The Ethics of AI Nudges

How AI Influences Decision-Making

And why decision-makers should care about it.

Artificial intelligence (AI) systems, through mechanisms like nudges and choice architecture, actively, yet often subtly, shape human decision-making in everyday life and professional settings.

AI systems can prioritise profit or efficiency at the expense of human agency, fairness, and well-being, highlighting the need to balance AI's capabilities with ethical considerations.

The EU AI Act is a landmark framework designed to curb manipulative AI practices, emphasising the protection of human autonomy and accountability in decision-making.

Picture Tom, a gig economy driver working for a major ride-sharing platform. After a gruelling 10-hour shift, he is ready to head home when he receives a notification, "You're just $15 away from earning your daily target of $300!" What Tom does not realise is that this seemingly helpful reminder is part of an AI-driven engagement system designed to maximise platform profits by keeping drivers on the road longer. The AI has analysed Tom's behavioural patterns, identified his personal financial goals, and calculated the precise moment when such a message would most likely trigger his loss-aversion bias–the psychological tendency to avoid 'missing out' on potential gains. Hence despite his fatigue and earlier decision to end his shift, Tom continues driving for another hour. This decision, subtly engineered by AI, prioritises the platform's need for driver availability over Tom's well-being and safety.

Here is another example. Consider your typical workday. Whether it is the embedded email assistant that suggests response times, the calendar that optimises meeting schedules, or the project management tool that prioritises tasks, each one of them represents an AI system that is quietly tweaking your behaviour. These digital nudges, while seemingly benign, accumulate to shape professional judgments, team dynamics, and ultimately, organisational outcomes.

As you can see, in the modern workplace and beyond, AI does not just automate tasks; it is a subtle but powerful form of influence that actively shapes our choices. While most executives understand AI's role in data analysis or process optimisation, few recognise how AI systems systematically guide decision-making through sophisticated behavioural interventions.

In this article, I examine the sophisticated mechanisms through which AI influences human decision-making, revealing patterns that often escape conscious awareness. By understanding how AI systems leverage behavioural psychology—from anchoring effects to choice architecture—leaders can better evaluate where algorithmic guidance ends, and human judgement begins.

Drawing on recent research in behavioural science and real-world cases, we will explore why humans are surprisingly susceptible to AI influence, even in high-stakes professional contexts. More importantly, we will examine how organisations can harness this understanding to design AI systems that enhance, rather than undermine, human agency in decision-making.

PERVASIVE AI PRACTICES THAT POSE ETHICAL CHALLENGES

Most commercially-developed AI systems are often programmed to modify user behaviour to either maximise conversion or increase engagement to drive behaviour that maximises profit. In some situations, the humans interacting with these systems might be aware of being influenced by these tools, but not about how the influence works or to what extent the influence will impact their decision-making outcomes. In most cases, users are not aware of the impact of these outcomes on their lives and well-being.

Unintended AI influence extends beyond the corporate domain; it can also be found in the use of AI-enabled decision support systems, such as judiciary processes. These AI systems process defendants' data to generate risk scores, presenting them as objective inputs for judicial consideration. However, this integration of AI into legal decision-making can create automation bias—an unconscious tendency to defer to automated assessments even when they conflict with professional judgment.

A 2013 criminal case in the US crystallised these concerns. When sentencing a defendant for theft, a judge doubled the recommended one-year sentence solely based on an AI risk assessment score.[1] This dramatic departure from both prosecutor and defence recommendations demonstrates how AI inputs can override experienced legal judgment, often without the decision-maker recognising the extent of this influence. Investigations conducted by a non-governmental organisation (NGO) named Pro-Publica revealed that AI scores in the US were biased against African Americans, leading to systematic biases in decision-making.[2]

In such situations of human-AI collaborations, an AI system's output creates an anchor point that can unconsciously skew perceptions, leading to decisions that may deviate from expert judgement. While judges might view the AI's input as just one factor among many, these algorithmic assessments can fundamentally reshape their evaluation of evidence and circumstances. This situation exemplifies a critical challenge in AI-assisted decision-making:

maintaining meaningful human agency when algorithmic influences operate below the threshold of conscious awareness.

AI assistants are increasingly used in medical diagnosis, particularly when analysing medical images like chest X-rays. However, research has shown that the AI models powering these systems demonstrate lower accuracy when evaluating women and people of colour.[3] Studies revealed that these systems were relying on demographic shortcuts to compensate for performance discrepancies, resulting in inaccurate diagnoses for these patient populations.

