

2022

Misleading AI: Regulatory Strategies for Algorithmic Transparency in Technologies Augmenting Consumer Decision-Making

Jeannie Marie Paterson
University of Melbourne

Follow this and additional works at: <https://lawcommons.luc.edu/lclr>



Part of the [Consumer Protection Law Commons](#)

Recommended Citation

Jeannie M. Paterson *Misleading AI: Regulatory Strategies for Algorithmic Transparency in Technologies Augmenting Consumer Decision-Making*, 34 *Loy. Consumer L. Rev.* 558 (2022).

Available at: <https://lawcommons.luc.edu/lclr/vol34/iss3/8>

This Feature Article is brought to you for free and open access by LAW eCommons. It has been accepted for inclusion in *Loyola Consumer Law Review* by an authorized editor of LAW eCommons. For more information, please contact law-library@luc.edu.

MISLEADING AI: REGULATORY STRATEGIES FOR ALGORITHMIC TRANSPARENCY IN TECHNOLOGIES AUGMENTING CONSUMER DECISION-MAKING

*Jeannie Marie Paterson**

ABSTRACT

Increasingly, consumers' decisions about what to buy are mediated through digital tools promoted as using "AI", "data" or "algorithms" to assist consumers in making decisions. These kinds of digital information intermediaries include such diverse technologies as recommender systems, comparison sites, virtual voice assistants, and chatbots. They are promoted as effective and efficient ways of assisting consumers making decisions in the face of otherwise insurmountable volumes of information. But such tools also hold the potential to mislead consumers, amongst other possible harms, including about their capacity, efficacy, and identity. Most consumer protection regimes contain broad and flexible prohibitions on misleading conduct that are, in principle, fit to tackle the harms of misleading AI in consumer tools. This article argues that, in practice, the challenge may lie in establishing that a contravention has occurred at all. The key characteristics that define AI informed consumer decision-making support tools —opacity, adaptivity, scale, and personalization — may make contraventions of the law hard to detect. The paper considers whether insights from proposed frameworks for ethical or responsible AI, which emphasise the value of transparency and explanations in data driven models, may be useful in supplementing consumer protection

* Professor of Law, Centre for Artificial Intelligence and Digital Ethics, The University of Melbourne.

law in responding to concerns of misleading AI, as well as the role of regulators in making transparency initiatives effective.

I. INTRODUCTION

In what seems to have been a relatively short period of time, the labels “artificial intelligence” or “AI”, “algorithm”, and “data” have moved into common parlance to describe an array of computational processes that may include expert systems, sensors, data analytics, machine learning, neural networks, natural language processing and computer vision. In consumer markets, the term “AI” is being used to promote the capacities of a range of consumer goods, from vacuum cleaners to share trading apps. Additionally, and the focus of this article, technologies described as using AI - such as recommender systems, comparison sites, virtual voice assistants, and chatbots - are being promoted as using AI to assist consumers in making decisions about what goods and services to buy, particularly online. The common feature of these kinds of technologies is that they provide services to consumers (as opposed to making decisions about consumers). The service they provide concerns information. They might therefore be described as digital information intermediaries because they find and present to consumers information consumers could not otherwise locate, obtain or use in a digital form.¹

Digital information intermediary tools may well benefit consumers, at least in some contexts. Choosing between different products, as well as finding and making use of the voluminous and often complex information relevant to these choices, is often hard for consumers. AI tools may assist or augment consumers making purchasing decisions in a variety of ways, for example, by curating relevant information, summarising the available options, synthesising salient factors, or making recommendations about the best fit for consumers’ preferences or budgets. Such processes may save consumers time and money, as well as improving the likelihood of consumers finding the products that are most suitable for them.

AI informed tools for augmenting or assisting consumers’ decision-making also carry various risks. The tools may replicate and amplify existing social biases or discriminatory practices. The tools may erode consumers’ privacy by collecting personal data about them to an extent

¹ Rory Van Loo, ‘Rise of the Digital Regulator’ (2017) 66(6) *Duke Law Journal* 1267, 1272.

that is disproportionate to the gain from the services offered, and for purposes that may provide detrimental to consumers or their communities interests in the longer term. Additionally, and the focus of this article, digital information intermediaries may profoundly mislead consumers. This is all the more likely currently, when consumers are still likely to be experienced in dealing with the bundle of technologies commonly covered by the descriptor “AI” or related terms.

There are at least three main ways in which the promotion of AI in digital information intermediaries may mislead consumers. A first form of misleading AI occurs when the tools promoted to assist or augment consumer decision-making provide only partial information in response to queries from consumers, while hiding other options, such as better products or prices in order to benefit the interests of the provider of the advice service. A second form of form of misleading AI occurs when consumers are misled about the capacities of the technology they are using. This may occur in circumstances where consumers think they are being advised by an agent with profound and human like intelligence, but are in reality being guided by much less sophisticated or nuanced processes. A third form of misleading AI occurs when consumers are misled about the identity of the entity they are dealing with, so that consumers think they are interacting with a human, but are instead conversing with a bot. This kind of misleading AI is enabled by the increasingly frictionless interactions now possible with AI, as well as the possible tendency of humans to anthropomorphise objects that speak to them.

Most statutory consumer protection regimes, as well as the general law, contain prohibitions on misrepresentation or misleading conduct. In principle, these prohibitions should be capable of responding to the risks of misleading AI. The three kinds of misleading AI identified above follow patterns of misconduct that exist in the bricks-and-mortar world. Human advisers sometimes fail to disclose information for self-interested reasons, over-exaggerate their level of skill, or pretend to be something or someone they are not. Additionally consumer protection prohibitions are typically expressed in an open textured form that allows them to respond to new kinds of behaviours or practices by traders that are harmful to consumers. In practice, the key characteristics of Digital information intermediary support tools may have features that make establishing misleading conduct contrary to these prohibitions more difficult that might at first appear. The tools are often characterised by being opaque as to the technology is being utilized and

how their recommendations are reached. They commonly purport to provide a service personalized to the consumers using it. AI tools may operate at scale and may be adaptive in the sense of learning from new data inputs. These characteristics may impede attempts to establish misleading conduct under general law and statutory regimes, giving rise to an effective accountability gap.

Growing understanding of the unique harms raised by AI has given rise to suggestions for a more targeted regulatory response, whether through new specifically targeted laws or soft law codes. There have been many proliferations of principles ethical, trustworthy, or responsible AI. These principles commonly emphasise the importance of transparency and explanations in AI technologies. This article agrees that transparency is a valuable precondition to preventing harms arising from misleading AI. However, it argues that insights from consumer protection law and policy can assist in understanding for who and in what form such demands may be useful. Simply providing more information to consumers about how an Digital information intermediary assistance tool works may benefit some consumers, will not be sufficient to improve the outcomes of the interaction across the market generally. Instead, in the absence of specifically focused regulation, it falls on regulators to be robust in demanding high levels of transparency in AI informed tools for assisting consumer decision-making in order ensure accountability from the firms deploying them.

