou, 2023), I believe that the first sentence in the device description from the ACM Ethics website, which states that it can be wirelessly monitored and controlled by a phone app, poses a potential threat. Can we be certain that a patient's well-being is secure if such a critical organ is controlled by an external mobile application? Have we exhaustively modeled potential scenarios to ensure that neither Principle 1.2 (Avoiding harm) nor Principle 1.6 (Respecting privacy) are breached?

In this case I can agree that for safety there should only be a one-way connection between the device, with the implant sending information to the app without the opportunity for vice versa.

**Peer responses from me 2:**

Thank you for the post – the Dark UI/UX Patterns topic seems very relevant, since each digital consumer has faced this problem at least once in their life. For example, some “leaveshaming” banners on the app/website like “Leaving so soon, \_username\_?” or hidden costs. There are a lot of examples on Twitter tagged as @darkpatterns to research. In fact sometimes user manipulations go so subtle that they cannot explain with one screenshot what is wrong.

Moreover, like Waldman (2020) noticed, “users routinely trade privacy for convenience”. Many consumers feel free with sharing their geolocation, plus the overall trend toward personalization, and that opens the way to weaponize dark patterns against users easily. Sell companies use AI methods to find users' weak spots, determining which interface, text, juxtapositions, and color maximize revenues (Luguri & Strahilevitz, 2021).

From this point of view, we might try to use AI to defend the users as well. According to the quick research, there is some work in progress. For instance, a group of scholars tried automated detection of five dark patterns (Nagging, Obstruction, Sneaking, Interface Interference, Forced action) by Machine Learning methods. (Soe et al., 2022). The best accuracy demonstrated by Nagging detection – 72%, the rest allowed a lot of room for improvement.

**References**

Lukashevich, V. (2023) Initial Post: Weapons and ethics. The University of Essex Online. Available from: https://www.my-course.co.uk/mod/forum/discuss.php?d=156259 [Accessed 20 June 2023]

Case: Automated Active Response Weaponry. Available from: https://ethics.acm.org/code-of-ethics/using-the-code/case-automated-active-response-weaponry/ [Accessed 5 May 2023]

Erden, Y. (2005) ICT Implants, nanotechnology, and some reasons for caution. BioCentre. Available from: https://www.bioethics.ac.uk/news/ICT-Implants-nanotechnology-and-some-reasons-for-caution.php [Accessed 11 May 2023]

Najibi, A. (2020) Racial Discrimination in Face Recognition Technology. Science in the News. Available from: https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/ [Accessed 5 May 2023]

Association of Computing Machinery Code of Ethics and Professional Conduct (2018) Available from: https://www.acm.org/code-of-ethics [Accessed 5 May 2023]

British Computer Society Code of Conduct (2022) Available from: https://www.bcs.org/media/2211/bcs-code-of-conduct.pdf [Accessed 5 May 2023]

Hansson, S.O., 2005. Implant ethics. *Journal of Medical Ethics, 31*(9), pp.519-525.

Harper, J. (2023) Pentagon requesting more than $3B for AI, JADC2. *Defensescoop.com*. Available from: https://defensescoop.com/2023/03/13/pentagon-requesting-more-than-3b-for-ai-jadc2/ [Accessed 5 May 2023]

Heikkilä, M. (2022) Why business is booming for military AI startups. *MIT Technology Review*. Available from: https://www.technologyreview.com/2022/07/07/1055526/why-business-is-booming-for-military-ai-startups/ [Accessed 5 May 2023]

https://neuralink.com/

Kyriacou, C. (2023) Initial Post: Medical Implants. [Collaborative Learning Discussion 1]. Research Methods and Professional Practice May 2023. The University of Essex Online. Available from: https://www.my-course.co.uk/mod/forum/discuss.php?d=157504 [Accessed 11 May 2023]

Luguri, J. & Strahilevitz, L.J. (2021) Shining a light on dark patterns. Journal of Legal Analysis, 13(1), pp.43-109.)

Maas, M.M. (2019) How viable is international arms control for military artificial intelligence? Three lessons from nuclear weapons. *Contemporary Security Policy, 40*(3), pp.285-311.

Soe, T.H., Santos, C.T. & Slavkovik, M. (2022) Automated detection of dark patterns in cookie banners: how to do it poorly and why it is hard to do it any other way. arXiv preprint arXiv:2204.11836.)

Waldman, A.E. (2020) Cognitive biases, dark patterns, and the ‘privacy paradox’. Current opinion in psychology, 31, pp.105-109.)