Taking the Person Seriously

Ethically Aware IS Research in the Era of Reinforcement Learning-based Personalization

Greene, Travis; Shmueli, Galit; Ray, Soumya

Document Version
Final published version

Published in:
Journal of the Association for Information Systems

DOI:
10.17705/1jais.00800

Publication date:
2023

License
Unspecified

Citation for published version (APA):
Greene, T., Shmueli, G., & Ray, S. (2023). Taking the Person Seriously: Ethically Aware IS Research in the Era of Reinforcement Learning-based Personalization. *Journal of the Association for Information Systems*, *24*(6), 1527-1561. Article 6. https://doi.org/10.17705/1jais.00800

Journal of the Association for Information Systems

Volume 24
Issue 6 *Special Issue: Data Science for Social Good (pp. 1439-1458, 1479-1499) Special Issue: Envisioning Digital Transformation: Advancing Theoretical Diversity (pp. 1459-1478, 1594-1798)*

Article 6

2023

Taking the Person Seriously: Ethically Aware IS Research in the Era of Reinforcement Learning-Based Personalization

Travis Greene
Copenhagen Business School, trgr.digi@cbs.dk

Galit Shmueli
National Tsing Hua University, galit.shmueli@iss.nthu.edu.tw

Soumya Ray
National Tsing Hua University, soumya.ray@iss.nthu.edu.tw

Recommended Citation

Greene, Travis; Shmueli, Galit; and Ray, Soumya (2023) "Taking the Person Seriously: Ethically Aware IS Research in the Era of Reinforcement Learning-Based Personalization," *Journal of the Association for Information Systems*, 24(6), 1527-1561.
DOI: 10.17705/1jais.00800
Available at: https://aisel.aisnet.org/jais/vol24/iss6/6

Journal of the Association for Information Systems (2023) **24**(6), 1527-1561
doi: 10.17705/1jais.00800

RESEARCH ARTICLE

ISSN 1536-9323

Taking the Person Seriously: Ethically Aware IS Research in the Era of Reinforcement Learning-Based Personalization

Travis Greene,[1] Galit Shmueli,[2] Soumya Ray[3]

[1]Department of Digitalization, Copenhagen Business School, Denmark, trgr.digi@cbs.dk
[2] Institute of Service Science, National Tsing Hua University, Taiwan, galit.shmueli@iss.nthu.edu.tw
[3]Institute of Service Science, National Tsing Hua University, Taiwan, soumya.ray@iss.nthu.edu.tw

Abstract

Advances in reinforcement learning and implicit data collection on large-scale commercial platforms mark the beginning of a new era of personalization aimed at the adaptive control of human user environments. We present five emergent features of this new paradigm of personalization that endanger persons and societies at scale and analyze their potential to reduce personal autonomy, destabilize social and political systems, and facilitate mass surveillance and social control, among other concerns. We argue that current data protection laws, most notably the European Union's General Data Protection Regulation, are limited in their ability to adequately address many of these issues. Nevertheless, we believe that IS researchers are well-situated to engage with and investigate this new era of personalization. We propose three distinct directions for ethically aware reinforcement learning-based personalization research uniquely suited to the strengths of IS researchers across the sociotechnical spectrum.

Keywords: Personalization, Reinforcement Learning, Sociotechnical, Data Protection, AI Ethics, Digital Platforms

Roger Chiang, Ahmed Abbasi, and Jennifer Xu were the accepting senior editors. This research article was submitted on April 14, 2021 and underwent three revisions. This paper is part of the Special Issue on Data Science for Social Good.

1 Introduction

Machine learning-based personalization shapes our digital experiences and social interactions in ethically important yet often imperceptible ways (Milano et al., 2020; Pariser, 2011). For consumers, personalization promises better preference matching, convenience, and reduced cognitive load (Aguirre et al., 2015). For businesses, personalization is a source of competitive advantage and a technological means to greater customer satisfaction, loyalty, and profits (Murthi & Sarkar, 2003). However, despite their outsized ethical, social, and political impacts (Floridi et al., 2018), platforms and the personalization algorithms used on them currently face relatively few regulatory and ethical constraints (Bak-Coleman et al., 2021).

These regulatory and ethical concerns have been amplified by the recent adoption of reinforcement learning-based algorithms on platforms. Reinforcement learning (Sutton & Barto, 2018) has played a pivotal role in advances in robotics, self-driving cars, video and board games, medicine and clinical decision-making, resource allocation and management, and interactive language agents such as the highly publicized ChatGPT (OpenAI, 2022). Indeed, eminent AI researchers have declared that reinforcement learning is the machine learning paradigm most likely to achieve artificial general intelligence (Silver et al., 2021). However, the narrow instrumental rationality of artificial agents worries some philosophers (Bostrom, 2012), who claim these agents have incentives to interact with their human counterparts in "pathological" ways involving manipulation and deception, the modification of users' beliefs, and even the

fostering of addiction (Burr et al., 2018; Kenton et al., 2021; Krueger et al., 2020; Russell, 2019).

Nevertheless, reinforcement learning is increasingly being used by commercial platforms to implement *autonomous experimentation systems* (Bird et al., 2016). These systems, particularly when deployed on social media platforms, are sociotechnical systems *par excellence*, as they substitute previously human actions—e.g., scientific experimentation, as in A/B testing—with an algorithmic mechanism, and thereby intervene in human social relations (Ropohl, 1999; van Dijck, 2013). While other fields may focus on the algorithmic subtleties of new personalization technologies, the information systems (IS) field, with its holistic blend of humanistic and technological perspectives, is uniquely qualified to evaluate the scientific, social, and technical aspects of such systems. For instance, autonomous experimentation systems on platforms have implications for the generalizability of empirical research, as they may unknowingly interfere with field experiments on platforms (Greene et al., 2022).

Although recent data protection laws are designed to protect persons from the undue influence of new algorithmic technologies, several ethically relevant technical aspects of reinforcement learning-based personalization remain unaddressed. These include its unprecedented speed, degree of interactivity, and automation based on large-scale, unspecifiable, and unpredictable personal data collection and processing, as well as its potential for nonconsensual human behavior modification and other pathological behaviors driven by engagement-based reward optimization. Collectively, these emergent features conflict with the normative foundations of a free and open democratic society and diminish our autonomy as persons by exploiting our evolutionary instincts, cognitive biases, and psychological vulnerabilities to manipulation and persuasion. At the individual level, this could result in fostering addiction and other psychological harms; at the social and political levels, this could entail the viral spread of misinformation and emotionally polarizing platform content.

To grapple with these issues, we propose three research directions designed to appeal to the IS research community, given its expertise in integrating data science, empirical research, social theory, and business and technology ethics. Each direction represents a research pathway aimed at supporting personal autonomy. The first direction is technological in nature and proposes the use of simulators to reduce the need for personal data collection and live interactions with human users while studying the properties and effects of reinforcement learning algorithms. The second is sociotechnical and addresses limitations in conducting platform-based field experiments and causal inference in the era of reinforcement learning-based personalization. The third is ethics oriented and involves the need for *critical IS* and other like-minded research communities to investigate how reinforcement learning personalization can be safely harnessed to promote personal autonomy, human flourishing, and social and political stability.

The paper is structured as follows. To better understand the ethical implications of reinforcement learning-based personalization on platforms, Section 2 describes the evolution and expansion of personalization technology, leading up to a new era of personalization aiming at the adaptive control of human user environments. Section 3 presents five emergent features of this new paradigm that introduce novel ethical, social, and political challenges that go above and beyond earlier paradigms of personalization. To address these challenges, Section 4 proposes three methodologically diverse research directions aimed at protecting our personal autonomy and the stability of social and political systems. Section 5 then contrasts the ethical landscape of reinforcement learning-based personalization on commercial platforms with the noncommercial domains of medicine, health, education, and social and public policy. Section 6 concludes the paper.

2 Losing the Person in Personalization: The Path to Deep Reinforcement Learning

Personalization research, if it mentions the concept of the *person* at all, typically views the person through the lens of orthodox economic theory. The goal of personalization is thus primarily hedonic, consequentialist, and utilitarian: the efficient satisfaction of a person's presumed stable and well-defined needs, tastes, and preferences. Little attention is paid, for instance, to how individuals' preferences can be influenced by personalization technologies toward platforms' economic interests (Hildebrandt, 2022). We take a different approach. Our guiding notion of the person stresses the capacity for conscious self-determination, free will, and autonomy from which persons derive their dignity (Floridi, 2011; Kant, 1785/1948). Moreover, we hold that both flourishing persons and thriving democracies require an inviolable private sphere from which they can freely develop their unique personalities and capacities safe from the control of others[1] (Mill, 2015). The exercise of personal autonomy and control is an intrinsically valuable and universal feature of psychological well-being (Helwig, 2006; Leotti et al., 2010; Ryan & Deci, 2017), and forms the ethical bedrock of seminal data

[1] Similar concerns about protecting individual liberties against the intrusions of government and tyranny of the majority underlay the inclusion of the Bill of Rights in the United States Constitution (Bodenhamer & Ely, 2008).

protection laws such as the European Union's (EU) General Data Protection Regulation (GDPR) (Floridi, 2016), which extends new rights protecting persons from increasingly invasive data collection practices by corporations and governments.

Personalization today is a catch-all term for what is variously called customization, behavioral profiling, algorithmic selection, computational advertising, or actuarial prediction. Traditionally, personalization is understood as an iterative process composed of learning, matching or recommendation, and measurement components (Adomavicius & Tuzhilin, 2005; Churchill, 2013; Murthi & Sarkar, 2003). We contend a paradigm shift is taking place in personalization as the person in personalization is quietly removed from the picture. The original democratic ethos of platform-based personalization (van Dijck, 2013) has been co-opted by engineering formalism, behaviorist philosophy of science, and economic orthodoxy. The resulting paradigm has potentially grave consequences for our autonomy as persons.

In order to highlight the novel technological and ethical aspects of reinforcement learning-based personalization on platforms, we contrast it with earlier paradigms of personalization. We start by briefly describing personalization prior to the introduction of reinforcement learning (Section 2.1). Then, we present an overview of reinforcement learning-based personalization (Section 2.2) and its most recent incarnation: deep reinforcement learning (Section 2.3). Figure 1 provides a schematic timeline of the evolution and related technological and conceptual paradigm shifts of platform-based personalization.