This article begins by discussing the promoted uses of AI in providing decision-making support to consumers, and the manifestations of misleading AI that may arise from these kinds of uses. Part III considers possible legal and regulatory responses to these concerns, and the practical challenges in applying existing law to emerging digital technologies. Part IV turns to the role for transparency in supporting beneficial consumer outcomes and enabling effective regulatory oversight of AI consumer products. Part V concludes.

II. MISLEADING AI IN CONSUMER MARKETS

AI technologies are rapidly expanding into consumer markets. The influence of these technologies is perhaps most prominent in consumers' interactions with digital platforms and targeted advertising but they are also influencing consumers' very act of exercising choice in the marketplace. Increasingly, consumers' decisions about what to buy are mediated through AI. Consumers are now able to rely on AI informed

tools to provide decision-making assistance in response to the many choices presented to them. Thus, for example, households are encouraged to use smart home devices to monitor energy use, virtual assistants (e.g. Alexa) to keep track of appointments or find information on household tasks, search engines (e.g. Google) for everyday information, and recommender systems (e.g. Amazon, Netflix, Facebook) to suggest shopping, music, viewing, and dining options.² Online market places provide reviews of available products and comparison sites rank possible options for products from discount flights or hotels to telco and energy providers, and even insurance. They may find themselves interacting with a chatbot when making inquiries to retailers or service providers.³ Consumers may obtain advice about their health, wellbeing, or legal issues through an app.⁴ Consumers may invest through a robo-advisers⁵ or manage their finances through automated budgeting tools.⁶ For consumers with limited mobility, smart devices are presented as a way of interacting with the outside world, such as through virtual assistants to provide reminders to take medication, sensors that monitor their movements in the home for falls, and even as robot companions.⁷

Many if not most of these products are promoted as powered by AI, algorithms, or machine learning. And certainly these uses of AI in this sense in augmenting consumer decision-making are potentially useful, and even beneficial. Digital information intermediaries of these kinds may be especially potentially beneficial in situations where the range of products choice is otherwise overwhelming, consumers lack experience or expertise about the products they are choosing between, and the information about different options is long, varied, or complex.

² See further Jennifer Cobbe & Jatinder Singh, *Regulating Recommending: Motivations, Considerations, and Principles*, 10 EUR. J.L. & TECH (2019).

³ See, e.g., *What Is a Chatbot and How Is It Changing Customer Experience?*, SALESFORCE BLOG (Jan. 20, 2022), <https://www.salesforce.com/eu/blog/2019/04/what-is-a-chatbot.html>.

⁴ See, e.g., Tom Baker & Benedict Dellaert, *Regulating Robo Advice Across the Financial Services Industry*, 103 IOWA L. REV. 713 (2018).

⁵ See e.g., Alana Benson, *11 Best Robo-Advisors of April 2022*, NERDWALLET (Mar. 19, 2022), <https://www.nerdwallet.com/best/investing/robo-advisors>.

⁶ See e.g., Q.ai, *How AI-Powered Investing Is Changing Wall Street for Millennials*, FORBES (Sep. 13, 2021, 1:06 PM) <https://www.forbes.com/sites/qai/2021/09/13/how-ai-powered-investing-is-changing-wall-street-for-millennials/?sh=2147bc6649fa>. Cf Wojtek Buczynski et al., *A Review of Machine Learning Experiments in Equity Investment Decision-Making: Why Most Published Research Findings Do Not Live up to Their Promise in Real Life*, 11 INT'L J. DATA SCI. ANALYSIS 221, 221–42 (2021).

⁷ See also Frank Pasquale, *NEW LAWS OF ROBOTICS* 49–54 (2020).

Consumers are busy and the volume of possible goods, services, and activities available to them, particularly online, may be overwhelming. Consumers find it difficult to manage large volumes of information, due to bounded rationality and information asymmetries.⁸ Inline notices and disclosures may be complex and difficult to navigate.⁹ Choice architecture in presenting information commonly steers consumers away from making choices that would restrict data flows and are counter to the interests of firms using that data.¹⁰ This all means that AI informed tools that help consumers find the right products, information or advice, whether it is through making recommendations, presenting preferred options or making delegated actions, may be very useful.

Yet AI tools used in these ways without proper oversight or regulation also raise significant risks of harm to consumer welfare. Given the extensive use of personal data collected by most Digital information intermediary assistance tools, these potential harms include biased or discriminatory outputs, and the erosion of personal privacy.¹¹ These are significant and serious risk that should be a reason for cautious and careful design before rolling out AI informed products. They have led in some jurisdictions to wide-ranging data protection laws, such as the European Union's ("EU") *General Data Protection Regulation* ("GDPR")¹², and the *California Consumer Privacy Act*,¹³ as well as laws targeting the use of unsafe AI, particularly in the EU.¹⁴

⁸ See, e.g., Russell Korobkin, *Bounded Rationality, Standard Form Contracts, and Unconscionability*, 70 U. CHI. L. REV. 1203 (2003).

⁹ See, e.g., NORWEGIAN CONSUMER COUNCIL, *OUT OF CONTROL: HOW CONSUMERS ARE EXPLOITED BY THE ONLINE ADVERTISING INDUSTRY* 60 (2020); Katharine Kemp, *Concealed Data Practices and Competition Law: Why Privacy Matters*, 16 EUR. COMPETITION J. 1 (2020); Lauren E. Willis, *Deception by Design*, 34 HARV. J.L. & TECH. 116, 115 (2020).

¹⁰ On the information asymmetry between consumers and digital service providers and the use of design strategies to nudge consumers away from privacy protection online, see Norwegian Consumer Council, *supra* note 7. See also Katharine Kemp, *Concealed Data Practices and Competition Law: Why Privacy Matters*, 16 EUR. COMPETITION J. 1 (2020).

¹¹ See generally Ryan Calo, *Digital Market Manipulation*, 82 GEORGE WASHINGTON LAW REV. 995 (2014); Gerhard Wagner and Horst Eidenmuller, *Down by Algorithms? Siphoning Rents, Exploiting Biases, and Shaping Preferences: Regulating the Dark Side of Personalised Transactions* 86 UNI. CHI. L.R. 581 (2019).

¹² Regulation (EU) 2016/679 of 27 April 2016 on the Protection of Natural Persons with regard to the Processing of Personal Data and on the Free Movement of Such Data and Repealing Directive 95/46/EC, art. 25, 2016 O.J. (L 119) 1. See generally Lee A. Bygrave, *Data Protection by Design and by Default: Deciphering the EU's Legislative Requirements*, 4 OSLO L. REV. 105 (2017).

¹³ 1.81.5 Cal Civil Code §§ 1798.135 (2018).

