Regulating Persuasive Technologies : the necessity of fighting fire with fire

Prof. Krishna Deo Singh Chauhan   July 15, 2024

-Prof. Krishna Deo Singh Chauhan*

Abstract

This post examines the challenges of regulating AI-driven persuasive technologies (PTs). It discusses existing regulatory approaches, their limitations, and why more comprehensive regulation is needed. The author proposes two strategies: expanded transparency obligations and enabling counter-nudging by regulators. The latter strategy, with transparency and accountability measures for its legitimate use, is likely to be both necessary and effective.

Introduction

Technologies are not only a subject matter *of* regulation, but they are also simultaneously a tool *for* regulation. And since regulation is "the sustained and focused attempt to alter the behaviour of others", it includes technologies that alter behaviour through technological management, as well as through technological persuasion. The latter iteration of these technologies is designed specifically for the purpose of persuasion, manipulation and nudging, as against for coercion or deception. These technologies are often referred to as 'Persuasive technologies' (PTs), and are designed to shape attitudes, behaviors, and decisions. Contemporary PTs are often powered by Artificial Intelligence (AI).

A question that appears naturally in the context of PTs is: how must these technologies themselves be regulated? PTs raise significant legal and ethical concerns. One major issue is the potential for these technologies to manipulate users without their knowledge or consent, undermining their autonomy and agency. This is particularly problematic when persuasive techniques are used to exploit psychological vulnerabilities (which is often the case) or promote addictive behaviors. Additionally, the use of personal data to tailor persuasive strategies raises privacy concerns and questions about data ownership and control. Furthermore, the use of these technologies in sensitive domains, such as healthcare or politics, raises questions about the appropriate boundaries of persuasion and the need for transparency and accountability. The question of the regulation of these technologies therefore ties-in with some core ethical and legal concerns.

In this post, I first discuss some prominent examples of PTs, followed by the existing approaches to regulation of PTs. Subsequently, I argue that while regulatory concern for PTs has grown, it has not grown enough, and that the current approaches are like to be ineffective as the capabilities of PTs grows. It is true that the concerns raised by PTs are not new *per se*. However, as Susser and others note, the issue has been metamorphosing – "Rather than condemning the particular harms wrought in particular contexts by strategies of online influence, scholars are beginning to turn their attention to the big picture". The evolving landscape of PTs necessitates a regulatory framework that encompasses not only immediate impacts but also far-reaching consequences. It also requires a more hands-on response. I conclude by suggesting two such responses.

AI-Driven Persuasion and Manipulation: Examples and Concerns

Before delving into the reasons why AI-driven behavioral change should be a bigger regulatory concern, it's crucial to understand the landscape of AI-driven persuasion and manipulation. This landscape is vast and complex, with numerous examples illustrating the power and potential risks of these technologies. But the growing prowess of AI is not just directly proportional to the capabilities of PTs, it is often directly responsible. As Floridi notes, AI has introduced a new form of persuasion called "hypersuasion," which leverages machine learning to process vast quantities of granular data on individuals and generate tailored content to influence beliefs and behaviors with unprecedented precision.

Some persuasive technologies are merely features of user interface design, such as notifications, infinite scroll, and social proof indicators, which can create a sense of urgency, scarcity, or popularity, ultimately driving user actions and choices. Another common set of examples are social media platforms that employ AI-driven algorithms to curate content, often leading to the creation of "echo chambers" that reinforce existing beliefs and limit exposure to diverse perspectives. These algorithmic echo chambers can significantly impact political polarization and the spread of misinformation. A related phenomenon are AI-powered advertising systems that analyze vast amounts of personal data to deliver highly targeted ads, enhancing relevance for consumers but also raising concerns about privacy and manipulation. Political microtargeting could also potentially undermine democratic processes by tailoring messages to exploit individual vulnerabilities.

Further, AI-driven gamification techniques are often employed to increase user engagement and shape behavior. While these can promote health and wellness, they can also be exploited to create addictive patterns of technology use. As with all these techniques, there is a fine line between persuasion and manipulation.

More recently, as conversational AI is advancing, chatbots are increasingly being used for customer service, therapy, and companionship. However, these systems can also be designed to subtly influence user behavior, and could in the long term prove to be much more potent PTs. Chatbots can change attitudes and behaviors related to climate change, which could be both pro- and anti- environment changes. They can also be the medium for launching cyberattacks.

These examples illustrate the diverse ways in which AI systems can influence human behavior, often in subtle and pervasive ways. The relentless nature of AI's 'hypersuasion', its magnitude, availability, affordability, and degree of efficiency based on machine-generated content tailored to individuals overshadow its precursors in terms of the depth of personalized influence and the potential scale of impact.