2.1 A Brief History of Personalization Prior to Reinforcement Learning

Personalization initially began as a top-down, static, knowledge-based technology, but has gradually been transformed into a bottom-up, dynamic, and machine learning-based technology whose quest for ever-greater scalability focuses on removing human biases from its design (see e.g., Sutton, 2019). Evolving out of research on recommender systems and the adaptive web (Brusilovsky & Maybury, 2002), a sea change occurred when technology corporations began embracing advertising-centric business models and expanding and networking digital platforms to harness their potential for the economic circulation of implicit behavioral data (Langley & Leyshon, 2017; Zuboff, 2019). The introduction of *persuasive technology* (Eyal, 2014; Fogg, 2003) to the platform concept marked a second major shift in personalization. Persuasive technology aims at changing users' attitudes and behavior by merging psychological behavior modification techniques with platform technology, thus allowing platforms to tailor persuasive strategies to individual user features (Berkovsky et al., 2012). Finally, machine learning techniques have also evolved in sophistication, from the once relatively static recommendation algorithms (e.g., collaborative filtering) trained using explicit user feedback data (e.g., ratings, surveys, or written text) to more adaptive algorithms trained using greater amounts of engagement-driven, implicit user feedback data in the form of clicks, scrolls, and dwell time[2] (Aggarwal, 2016; Ekstrand & Willemsen, 2016). Moreover, advances in neural networks (Hinton & Salakhutdinov, 2006) and platform data infrastructure (Helmond, 2015) spurred the collection and analysis of massive volumes of unstructured user-generated text, images, and videos without human-labeled features (Krizhevsky et al., 2012).

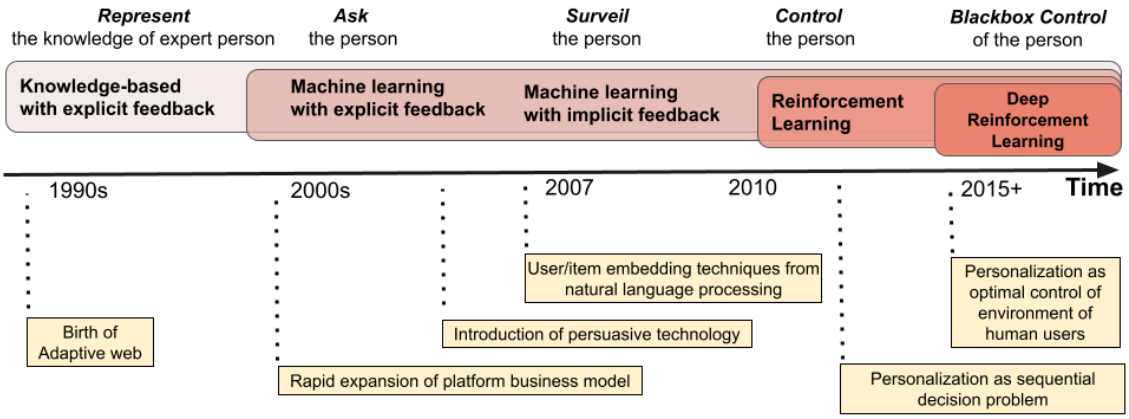

Figure 1. Schematic and Thematic Timeline of Notable Paradigms of Personalization, with Key Change Points in Algorithms and Approaches to Data Collection

[2] Dwell time is defined as time spent on a clicked search result and is considered one of the most important implicit measures of web search quality and user satisfaction. For instance, a "satisfied click" is defined as a dwell time of greater than 30 seconds (Kim et al., 2014).

Similar to the *kluge*-like evolution of the human brain (Marcus, 2009), each advance in platform-based personalization technology rests on top of earlier methods, allowing for sophisticated hybrid systems that repurpose the functionality of earlier methods. As shown in Figure 1, personalization originally involved explicitly representing the knowledge of expert persons ("Represent experts' knowledge"), moving to explicitly asking people what they liked ("Ask the person"), then to using platform infrastructure to collect implicit behavioral data ("Surveil the user"), and finally to adaptively controlling human platform environments to achieve business goals ("Control the person" and "Blackbox control the person").

2.2 Reinforcement Learning-Based Personalization on Platforms

Reinforcement learning-based personalization is an emergent cybernetic [3] paradigm of personalization (Cristianini et al., 2021) that replaces traditional notions of the autonomous person with the more general and scalable engineering formalism of *controller* and *controlled system*. Passive recommender systems are now transformed into active and adaptive artificial agents interacting with—controlling—environments of human platform users—the controlled system. By combining research in animal learning and behaviorist psychology with statistics, neuroscience, robotics, and feedback control theory, personalization has become the problem of learning an optimal recommendation or *control policy*—matching content to users—while maximizing a reward metric over a finite horizon of interactions.

The new era of personalization is rooted in a mélange of algorithms, methods, and problem formulations. Cai et al. (2017) describe *reinforcement learning* as a "family of methods for addressing sequential decision-making problems characterized by non-deterministic dynamics, delayed decision-outcome pairings, and a lack of ground truth regarding optimal decisions." Depending on the complexity and assumptions of the problem, such as whether the reward is delayed or environment states are fully or only partially observable, a *multi-armed bandit*, *contextual bandit*, or more generally, a *Markov decision process* (MDP), can be used to represent the task of reinforcement learning-based personalization. Once formalized, the agent's goal is to learn how to behave in its human environment (see Appendix A). Seen from an optimization or engineering perspective, the agent is a "solution" to the "problem" posed by the environment of human users (Silver et al., 2021).

A variety of algorithms exist for learning an optimal policy (see Appendix B for key terms). For now, we point out two technical distinctions that will be revisited in Section 3. First, some reinforcement learning algorithms require live interactions with the environment in order to find an optimal policy (known as *online learning*), while others only need stored interactions (known as *offline learning*) (den Hengst et al., 2020). Secondly, some algorithms rely on a prespecified model of environment dynamics (*model-based*), while others treat the environment as a blackbox and learn the optimal policy or value function by directly interacting with it (*model-free*) (Ibarz et al., 2021). Due to the complexity of typical platform environments, model-free methods are commonly used. While more general, scalable, and less prone to human bias, they require much more implicit interaction data from users.

Facebook, Instagram, Spotify, Amazon, Pinterest, and TikTok (Zhou et al., 2020; Zou et al., 2019) employ model-free reinforcement learning because it offers a general and scalable framework for dealing with new users, items, features, and users' nonstationary preferences (McInerney et al., 2018). Interactions between an agent and environment are steered by evaluative feedback in the form of rewards and punishments, not by explicit ground truth labels as in supervised learning (Littman, 2015). Autonomous and adaptive agents on platforms can interact with users and learn to determine the content, timing, and delivery method of information to platform users. In doing so, the artificial agent's goal is to find an optimal interaction policy as it continually selects actions to explore or exploit its human user environment on the basis of past observations. This trade-off is known as the *exploration-exploitation dilemma* (Sutton & Barto, 2018). Over time, increasingly personalized ads, notifications, interventions, or recommendations can be tailored to their (and other similar platform users') behavioral interaction sequences, demographic characteristics, and item features in order to maximize the reward accumulated by the artificial agent.

2.3 Fusing Deep Learning with Deep Reinforcement Learning

Deep neural networks (LeCun et al., 2015) provide a major advance in personalization, as they can automatically discover latent structure in multimodal platform data, such as user-generated images, text,

[3] The English "cybernetics" stems from the Greek *kybernetes*, meaning "pilot" or "steersman" of a ship, from which "governor" and "government" are also derived, thus connoting the act of lawlike control (Wiener, 1988, p. 15). Such control conflicts with the Kantian notion of the *autonomy* of persons, which literally means "self-legislating" or "self-lawgiving" and refers to the reflexive ability of persons to choose to bind themselves (i.e., not to be externally steered) to universal laws of moral action valid for all rational beings (O'Neill, 2002).

videos, and behavioral data (Covington et al., 2016). Yet these data-hungry "blackbox" (Rudin, 2019) methods of personalization have the potential to not only further erode our capacity for self-determination via the strategic sequences of algorithmic interventions operating beneath our conscious awareness but they can do so in opaque, unpredictable, and potentially discriminatory ways that violate democratic values of transparency and accountability (Citron, 2007). Even when an artificial agent's goals are known and set by algorithm designers, it may be impossible to predict what specific sequence of actions a smarter-than-human system will take to achieve its objectives (Yampolskiy, 2020).

The power of deep nets derives from their ability to automatically derive complex, nonlinear hierarchical features from raw input data, a process known as *representation learning* (Bengio et al., 2013). When augmented with unsupervised and self-supervised learning techniques, deep networks can be trained even faster by encoding raw input data into a more informationally compressed form (Schmidhuber, 2015). Highly scalable, self-supervision works to pretrain networks by learning to predict unobserved or hidden parts of the input data from observed parts of the input data (Zbontar et al., 2021). Consequently, self-supervision is well-suited to harness the masses of unstructured data collected by platforms and thereby reduces the need for human data scientist oversight and labor.

Another direction in which deep learning has influenced personalization is via sequential neural net architectures that treat personalization as a complex sequential prediction task. Mirroring the language modeling problem of predicting the next word given a sequence of previous words, sequential personalization methods can predict the next item clicked or purchased given a user's past interactions (Xin et al., 2020). Inspired by the analogy with machine translation, new attention-based *transformer* (Vaswani et al., 2017) architectures are now being used for personalization. These attention-based systems work by generating the next item in a long sequence, essentially turning personalization into a very complex supervised learning problem. Put simply, attention mechanisms improve personalization by automatically selecting the most representative items for a given user and removing unwanted "noisy" elements in a user's interaction history (Zhang et al., 2017).

Arguably, however, the most consequential advance in personalization comes from *deep reinforcement learning* (Mnih et al., 2015). Deep reinforcement learning combines deep neural net architectures with the reinforcement learning formalism of an environment and artificial agent that perceives states, selects actions, and receives rewards. Deep nets take very complex, nonlinear, and high-dimensional state representations as input and output an action selection probability to guide the agent in controlling its environment. What is new, however, is that deep reinforcement learning ties the learning of complex, temporally extended representations of states and actions to their functional ability to promote the accumulation of reward. In other words, deep reinforcement learning automatically discovers task-relevant, goal-directed abstractions by embedding them in a space where inputs serving similar functions are mapped closely together. To illustrate this concept, consider the task of going to work. Both walking and riding a bike—while different actions—are functionally similar ways of going to work and hence should be embedded near one another.

