Autonomous nudges and Ai Choice Architects – Where does responsibility lie in computer mediated decision making?

AI and algorithms shape many aspects of our everyday life, from the familiar algorithms structuring our social media feeds, to those subtly transforming more complex fields, such as policymaking and commerce. **Stuart Mills** *argues that as these choice architects become increasingly autonomous and automatic, and produce nudges that are difficult if not impossible to explain, there is a need to reassess the ethical limits underpinning how and who is nudged.*

Nudges have been around for over a decade, and some might say *nudging* has been around for as long as humans have been trying to influence one another. These subtle changes to the environments in which we make decisions – so-called *choice architecture* – have proven popular amongst policymakers and the private-sector as unobtrusive and liberty-preserving means of influencing behaviour. Be it behavioural scientists changing default options on forms, or Frederick the Great feigning a great love of potatoes to encourage his subjects to eat this wonder vegetable, we have all been nudged at some point.

In my article with Henrik Skaug Sætra, recently published in *AI and Society*, we explore a new phenomenon – artificial intelligence (AI) nudging humans. This phenomenon is more common than one might immediately believe.

But what exactly is AI? One recent review for example identified 55 different applications of the term. We define AI – at least in its *current* form – as consisting of two features. *Firstly*, AI is an algorithm or set of instructions designed to exhibit actions which it predicts will accomplish whatever function it has been programmed to achieve. *Secondly*, AI possesses its own *motive power*, meaning it can act *independently* of humans. With this definition, we ask a simple question: can an AI nudge a human? Our answer is firmly: *yes*. And, it is probably happening to your right now.

> The Facebook News Feed is a classic example of nudging, just as other 'feeds' such as Netflix's homepage are, while the algorithms which power these feeds are choice architects.

Take for example the Facebook News Feed algorithm, it will select approximately 300 posts to appear on a person's Facebook feed every day, out of a possible pool of around 1,500 posts. The algorithm selects these posts based on what it predicts will maximise a user's *click-through-rate* – the number of posts a person clicks on, given the number of posts a person is shown. A higher click-through-rate implies the user is enjoying what Facebook is showing them, and leads to more eyeballs of advertisements (which Facebook wants). Furthermore, the algorithm curates a user's feed *automatically* – no human could possibly design a Facebook feed for every Facebook user. The algorithm therefore *must* have its own motive power, and is therefore what we would call *autonomous*.

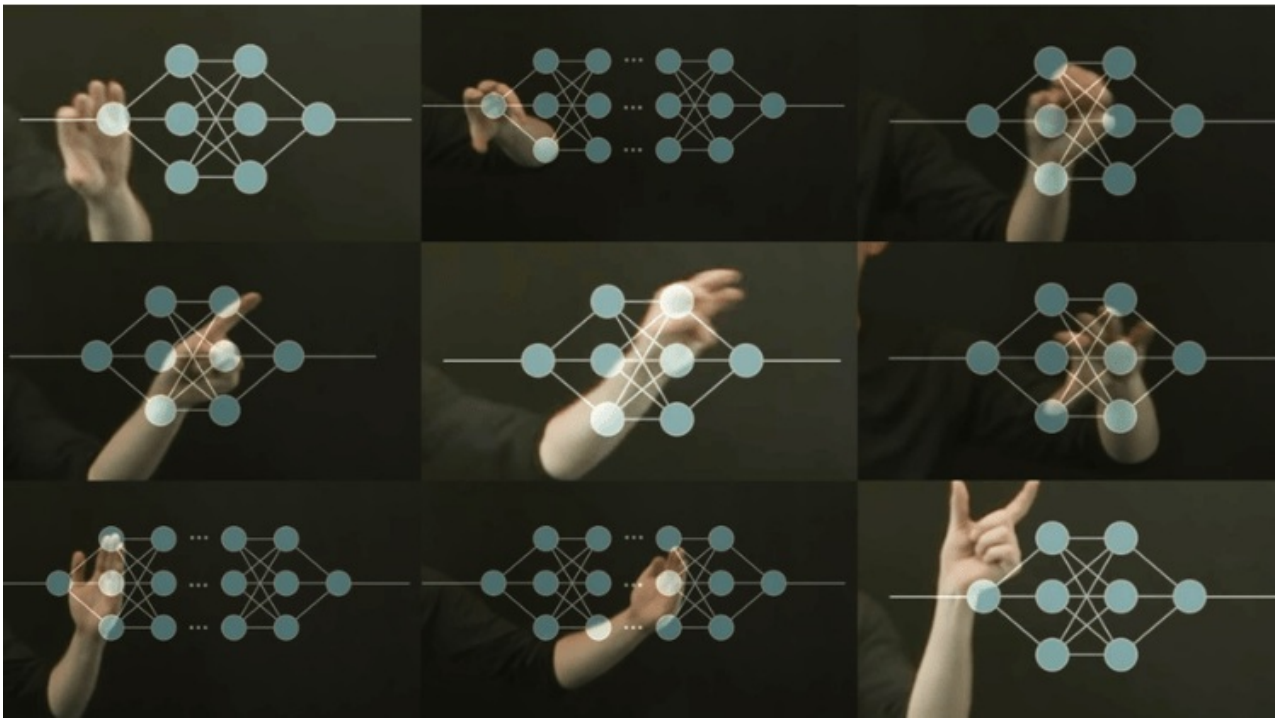
By our definition, the Facebook algorithm is an AI. But it is also nudging. Nudges should not mandate or ban any options, nor should they impose significant economic incentives. The Facebook algorithm follows these principles. The 1,500 posts which the algorithm selects from are all available for the user to view; the 300 selected are merely *easier* to view. Furthermore, an algorithm is used because of human cognitive limitations – it would be harder for us to wade through 1,500 posts everyday to find what we're looking for. Not only does the algorithm make the 300 posts *easier* to view, but it also makes the platform *as a whole* easier to use. The Facebook News Feed is a classic example of nudging, just as other 'feeds' such as Netflix's homepage are, while the algorithms which power these feeds are choice architects.