¹⁴ See e.g. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM (2021) 206 final (Apr. 21, 2021) [hereinafter *EU AI*

Another risk that arises through the uses of Digital information intermediary assistance in consumer markets is the risk of misrepresentation or misleading conduct. Misleading conduct is never morally or commercially acceptable. It offends human dignity and undermines agency. It undermines the base assumptions of efficient markets because consumers who are misled are unable to make informed decisions.¹⁵ Misleading conduct is also usually manipulative, as it typically has the purpose or effect of steering consumers to outcomes they have not chosen and may not benefit them. Traders who engage in misleading conduct unfairly and inefficiently obtain an advantage over other more honest traders. In the context being considered here, consumers who are misled about the capacities of an AI tool may place unjustified faith in its recommendations or forego other more reliable opportunities, and thereby suffer harm.

Misleading conduct arising from the technology of Digital information intermediary tools may manifest in at least three ways.

First, (*category 1*), consumers may be misled about the selection, ranking, or fit of the recommendations, information or advice being presented to them by the tool. Of course, some platform recommendation systems are premised on promoting their own content, primarily by applying criteria conducive to sustaining consumer engagement.¹⁶ Most consumers will understand this to some extent. At some point, however, recommendation and comparison sites may cross the line from self-promotion to misleading consumers about the independence or merits of their recommendations. For example, consumers making use of comparison or recommender sites may assume they will be presented with a selection of options, ranked according to their search parameters, whereas in fact the options may be presented, curated, or priced on quite different grounds. Recommendation systems may, without making it apparent to the consumer, respond to a consumer query by promoting their own products or those of paid advertisers.¹⁷

Act]; Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC [hereinafter *EU Digital Services Act*].

¹⁵ See Robert Sparrow & Linda Sparrow, *In the hands of machines? The future of aged care* 16 MINDS & MACHINES 141, 155-156 (2006).

¹⁶ Cobbe & Singh, *supra* note 2.

¹⁷ Maurice E. Stucke and Ariel Ezrachi, *How Digital Assistants Can Harm Our Economy, Privacy, and Democracy*, 32 BERKELEY TECH. L.J. 1239, 1258-62 (2017). See also Greg Sterling, *Google Takes Baby Steps to Monetize Google Assistant, Google Home*, SEARCH ENGINE

Comparison websites may provide rankings that are not based on the best deal for a consumer but instead on self-interested factors, such as commissions or business relationships.¹⁸

Secondly, (*category 2*), consumers may be misled about the capacities of the Digital information intermediary tool. At this point in time, most consumers probably have a poor understanding of AI. Thus, in dealing with so-called AI in decision-making support tools, consumers may assume the AI has a level of competence or expertise that does not in fact exist. This might be described as thinking that the tool represents ‘general’ AI,¹⁹ capable of outperforming humans at any intellectual task, whereas it is in fact it is merely ‘narrow’ AI, capable of performing a specific task, or perhaps not even AI at all but some simpler decision tree or data processing technology.²⁰ Thus, consumers seeking advice from an AI tool may assume the recommendation or advice may appear more accurate or informed than is the case because it is provided through an “algorithm” or machine learning technology.²¹ Or consumers may not recognise that a poorly trained machine learning model may make less accurate predictions than a well-designed expert system built from human experience.

Firms providing digital information intermediaries will not be responsible for all misapprehensions of their consumer clients. However, in some instances, firms’ conduct will have contributed to that misunderstanding. For example, expectations of high expertise and insights from AI tools is likely to be reinforced in circumstances where consumers are asked to enter a significant degree of personal information about themselves to achieve a recommendation. This is because consumers asked to enter significant personal information may reasonably

LAND (Apr. 22, 2019, 4:53 PM), <https://searchengineland.com/google-takes-baby-steps-to-monetize-google-assistant-google-home-315743>.

¹⁸ Stucke & Ezrachi, *supra* note 17, at 1258 (discussing manipulation of results in online air and hotel bookings). See, e.g., Press Release, Australian Competition & Consumer Commission, iSelect in Court for Alleged Misleading Conduct and Claims About Energy Plan Comparisons (Apr. 12, 2019), <https://www.accc.gov.au/media-release/iselect-in-court-for-alleged-misleading-conduct-and-claims-about-energy-plan-comparisons>.

¹⁹ See Mark Coeckelbergh, *How to describe and evaluate “deception” phenomena: recasting the metaphysics, ethics, and politics of ICTs in terms of magic and performance and taking a relational and narrative turn*, ETHICS INF. TECHNOL. 71 (2018).

²⁰ On general and narrow AI, see Simon Chesterman, WE, THE ROBOTS?, 1–2 (Cambridge, 2021)

²¹ Lydia Kostopoulos, DECOUPLING HUMAN CHARACTERISTICS FROM ALGORITHMIC CAPABILITIES 3 (2021) (“The term itself already has an important human characteristic – intelligence. Terming it as artificial creates expectations and assumptions about the intelligence capabilities in relation to human intelligence”).

expect that they will be provided with a personalised recommendation that uses that information meaningfully. A tool that fails to make any use of that information in producing its outputs, may, in addition to acting as a blatant data grab, be misleading. Equally, a tool that refers to itself as using AI, while hiding behind its interface a random selection-based of outputs generated by the technical equivalent of flipping a coin, may also mislead.

Further examples of this second category of misleading AI can be imagined. It may be misleading to promote insights obtained from an AI informed decision assistance tool that go beyond what is credibly possible at this point in time, particularly as this may result in consumers relying on outputs that are baseless, harmful, or discriminatory. For example, there are efforts to develop tools that use biometric makers to allegedly provide insights into human character or behaviour, including emotional state, mental health, intelligence, sexual proclivity, learning capacity and attention/distraction.²² These kinds of uses of AI have been described by former Australian Human Rights Commissioner, Edward Santow, as “junk science”.²³ Attempts to package these techniques into a consumer tool are highly likely to mislead.

AI informed tools to augment or assist consumer decision-making may further be designed to manipulate consumers. For example, think of a budgeting or investment tool that nudges consumers towards investments that benefit the firm providing the service, or a video game that targets addictive behaviour in users in order to increase the amount of money spent on loot boxes.²⁴ Alternatively, consider a companion bot that uses extreme flattery to guarantee its place in the affections of lonely consumers that use it or promotes drug use to ensure compliance.²⁵ These tools will be misleading insofar as they promote themselves as benefiting consumers, and they may also be unfair.

²² See e.g., Pasquale, *supra* note 7, at 60. Also Mark Purdy, John Zealley & Omaro Maseli, *The Risks of Using AI to Interpret Human Emotions*, *HARVARD BUSINESS REVIEW* (Online, 18 November 2019); Alexandrine Royer, *The wellness industry's risky embrace of AI-driven mental health care*, *TECH STREAM* (online, 14 October 2021).

²³ Edward Santow, *Can Artificial Intelligence Be Trusted with Our Human Rights?*, 91 *AUSTRALIAN QUARTERLY* 10, 17 (2020).

²⁴ Peter Cartwright & Richard Hyde, *Virtual Coercion, and the Vulnerable Consumer: 'Loot Boxes' as Aggressive Commercial Practices*, *LEGAL STUDIES* 1–21 (2022).