Fusing deep learning with reinforcement learning can synergistically produce never-seen-before computational phenomena (Botvinick et al., 2020). Take OpenAI's recently released ChatGPT, for example. Interactive language agents such as ChatGPT can be constructed by combining deep learning-based supervised learning with reinforcement learning in a framework known as *reinforcement learning from human feedback* (Christiano et al., 2017). Briefly, the framework proceeds in three steps (Ouyang et al., 2022). First, a pretrained language model is fine-tuned using supervised learning from a dataset of text prompts and desired output text written by teams of trained humans. Then, another deep neural network learns an associated reward model using a comparison dataset of human annotators' rankings of generated text quality. Finally, example text is generated from prompts using the fine-tuned language model and scored with the reward model. These predicted rewards act as the reward signal used to update the parameters of the language model. Using this general reinforcement procedure, language models can generate high-quality text with (or without) complex, abstract properties such as truthfulness, creativity, or dangerousness that cannot easily be specified in traditional reward or loss functions.

Returning to the implications of deep reinforcement learning for personalization, framing personalization as a sequential decision-making problem makes sequential neural network architectures that capture temporal dynamics of user behavior very useful. Numerous variations of sequential architectures exist and can be used to generate complex user-item state embeddings capturing *long-term* dependencies relevant to the personalized prediction and control of platform users (Zou et al., 2019). As one example, Chen et al. (2019) used a recurrent architecture to create a personalized slate of YouTube videos from among millions, given a user interaction history and user and item features. Newer, more advanced architectures further extend the sequential processing, memory, and state-tracking capabilities relevant to personalization. Yet the complex sequential

structures discovered may have no human-understandable analog, making it difficult to give intelligible explanations for an artificial agent's decisions. Moreover, deep neural nets may also harbor latent discriminatory or spurious associations learned from undetected biases in the data collection process. These aspects worry legal scholars who view blackbox technologies as eroding the basic norms of constitutional democracies, such as the rule of law, due process, and equal protection, which in part ensure against unfair treatment and discrimination (Pasquale, 2015).

Ultimately, the potential of combining behavioral data from different platform sources of unstructured text, audio, and video, including on-ground sensors (IoT) and wearables, to create highly complex, nonlinear, and abstract representations useful for long-term prediction and adaptive control marks the emergence of a new cybernetic paradigm of personalization. When behavior-based feedback control is implemented on platform ecosystems with billions of networked human users, the result is a highly profitable—but opaque—sociotechnical system capable of influencing persons and societies at scale. This automated, adaptive, and sequential approach to personalization threatens users' personal autonomy, not only by the increased emphasis on collecting implicit behavioral feedback but also by the way in which artificial agents can evolve complex interactive behaviors to strategically modify the attitudes and behaviors of platform users in ways conducive to platforms' financial goals.

As illustrated in the previous three sections, the introduction of reinforcement learning has transformed platform-based personalization. But what are the unique features of this technology that potentially threaten our autonomy as persons and the stability of our social and political systems? The next section discusses five emergent features of reinforcement learning-based personalization and their ethical, social, and political implications.

3 Five Emergent Features of Reinforcement Learning-Based Personalization

Philosophers and psychologists have argued that we require a protected sphere safe from the control of external influences in order to choose freely, develop into healthy and well-functioning persons, and serve as the authors of our own lives. Indeed, belief in the inherent value of personal autonomy is embedded in the concept of constitutional democracy itself, which aims to protect citizens' basic rights and freedoms from the ever-present dangers of coercive power (Winick, 1992). Yet our autonomy and the stability of our societies appear increasingly threatened by a new cybernetic paradigm of reinforcement learning-based personalization of unprecedented speed, automation, and interactivity. Moreover, new training and testing methods may incentivize platforms to hoard users' implicit behavioral data in order to develop newer and more effective control strategies or collect and process personal data in unforeseeable ways for unspecified future control purposes. In addition, this new engineering-centric paradigm raises ethical issues related to human behavior modification via persuasive technology and pathological agent behavior due to optimizing engagement metrics.

Below we describe five emergent features of reinforcement learning-based personalization on platforms that pose novel challenges to personal autonomy and social and political stability. These include its unprecedented speed and degree of interactivity and automation (Section 3.1) based on large-scale, unspecifiable, and unpredictable personal data collection and processing (Sections 3.2 and 3.3), and its potential for nonconsensual human behavior modification (Section 3.4) and other pathological behavior driven by engagement-based reward optimization (Section 3.5). At the end of each subsection, we consider the GDPR's key principles, ambiguities, and limitations in addressing the dangers associated with each emergent feature. In short, while the rights and principles enshrined in the EU's GDPR—arguably the most influential of modern data protection laws—may help mitigate some of these issues, exceptions, and ambiguities in the law make it an incomplete regulatory solution, as we summarize in Table 1 and describe in the following subsections.

3.1 Unprecedented Interactivity, Speed, and Automated Adaptivity

Reinforcement learning is a natural choice for personalization due to its inherent adaptivity. In a process resembling evolution by natural selection, an artificial agent's behavior can evolve over time to better fit its environment. Reinforcement learning agents can randomly vary their behavioral strategies in response to differences in reinforcement (reward or punishment) received during interactive feedback with human user environments. Depending on how users react to the agent's actions, the agent can modify its strategy—i.e., *learn*—for the next interaction. This adaptability allows artificial agents to evolve anticipatory behaviors and sophisticated interaction patterns to fit specific users and optimize for a platform's goals. As with any evolutionary process involving randomness, however, it may be impossible to predict which behaviors the artificial agent will acquire. Besides making external platform-based research increasingly difficult (Greene et al., 2022), automated adaptability opens the door to potentially dangerous and unethical means of achieving the ends of environmental control.

Table 1. Five Emergent Features of Reinforcement Learning-Based Personalization, Their Respective Potential Dangers to Persons and Society, and the Limitations of the GDPR

Emergent feature of personalization	Dangers to persons and society	Ambiguities and limitations of the GDPR
Unprecedented interactivity, speed, and automated adaptivity	• *Political and social instability:* Networked agents and adversarial bots on platforms may quickly propagate misinformation and sow social, financial, and political discord • *Addiction:* Agents may associate more rewards with vulnerable user populations leading to addiction • *Exploitation:* Agents can take advantage of users' cognitive biases, inattention, and lack of self-control	• No public disclosure of DPIA results • Technical information may not be intelligible to ordinary users • "Human-in-the-loop" review may not be possible with online adaptive algorithms
Excessive volumes of interactive behavioral training data	• *Mass surveillance:* Online and newer offline learning methods create opportunities for novel forms of surveillance and social control • *Privacy invasions:* Complex networked platform environments require more interactive data • *Reduced accountability, transparency, and trust:* Platforms have incentives to hoard behavioral data for offline learning and improved user interaction strategies, including new and cheaper means of data collection (e.g., nudging users to label data)	• Unclear if user interactions with platforms constitute personal data • Small corporations exempt from record-keeping rules around data collection
Unspecifiable data collection and processing	• *Unfairness and safety:* Balancing exploitation with exploration requires random state-action selection with unforeseen consequences on "unlucky" social groups and individuals • *Discrimination:* Blackbox algorithms can discriminate on the basis of morally and legally objectionable correlations discovered in unstructured data	• Unpredictable agent behavior makes regulatory oversight difficult • Platforms can use other legal grounds besides explicit consent for data collection and processing (e.g., "contractual necessity" and "legitimate interest")
Nonconsensual human behavior modification	• *Diminished personal autonomy:* Adaptive algorithms can be combined with psychological principles to modify user behavior without explicit consent or reflective endorsement • *Reduced political legitimacy and erosion of trust:* Blackbox algorithmic nudging undermines the normative foundations of democratic authority	• Member states can create exemptions for automated profiling • Platforms do not need to notify users of data processing if "too burdensome"
Pathological agent behavior via business metric optimization	• *Deception and manipulation:* Agents have incentives to modify their human environments to better achieve long-term reward maximization • *Value misalignment:* Agents can exhibit pathological behaviors that endanger personal and social well-being	• Data processing can still occur if consented to or in the "legitimate interests" of the data controller • Unclear how to balance business optimization goals with human rights and freedoms in a principled way

For example, consider Facebook's News Feed system (Gauci et al., 2019). A sequential human-machine interaction drives the learning and adaptation process of News Feed, forming a cause-effect feedback loop. News Feed initially begins with generic, hard-coded rules—analogous to innate animal reflexes—for displaying and ranking content on a user's newsfeed. After collecting implicit feedback on how users collectively interact with a particular display combination, the algorithm's parameters are updated. During the next interaction with a particular user, News Feed will display a personalized set of items associated with improved clickthrough or dwell time as inferred from its large-scale understanding of user behavior. Alternatively, for some small subset of users, News Feed can randomly select other "suboptimal" items to display in order to gather more information about a user's underlying preferences. These options illustrate the explore-exploit trade-off faced in online learning. In the personalization context, exploitation refers to recommending content currently estimated to lead to the highest reward, while exploration involves recommending content for the purpose of reducing the uncertainty regarding an action's long-term reward potential. In short, News Feed shows what it predicts will impact users' behavior (what they read or click on), which in turn impacts its future predictions that drive its actions, ad infinitum.

The speed and automated adaptability of reinforcement learning algorithms offer exciting new applications for computational advertising, but also may undermine a person's autonomy by encouraging behavior a user would otherwise not *reflectively endorse.* [4] For example, the iterative generation of real-time images and videos to varying audiences is a nascent but important aspect of online advertising that benefits from the ability of reinforcement learning agents to learn to dynamically match users to ads for maximal rewards (Liu & Chao, 2020). An artificial agent on a platform might dynamically modify the logo, content, font, and call to action for particular users after observing their behavior when interacting with different configurations. In one real-world application, an advertisement featuring a female model's face—her eye gaze, in particular—can be synthetically adjusted in real time to influence where users look and click, thereby leading to higher conversion rates (Liu & Chao, 2020). Yet the data generated by these nonconscious, implicit feedback behaviors are typically not the result of conscious, deliberative processes and thus may reduce users' autonomy (Prunkl, 2022).