Existing Regulatory approaches

Data Protection and Privacy Regulations

Persuasive technologies nearly always rely heavily on the collection and processing of personal data to enable personalized influence. As such, existing data protection and privacy regulations, such as the European Union's General Data Protection Regulation (GDPR), Digital Personal Data Protection Act 2023 in India and various national laws, provide a potential foundation for regulating these technologies. By mandating transparency about data collection, usage, and sharing, they empower individuals with some awareness about how their personal information might be used to shape their attitudes and behaviors through targeted persuasion.

However, privacy regulations alone are insufficient to fully address the challenges of AI-driven persuasion. While they empower individuals with some control over their data, they do not directly regulate the persuasive techniques themselves or the societal-level impacts of these technologies. Further, the onus of exercising certain rights on users reduces their effectiveness and there is often sufficient scope for legitimate processing of data that can still power PTs.

The Right to Mental Self-Determination

Faraoni has proposed the concept of a "right to mental self-determination" as a framework for regulating persuasive technologies. This right encompasses the ability to make autonomous decisions without undue external influence, including from AI systems designed to shape attitudes and behaviors.

Recognizing mental self-determination as a fundamental right could provide a legal basis for regulating persuasive technologies. However, operationalizing this right in practice presents challenges, such as defining the boundaries between acceptable persuasion and undue manipulation and enforcing protections without unduly restricting beneficial applications. Similar to data protection laws, this would also be in principle a subject of private enforcement, thereby reducing its effectiveness.

Transparency Obligations for Recommender Systems

Some jurisdictions have introduced specific transparency requirements for AI-driven recommender systems, such as those used by social media platforms and e-commerce sites. For example, the European Union's Digital Services Act includes obligations for very large online platforms to provide transparency about the main parameters used in their recommender systems and options for users to modify those parameters.

While these transparency measures are important, they primarily focus on empowering individual users rather than addressing broader societal impacts. Moreover, the technical complexity of AI systems can make meaningful transparency challenging, as simply disclosing parameters may not provide sufficient insight into how persuasive techniques are being applied.

Educational Approaches

Education is often cited as a key tool for mitigating the risks of persuasive technologies. By equipping individuals with critical thinking skills and awareness of persuasive techniques, educational efforts aim to empower people to navigate these influences more effectively. However, the scale and pervasiveness of AI-driven persuasion raise questions about the sufficiency of education alone as a regulatory approach. Even well-informed individuals may struggle to resist carefully crafted persuasive appeals delivered with high frequency and precision. Moreover, educational initiatives may struggle to keep pace with rapidly evolving technologies.

Why More Needs to Be Done

Some drawbacks of existing regulatory approaches have been identified above. However, beyond specific drawbacks, there are certain overarching issues that give peculiar characteristics to AI driven hypersuasion.

The Transformative Scale of AI-Driven Persuasion

Scale changes the kind, not just the degree. We have seen this problem in the context of misinformation on social media, where the falsehood propagated at scale has intrinsically different characteristics and challenges than misinformation doled out individually. As a corollary, the scale at which AI-driven persuasion can be deployed fundamentally changes the nature of the challenge. Whereas traditional forms of persuasion, such as advertising and political campaigns, have always sought to influence attitudes and behaviors, AI systems can do so with unprecedented precision, personalization, and persistence. The combination of big data, machine learning, and digital platforms enables persuasive technologies to reach individuals with tailored appeals at a frequency and depth that was previously impossible. This scale transforms persuasion from an occasional encounter to a pervasive feature of the digital environment, making it more difficult for individuals to resist or even recognize.

Moreover, the scale of AI-driven persuasion means that even small influences on individual behavior can aggregate into substantial effects at the societal level. Slight nudges toward particular products, ideas, or actions, when applied to large populations, can shape broader social, economic, and political trends in ways that may not align with the public interest.

The Dual-Use Nature of Persuasive Technologies

Persuasive technologies are inherently dual use in nature, meaning they can be applied for both beneficial and harmful purposes. The same techniques that can be used to encourage healthy behaviors, promote sustainability, or enhance education can also be exploited to spread misinformation, exacerbate polarization, or manipulate individuals for commercial or political gain. This dual-use potential complicates the regulatory landscape, as simply banning or drastically restricting persuasive technologies could forgo significant benefits. At the same time, allowing their unrestricted use creates risks of malicious actors coopting these tools for nefarious ends.

Existing regulations, even when combined, struggle to navigate this tension. Data protection laws, rights frameworks, and transparency requirements each address important pieces of the puzzle, but none provide comprehensive guidance on how to harness the benefits of persuasive technologies while mitigating their risks. Educational approaches, while valuable, cannot fully inoculate individuals and societies against the effects of AI-driven persuasion at scale.

Bridging the Gap: Potential Regulatory Strategies

1.   Expanded Transparency Obligations

Transparency is a cornerstone of many existing approaches to regulating persuasive technologies, but current requirements often fall short in capturing the full range of relevant information. Expanded transparency obligations could help bridge this gap by mandating more comprehensive and accessible disclosures about the use of persuasive techniques.