²⁵ Cf. Göran Wågström, *Why Behavioural Advertising Should be Illegal*, *FORBES* (Mar. 5, 2019, 7:45 AM), <https://www.forbes.com/sites/forbestechcouncil/2019/03/05/why-behavioral-advertising-should-be-illegal/?sh=40dd9c065b89>. See also Sam Levin, *Facebook Told Advertisers It Can Identify Teens Feeling "Insecure" and "Worthless"*, *THE GUARDIAN* (May 1, 2017, 5:01 PM),

Thirdly (*category 3*), consumers may be misled about the identity of the AI informed tool. Misleading conduct of this kind is most obviously manifest where humans are dealing with robots that resemble humans, animals, or some other sentient being, long an anxiety expressed in movies such as *Metropolis*, *Bladerunner* and *Ex Machina*. The possibility also arises with chat or voice bots that converse in a naturalistic manner, and indeed are designed to replicate the often imprecise or flawed modes of human conversation.²⁶

The concern with these naturalistic devices is that consumers may be misled into thinking that they are interacting with a human or other living thing,²⁷ when in fact they are dealing with an automated process.²⁸ We might regard presenting robots or bots that seem “alive” to unsuspecting humans as morally wrong for the very reason that it is misleading them about the sentient status of the technology. Certainly, modern laws of robotics include a requirement that the bot will not deceive humans as to its identity.²⁹ There is additionally the concern that life-like AI-based voice assistants or robots may be introduced to care for vulnerable people as cost savings measures, rather than for beneficent reasons, leaving people who may already be isolated, such as elderly or chronically ill people, devoid of human interaction and care.³⁰

Human-looking bots seem more likely to mislead consumers as to their “real” existence than those represented as stereotypical robots. This means that a partial solution to the problem may lie in ensuring the bot

<https://www.theguardian.com/technology/2017/may/01/facebook-advertising-data-insecure-teens>; Sidney Fussell, *Facebook Allow Drug Ads to Target Teens, Activists Say*, WIRED (May 4, 2021, 9:00 AM), <https://www.wired.com/story/activists-facebook-allows-drug-ads-target-teens/#:~:text=The%20social%20network%20has%20rules,they%20aren't%20being%20enforced.>

²⁶ See Judith Shulevitz, *Alexa, How Will You Change Us?*, 322 THE ATLANTIC MAGAZINE, 96, 96 (2018) <www.theatlantic.com/magazine/archive/2018/11/alexa-how-will-you-change-us/570844/> accessed 5 April 2022.

²⁷ Cf. Simon Coghlan, *Robots and the Possibility of Humanistic Care*, INT’L J. SOC. ROBOTICS (SPECIAL ISSUE) (2021).

²⁸ Jacob Kastrenakes, *Google Starts Rolling out Duplex Feature that Can Call Salons to Book a Haircut for You*, THE VERGE (Oct. 13, 2020, 3:09 PM), <https://www.theverge.com/2020/10/13/21514427/google-duplex-haircut-booking-feature-rolling-out-robot-natural-voice>.

²⁹ See Pasquale, *supra* note 7, at 7; Engineering and Physical Sciences Research Council, *Principles of Robotics*, UK RSCH. & INNOVATION (July 1, 2021), <https://webarchive.nationalarchives.gov.uk/ukgwa/20210701125353/https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/>.

³⁰ Pasquale, *supra* note 7, at 52.

is encased or represented by artificial, and not human-like, form. However, this may not avoid the problem entirely. Joanna J. Bryson alerts us to the tendency in humans to ascribe human-like qualities to robots, and presumably also to disembodied bots such as chat bots or voice assistants.³¹ In some instances, this problem may be harmless, but we do not know enough about the human-computer interaction in these scenarios to dismiss the concern entirely.³² Pasquale describes the problem as “counterfeiting” distinctively human characteristics.³³ Even where consumers know they are dealing with a technology, they may nonetheless ascribe consciousness or empathy to that entity if it appears to converse in a naturalist or appealing manner. It is then possible to envisage consumers being manipulated as a result of this counterfeited relationship.³⁴ Problems of misplaced trust may occur if consumers wrongly assume the AI understands or cares about them, or overidentify with the AI,³⁵ when in fact there is simply no sentience or ability to care at all.³⁶

III. RESPONDING TO MISLEADING AI: PRINCIPLES AND PRACTICE

The potential for harm arising from these various categories of misleading AI is still underexplored. However, given the relatively recent emergence of the technology, consumers are unlikely to be alert to the risks of misleading conduct, or indeed other harms, that arise from Digital information intermediary tools, or have adequate strategies for self-help.³⁷ There is therefore a strong case for regulators and

³¹ Joanna J. Bryson, *Robots Should Be Slaves*, in CLOSE ENGAGEMENTS WITH ARTIFICIAL COMPANIONS: KEY SOCIAL, PSYCHOLOGICAL, ETHICAL AND DESIGN ISSUES 63, 63–74 (Yorick Wilks ed., 2010).

³² See also Jeannie Marie Paterson & Yvette Maker, *AI in the Home: Artificial Intelligence and Consumer Protection*, in THE CAMBRIDGE HANDBOOK OF PRIVATE LAW AND ARTIFICIAL INTELLIGENCE (Ernest Lim & Phillip Morgan eds., forthcoming 2022).

³³ Pasquale, *supra* note 7, at 7.

³⁴ Eleni Adamopoulou & Lefteris Moussiades, *Chatbots: History, Technology, and Applications*, 2 MACH. LEARNING WITH APPLICATIONS (2020); Xueming Luo et al., *Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases*, 38 MKTG. SCI. 937 (2019)

³⁵ Bryson, *supra* note 31, at 63–74.

³⁶ E.g., Chesterman, *supra* note 20, at 114 (discussing Eliza the chatbot psychotherapist – “Even when they were told how it worked, some users insisted that Eliza has ‘understood’ them.”)

³⁷ See Skander Bennis, *AI and the Consumer*, in ARTIFICIAL INTELLIGENCE AND THE LAW 461, 469–72 (Jan De Bruyne & Cedric Vanleenhove eds., 2021).

consumer advocates to be proactive in seeking strategies to deter such conduct.

Principle-based prohibitions on misleading conduct

Concerns about misleading AI should, in principle, be amenable to redress through the prohibitions on misleading or deceptive practices commonly found in consumer protection regimes, such as the Federal Trade Commission Act,³⁸ the EU's Unfair Commercial Practices Directive,³⁹ or the Australian Consumer Law.⁴⁰ These core "safety-net" prohibitions are typically expressed as open-textured standards, which makes them capable of adapting to new manifestations of the prohibited conduct, including those raised by emerging technologies.⁴¹ The prohibition on misleading conduct has already been used in several jurisdictions to respond to regulatory concerns about the conduct of digital platforms, including to hold digital platforms to their stated privacy policies, and to ensure greater clarity in those policies.⁴² The Federal Trade Commission has signalled a preparedness to use its regulatory powers to address concerns about bias and unfairness in consumer AI applications.⁴³

³⁸ 15 U.S.C. § 45(a)(1). *A Brief Overview of the Federal Trade Commission's Investigative, Law Enforcement, and Rulemaking Authority*, FEDERAL TRADE COMMISSION (last updated May 2021), <https://www.ftc.gov/about-ftc/mission/enforcement-authority> ("Deceptive" practices are defined in the Commission's Policy Statement on Deception as involving a material representation, omission or practice that is likely to mislead a consumer acting reasonably in the circumstances. An act or practice is 'unfair' if it 'causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition.' 15 U.S.C. Sec. 45(n).