As the above example illustrates, reinforcement learning algorithms can exploit regularities—both "features" and "bugs"—in our evolutionary instincts, cognitive biases, perceptual faculties, and propensities for addiction to induce target behaviors that maximize its reward signal, all without explicit instruction. Take TikTok's adaptive recommender algorithm, which relies on a host of implicit feedback cues designed to exploit our "tendencies toward boredom" and our "sensitivity to cultural cues" to maximize time spent and user retention on the platform (Smith, 2021). The young, the elderly, and the distracted (Pennycook & Rand, 2019) may become "motivational magnets" (Tindell et al., 2009) for artificial agents capable of learning complex reward cues related to engagement metrics and adjusting their behavior in response to behavioral regularities in individual persons and user subpopulations. Su et al. (2021), for instance, conducted a neuroimaging study and found that personalized TikTok video recommendations led to higher activation of the brain's default mode network compared to nonpersonalized recommendations in young adults. TikTok users with lower levels of self-control are particularly susceptible to addictive video consumption on that platform, according to the study. An instrumentally rational agent (Bostrom, 2012) with the goal of maximizing engagement could learn this association by trial and error and leverage it by, for example, selecting actions correlated with inducing "ego depletion" effects in users (Baumeister et al., 1998), or by recommending mood-altering content to users determined to have low levels of self-regulatory control (Ryan et al., 2014). Scaled across billions of user interactions, these subtle effects may have tangibly negative impacts on democratic social and political systems.

Besides the possibility of subconscious manipulation and exploitation of the cognitively vulnerable, one real-world consequence of speed, automation, and adaptivity is the potential to quickly propagate misinformation across platform networks. Owing to their ability to discover latent sequential structure hidden in implicit behavioral data, autonomous agents could learn to exploit subtle correlations between user inattentiveness and public health misinformation when platform sharing behavior is rewarded. An algorithm's learning rate—i.e., the speed at which autonomous agents adapt their behavior policy in response to observed reward—may also be problematic. The algorithm's speed in adapting the form, content, and delivery method of information may limit users' ability to exercise and develop important reflective

[4] Reflective endorsement is the ability to change one's beliefs, desires, or actions in response to reasoned argument and conscious deliberation. This voluntary, higher-order cognitive capacity is considered by many philosophers to be a uniquely human ability separating persons from "mere" animals (Kornblith, 2010).

metacognitive abilities, such as scrutinizing the reliability of information sources (Rollwage & Fleming, 2021), thus reducing their autonomy as persons and leaving them susceptible to political polarization and dogmatism. Indeed, Avram et al. (2020) conducted an experiment on the viral spread of misinformation on social media feeds and concluded that adding "intermediary pauses" limiting "automated or high-speed sharing" can dampen the spread of misinformation.

Another issue concerns the potential for adversaries to manipulate the algorithm's input data by deploying social bots or "spambots" to give thousands of fake "likes" or followers in order to change the behavior of the system (Badri et al., 2016), or encourage offline political effects such as "sowing discord" among citizens (Dutt et al., 2018). Also, because platforms themselves can be connected into platform ecosystems (van Dijck, 2013) the speed and complexity of *networks* of autonomously adaptive decision-making agents raise issues of volatility and the potential for "flash crashes" seen in high-frequency trading (Karppi & Crawford, 2015). Trading algorithms often rely on sentiment analysis performed on social media platforms, thereby coupling their behavior with global financial markets (Sharma, 2020). Alarmingly, viral misinformation on social media platforms can affect financial markets almost immediately.

The adaptability of reinforcement learning agents leads to uncertainty surrounding their cognitive effects on individuals, society, and even future generations. For that reason, the GDPR introduces Data Protection Impact Assessments (DPIAs), modeled on impact assessments made in environmental protection law (Costa, 2012). Crucially, however, DPIAs do not require public disclosure of assessment results and the information they contain may not be intelligible to those arguably most vulnerable to deception or manipulation. Second, although the GDPR offers data subjects rights to contest algorithmic decisions, the actions of blackbox deep reinforcement learning systems may not be explainable or intelligible to experts or ordinary users. They may also be based on discriminatory associations learned from undetected biases in automated data collection processes. Finally, in earlier, more static paradigms of personalization, the GDPR's right to human intervention (i.e., putting a "human in the loop"), might have been feasible. But the nature of online learning means that putting a human in the loop may be impossible. Only post hoc analyses and explanations of decisions may be possible after harms have transpired. Section 4 thus discusses ethics-based simulators and experimental sandboxes as future research avenues that limit the need for live online interactions with masses of human users.

3.2 Excessive Volumes of Interactive, Implicit Behavioral Training Data

Earlier supervised approaches to personalization focused on exploiting short-term gains in accuracy or ranking-based metrics but did not consider how these changes may affect other business and platform-related performance metrics in the long term (e.g., user churn or daily active users). Reinforcement learning agents, however, particularly when implemented via recurrent architectures or attention-based transformers, can better capture long-term dependencies between past state-action pairs and later engagement-based rewards. Doing so requires vast quantities of behavioral data compared to earlier paradigms. One reason is that reinforcement learning agents receive correlated—as opposed to independent—samples of experience (Mnih et al., 2015). Another reason is that feedback about the quality of vast combinations of state-action pairs is often sparse, so most of the information the agent receives relates to what does not work (Hutsebaut-Buysse et al., 2022). Finally, platform environments with millions of users interacting are extremely complex, which drastically increases the "sample complexity" of the learning task (Gu et al., 2016). Indeed, state-of-the-art transformer-based architectures can have hundreds of billions of parameters and require training on "internet-sized" datasets (Borgeaud et al., 2021). There is thus worry that in the quest for scalability, data-hungry reinforcement learning methods could exacerbate opportunities for mass surveillance and social control by corporations or governments (Whittlestone et al., 2021). By treating personalization as a sequential prediction problem and taking advantage of the analogy between personalization and language translation, major social media platforms could leverage their global user bases to develop massive databases of "behavioral tokens" in order to predict and control user behavior across a variety of geographical, cultural, and temporal contexts.

In Section 2.2, we mentioned that reinforcement learning can be done online or offline. Online learning requires maximizing reward while interacting with human users. Offline learning does not require interacting with human users and can be done using a fixed dataset of experiences collected by a preexisting interaction policy and stored in a replay buffer. Offline learning methods are popular for safety, user experience, and data efficiency reasons. But sparser, delayed reward and more variable environments require platforms to collect more interaction data to learn safe and robust policies. Consequently, the ability to repurpose earlier experiences via replay buffers creates incentives for platforms to hoard implicit feedback data in as many contexts as possible in order to promote learning sub-behaviors or "skills" useful for future downstream tasks.

Locked in a competitive struggle to discover profitable new personalization opportunities, corporate platform ecosystems and governments with unified behavioral collection and processing architectures are likely poised to benefit the most from offline learning. For example, relying on large user bases and relatively lax regulation, major corporations with platform ecosystems—such as Baidu, Alibaba, Tencent, and Bytedance in the Chinese market—have invested immense sums of money in building powerful data collection systems to better sense their human user environments and facilitate enhanced targeted intervention and control possibilities. Tencent's messaging app WeChat alone allows for the collection and analysis of behavioral data on over 800M daily users (Jia et al., 2018; Tang, 2017). With offline learning, platforms can use the collected interaction data to update existing agents' policies and promote the learning of new skills to help the agent—or other agents on the platform—optimize future data collection policies (Riedmiller et al., 2021) or accumulate greater future reward in increasingly complex and unpredictable ways.

In light of the growing amount of personal data collected, stored, and processed by corporations and governments, European lawmakers have incorporated basic principles of data minimization, data storage limitations, and privacy by design into the GDPR (Greene et al., 2019). In the European view, data protection and privacy are legal tools aimed at preserving human dignity (Floridi, 2016), ensuring a stable and vibrant democracy (Rouvroy & Poullet, 2009), and reducing the power asymmetries generated by personal data processing (Lynskey, 2015). Mirroring philosophical points made by both Kant and Mill, these normative ethical and political goals require that citizens possess a sufficient degree of self-determination and freedom from external control (Hildebrandt & de Vries, 2013). Indeed, privacy in the European context has been defined as "the freedom from unreasonable constraints on the construction of one's own identity" (Hildebrandt, 2015, p. 80). The GDPR thus makes clear that the potential benefits and harms of large-scale data processing must be considered and, further, personal data should only be adequate, relevant, and limited to what is necessary for processing. We note, however, that there are conflicting fundamental rights to conduct business (Zuiderveen Borgesius, 2015) and it is not clear how to manage conflicts between the advancement of human autonomy and business innovation, particularly for businesses with strong incentives to modify human behavior in profitable ways[5] (Shmueli & Tafti, 2023; Susser et al., 2019).

Finally, there is a major ambiguity concerning the legal status of sequential interactions between individual users and the agent—do these constitute personal data? Although it might be possible to single out individual users by their unique interaction histories, it is not clear that the GDPR's definition of personal data applies to sequential interaction data used by reinforcement learning algorithms. Therefore, the rights of data subjects to remove, edit, and obtain copies of personal data, or even opt out of automated decision-making, may not apply.

3.3 Unspecifiable Data Collection and Processing

The adaptability of reinforcement learning algorithms requires a different approach to training and evaluation compared to supervised approaches to personalization. As noted, online reinforcement learning algorithms adaptively determine their own training data, while offline algorithms learn from previously collected data. Both learning methods, however, can generate unpredictable agent behavior, albeit in different ways. Offline learning, for instance, opens the door to *self-play* techniques (i.e., simulating counterfactual interactions using previously stored experiences or "memories"), whereby agents learn to control human platform environments in as many ways as possible (Levine, 2021). But this unpredictability and adaptability—while a boon to platforms with shifting user and content bases—make it difficult to specify in advance the purpose and nature of data collection and processing. Together, these emergent technological features can conflict with demands for regulatory oversight and transparency (Rahwan et al., 2019) and traditional informational norms of confidentiality and consent (Nissenbaum, 2004). Preexisting AI safety and governance mechanisms may therefore no longer be reliable or useful in this new era of platform-based reinforcement learning (Yampolskiy, 2020).