Credit scoring firms in Vietnam have also turned to digital footprint, such as data from smartphone use and social media activities, to evaluate the creditworthiness of individual applicants. While this approach may enable financial institutions to reach out to unbanked individuals, it could also result in the unintended consequence of reinforcing financial exclusion for those with limited or no access to digital technology.[4] This is because the creditworthiness assessment may be biased or it could discriminate against specific groups, such as rural applicants who leave less of a digital footprint than those residing in bigger towns and cities.

These scenarios underscore the pervasive ethical challenges posed by AI applications and their potential to negatively impact our businesses, lives, and society at large. The AI tools of today are different from other technologies as they are fully capable of acting

as autonomous decision-making agents, albeit with unforeseen consequences. Unethical outcomes such as infringement of privacy, leaking of proprietary and private information, biased outcomes, and propagation of misinformation are just some of these possible issues.

Despite such ethical concerns, recent studies and reports highlight the growing adoption of AI tools by businesses to enhance productivity and decision-making. A Deloitte Insights report released in May 2024 highlighted that 67 percent of employees in Singapore are utilising generative AI tools, surpassing the Asia Pacific average of 62 percent.[5] It also highlighted that employees in the country believe that 64 percent of their tasks will be automated or augmented by AI within the next five years, underscoring the technology's significant impact on productivity and decision-making.

Another survey in 2024 revealed that 68 percent of business leaders in the Asia Pacific region, including ASEAN countries, agree that emerging technologies like AI are crucial for driving innovation, creativity, and productivity.[6] AI tops the list of technologies deemed most important for businesses over the next five years, followed by cloud computing and robotic process automation.

Even our daily interactions with technology are increasingly shaped by subtle AI influences that we may not fully recognise. For example, social media platforms employ sophisticated recommendation algorithms that shape our information consumption and social interactions. These systems analyse our behavioural patterns, emotional responses, and social connections to present content that maximises engagement, often without consumers understanding the extent of this curation.

E-commerce platforms use AI-driven pricing and presentation strategies that influence purchasing decisions. For instance, dynamic pricing algorithms might adjust prices based on user behaviour patterns, while product recommendation systems employ psychological targeting to present items in ways that maximise purchase likelihood.

Search engines represent another pervasive form of AI influence in daily life. Studies show that the ordering and presentation of search results can significantly impact decision-making, from consumer choices to political opinions. Researchers confirmed that search result rankings can influence voting preferences by more than 20 percent, often without users being aware of this influence.[7]

Digital wellness and productivity apps increasingly use AI-driven nudging techniques to influence behaviour. While often well-intentioned, these systems can sometimes override user autonomy through persistent notifications and psychologically-optimised messaging and notifications.

SHAPING AI INFLUENCE THROUGH CHOICE ARCHITECTURES

Behavioural scientists in recent decades have found that decision-making deviates significantly from the rational ideal depicted by classical economic theory where humans are expected to consistently evaluate all available options by carefully weighing costs and benefits to maximise value. In reality, humans are

While most executives understand AI's role in data analysis or process optimisation, few recognise how AI systems systematically guide decision-making through sophisticated behavioural interventions.

cognitive 'misers'. To cope with the overwhelming complexity of daily decisions, we choose to 'satisfise' by employing mental shortcuts and simplified strategies that, while not mathematically optimal, allow us to navigate and make 'good enough' choices effectively.

This understanding has given rise to the concept of 'bounded rationality' in economics, which acknowledges that our capacity for rational analysis has clear constraints. This tendency is compounded by inherent limitations in attention spans and self-regulation, particularly when we are confronted with the allure of immediate rewards.

Human decision-making is also profoundly shaped by the context in which choices are presented, along with the accessibility of different options. These seemingly minor factors, such as how alternatives are arranged, the format in which information is displayed, and which options are set as defaults, can significantly influence final choices. The environmental context surrounding a decision acts as a powerful force that can either facilitate or impede certain choices.

Considering both bounded rationality and the environmental context, designers of AI systems can then deliberately shape human choices by thoughtfully redesigning the environment in which decisions are presented. This strategic modification of the social, physical, and cognitive landscape surrounding decision points is known as 'choice architecture intervention' or 'nudges'.[8] Just as an architect designs physical spaces to

Studies show that the ordering and presentation of search results can significantly impact decision-making, from consumer choices to political opinions.

guide movement and interaction, choice architects craft decision environments to guide people toward certain behaviours while preserving their freedom to choose.