³⁹ Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council, Article 6 [hereinafter Directive 2005/29/EC].

⁴⁰ *Competition and Consumer Act 2010* (Cth) sch 2 (Austl) (*Australian Consumer Law*) s 18.

⁴¹ Jeannie Marie Paterson & Elise Bant, *Misrepresentation, Misleading Conduct and Statute Through the Lens of Form and Substance*, in *FORM AND SUBSTANCE IN THE LAW OF OBLIGATIONS* 403 (Andrew Robertson & James Goudkamp eds., 2019).

⁴² See Lesley Fair, *FTC's \$5 Billion Facebook Settlement: Record-Breaking and History-Making*, FEDERAL TRADE COMMISSION: BUS. BLOG (July 24, 2019), <https://www.ftc.gov/business-guidance/blog/2019/07/ftcs-5-billion-facebook-settlement-record-breaking-history-making>; *Australian Competition and Consumer Commission v Google LLC [No 2]* (2021) 391 ALR 346; *Australian Competition and Consumer Commission v HealthEngine* [2020] FCA 1203 (20 August 2020) (Austl.). Also, Jeannie Marie Paterson et al., *Australian Competition and Consumer Commission v Google: Deterring Misleading Conduct in Digital Privacy Policies*, 26 *COMM'N L. - J. COMPUT. MEDIA & TELECOMM. L.* 136, 136-48 (2020).

⁴³ See, e.g., Elisa Jillson, *Aiming for Truth, Fairness, and Equity in Your Company's Use of AI*, FEDERAL TRADE COMMISSION: BUS. BLOG (Apr. 19, 2021), <https://www.ftc.gov/business->

In principle, using statutory prohibitions on misleading conduct to respond to categories of misleading AI may be relatively straightforward. The prohibition's intended purpose is to address interactions between firms and consumers in which information is distorted.⁴⁴ AI tools will, contravene these prohibitions, in the scenarios identified above, where they lead consumers to believe they are obtaining a loyal as opposed to self-interested recommendation (category 1), obtaining an expert service individually tailored their needs to rather than an off-the-shelf product with only three kinds of rotating recommendations (category 2), or dealing with a real person as opposed to a bot (category 3). All of these examples have analogues in the real world and established case law. For example, promoting a "best price guarantee" may be misleading if different consumers are, for no good reason, presented with different prices for the product, or the outputs are ranked by commission not fit or value, whether that conduct occurs online or otherwise.⁴⁵

Typically, positive misstatements are more likely to offend prohibitions on misleading conduct than failures to act, such as failing to provide salient information. Nonetheless, prohibitions on misleading conduct extend to omissions, particularly where non-disclosure is accompanied by conduct that creates a misleading impression, or where there is a reasonable expectation of disclosure.⁴⁶ Thus, some manifestations of digital information intermediaries may mislead simply through the context or form in which they are presented, without the need for expressly wrongful statements. For example, a recommendation system that solicits extensive information about consumers' preferences, and then does not utilize these in producing results may be misleading, because the impression has been created that the information will be used in providing the advice (category 2). An AI tool

guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai (warning companies to ensure against bias in algorithms as contrary to fairness in trading from the Federal Trade Commission). See also AUSTRALIAN COMPETITION AND CONSUMER COMMISSION, DIGITAL PLATFORMS INQUIRY: FINAL REPORT (2019); *Strengthening Privacy for the Digital Age: Proposals to Modernize the Personal Information Protection and Electronic Documents Act*, GOVERNMENT OF CANADA, https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00107.html (last modified May 21, 2019); COMPETITION AND MARKETS AUTHORITY, ONLINE PLATFORMS AND DIGITAL ADVERTISING: MARKET STUDY FINAL REPORT (2020).

⁴⁴ Lauren E. Willis, *supra* note 9, at 169.

⁴⁵ See, e.g., *Australian Competition and Consumer Commission v Trivago N.V.* (2020) 142 ACSR 338.

⁴⁶ See in particular Directive 2005/29/EC article 7. Also generally Elise Bant Paterson and Jeannie Marie Paterson, *Misleading Silence* (Hart, 2019).

that presents with a name or other human characteristics, may mislead if it does not disclose its artificial status, because the use of a name or face suggests a human interaction (category 3).

Practical hurdles to establishing misleading AI

In practice, there may be real hurdles for consumers (or regulators) trying to prove a contravention in of the prohibitions on misleading conduct by Digital information intermediary assistance tools.⁴⁷ Primarily, these difficulties are likely to arise from the typically distinguishing features of the relevant technologies: opacity, personalisation, adaptivity, and scale.

Problems of opacity, or the “blackbox” effect, may arise in a variety of ways.⁴⁸ The operation of the tool may be opaque due to the demands of firms to protect the confidentiality of the algorithms they are using. Or the tool may be opaque because combinations of different algorithmic processes are used. For example a consumer chatbot may rely on natural language processing for interacting with consumers, and an expert system for producing recommendations based on the information entered. The operation of the tool may be opaque where it is based on models derived from machine learning or neural networks, which find correlations in huge numbers of data points, in ways that are not necessarily explicable or causally coherent to the firms making the tool available to consumers.

Opacity of these kinds in digital information intermediary tools may make identifying misleading conduct more difficult for consumers than in “bricks and mortar” transactions. Consumers may simply not be able to interrogate the tool sufficiently to show that they did not get what was represented to them in using the tool. For example, consumers may struggle to show they were misled by a self-serving recommendation (category 1) because influence of a commission payment may be difficult to extract, given the range of other factors involved in processing the output. Recommendations, predictions, or advice are usually not misleading merely because they prove to be wrong or inaccurate. Generally, however, there is an implicit representation that

⁴⁷ See Kayleen Manwaring, *Emerging Information Technologies: Challenges for Consumers*, 17 OXFORD U. COMMONWEALTH L.J. 265, 285 (2017).

⁴⁸ Jenna Burrell, *How the Machine “Thinks”*: *Understanding Opacity in Machine Learning Algorithms* 3 BIG DATA & SOCIETY 1 (2016); Cynthia Rudin, *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*, 1 NAT.MACH. INTELL. 206, 207 (2019).

recommendations are based on reasonable grounds and the result of reasonable care. Consumers given a poorly fitting as opposed to an expert recommendation may not be able to establish the expertise of the tool was misrepresented (category 2) because they are unable to access and therefore establish any flaw in its operation.