In environments undergoing constant change, effective online learning requires balancing exploration with exploitation of the environment. New users, new items, nonstationary preferences, and new item attributes make the agent's human environment on platforms especially dynamic and complex, thus making exploration—either performed randomly or guided by heuristics—increasingly important for platforms. One common exploration scheme is the "ε-greedy" approach, where the agent chooses the action with the currently highest action value (1- ε)% of the time and chooses random actions ε% of the time (Sutton & Barto, 2018). Artificial agents can also be incentivized to explore more of their environments by adding curiosity or exploration bonuses to reward signals,

[5] Due to exemptions such as the GDPR's Article 30, small corporations with fewer than 250 employees may be able to deploy such agents so long as the data they process does not pose a risk to the "rights and freedoms of data subjects" or involve "special categories of data" (Official Journal of the European Union, 2016).

corresponding to play-like behavior in animals and humans (Gottlieb et al., 2013). Yet another approach involves *optimism in the face of uncertainty:* among the actions whose values are being estimated, the one with the highest upper bound should be selected (Littman, 2015). Unsupervised pretraining (Liu & Abbeel, 2021), particularly when combined with regularization techniques, can also be used to encourage artificial agents to explore areas of the state space where it might be highly rewarded in later downstream reinforcement learning tasks or develop interesting new sub-behaviors to promote future reward accumulation. The essence of these approaches is that they produce—indeed encourage—probabilistic on-the-fly data collection decisions (i.e., stochastic behavior policies) that may not be specifiable in advance or intelligible in retrospect.

Random exploratory actions make it difficult to specify the learning trajectory of the agent in advance and may conflict with provisions of data protection laws designed to ensure transparent, accountable, and trustworthy algorithmic systems that increasingly impact people's lives in significant ways (Rahwan et al., 2019). Indeed, partly stemming from these concerns, and following the privacy rights found in the Charter of Fundamental Rights of the European Union, the GDPR emphasizes the need for *data minimization* and *data specification*. That is, platforms (as data controllers) must specify in advance the purpose of data collection (Zarsky, 2016). Data specification is a principle that reflects the rule of law and due process—foundational concepts of constitutional democracies requiring laws affecting the basic interests of citizens to be transparent, nonarbitrary, and fairly applied (Citron, 2007). In personalization, due process means notification that one is being algorithmically profiled and requires some procedure through which one can contest algorithmic decisions. This is important because arbitrary and irrelevant factors may be included (or automatically "discovered" in the case of deep reinforcement learning) in the state space. Yet the stochastic actions of black-box, exploratory algorithms would seem to violate the expectation of notification as their behavior cannot be determined in advance of running them on the platform. Further, when applied to millions of platform users, by sheer chance some individuals or groups will inevitably be on the receiving end of a long sequence of "suboptimal" random actions, raising issues of algorithmic harm and fair allocation of technological risks across individuals and social groups (Rhoen, 2017).

Offline learning also challenges the GDPR's requirement of explicit, prespecified data collection, one of several principles designed to ensure that personal data processing and algorithmic decision-making ultimately "serve mankind" (Greene et al., 2019). Offline learning techniques offer new opportunities to outsource data processing to third parties and can contribute to new, untraceably complex flows of implicit data that violate nondigital information transmission norms (Nissenbaum, 2004). But as Zarsky (2016) points out, the value and innovation of big data come from using it in new ways unintended and unspecified by the original collector. Although offline learning resembles supervised learning in relying on data that are pre-gathered by a platform, the data themselves could have been collected using a highly exploratory interaction strategy in the past. Moreover, major platforms with vast stores of previously collected behavioral data can exacerbate violations of data specification and minimization by developing and even licensing reinforcement learning-based personalization systems to other platforms for further online learning or fine-tuning. Similarly, policy learning could be outsourced to countries where data protection laws do not restrict the storage, collection, or processing of personal data, or where existing data laws are either not enforced or their enforcement is not subject to outside inspection.

Furthermore, data specification and transparency of purpose are important given the potential for strategic manipulation of users and the economic imperatives of platforms. Firms could deploy techniques to improve existing behavior policies or develop new ones aimed at manipulating and exploiting the various decision-making vulnerabilities of platform users (Susser et al., 2019). As major platforms increasingly look to self- and semi-supervised learning techniques to avoid having to pay humans to label vast quantities of unstructured data used in early-stage platform services, platforms have economic incentives to harness the power of reinforcement learning algorithms to learn effective strategies aimed at getting users to label implicit data (e.g., emotion tagging or object labeling) (Posner & Weyl, 2019). The unspecifiability of these new data flows and purposes makes it difficult to trace the provenance of data and can violate implicit (social) contextual norms around consent to data collection and processing. Ultimately, unspecifiability reduces platform users' trust that their interactions with the platform will not be used against them to either diminish their autonomy (Kane et al., 2021) or exploit them as a source of free labor (Couldry & Mejias, 2019).

Regarding the specifiability of data collection and processing under the GDPR, if platforms make it clear that personal data may be collected for purposes of algorithmic improvement and obtain explicit user consent, exploratory learning may still be permissible. But algorithmic improvement (i.e., greater reward accumulation) may itself result from learning more complex, deceptive, and manipulative interaction policies. Despite this issue, the GDPR gives platforms a variety of legal justifications for processing and collecting personal data in ways that could be

construed to support both online and offline learning. Beyond explicit user consent, *contractual necessity* and *legitimate interest* are frequently cited as legal grounds for processing personal data (Official Journal of the European Union, 2016), as Facebook does for example. It may thus be possible for data controllers to use legitimate interest as legal grounds for online data collection (i.e., exploration) and accumulating massive replay buffers of interactive data useful for later offline policy learning and evaluation.

3.4 Nonconsensual Human Behavior Modification

Research has revealed that when Facebook *likes* and psychometric tests are combined, Facebook users' IQ, political preferences, and openness can be accurately and thus profitably predicted (Matz et al., 2017). Yet this initial research manually targeted Facebook users and manually designed advertisements to appeal to specific psychological profiles. Platform-based reinforcement learning enticingly offers a new, scalable means of automating this strategy by fusing state-of-the-art models of brain function with engineering formalism (Botvinick et al., 2020). Indeed, using *likes* as rewards, neuroscientists have demonstrated that behavioral predictions made from reinforcement learning models are nearly indistinguishable from actual human social media engagement behavior, suggesting that platform algorithms can control user engagement simply by varying the rate of associated rewards (Lindström et al., 2021).

The new era of reinforcement-learning-based personalization has the potential to scale and automate the strategic process of psychological targeting and behavioral modification beneath the level of users' conscious awareness, particularly as novel and realistic text, audio, image, and video content can be generated by deep neural networks (le Moing et al., 2021). Platforms can also leverage newly formalized *Markov persuasion processes* (Wu et al., 2022)—which combine the economic theory of sequential information design with reinforcement learning—to strategically exploit information asymmetries in favor of arbitrary platform interests. In light of these and other developments, Wertenbroch et al. (2020) warned that automated and personalized interventions threaten consumer autonomy, particularly when consumers lack the "persuasion knowledge"[6] regarding the possibility of sophisticated persuasive behavior by artificial agents deployed on commercial platforms. In situations marked by such power and information asymmetries, data protection laws like the GDPR may provide some protection of an individual's capacity for self-determination by granting users the right to opt out of automated data processing and profiling when it has significant legal effects (Greene et al., 2019).

In any case, platforms can now combine the automated control of reinforcement learning with the science of behavior modification to modify both digital content and delivery methods. Behavior modification is traditionally defined as "an intervention designed to alter or redirect causal processes that regulate behavior" (Michie et al., 2013). Historically, behavior modification techniques were used by professionally certified therapists in clinical settings or in animal training contexts (Baum, 2017). Importantly, human behavior modification in scientific, clinical, and academic settings must be carried out with full transparency and the consent of participants and usually involves approval by an Ethics or Institutional Review Board in the United States. When experimental designs require deception, a detailed cost-benefit calculation considers possible short-term and long-term physical and psychological harms to participants (Sell, 2008). Currently, however, autonomous experimentation systems can be deployed on platforms without any independent ethics board approval and may exhibit deceptive or coercive behavior as they interact with human users (Kenton et al., 2021), potentially modifying users' beliefs and fostering addiction (Russell, 2019) in a process resembling *operant conditioning.*[7] However, one cannot meaningfully consent to habit formation via conditioning if one is consciously unaware of it (see e.g., O'Neill, 2002). Indeed, the European Data Protection Board has stated that "scrolling or swiping through a webpage or similar user activity" does not constitute clear and unambiguous consent to data processing (European Data Protection Board, 2020). These instances of personalization aimed at the nonconsensual but gradual formation of new and profitable habits go against the intentions of the pioneers of persuasive technology, who explicitly condemn the use of deception or coercion (Fogg, 2003).

Just as humans learn that money can be used to buy food, artificial agents can not only learn which behaviors are associated with long-term primary rewards specified by the algorithm designer but they can also learn secondary rewards: rewards that act as reliable predictors of primary rewards (Singh et al., 2009). In other words, for the artificial agent, an action in some states can become instrumentally valuable by allowing it to reach more valuable states associated with greater expected future rewards. For example, if

[6] Wright (2002) defines *persuasion knowledge* as "an individual's beliefs about the mental states and psychological change processes that operate as mediators of persuasion or intentional social influence."

[7] In behaviorist psychology, operant conditioning refers to the acquisition and performance of a behavior in response to its observed effects on the environment (Touretzky & Saksida, 1997).

click behavior (an implicit form of feedback) is included in the agent's reward function, artificial agents can learn that emotionally polarizing and morally divisive content is associated with a higher probability of clicks for certain users. Over time, human platform users may develop behavioral and attitudinal changes by constantly being recommended emotionally and politically polarizing content. In fact, the click behavior of such users may be easier to predict, further incentivizing human behavior shaping. Even though human platform users may not consciously recognize these subtle secondary or even tertiary rewards driving the personalization process, the individual effects of being exposed to morally outrageous content, when aggregated across millions of users, can affect public discourse at scale (Brady et al., 2021).

