The gap between AI practitioners and ethics is widening – it doesn't need to be this way

The application of AI technologies to social issues and the need for new regulatory frameworks is a major global issue. Drawing on a recent survey of practitioner attitudes towards regulation, **Marie Oldfield** *discusses the challenges of implementing ethical standards at the outset of model designs and how a new Institute for Science and Technology training and accreditation scheme may help standardise and address these issues.*

Practitioners face overwhelm and indecision when incorporating standards and ethics in the development of machine learning models or 'AI'. Regulation, working groups, guidance and training come from a multitude of sources and it is hard to know which one to rely on. A 2021 House of Lords Enquiry on analytical modelling in the UK raised wide ranging issues for practitioners around lack of respect for modelling professionals, policy not being evidence based, lack of resource and lack of leadership. Similarly, in 2019 [Holstein et al.](#) found and made recommendations to address issues faced by practitioners, with a particular emphasis on gaps in legislation and regulation.

To see if any progress had been made in recent years in addressing these issues and to explore how recent developments in machine learning and AI were reshaping perceptions, I approached practitioners in the field, to effectively rerun Holstein et al's 2019 survey and conducted a follow-on survey exploring respondents' perceptions of AI were developing.

We found that not only has the situation worsened since 2019, but there is a propensity to avoid addressing ethical issues within the model build by utilising PR to improve the image of their products in the public view. What fairness and bias means to the public seems, in some cases, to be more important than the regulation itself. If the issues that come with a hastily implemented model can be glossed over with PR, there was an attitude that saving money on due process and then paying for PR if there is an issue

was acceptable.

> What fairness and bias means to the public seems, in some cases, to be more important than the regulation itself.

For example, the way in which Amazon positioned their virtual assistant Alexa in an anthropomorphic manner to encourage a lower risk perception, whilst at the same time the device was at the centre of a number of scandals due to it 'hoovering up' data and collecting shadow data and then storing it without informing the user. This leads to a discrepancy in the sector whereby regulation and research focus on ethical issues of 'de-biasing' models, but practitioners are more focused on avoiding bad perceptions, or 'PR disasters'.
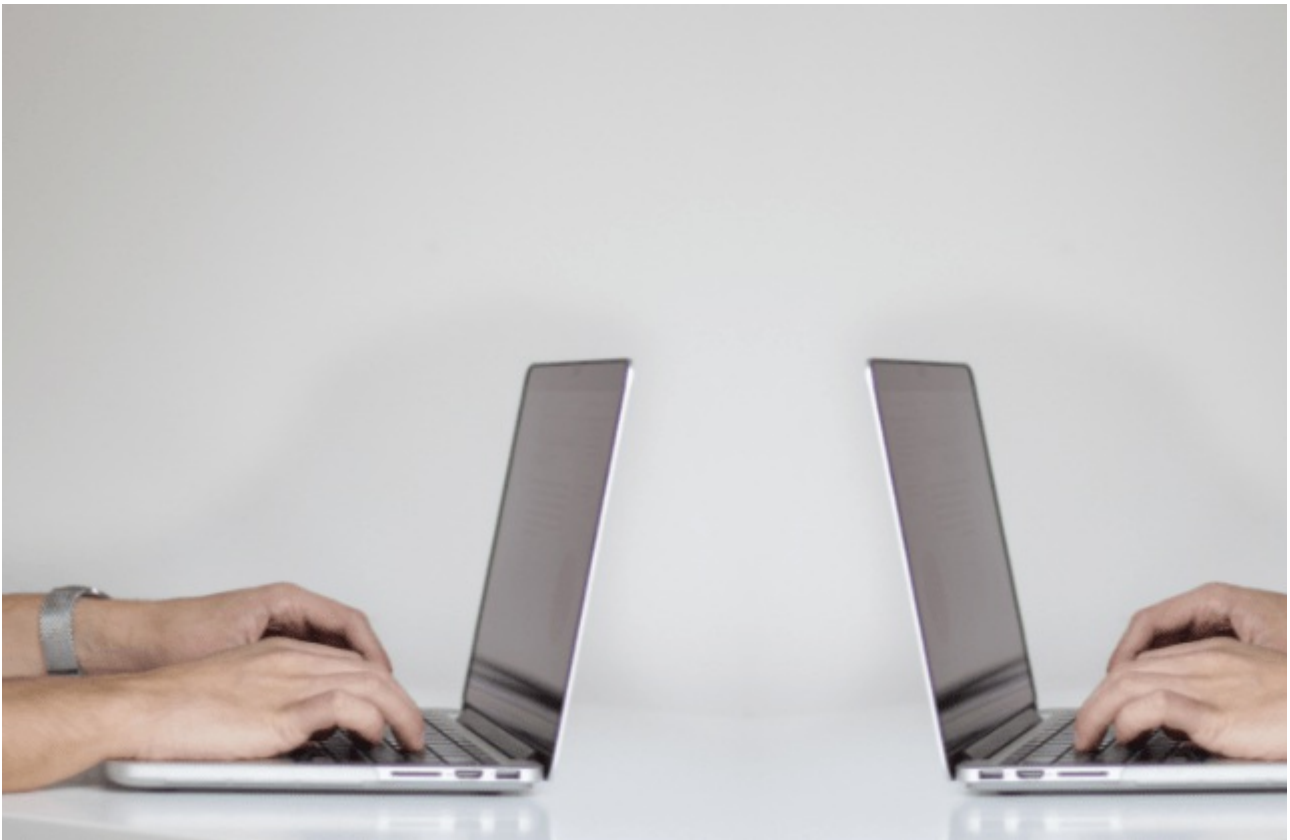
For its promoters, AI is an attempt to create 'human consciousness' and decision making via technology. Not only is this not currently possible, not being able to code the masses of information or context that a human would use to make a decision means we will have poorly built models. If you want to code an autonomous vehicle you have to examine

