**Name:** Eamonn Herlihy **Student ID:** 14155605

**Module:** IN5103 - Risk, Ethics, Governance and AI

**Assignment:** Week 2: The Politics of Innovation and Design (AI)

**Essay**

Langdon Winner produced the landmark paper ‘Do Artifacts have Politics?’ in 1980. It is amazing that over 40 years later, the concepts he describes remain relevant today. I would argue that this paper can be applied and is of great consequence to the field of Artificial Intelligence.

Throughout this seminal article, Winner (1980) makes a compelling argument that technological artifacts indeed have political properties and therefore assessment via traditional metrics of productivity and efficiency is simply not sufficient. Winner (1980) compares technological innovations to legislative acts (that establish a framework for public order) and suggests that critical analysis of societal impacts and embodiments of power is necessary to derive a complete picture. By coupling the doctrine of social determinism and the notion that certain technologies should be viewed as political phenomena in their own right, Winner (1980) identifies two schemas by which technologies contain political properties i.e., ‘Technical arrangements as forms of order’ (by decision) and ‘Inherently political technologies’ (by necessity).

The focus will be on the former schema (by decision) as I believe it to be more prevalent in the field of AI. However, note that the latter (by necessity) also exists. For example, China’s mass surveillance/facial recognition system is an example of an inherently political technology that seems to serve as a tool to help maintain the autocratic state (Leibold, 2020). Winner (1980) suggests that the key question to an inherently political technology is, whether to adopt the technology at all? In contrast, the key question to technologies defined by ‘Technical arrangements as forms of order’ is, what form the technology should take?

‘Technical arrangements as forms of order’ refers to technologies that have flexibility in the way they are constructed, and therefore the subsequent decisions may have some political tendencies. Winner (1980) defines several examples of this mechanism in action and distinguishes between political biases that are intentional and those that are not. One such example Winner (1980) defines is the height of bridges over Long Island. Here, there was flexibility in the construction of this technology and the subsequent decision to build these bridges unusually low has political consequences i.e., the low bridges discourage the presence of buses. Moreover, Winner suggests that this was done to serve an intentional racial bias by ‘master builder’ Robert Moses. Although Joerges (1999) and Woolgar and Cooper (1999) are critical of Winner’s recounting of this story due to the lack of evidence surrounding Moses's definitive motivation, I believe this proves the point that intentional or not, technological decisions can have substantial consequences. Although AI was a relatively new field and experiencing a ‘winter’ (Hendler, 2008) at the time of publication, many of today’s AI systems have the potential to fall into this category.

Amazon's 2014 AI recruiting tool is one such example of unintentional technological bias. Regrettably, this AI-based algorithm had taught itself that male candidates were preferable i.e., resumes that included the word “women’s” were penalised (Dastin, 2018). This is a clear example of technological bias with real-life consequences, and unfortunately, the literature suggests that such bias is quite prevalent (Leavy, 2018). Unsurprising, when talking about AI and its complex algorithms (black boxes), the solution is not as easy as building higher bridges. In fact, the flexibility and decision components of Winner’s theory are much more subtle. However, examples such as this one expose the underdeveloped field of explainable AI (Deeks, 2019) and the requirement for the international community to come to an agreement on how to regulate it (Lee, 2022).

Applying Winter’s theory to the field of AI exposes the requirement for change. My hope is that the recently proposed ‘EU Artificial Intelligence Act’ (Kop, 2021) will act as a catalyst to kickstart a new age responsible AI with strict, fully enforced regulation. Such a shift could allow us to continue advancing society with remarkable AI innovations while limiting any of the unjust, inequitable societal consequences this technology has the potential to produce.

**Bibliography**

Dastin, J., 2018. Amazon scraps secret AI recruiting tool that showed bias against women. In Ethics of Data and Analytics (pp. 296-299). Auerbach Publications.

Deeks, A., 2019. The judicial demand for explainable artificial intelligence. Columbia Law Review, 119(7), pp.1829-1850.

Hendler, J., 2008. Avoiding another AI winter. *IEEE Intelligent Systems*, *23*(02), pp.2-4.

Joerges, B., 1999. Do politics have artefacts?. Social studies of science, 29(3), pp.411-431.

Kop, M., 2021, September. EU Artificial Intelligence Act: The European Approach to AI. Stanford-Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue.

Leavy, S., 2018, May. Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning. In Proceedings of the 1st international workshop on gender equality in software engineering (pp. 14-16).

Lee, J., 2022. Legal Vacuum to Regulate AI. In Artificial Intelligence and International Law (pp. 219-257). Springer, Singapore.

Leibold, J., 2020. Surveillance in China’s Xinjiang region: Ethnic sorting, coercion, and inducement. Journal of Contemporary China, 29(121), pp.46-60.

Winner, L., 2017. Do artifacts have politics?. In Computer Ethics (pp. 177-192). Routledge.

Woolgar, S. and Cooper, G., 1999. Do artefacts have ambivalence: Moses' bridges, winner's bridges and other urban legends in S&TS. Social studies of science, 29(3), pp.433-449.