From a philosophical and psychological perspective, implicit feedback is morally distinct from explicit feedback—the kind needed for meaningful consent. More often than not, implicit data reflect "subpersonal" behaviors (Bermúdez, 1995; Dennett, 2002) lacking intentionality or conscious awareness (Bargh & Chartrand, 1999), akin to the distinction philosophers make between a *blink* and a *wink* (Wegner, 2004). Implicit data typically result from impulsive, habitual, and instinctive behaviors requiring no conscious planning or reflection (Lyngs et al., 2019)—hence the oft-used phrase "mindless scrolling." Implicit data thus arguably reveal more about what primitive reward systems in our brains *want* than what we, as persons,[8] *like* (Berridge, 2009). When combined with reinforcement learning, personalization driven by implicit behavioral data can undermine our capacity for autonomy, which requires consciously identifying with a unique set of personal values that properly motivate our actions by passing a test of higher-order reflective endorsement (Frankfurt, 1971; Korsgaard, 1989). Incentivized by the goal of maximizing engagement-based implicit feedback, the narrow instrumental rationality of artificial agents may slowly but surely steer us toward lower or "alienated" forms of ourselves (Prunkl, 2022; Wolf, 1993) through nonconsensual forms of behavior modification.

Artificial agents can be augmented with—or learn implicitly via trial and error—empirically validated principles from social psychology as well. For instance, they may autonomously learn to take advantage of heterogeneity in user click and dwell time responses to nonconscious priming effects of personalized content (see, e.g., Wheeler & Berger, 2007). Artificial agents may also learn to exploit more obvious persuasive communication strategies to manipulate the attributions users make about themselves, thus modifying their subsequent behavior (see, e.g., Miller et al., 1975). Interactive agents may also be designed (or learn) to exploit the "identity-relevance" of persuasive messaging to individuals' self-ascribed identities (Bhattacharjee et al., 2014). For instance, Cai et al. (2017) suggest pairing "behavior change interventions" with reinforcement learning-based personalization to influence "underlying behavioral determinants, such as perceptual and self-efficacy biases" in mobile settings. Nevertheless, from the perspective of AI researchers, an artificial agent's deceptive or manipulative behavior may be an exciting expression of "social intelligence" in interactions with humans (Silver et al., 2021).

Another example of social psychology research explicitly being applied to reinforcement learning-based personalization involves *persuasion profiling*. Persuasion researchers imagine massively distributed reinforcement learning agents embedded in digital and real-world environments that learn users' persuasion profiles and intelligently tailor specific messages and delivery formats to them (Kaptein et al., 2009). Examples include ads designed using principles from the psychology of influence, which appeal to authority, social proof, scarcity, or reciprocity, for instance. Reciprocity strategies can make consumers more likely to self-disclose intimate information about themselves as they interact with computer-based agents, and the tendencies of intimate self-disclosure may depend on the particular sequencing of these disclosures (Moon, 2000). An autonomous reinforcement learning agent may thus learn the most effective disclosure-escalation patterns conducive to maximizing its reward signal, resulting in behaviors appearing pathological or manipulative to human users.

A more sinister use of reinforcement learning involves the intentional steered control of user choice behavior. Empirical research in cognitive neuroscience has shown that deep neural networks can be trained on users' past choices and rewards received to dynamically allocate future rewards in order to "maximally bias" user choice toward the designer's goal (Dezfouli et al., 2020). This ability extends the concepts of nudging and choice architectures to an algorithmically specified form of "choice engineering" (Dan & Loewenstein, 2019) and raises new possibilities for marketers and advertisers to influence vulnerable platform users, especially during multiple interaction sessions. AI researchers are currently studying how to incorporate human biases and

[8] We use the term *person* in a technical sense that implies moral agency and responsibility. For Kant, a person is a "subject whose actions can be imputed to him" (Kant, 1797/2017, p. 50); actions are "voluntary in the sense of not determined by natural impulses" (p. 30). Persons are thus defined, in part, by their capacity to resist the impulsive, stimulus-response behavior of "lower" animals.

deviations from rational choice theory into reinforcement learning-based personalization (Theocharous et al., 2019). For instance, the task of purchasing a product on a website might be broken down into various subtasks, each concretely operationalized and rewarded as it contributes to the goal of conversion (Wiesel et al., 2011). Artificial agents can be trained to optimize the timing, placement, and presentation of content so that platform users are unlikely to be consciously aware of its influence on their own behavior. Personalized offers and notifications could then be delivered to users at key "navigational touchpoints" discovered by the agent in its interactions with users in order to maximally influence the probability of a browsing session ending in a purchase.

As more behavioral data are collected and ambient sensor technology (IoT) expands, governments—likely in cooperation with major technology companies—may also look to deep reinforcement learning as a powerful kind of *hypernudge* (Yeung, 2016) aimed at implementing new forms of governance and public policy of dubious political and moral legitimacy. Deep reinforcement learning can already help manage the data generated by "smart cities" and large-scale resource management issues related to power usage, traffic signal control, and agricultural productivity (Whittlestone et al., 2021). Governments may also be interested in the ability of deep reinforcement learning agents to carry out adaptive experiments aimed at modifying citizens' behavior over time to promote social or public policies. However, more concerning than traditional modes of behavior modification that focus on consensual, static, one-shot behavior change interventions, deep reinforcement learning can generate elaborate sequences of causally opaque hypernudges whose logic may be unintelligible to policymakers or citizens. To make matters worse, citizens cannot meaningfully consent to something they cannot understand.

All the same, the use of black-box technologies may undermine traditional justifications for the authority of government and law, especially as social and health policies of governments increasingly require the kind of data infrastructures and algorithmic expertise that only major technology firms can provide (see e.g., Marhold & Fell, 2021). Rule by consent is essential to the political legitimacy of governments as it obligates citizens to respect the power of the government to intervene in their lives (Raz, 1987). Further, corporate interests in intellectual property and profit may conflict with democratic norms of public accountability and transparency. In the absence of widespread public support for and greater education on the persuasive power of these new black-box technologies of social control, citizens have little reason to trust that government institutions and public policy align with and advance their interests, and thus little reason to obey governmental authority. This erosion of governmental trust and authority could further contribute to political and social instability.

Despite these social and political concerns, policymakers in various countries are now considering personalized nudges and choice architectures that vary both the content and method of delivery to individuals (Mills, 2022). Lawmakers in the United States, Australia, and Singapore, for example, have teamed up with behavioral scientists to implement behavior change techniques. The United Kingdom even has a dedicated Nudge Unit in government (Soman & Yeung, 2020). In China, the government is famously piloting a social credit scoring system that rewards or punishes citizens based on their workplace performance, health, consumer purchases, interpersonal relationships, and political activities (Wong & Dobson, 2019). When combined with advances in ubiquitous, interconnected sensors and devices, the specter of applying black-box reinforcement-learning methods to implement large-scale social nudging policies remains a possibility that threatens the basic underpinnings of constitutional democracies and the rule of law (see e.g., Binns, 2018; Nemitz, 2018; Pasquale, 2015).

We note that behavior modification may still be possible under the GDPR, as it offers member states exemptions to carry out automated profiling related to preferences, behavior, or interests; for purposes of contractual necessity; or given a user's explicit consent for doing so. This means that certain EU countries may choose to allow users to opt-in to using reinforcement learning-based systems. Ideally, as in the case of clinical and therapeutic applications of behavioral modification, users should be aware of the short-term and long-term risks involved. Yet platforms may not necessarily be obligated to inform users about the use of reinforcement learning. Under the GDPR's Recital 62, platforms could still perform nonconsensual behavior modification by arguing that "the provision of information to the data subject proves to be impossible or would involve a disproportionate effort" (Official Journal of the European Union, 2016). This may, however, change with the introduction of risk tiers used to classify AI systems, as suggested in the EU's AI Act. The AI Act bans systems that pose "unacceptable risks" and use "subliminal manipulation" techniques (European Commission, 2021). It is not clear whether this covers explicit applications of social psychology research to reinforcement learning or the subtler reward associations learned by autonomous agents.

3.5 Pathological Agent Behavior via Business Metric Optimization

Equal attention must be paid to the other side of platform interactions, i.e., the artificial agent's behavior. Compared to earlier paradigms, reinforcement learning-based personalization requires greater clarity and focus on selecting the right performance metrics and abstract system goals. The unique features of reinforcement learning, such as its speed, adaptability, and closed-loop nature, violate traditional statistical assumptions used in machine learning and thereby closely relate the choice of metrics to be optimized with the resulting algorithmic behavior (Rahwan et al., 2019). As the previous section explains, if these relationships are not carefully monitored and corrected, dangerous and pathological agent behavior may ensue.

In earlier machine learning-based paradigms of personalization, "static" recommender systems such as those using collaborative filtering, typically aimed to optimize generic performance metrics such as precision, recall, or normalized discounted cumulative gain for ranked lists (Pei et al., 2019). These metrics were chosen because they were presumed to correlate with key business objectives (Adomavicius & Tuzhilin, 2005). Reinforcement learning-based personalization, however, offers a more principled approach for balancing both short-term and long-term interests of platform stakeholders in a way naturally aligned with business goals. The reward function can, in theory, be designed to solve any conceivable task or, equivalently, control any environment. In fact, the pioneers of reinforcement learning have put forth what they call the "reward hypothesis" (Sutton & Barto, 2018, p. 53). In practical terms, this means an agent's reward function can be custom made to optimize for the business goals of different stakeholders and maximize various engagement metrics. Indeed, platform-based personalization systems are increasingly viewed as algorithmic matchmakers distributing "utility" according to "differing configurations of interests among stakeholders," stakeholders being platform users, corporate owners, and interested third parties such as advertisers (Abdollahpouri et al., 2017).

The shift in personalization from short-term to long-term optimization mirrors an insight from marketing science that a firm's long-term competitive advantage depends on its ability to capture, retain, and nurture a customer base (van Doorn et al., 2010). For example, TikTok's business goal is to add daily active users. Its personalization algorithm therefore optimizes both for time spent and user retention (Smith, 2021). With the right kind of reward function, platforms can jointly optimize for instant engagement and multiple long-term goals, e.g., clicks, infinite scrolling behavior, and user retention (Zou et al., 2019). This is an important advantage over simpler bandit algorithms that optimize for rewards in the current iteration, potentially leading to user churn in the long term (Theocharous et al., 2015).