Potential elements of expanded transparency could include:

- Clear and more conspicuous labeling of content that has been personalized or generated by AI systems to influence attitudes or behaviors

- Detailed information about the specific persuasive techniques being employed, such as tailored messaging, emotional appeals, or gamification elements

- Disclosure of the intended outcomes or objectives of the persuasive technology, whether commercial, political, or otherwise

- Ongoing reporting on the aggregate impacts of persuasive technologies, such as changes in user behavior patterns or effects on public discourse

Importantly, transparency obligations should be designed with usability and accessibility in mind. Complex technical disclosures that are difficult for the average person to understand or engage with are unlikely to meaningfully empower individuals. Instead, transparency requirements should prioritize clear, concise, and actionable information that enables users to make informed choices.

One promising approach to enhancing transparency (without falling into the trap of transparency paradox) is the use of *smart disclosures*. In their seminal book "Nudge: Improving Decisions About Health, Wealth, and Happiness," Thaler and Sunstein introduce the concept of nudges—subtle interventions that steer people toward better decisions without restricting their freedom of choice. Smart disclosures function as a type of nudge by presenting information in a way that helps guide individuals towards more

informed and beneficial decisions. For example, a smart disclosure might involve presenting energy consumption data in a visual format that highlights high usage periods, encouraging users to adjust their behavior to save energy.

While smart disclosures are themselves a type of persuasive technique, they persuade to act by enhancing the agency of the individual user, instead of by diminishing it. And when the individual user encounters other PTs, smart disclosures can provide users with timely, relevant, and easily understandable information about how their data is being used and how they are being influenced.

For example, a social media platform could display a prominent notification to users when their feed has been personalized based on their browsing history, explaining how this personalization may affect the content they see and offering options to adjust their preferences. Similarly, a mobile app using gamification techniques to encourage certain behaviors could provide users with a clear dashboard showing how these techniques are being employed and what goals they are intended to achieve.

Yet, even expanded transparency alone is not a panacea, as it relies on individuals having the time, motivation, and cognitive resources to process and act upon the provided information. However, when combined with other regulatory approaches and accountability measures, robust transparency can help create a more informed and resilient public.

2.   Enabling Counter-Nudging by Regulators

This last suggestion takes a bigger leaf out of the work of Thaler and Sunstein. A more controversial but potentially powerful regulatory strategy is to enable regulators to engage in counter-nudging techniques to mitigate the effects of harmful persuasive technologies.

For example, if a social media platform's algorithms were found to be amplifying misinformation or encouraging polarization, regulators could require the platform to deploy counter-nudges, such as prompts encouraging users to fact-check information or engage with diverse perspectives. Similarly, if a mobile app used gamification to promote addictive usage patterns, regulators could mandate the inclusion of persuasive elements that encourage healthy boundaries around screen time.

Counter-nudging by regulators is controversial because it raises concerns about government overreach and the manipulation of public opinion. To address these concerns, any use of persuasive techniques by regulators would need to be subject to strict transparency and oversight mechanisms. This could include public disclosure of the specific techniques being used, the intended outcomes, and the empirical evidence justifying the intervention.

Moreover, the use of counter-nudges should be grounded in democratic principles and subject to public input and debate. Regulators should not have carte blanche to manipulate individual behavior, but rather should use persuasive techniques judiciously and in alignment with established public policy goals. When implemented with appropriate safeguards, counter-nudging has the potential to provide a more agile and adaptive response to the challenges of AI-driven persuasion than traditional regulatory tools alone. By leveraging the same powerful techniques used by persuasive technologies, counter-nudges can help shape the digital environment in ways that promote individual and societal well-being.

Conclusion

As persuasive technologies, powered by artificial intelligence, become increasingly prevalent and potent, the regulatory landscape must evolve to keep pace. Existing approaches, while valuable, have struggled to address the full scope of the challenge, particularly the transformative scale of AI-driven persuasion and the dual-use nature of these technologies.

To bridge this gap, policymakers and stakeholders should consider expanded transparency obligations that provide comprehensive and accessible information about the use of persuasive techniques. Additionally, exploring the potential for regulators to engage in counter-nudging, with robust transparency and accountability measures, could provide a more adaptive response to the ever-changing landscape of persuasive technologies.

Ultimately, the goal of regulating persuasive technologies should be to harness their benefits while mitigating their risks. By extending regulatory concern beyond immediate impacts to encompass far-reaching consequences, we can work towards a future in which these powerful tools are developed and deployed in service of individual autonomy, social cohesion, and the greater good.

*Prof. Krishna Deo Singh Chauhan is an Associate Professor at Jindal Global Law School. Specializing in Technology regulation, he is currently writing his doctoral thesis on the regulation of personalized AI-driven companions, focusing on their role as choice architects. He was a member of the Jean Monnet Chair established by The European Union at the OP Jindal Global University, under which he taught courses on Privacy and data protection in the context of new technologies. His recent paper published in the 'International Review of Law, Computers & Technology' criticizes the existing approaches on the concept of beneficence in ethical principles of AI.