The hurdles to establishing misleading AI may be compounded by the promise of personalization. One of the key promoted attractions of digital information intermediaries is that they will provide personalised recommendations, decisions, or predictions. They purport to do this on the basis of the personal data entered by consumers, and also often gathered from online interactions. Personalisation may ensure more nuance and a better fit for consumers than generalized suggestions. However, personalisation may also make errors in the process used by digital information intermediary tools more difficult to identify and remedy, particularly when combined with the opaque processes underlying such decisions. This is because their recommendations are personalised, each consumer will receive a different recommendation, purportedly based on their own circumstances. This may make it difficult to show that a recommendation derived from a digital information intermediary tool that proved to be a poor fit for the needs or circumstances of the consumer (scenario 1) was the result of a lack of care, deliberate action, as opposed to bad luck given the peculiar circumstances of the consumer. Personalization may also make it more difficult to identify trends or patterns across the market that might otherwise have acted as a red flag warning of problems of misleading conduct or other consumer law breaches.

A further feature of digital information intermediary tools that may, in conjunction with opacity and personalization, create hurdles in establishing misleading conduct is adaptivity. AI-informed technologies may evolve as they “learn” from new data. As Lauren Willis has identified in the context of dark patterns, adaptivity may escalate both problems of misrepresentation and the challenges for consumers seeking to establish misleading conduct.⁴⁹ It is conceivable that a digital information intermediary tool may learn to manipulate or mislead consumers in order to produce better outcomes or higher profits for the firm deploying it. Consumers may struggle to show they were misled in promotional material or the outputs of an AI information decision making support tool because the material provided to consumers at any

⁴⁹ See Willis, *supra* note 9, at 166.

particular point in time may not be identifiable in the multitude of variations being distributed by the system.⁵⁰ A specific kind of interaction between consumers and an advisory tool may be lost as the tool “learns” to improve its naturalistic or seamless interactions with consumers.

IV. THE VALUE OF TRANSPARENCY IN AI CONSUMER TRANSACTIONS

Promoting AI transparency

In response to the unique risks of harm arising from AI, some jurisdictions have introduced law directed squarely at the technology itself. Currently, the law applicable to new technologies is usually identified by a process of characterisation, which involves matching the problem to existing legal categories (i.e. negligence, contract, consumer protection etc). A different approach is for the law to be specifically directed to the technology of concern (i.e. AI), or equally to develop a specific soft law code complementing the application of more general law. In AI specific regimes, transparency is commonly a core element.

The California Bot Disclosure Law (the *Bolstering Online Transparency Act*) requires bots used to influence a vote or incentivize a transaction to be expressly identified as such.⁵¹ The EU already requires traders to inform consumers about differential pricing.⁵² New EU regulations on consumer protection require greater transparency from platforms as to the key criteria used to rank search results⁵³ and whether traders have paid for higher rankings.⁵⁴ Under the new EU

⁵⁰ Compare the transparency requirements in the *EU Digital Services Act*, *supra* note 14.

⁵¹ CAL. BUS. & PROF. CODE § 17940 (Deering’s).

⁵² Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on Consumer Rights, art 6(1)(ea), 2011 O.J. (L 304) 64, as amended by Directive (EU) 2019/2161, 2019 O.J. (L 328) 7.

⁵³ Directive 2005/29/EC, *supra* note 37, at art 7(4a), as amended by Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019, 2019 O.J. (L 328) 7. See also Org. for Econ. Co-operation & Dev. [OECD], *Good Practice Guide on Online Consumer Ratings and Reviews*, OECD Digital Economy Papers no. 288 (2019) [https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/CP\(2019\)5/FINAL&docLanguage=En](https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/CP(2019)5/FINAL&docLanguage=En).

⁵⁴ Directive 2005/29/EC, *supra* note 37, Annexure 1, 11a, as amended by Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019, 2019 O.J. (L 328) 7.

Digital Services Act, greater transparency measures will apply to online advertising, including a requirement for platforms to display “meaningful information about the main parameters used to determine the recipient to whom the advertisement is displayed”.⁵⁵ The draft EU *AI Act*⁵⁶ adopts a risk-based approach to regulating AI, based on the likely effect of the system on health, safety and fundamental human rights,⁵⁷ and also imposes transparency obligations when AI systems interact with humans, unless this is obvious from the circumstances and conditions of use.⁵⁸

The soft law response to specific concerns about increasing human reliance on AI technologies is found in codes or frameworks of AI ethics, or responsible or trustworthy AI.⁵⁹ These codes typically emphasise the value of transparency, complemented by explanations or explicability in AI recommendations and decisions. Codes of AI ethics have been subject to extensive criticism, including that they may be co-opted by firms deploying AI to further entrench their market power and cloak the need for stronger regulation.⁶⁰ These kinds of criticism should not mean abandoning codes of AI ethics. Codes of AI ethics may be useful as a way of prompting ex-ante and technologically-based responses to the risks of harm from AI. They also draw attention to the need for accountability by firms that develop and deploy such technologies, which may be given force through complementary legal rules and enforcement strategies..

⁵⁵ *EU Digital Services Act*, *supra* note 14.

⁵⁶ *EU AI Act*, *supra* note 14.

⁵⁷ Lillian Edwards, *Expert Opinion: Regulating AI in Europe* (Ada Lovelace Institute, 2022); Nathalie A Smuha, Emma Ahmed-Rengers, Adam Harkens, Wenlong Li, James MacLaren, Riccardo Piselli, and Karen Yeung, *How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act* (August 5, 2021). Available at

SSRN: <https://ssrn.com/abstract=3899991> or <http://dx.doi.org/10.2139/ssrn.3899991>.

⁵⁸ *EU AI Act*, *supra* note 14, art. 52.

⁵⁹ See e.g., *Commission White Paper on Artificial Intelligence*, COM (2020) 65 final (Feb. 19, 2020); *Australia's Artificial Intelligence Ethics Framework*, AUSTRALIAN GOVERNMENT DEPARTMENT OF INDUSTRY, INNOVATION & SCIENCE, <https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework> (last visited April, 2022); SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE, *AI IN THE UK: READY, WILLING AND ABLE?*, 2017–9, HL 100, at 38 (UK).

⁶⁰ Luke Stark, Daniel Greene & Anna Lauren Hoffmann, *Critical Perspectives on Governance Mechanisms for AI/ML Systems*, in *THE CULTURAL LIFE OF MACHINE LEARNING: AN INCURSION INTO CRITICAL AI STUDIES* 257 (Jonathan Roberge & Michael Castelle eds., 2020) 257; AUSTRALIAN HUMAN RIGHTS COMMISSION, *HUMAN RIGHTS AND TECHNOLOGY DISCUSSION PAPER 54* (2019). See also Carly Kind, *The Term “Ethical AI” is Finally Starting to Mean Something*, VENTUREBEAT (Aug. 23, 2020, 10:25 AM), <https://venturebeat.com/2020/08/23/the-term-ethical-ai-is-finally-starting-to-mean-something/>.