Reward engineering is difficult—possibly more art than science—and, when not done carefully, may lead to pathological outcomes over the long term (Everitt et al., 2021). This problem is known as *value alignment* and concerns the technical and normative aspects involved in aligning an artificial agent's values with human values (Gabriel, 2020). Platform personalization further complicates the alignment problem. As the economic interests of an increasing number of stakeholders, each with different goals, are encoded in the reward function, specifying complicated reward functions that result in safe and ethical agent behavior may become even more difficult. Despite this, or perhaps because of this, AI researchers and roboticists are exploring ways to automate reward function design through self-supervised learning, thus removing the human bottleneck (Levine, 2021). But besides possibly making artificial agents' behavior even more unpredictable, misalignment due to improperly specified reward functions can lead to pathological outcomes, such as promoting addiction (Burr et al., 2018).

Pathological algorithmic behavior of artificial agents has been described as a failure mode of overoptimization (Manheim & Garrabrant, 2018). Recall that the reward function is defined by algorithm designers and results in the artificial agent receiving a reward signal—reinforcement—after performing "good" or "bad" actions in certain states. In practice, the algorithm's designer selects a reward function that correlates highly *but not perfectly* with the intended business goal (e.g., rewarding the agent for increased clicks with the goal of increasing revenues). In this sense, reinforcement learning is a powerful optimization method using the artificial agent's capacity for trial-and-error learning to find creative solutions to problems. Problematically, however, "gaps" between the proxy reward signal and the true goal can be exploited by an instrumentally rational agent in ways unforeseen by the designer, resulting in behavior that may appear to humans as Machiavellian (Bostrom, 2012). In Section 4, we address this issue by suggesting research on simulators and experimental sandbox environments.

There are several known failure models of metric-driven overoptimization, arguably the most famous being Goodhart's law. This "law" states that "any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes" (Manheim & Garrabrant, 2018). In a causal variant of Goodhart's law, artificial agents may even learn to manipulate or directly control their reward signals as a side-effect of pursuing their goals (Russell et al., 2015). For instance, if an agent's performance is measured solely by metrics such as accuracy or recall,

an instrumentally rational agent will seek the greatest improvement in accuracy with the least possible effort. Rather than making better predictions (the ostensible intention of the designers) to improve accuracy, the agent may instead learn to shape or causally intervene in the distribution of users (its environment) such that predictions are easier to make. This tactic is known as *auto-induced distributional shift* (Krueger et al., 2020), or alternatively, *user tampering* (Evans & Kasirzadeh, 2021), and reflects the inability of artificial agents to distinguish between satisfying a user's preferences and influencing a user to have preferences that are easier to satisfy (Krakovna et al., 2020). A practical illustration of the difference would be an artificial agent encouraging addictive behavior because addicted human users are more likely to behave in ways specified by its engagement-driven reward function.

But this Machiavellian strategy violates the Kantian duty to respect the autonomy of persons in freely choosing their preferences (Frankfurt, 1971). In the extreme, platform users exposed to such artificial agents may develop paradoxically positive relationships with their platform-based algorithmic oppressors—a kind of AI-induced Stockholm syndrome—electing to view their domination or addiction in a positive, or at least neutral, light. Yet few users would reflectively endorse such a dystopian situation of algorithmic domination and learned helplessness (Kane et al., 2021). Indeed, concerns about preference manipulation by artificial agents resemble critiques of utilitarianism in which the narrow focus on preference satisfaction can result in the "cheerful endurance" of suffering and unjust social and economic circumstances (Sen, 1984). That is, as a matter of survival, persons can and often do adapt to—and even rationalize—situations of unjust power asymmetries and exploitation. Relying on utilities as a measure of subjective well-being overlooks this curious fact of human psychology and its impact on autonomy and social and political self-determination (Floridi et al., 2018). Given the Machiavellian potential of artificial agents,[9] we suggest that more philosophical attention be paid to the adaptive, dynamic, and temporal nature of human preference formation (Christman, 1991; Elster, 2016) and more empirical consideration be given to alternative measures of user well-being that take into account the exercise of basic universal human capabilities needed for a good life (Nussbaum, 2001; Teschl & Comim, 2005).

Unfortunately, even though the GDPR is founded on the basic human rights to privacy and data protection, the GDPR is limited in preventing such pathological outcomes due to failures of business optimization and reward maximization. According to the Charter of Fundamental Rights of the EU, platform data controllers also have the "freedom to conduct a business," and processing personal data may be inherent to the business, e.g., behavioral targeting for an ad network (Zuiderveen Borgesius, 2015). In other words, data collection and processing can still occur if it is in the "legitimate interests" of the data controller. The determination of legitimate interest, however, depends on the type and extent of the relationship between the platform and its users. For example, banks need to know certain details about their customers in order to fulfill contractual obligations. For platforms, personalized algorithmic services may be considered an essential aspect of the business, thus justifying their legitimate interest in processing logged implicit feedback in order to adapt to new platform content and user preferences over time. A major source of contention, however, is how commercial interests in profit maximization and engagement-centric business metrics, such as clicks or dwell time, can be balanced with fundamental human rights and freedoms, such as the right to freely develop one's personality (Coors, 2010). One issue is that corporate-owned platforms have incentives to hide or obscure any research potentially revealing negative psychological outcomes from interactions with their artificial agents (see e.g., Haugen, 2021). For this reason, we support research using simulators and sandboxes to verify platform claims, study the process of adaptive (or manipulative) preference formation, and examine the extent to which the rights and freedoms regarding personality development might be infringed upon (see Section 4). We note that provisions in the EU's proposed AI Act may more effectively address some of these issues, particularly those related to algorithm-induced behavior modification and user manipulation.

4 Ethically Aware Research Directions for Reinforcement Learning-Based Personalization

Building on the IS field's strengths in systems design and development, technology ethics and social theory, experiments and causal inference, user behavior models, and simulation, we propose three future research

[9] Readers familiar with inverse reinforcement learning (Ng & Russell, 2000) and its variations may argue that pathological agent behavior results from misalignment and uncertainty regarding users' underlying reward functions. The implication is that learning users' reward functions might solve the problem of pathological agent behavior. Besides the questionable assumption that users' preferences are fixed and that their observed behavior represents "optimal policies," there remains the problem of reward function unidentifiability from observed behavior (Armstrong & Mindermann, 2018). Even granting these assumptions and issues, notable philosophers (Rawls, 1971) and economists (Sen, 1984) have argued against utilitarianism and consequentialism as ethical systems conducive to either flourishing human lives, right conduct, or just political systems.

directions to address the ethical challenges posed by reinforcement learning-based personalization. These suggestions are technological, sociotechnical, and ethical research directions designed to appeal to IS researchers with expertise in data science, empirical research, and philosophy, ethics, and social theory, respectively. As a bonus, they may help overcome related scientific challenges of data collection and reproducibility raised by platform-based reinforcement learning.

4.1 Simulators as Technological Tools for Safe and Ethical Reinforcement Learning

Simulation has a long history in reinforcement learning research and is commonly used in the domains of robotics and autonomous driving, where systems are trained in simulated environments before fine-tuning and deploying them in the real world (Tobin et al., 2017). Simulated data can be generated at nearly no cost, and no potentially dangerous interactions with humans or physical machines are required. Simulation is now also being used in personalized medicine to help run "in silico" clinical trials, where "digital twins" can be used to help select and optimize treatments for individual persons (Corral-Acero et al., 2020). The digital twin concept is also increasingly applied in industries as diverse as aviation, hospital management, and manufacturing, although much research remains to be done to tackle the design of such systems and their technical limitations (Barricelli et al., 2019). In the case of personalization, digital twins could be representations of actual platform users built from historical and simulated interaction data and used to facilitate real-time prediction, optimization, and monitoring (Rasheed et al., 2020). We believe the digital twin design concept can be valuable in advancing applications of safe and ethical reinforcement learning-based personalization and could draw on the IS community's existing strengths in systems design, simulation, and modeling.

Several existing simulator platforms, such as DeepMind Control Suite, OpenAI Gym, and the Arcade Learning Environment, provide reproducible environments for researchers to experiment with, develop, and evaluate reinforcement learning agents without requiring interaction with real humans. But these systems are designed primarily with engineering goals in mind. New "recommender gyms" focusing on reinforcement learning, such as Virtual TaoBao (Shi et al., 2019), RL4RS (Wang et al., 2021), RecoGym (Rohde et al., 2018), and RecSim (Ie et al., 2019), allow for greater customization but are generally focused on the domain of personalized advertising. This gap leaves open a space for IS researchers to make novel research and design contributions to reinforcement learning-based personalization that go beyond for-profit applications.

Currently, most public datasets for training and evaluating machine learning algorithms are not set up for sequential reinforcement learning problems, and it is not trivial to manually convert existing datasets into a sequential format (Wang et al., 2021). Fortunately, simulation gyms allow us to simulate arbitrary logging policies representing the platform's current personalization system. Simulators can then evaluate how the speed and automated adaptivity of reinforcement learning-based personalization affect the behavior of human users on platforms. In this respect, simulators offer a valuable methodology to advance various areas of IS theory (Dong, 2022) traditionally hampered by data availability and quality issues, not to mention the emerging problem of algorithmic confounding on platforms (Chaney et al., 2018). Relevant additional uses of simulation include studying the diffusion of misinformation, verifying a personalization system's robustness to adversarial attacks, and identifying pathological agent behaviors by providing an early warning system to detect failure modes of reward-driven optimization. [10] Lastly, simulators mitigate the degree to which excessive volumes of implicit data are needed on human subjects, thus addressing the problem of nonconsensual behavior modification and respecting the GDPR's principle of data minimization.[11]

Simulators can also contribute to interesting cross-disciplinary research. For instance, simulators could be used to study the effects of modifying the reward function on various outcomes of interest related to aspects of personal autonomy and human flourishing, or compare new behavior policies against baselines with respect to ethical dimensions of personalization, such as safety or unhealthy addictive behaviors. We envision IS researchers leveraging their knowledge of user behavior models and experimental strengths in operationalizing theoretical constructs to create custom "ethics-based" reward functions inspired by philosophy and the psychology of self-determination and well-being. In short, simulation is an emerging methodology offering IS researchers greater ability to proactively ask and investigate IS-specific research questions—independent of platform data access and quality—while also advancing IS theory (Grover et al., 2020).