something like the trolley problem, i.e if you have a child on your left and 2 adults on your right which way do you swerve to avoid a careering car? We not only need a lot more context to be coded (what are the weather conditions, are there any other obstacles, if we just stop what happens etc etc), but we need to prioritise who dies. This might not be a palatable discussion, but it is the type of decision we make every day as a humans without really realising it. We take a lot of things for granted in the world and focus on those things more out of the ordinary. However, the part we take for granted also needs to be coded into models, not just the event we need to address. This makes it difficult for practitioners to code complex models and unfortunately the speed with which technology has advanced has outstripped any current guidance and we are playing catch up.

> without funding ethical gateways run the risk of stalling innovation

Other practitioner problems highlighted by the survey centred around soft skills and human factors, which had been raised in the [2012 Macpherson Report](#), but have not subsequently been rectified in practice. A situation that compounds emerging challenges for practitioners to apply ethical standards to their work. Perceptions of AI can also skew the attitudes of developers and the public. Those who work in AI seemed to be more optimistic about the technology in general. Those who do not work in AI may tend to be more pessimistic in general. This could reveal an optimism bias in working in AI and a mistrust bias amongst those who do not work in the field.

This presents a particular challenge. Innovation is driven by a race for priority in developing new products or securing funding to do so. Ethics are often perceived to be a barrier to this process. Putting in the effort to understand the effects of a model and involving users at the correct time to understand the impact is not something most small companies want to pay to do, as without regulation or legislation, they can launch new technology without considering these impacts. The Institute of Science and Technology (IST) goes some way to filling this gap with cutting edge [research and professional accreditation](#) for AI practitioners. However, without funding ethical gateways run the risk of stalling innovation. To address this gap I have created a software system to provide the support to practitioners to develop robust and ethical software and with the IST I have created [academic and professional accreditations](#). This framework aims to support practitioners and provide a methodological centre of excellence so that we can move forward with more ethical modelling that is safe for society.

This post draws on the author's article, [Technical challenges and perception: does AI have a PR issue?](), published in AI and Ethics.

The content generated on this blog is for information purposes only. This Article gives the views and opinions of the authors and does not reflect the views and opinions of the Impact of Social Science blog (the blog), nor of the London School of Economics and Political Science. Please review our [comments policy]() if you have any concerns on posting a comment below.

Image Credit: [NordWood Themes]() via Unsplash.