As yet, there has been little work on the relationship between codes of AI ethics and consumer protection statutory regimes. The regimes are possibly complementary, with each providing insights for the operation of the other. In particular, technical advances in AI transparency and explanations may be useful in overcoming some of the hurdles to establishing misleading conduct and other contraventions of consumer protection law by digital information intermediary tools. Consumer protection scholarship may provide insights into the ways in which transparency may be useful, and also its limitations, in protecting consumers from the risk of harms arising from AI.⁶¹

Transparency in relation to AI products is not merely a matter of making available the detail of the algorithms providing the decision, prediction, or recommendation. Algorithmic transparency operates at greater or lesser levels of detail. Transparency may include information about the use of AI, the kinds of algorithmic process being used, the nature and sources of the data used to develop or train the underlying model, the efficacy or accuracy of the AI in performing its allocated tasks, and the steps that should be taken by those using AI informed tools to safeguard their interests. A commitment to AI transparency in this sense would, regardless of the existence of specific law requiring such processes, clearly disclose at the outset any commissions or other payments informing recommendations made to consumers (category 1), the limitations of the tool (category 2), and when human consumers are conversing with a bot (category 3). Such information would potentially be useful for consumers using AI-informed tools in improving their baseline understanding of the technology. It may also alert them to the possibility of recommendations that are not neutral. It may prompt consumers to think about the expertise behind the recommendations, or who they are communicating with.

Transparency in AI informed tools should be tailored towards the information needs of the recipient.⁶² Thus, in the consumer context, short, targeted information will usually be better than long explanations in clarifying the character of the entity consumers are dealing

⁶¹ Lauren E. Willis, *Performance-Based Consumer Law*, 82 U. CHI. L. REV. 1309 (2015) (listing the “meta-goals” of consumer law as “enhancing consumer decisional autonomy in the marketplace, encouraging market transactions that optimize consumer welfare, and fairness in outcomes as between consumers, particularly for disadvantaged consumers”).

⁶² See Sandra Wachter et al., *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR* 31 HARV. J.L. & TECH. 841, 843 (2018); Tim Miller, *Explanation in Artificial Intelligence: Insights from the Social Sciences*, 267 A.I. 1 (2019).

with or the processes involved in producing recommendations or advice.⁶³ For example, effective transparency in recommender systems might simply reveal that “this recommendation is based on the commission paid to us – other consumers may see different options or different prices”. A different kind of response might be required in contexts where the concern is that consumers may be misled into thinking a bot has human or life-like qualities. This is a field where more research is required, but it is possible that reminders of the status of the entity might assist, such as “I am a computer program” as opposed using a robot-like avatar but which uses a human name. Design features may also be relevant. It is possible that a frictionless or naturalistic interaction with a bot may be more likely to lead a human into thinking they are dealing with a being, than a more formal or artificial conversation style.

Transparency is closely associated with initiatives in explainable AI.⁶⁴ The computer science field of explainable AI aims to provide greater transparency into the basis for automated decisions, predictions and recommendations.⁶⁵ One mechanism proposed for providing meaningful explanations to non-technical audiences involves the use of counterfactuals.⁶⁶ Using insights from social sciences, Miller argues that meaningful explanations in explainable AI should be contrastive, rather than causal.⁶⁷ Lawyers are familiar with causal counterfactuals, which may be used to establish causation in civil litigation. Causal counterfactuals involve asking whether a different result would have ensued if the factor being considered was changed. For example, in establishing whether a financial adviser was liable for negligence, a court might consider whether, but for the adviser’s failure to consider the plaintiff’s health insurance costs, the client would have become bankrupt.

In contrast to causative counterfactuals, Miller explains that contrastive explanations explain an event relative to a different counterfactual

⁶³ See Anthony Duggan and Iain Ramsay, *Front End Strategies for Improving Consumer Access to Justice*, in Michael Trebilcock, Anthony Duggan and Lorne Sossin (eds), *MIDDLE INCOME ACCESS TO JUSTICE* (University of Toronto Press, 2012).

⁶⁴ See Select Committee on Artificial Intelligence, *supra* note 59, at 40. See also Australian Human Rights Commission, *supra* note 60, at 75.

⁶⁵ Tim Miller, *Explainable Artificial Intelligence: What Were You Thinking?*, in *ARTIFICIAL INTELLIGENCE FOR BETTER OR WORSE* 19, 21 (Future Leaders, 2019); Wachter et al., *supra* note 59, at 844.

⁶⁶ Miller, *supra* note 65; Wachter et al., *supra* note 62.

⁶⁷ Miller, *supra* note 62, at 6 and 16.

event or “foil” that did not occur.⁶⁸ The key is to identify the variables that are meaningful or salient in terms of allowing the person seeking the explanation to understand the decision, as opposed to listing the entire causal chain or responding to all alternative foils.⁶⁹ For example, a contrastive explanation might seek to answer the question “why did you recommend product A and not product B?”. A compelling answer might be that 90% of consumers who had purchased product C also purchased product A.⁷⁰ Contrastive counterfactuals used in this way may assist consumers to understand more about the basis for a decision or recommendation made by an AI-informed tool, and how changes in their behaviour may produce a different outcomes. For example, an explanation for why an AI informed tool recommended one particular credit card or phone plan to a consumer, as opposed to another, might be that consumers offered different options had lower peak data use or more timely payment practices. Such an explanation may assist the consumer by revealing the kinds of behaviour that would have produced different product choices being available to them.

The limits of AI transparency in consumer protection

AI transparency or explanations are unlikely to be complete or fully effective responses in counteracting the effects of misleading AI in the sense discussed earlier. AI transparency is not quite the same as the disclosure requirements often seen in consumer markets, such as for credit, insurance, and pharmaceuticals. Disclosure operates to inform consumers of the risks inherent in particular products. The role of transparency promoted by codes of ethical AI is to alert users to the use of the technology in any particular scenario, while explanations help consumers understand the basis on which AI decisions or recommendations are made. Nonetheless, transparency measures overlap with disclosure in that ultimately the purpose of AI transparency/explanations directed at consumers is to improve their understanding of the technology and thereby to build trust and confidence in that technology.

⁶⁸ *Id.*

⁶⁹ *Id.* at 11. See also Wachter et al., *supra* note 62, at 851.

⁷⁰ Miller, *supra* note 65, at 28.

Consumer protection scholars and advocates have criticised policy measures that make disclosure a central tool for consumer protection.⁷¹ They have argued that disclosure-based initiatives for consumer protection ignore factors such as the bounded rationality of consumers and the inequality of bargaining power in most transactions. These factors substantially limit the ability of consumers to use the disclosed information by selecting better consumer products or bargaining for fairer contract terms.⁷² If these insights are applied to AI transparency measures, they suggest that merely providing information about the use of AI in decision making support tools, or explanations of how recommendations are arrived at, may have limited salience in improving consumers understanding of the technology, ability to assess the cogency of the recommendation (category 1), or the tendency to ascribe super-human capacity (category 2) or to anthropomorphise those tools (category 3). But these limitations on the effectiveness of transparency in improving the welfare of individual consumers is not the end of the story for transparency.