While behavioural scientists have extensively studied how choice architecture affects behavioural change, context dependency makes it difficult to draw universal conclusions about the impact of choice architecture, as its effectiveness often depends on subtle interactions among the decision environment, the nature of the choice itself, and the characteristics of the decision-makers involved.

Sophisticated AI applications can influence decision-making through strategic designing of the decision environment. Below are three key strategies which are most commonly used by AI systems using the choice architecture approach.

Designing the information presentation

Given that decision-makers primarily rely on immediately available information rather than conducting exhaustive analyses, AI systems can personalise and present information in ways that influence decision-making. This can be done by providing social reference points (comparing individual behaviour to that of peers), making hidden information visible through feedback mechanisms, and translating complex data into comprehensible formats. For instance, streaming platforms do not just recommend what to watch next—they carefully curate thumbnails, descriptions, and timing of recommendations based on your psychological profile and viewing patterns. When Netflix shows different artworks for the same movie to different viewers, it is leveraging AI to present information in the most persuasive way possible for each individual.

Altering the structural design of choices

Given that the arrangement and format of options significantly impact choices, AI systems use techniques such as setting strategic defaults (pre-selected options), adjusting the effort required to choose different options, and carefully curating the range and composition of choices. For example, e-commerce platforms modify the sequence and presentation of products based on an AI analysis of your browsing patterns and psychological profile.
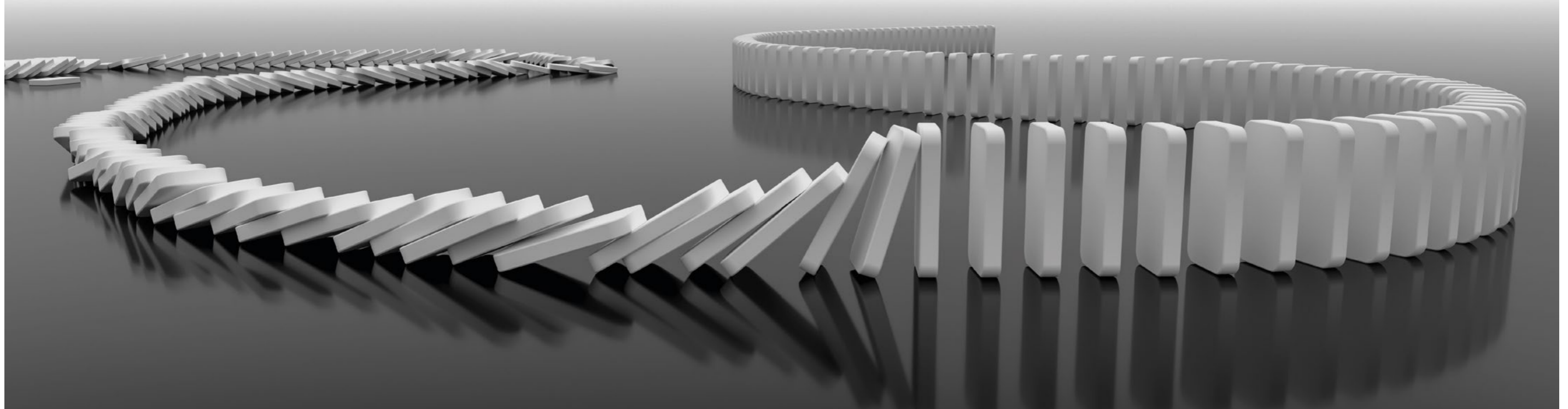
Providing strategic decision assistance

Even when people intend to make certain choices, limited attention and self-control can prevent them from following through on their original choices. AI systems influence this intention-behaviour gap through commitment devices that lock in future decisions, timely reminders that increase behavioural salience, and the removal of psychological barriers such as a call to take immediate action to prevent procrastination.

Consider how fitness apps use personalised goal-setting and reminders. These seemingly helpful nudges are often powered by AI systems that analyse patterns to determine the most effective timing and framing of interventions. Banking apps similarly use AI to suggest spending limits or savings goals, presenting them as helpful tools while potentially influencing financial behaviour.

> The EU AI Act highlights how sophisticated AI techniques can exploit human psychological vulnerabilities in ways that might be visible, but whose manipulative mechanisms remain hidden from human awareness.

COUNTERING UNETHICAL AI INFLUENCE: THE SIGNIFICANCE OF THE EU AI ACT

As AI systems increasingly shape human decision-making, there is a need to carefully distinguish between the ethical and unethical forms of influence of AI systems. Countries and regulatory bodies are trying hard to keep pace with the state of progress of AI capabilities to address the complex nuances of its influence, which can bring about widespread impact on the well-being of humans and society at large.