Specifically, we call AI systems which nudge *autonomous choice architects*. Many more examples of autonomous choice architects exist, beyond the comparatively well known examples of social media. For instance, AI is increasingly being used to evaluate vast corpuses of behavioural science literature to design effective policy quickly. AI is also driving many ecommerce recommendations tabs. And in areas such as finance, so-called *robo-advice* is an example of AI nudging through information filtering. I would be so bold as to predict that it is more likely that you were most recently nudged by an *algorithm*, rather than a human.

Date originally posted: 2022-08-02

Permalink: https://blogs.lse.ac.uk/impactofsocialsciences/2022/08/02/autonomous-nudges-and-ai-choice-architects-where-does-responsibility-lie-in-computer-mediated-decision-making/

Blog homepage: https://blogs.lse.ac.uk/impactofsocialsciences/



This raises several important questions, chief amongst them being *who is responsible for autonomous choice architects*? Responsibility is, *ethically speaking*, a tricky idea. If you are driving your car, and you choose to run a red light, most people would assume you are responsible for any traffic collision which then follows. But what if you run the red light, despite desperately stomping on your brake peddle? If your brakes fail, and *that* causes a collision, should you the driver be held responsible? The best answer is probably: *it depends*. Did you neglect to get your car serviced, or did the manufacturer overlook something important when they made your car? Perhaps the worst possible answer for this philosophical quagmire is: *it was just dumb bad luck*.

> the responsibility gap for autonomous choice architects is an illusion; the product of a 'veil of complexity'

It is important to think about who is responsible for autonomous choice architects because AI can very easily create these sort of *responsibility gaps*. It is famously – or perhaps *infamously* – difficult to explain *why* modern AI systems such as deep learning systems do what they do. Trained on cascades of data often taken from many people (in the case of Facebook, literal *billions* of people), and programmed to respond to dynamic and often novel situations, it is perilously difficult for *anyone* to know why, say, Facebook post A was shown over Facebook post B. When no one is in control – just as the driver is not in control of their broken brake – it is extremely difficult to attribute responsibility.

Yet, the responsibility gap for autonomous choice architects is an illusion; the product of a 'veil of complexity.' While those who design autonomous choice architects may not know *why* post A was shown over post B; these people still *control* the AI making the selection, and still *choose* to use an AI in the first place. The apparent complexity of autonomous choice architects should not allow those who implement them to shirk responsibility, and thankfully, the public seems to agree, as examples such as the infamous Facebook mood experiment demonstrate.

But all of this could perhaps be ignored in the case of autonomous choice architects. Autonomous choice architects are not really a unique example of AI, while ethical questions about responsibility and AI are not unique to autonomous choice architects. But autonomous choice architects *are* ethically challenging. Nudges are meant to allow individuals to 'go their own way;' they are *suggestions*, rather than commands. When dealing with a human-implemented nudge, one can quite easily choose to, say, select a different option on a form, or snub their nose at a potato. But autonomous choice architects, possessing their own motive power, are different. These systems constantly follow and learn from individuals in the form of data, and *constantly* nudge individuals. They are a kind of choice architect we cannot easily escape. As such, questions such as *who is responsible*? and *why am I being nudged this way*? become more important, as technology challenges the limits of nudging itself.

Date originally posted: 2022-08-02

Permalink: https://blogs.lse.ac.uk/impactofsocialsciences/2022/08/02/autonomous-nudges-and-ai-choice-architects-where-does-responsibility-lie-in-computer-mediated-decision-making/

Blog homepage: https://blogs.lse.ac.uk/impactofsocialsciences/

Image Credit: Alexa Steinbrück / Better Images of AI / Explainable AI / CC-BY 4.0

Date originally posted: 2022-08-02

Permalink: https://blogs.lse.ac.uk/impactofsocialsciences/2022/08/02/autonomous-nudges-and-ai-choice-architects-where-does-responsibility-lie-in-computer-mediated-decision-making/

Blog homepage: https://blogs.lse.ac.uk/impactofsocialsciences/