What is needed are interventions that assist regulators and consumer advocates in responding to systemic issues of opacity, personalisation, and adaptivity at scale, rather than misleading AI being addressed on a piece-meal basis through individual actions. This means it will be important to think about transparency in terms of allowing regulators to oversee the operation and performance of AI tools in consumer markets and to verify, or disprove, claims about the operation of those tools.⁷³ The kinds of information about AI informed tools useful to regulators are likely differ from those aimed at consumers.

As we have seen, counterfactual explanations, such as key features of those who received different recommendations, may help consumers understand more about the recommendation provided to them. Regulators may be interested in the detail of the algorithms being used in AI informed tools,⁷⁴ although this is not the only way for regulators to gain the necessary oversight of such products. Regulators are likely to be concerned with the overall patterns and trends in the outputs of AI

⁷¹ See Geraint Howells, *The Potential and Limits of Consumer Empowerment by Information*, 32 J.L. & SOC'Y 349, 359 (2005); Omri Ben-Shahar & Carl E. Schneider, *The Failure of Mandated Disclosure*, 159 U. PA. L. REV. 647, 679–729 (2011).

⁷² See generally Robert A. Hillman & Jeffrey J. Rachlinski, *Standard-Form Contracting in the Electronic Age*, 77 N.Y.U. L. REV. 429 (2002); Korobkin, *supra* note 8.

⁷³ See Wachter et al., *supra* note 62, at 843.

⁷⁴ See, e.g., *Australian Competition and Consumer Commission v Trivago N.V.* (2020) 142 ACSR 338.

informed tools, which may identify trends of self-serving or biased recommendations, or patterns of results that undermine claims to personalisation in AI informed recommendations or advice. Regulators may also be interested in some form of feature analysis, which list the factors and their weightings in producing recommendations or decisions as these may also provide insights into whether the tool's output was tailored to an individual consumer's inquiry, generic to all consumers, self-serving or biased.⁷⁵ Going further, regulators may be interested in promoting the possibility of simpler design strategies in developing AI informed tools to assist consumer decision-making. Certainly, Cynthia Radin argues that so-called black box AI may not be necessary for accurate predictions and that the emphasis for promoting accountability should be on interpretable models.⁷⁶

Incentives to greater transparency

Assuming the value of AI transparency in promoting good consumer protection outcomes in AI informed tools for consumer decision-making is recognised, there is a subsequent issue of how to encourage firms to provide reasonable levels of transparency around the operations of their product. Arguably, some level of transparency is required as part of a firm's general law obligations of care and skill in providing the service of digital information intermediary tools. Such measures are necessary for the oversight and governance of AI tools. Without some level of transparency as to the processes and outcomes of the AI tools they are deploying, firms cannot assess the value of the benefit provided by those tools or the risks of error. Thus, the kind of transparency being discussed here may be seen as a pragmatic risk management tool by firms to reduce their potential exposure to litigation arising from misleading consumer users, or regulatory enforcement related to this conduct. Regulators should then be able to exercise enforcement powers to scrutinise these materials where there are doubts or concerns about the fair operations of a product or tool in the marketplace.

It is also reasonable to suggest that firms should not be left to self-regulate on these matters and that more direct regulatory intervention is required to ensure the transparency required for effective internal and external oversight of digital information intermediary tools offered

⁷⁵ See also discussion in Miller, *supra* note 65, at 30. Also *AI Explainability 360 – Resources*, IBM RESEARCH: TRUSTED AI (last visited April 2022), <https://aix360.mybluemix.net/resources#guidance>

⁷⁶ Cynthia Radin, *supra* note 48.

to consumers. This approach is consistent with the aims of consumer protection law, which are not only providing ex-post redress for harms that are occasioned by mis-conduct, but also ex-ante deterrence to prevent harmful conduct arising at all.

In the US, as in Australia, there is no specific AI law on the agenda that might demand transparency from firms in the manner seen in the EU. It is possible that existing consumer law, and the central prohibition on misleading conduct, might be leveraged to promote the kind of transparency in AI needed by regulators to protect consumer interests. This might be achieved by regulators taking a robust approach to the prohibition in its application to AI informed tools on the basis that, at least currently, consumers are likely to be unaware of how the relevant technologies function, and hence more vulnerable to their operation. This regulatory stance might reasonably include asking include firms to show audit trails for verifying the accuracy and neutrality of the recommendations being provided through digital information intermediary tools (category 1), verify their own claims about the capacity of the tools they offer (category 2), and to pre-emptively respond to the tendency of humans to anthropomorphise avatars with voices or faces, even where they do not overtly resemble humans (category 3).

V. CONCLUSION

There are potential benefits to consumers in the increasing availability of so-called AI informed, digital information intermediary tools. Such tools, in the form of recommender systems, comparison sites, apps, virtual voice assistants and chatbots may assist consumers in making more effective and welfare-enhancing decisions, especially in the face of multiple available products and voluminous amounts of complex, relevant information. But there are also undeniable risks in such tools. The tools may contain imbedded biases, misuse consumer data, and erode privacy. The tools further hold the capacity for misleading consumers, including about the scope of the assistance provided, the quality of the advice or recommendation, and who they are getting assistance from. The risk is amplified by the relative inexperience of most consumers in dealing with AI. Existing consumer protection law, especially prohibitions on misleading conduct, can in principle provide an effective response to such concerns. However, as we have seen, the

unique characteristics of digital information intermediary tools may make misleading conduct difficult to detect and establish in practice.

One initiative that may complement the work done by existing consumer protection regimes is AI ethics frameworks, which emphasise the importance of transparency in the design and use of AI, including through explanations of the process that produces the relevant output. Greater AI transparency may lead to better informed consumers, both about the technology being used and the decision-making support they are being offered. However, there are limits to the role of disclosure and explanations in informing the behaviour of already time poor and information overloaded consumers. Nonetheless the demands of transparency in digital information intermediary tools remain important. Initiatives in transparent and explainable AI are key to allowing regulators more effectively to monitor the performance of AI informed tools in consumer markets, and respond to transgressions of consumer protection law.

On their own, these measures are unlikely to be sufficient in providing adequate consumer protection for digital information intermediary tools, or indeed AI informed consumer products generally. They should be seen as one element in a combination of initiatives aimed at better consumer protection in the face of AI informed consumer products, working in conjunction with initiatives in data protection and measures for improving the fairness and safety of such products. Nonetheless, while transparency is only one part of the consumer protection toolkit, it is a necessary precondition for the greater accountability of firms deploying AI, including AI informed tools for assisting and augmenting consumer decision-making.⁷⁷

⁷⁷ Tim Miller, *But Why? Understanding Explainable Artificial Intelligence*, 25(3) XRDS: CROSSROADS, THE ACM MAGAZINE FOR STUDENTS 20, 25.