The European Union's AI Act, which was passed in May 2024, establishes clear parameters for protecting the autonomy of human decision-making from AI manipulation.[9] It acknowledges that when AI systems deploy deceptive techniques at scale, they can significantly alter societal behaviour patterns and reshape cultural norms. Article 5 of the Act in particular identifies and prohibits AI systems that are designed to exploit psychological vulnerabilities, creating a distinct boundary between ethical AI applications and those that deploy manipulative techniques to influence human choices. By prohibiting such systems, the provision serves as a safeguard against the potential for AI to undermine collective decision-making autonomy. As a result, the EU AI Act also ensures that AI systems respect the fundamental human rights of integrity and autonomy in all instances.

The definition of prohibited AI influences in the EU AI Act is based on three key dimensions: the nature of influence, its intended effects, and its underlying mechanisms. The nature of influence exists on a spectrum ranging from the subliminal to the supraliminal. Subliminal influences operate below the threshold of human consciousness, where individuals are unaware of either the influence itself or its potential impact on their behaviour. In contrast, supraliminal influences operate above the perceptual threshold, allowing individuals to consciously recognise and evaluate the AI's impact on their decision-making process.

While the concept of subliminal influence has been studied for decades, its regulation in the context of AI systems only received formal recognition through the EU AI Act. This Act marks one of the first major regulatory frameworks to explicitly address this form of algorithmic manipulation.

The second dimension examines the intention and effects of AI influence. Systems designed with malicious intent and a deliberate disregard for user well-being should be considered harmful and hence must be prohibited. This category includes AI systems that cause or risk causing physical, psychological, or financial harm to individuals or groups, regardless of their stated intentions.

The third dimension focuses on the mechanisms through which AI systems influence human decision-making. The EU AI Act specifically prohibits systems that impair an individual's ability to make informed decisions, leading them to choices they would not otherwise make. This aspect emphasises the need to preserve autonomous decision-making capacity. The underlying fundamental principle is that ethical AI systems should enhance, rather than diminish human agency.

These three dimensions—nature, intention, and mechanism—together can help to create a comprehensive framework for evaluating the ethics of AI influence. By understanding these distinctions, we can better identify and promote AI systems that support positive behavioural changes while protecting against manipulative or harmful influences that undermine human autonomy and well-being.

For business leaders in Asia, this regulatory development marks a critical juncture in the ethical deployment of AI systems. While the potential for AI to optimise business operations is immense, the EU AI Act highlights how sophisticated AI techniques can exploit human psychological vulnerabilities in ways that might be visible, but whose manipulative mechanisms remain hidden from human awareness. As Asia continues to lead in digital innovation and AI adoption, understanding these influence mechanisms becomes crucial not just for regulatory compliance, but also for building sustainable business models that balance profit optimisation with ethical considerations and genuine user well-being.

CONCLUSION: PROTECTING HUMANS DURING AI-HUMAN INTERACTIONS

The age of sophisticated AI is on us now and we find ourselves in an environment saturated with AI nudges. In this article, I aimed to draw your attention to the subtle algorithmic interventions that shape our daily choices, habits, and ultimately, life outcomes. While we cannot completely insulate ourselves from AI nudges, awareness of their presence and mechanisms empowers us to maintain autonomy in our decisions.

It is also clear that for executives navigating the AI revolution, recognising AI's subtle influences is not only about maintaining decision autonomy. It also involves ensuring that AI augments, rather than supplants, human judgment in shaping organisational strategy and culture. Understanding the psychology behind AI influence has therefore become as crucial as understanding the technology itself.

Executives also need to grapple with the issue of accountability for AI-augmented decisions. We need to figure out how to parse out the distribution of responsibility when outcomes are increasingly shaped by both human judgement and algorithmic influence. The answer must emerge from a nuanced understanding of human-AI interaction.

Lastly, the EU AI Act marks a significant milestone by explicitly prohibiting AI systems that employ manipulative techniques. In doing this, the EU affirms a fundamental principle: human agency must be preserved in an AI-augmented world. As various legal regimes in Asia develop their own regulatory responses to these challenges, the EU AI Act provides a valuable blueprint for balancing technological innovation with the preservation of human agency. The future of ethical AI deployment lies not in eliminating algorithmic influence, but in ensuring that it operates transparently and respects human autonomy in decision-making. **AMI**

DR SEEMA CHOKSHI
is a researcher and speaker on AI ethics, and also founder and CEO of DataWyz.ai

For a list of references to this article, please visit https://smu.sg/z9fc or scan the QR code below.