[10] Deepmind researchers are now investigating the problem of reward tampering through the use of specialized simulator software (Kumar et al., 2020).

[11] We note, however, this is a double-edged sword. In theory at least, a platform could nefariously use digital twin simulations to *improve* control policies over individual human users in the face of reduced personal data. This problem parallels the ethical "dual-use" concerns of AI: just as AI can be used to predict the structure of useful materials and health-promoting molecules, so too can it be used by bad actors to predict and synthesize deadly chemicals, compounds, and toxins (see e.g., Urbina et al., 2022).

4.2 Sandboxes as Sociotechnical Enablers of Platform-Based Field Experiments

The IS discipline has a strong tradition of randomized field experiments on digital platforms for evaluating the causal effects of interventions relevant to personalization. Nevertheless, corporate platforms are increasingly unlikely to allow external academics to perform experiments on their platforms for fear of harming the user experience or because their work may reveal potentially damaging psychological and social effects of using their platforms (Haugen, 2021). Worse still, the speed and adaptability of personalization algorithms on platforms makes disentangling the causal effect of a single intervention nearly impossible for internal platform data scientists, let alone external academic researchers (Greene et al., 2022). We propose sandbox environments as a means to mitigate some of these issues and advance IS research in this new age of reinforcement learning-based personalization.

In software safety and security testing, sandboxes are staging environments that limit exposure to an action in accordance with a specific security policy (Bishop, 2002). Sandboxes are also used as experimental tools by regulators to evaluate the safety and reliability of emergent financial technologies before releasing them to market (Allen, 2020). Sandboxes typically involve "live" human users, and their applications are thus generally more expensive and limited to later stages of testing and experimentation. Prior to this "online" experimental stage, simulations can fill this gap by generating a candidate set of the most promising policies to explore further in live settings. Surprising results from simulations can also direct the attention of researchers and help them identify signs of human behavior modification or pathological agent behavior when interacting with live humans in sandbox environments.

The development and design of sandboxes designed to mimic platform environments can bolster both the science and technology around reinforcement learning personalization. By providing controllable and safe digital environments, they offer the ability to conduct field experiments on platforms and study the causal effects of digital interventions while avoiding excessive and unspecified data collection irrelevant to the research question of interest. For example, although a single sandbox may involve a relatively small number of real users, several concurrent sandboxed experiments can be pooled and analyzed together to detect the relatively subtle effects of adverse events—a technique also used in multisite clinical trials to improve the statistical power of experiments (Meinert, 1980). On the technical, systems-oriented side, the sandbox can crucially help to arrange treatment exposure restrictions (i.e., consenting platform participants in studies) and provide a general model for auditing reinforcement learning algorithms by setting standardized disclosure and transparency requirements, perhaps for sharing with regulatory bodies or other researchers.

Besides advancing the practical knowledge involved in designing and implementing a sandbox, future sandbox-based research could be an important source of insights for policymakers, non-governmental organizations, and professional societies, as they develop legal and ethical principles around reinforcement learning applications. We note that the European Union's proposed AI Act utilizes regulatory sandboxes to test, evaluate, and monitor AI systems before releasing them to market, particularly those deemed "high risk" (European Commission, 2021). The United States Government Accountability Office is also investigating sandboxes for testing emerging AI applications (Persons, 2018).

By combining simulators with the sandbox concept, empirical IS researchers—with varied expertise in field experiments, causal inference, and econometric techniques—can play an active and important role in studying and producing independent evidence for the safety and reliability of reinforcement learning-based systems. In short, we see simulators and sandboxes as scientific and technical enablers in developing responsible AI innovation ecosystems (Stahl, 2022) in this new era of reinforcement learning-based personalization.

4.3 Engaging Critical IS, Business Ethics, and AI Ethics to Harness Reinforcement Learning for Human Flourishing

Reinforcement learning-based personalization abstracts human persons as an environment to be shaped and modified in a way conducive to platform business goals. Autonomous experimentation systems increasingly intervene in our digital lives without our consent, potentially altering our attitudes and behavior in the process. Platforms collect and use massive amounts of implicit behavioral data—often without our conscious awareness—to personalize services and content with increasing speed and adaptivity, possibly removing opportunities to reflectively endorse such changes, and thereby reducing our autonomy as persons. Even when individual effects are small, these automated adaptations may have large social and political effects when networked and scaled on global platforms.

Early IS research on platform-based personalization assumed managerial perspectives aimed at developing methods to counter "strategic customer behavior" and "help shift the power back in favor of the firm" (Murthi & Sarkar, 2003). Yet, in light of the ubiquity of personal data collection (Leidner & Tona, 2021) and the ethical and social impact of personalization (Milano et al., 2020), the pendulum today appears to be swinging in the opposite direction. Reinforcement learning provides an

instructive example of this trend. An emerging literature deals with its safety, value alignment, containment, governance, and even its potential ability to overtake and destroy the human species (Bostrom, 2014).

While the fields of AI and machine ethics focus on building reinforcement learning agents that can—either explicitly or implicitly—follow ethical principles and rules (Abel et al., 2016; Briggs & Scheutz, 2015; Russell et al., 2015; Wu & Lin, 2018), this rather technical research niche generally focuses on avoiding or minimizing harm to humans. It neglects the question of whether and how personalized reinforcement learning agents can actively promote human flourishing and autonomy. Future IS research on personalization may thus benefit from incorporating aspects of value-sensitive design (Friedman & Hendry, 2019), participatory design (Schuler & Namioka, 1993), and human rights-based AI frameworks (Aizenberg & van den Hoven, 2020). Might platform users even play a role in deciding—reflectively endorsing—their reward function specification?[12] We leave this as an intriguing topic for further study.

Clarifying and discovering the ethical "unknown unknowns" of reinforcement learning-based personalization will require engagement with legal scholars and AI, technology, and business ethicists. Fortunately, IS scholars are increasingly focused on conducting holistic, cross-disciplinary research examining the interplay between the technical and ethical aspects of big data and machine learning applications (Abbasi et al., 2016). Stahl et al. (2021) is a promising start in this direction but does not discuss reinforcement learning specifically. Du and Xie (2021) draw on institutional and stakeholder theory to illustrate how ethical principles around AI can be framed within a corporate social responsibility perspective but are not focused on the emergent technical features of reinforcement learning. Lastly, Prunkl (2022) provides a concise summary of the dangers of AI to human autonomy, with many examples drawn from reinforcement-learning applications.

The corporate platform context of reinforcement learning further offers interesting research directions involving law and business ethics. For example, can exploratory "random" agent behaviors resulting in unspecifiable data collection be understood through the lens of public or private nuisance law (Balkin, 2017)? What are the legal liabilities and moral responsibilities of artificial agents deployed on corporate platforms(Asaro, 2007)? And to what extent could research on incentives and principal-agent problems be usefully applied to artificial agents on platforms (see e.g., Hadfield-Menell et al., 2016)? Fruitful analogies might also be drawn between environmental justice (Mohai et al., 2009) concerns—i.e., the siting and distribution of environmental pollutants, landfills, and waste sites in areas primarily inhabited by minority groups without political power or representation—and the exploitation of (cognitively) vulnerable populations in digital platform environments. Building on the IS field's strength in platform-focused research (Tiwana et al., 2010), extending and linking ideas from law, economics, business ethics, and environmental justice to reinforcement learning-based personalization could be a novel source of contributions from the IS community.

Regarding the normative and social implications of reinforcement learning, we believe the critical IS (CRIS) research community has a key role to play and call on CRIS researchers to help further the development and understanding of reinforcement learning-based personalization. CRIS primarily adapts the social theory and philosophy of Jürgen Habermas to analyze and evaluate how IS can be used to promote or hinder the goal of human emancipation (Lyytinen & Hirschheim, 1988; Myers & Klein, 2011; Stahl, 2008). Drawing on and extending its foundation in European philosophy, CRIS researchers understand emancipation as freedom from ideology, power asymmetries, psychological compulsions, and social constraints (Hirschheim & Klein, 1994). CRIS can guide answers to questions about how to do data science for social good and reimagine reinforcement learning-based personalization beyond its role as an amoral technological instrument for human control and engagement optimization. As a recent example of research in the critical spirit, Kane et al. (2021) adapted Paulo Freire's notion of emancipation to analyze the ways in which machine learning can contribute to human oppression. In short, CRIS provides a strong normative and theoretical lens—unique to the field of IS—through which to view the governance, social, legal, and ethical issues raised by reinforcement learning-based personalization.

We see our proposed research directions as both a starting point and a call to action for the IS community. To ensure that reinforcement learning enhances personal autonomy and human flourishing, we urge IS researchers to apply their unique knowledge of systems design and development, technology ethics and social theory, experiments and causal inference, user behavior models, and simulation to the novel social and ethical problems posed by emerging personalization technologies.

[12] The literature on human-centered or interactive reinforcement learning (Li et al., 2019) reveals mixed results in getting nonexpert humans to teach artificial agents and robots viable behavioral policies using "social" or evaluative feedback (e.g., punishments and rewards). Failure often results from a mismatch between human teachers, who tend to treat evaluative feedback as signaling communicative intent, and artificial agents, who treat it as a reward to maximize (Ho et al., 2017).

5 Reinforcement Learning Beyond Platforms: Contrasting Ethical Landscapes in Noncommercial Domains

Reinforcement learning has found rapid uptake in medicine, health, education, and social and public policy contexts—areas the IS field has studied extensively. To jumpstart further research and reflection, we offer a preliminary analysis of ethical differences among these domains relevant to personal autonomy and political and social stability in light of the features of reinforcement learning-based personalization.

Reinforcement algorithms play an important role in developing personalized intervention sequences to automate complex decision-making. For instance, in online education, Bassen et al. (2020) used reinforcement learning to personalize the scheduling of educational activities based on individual learning performance. In medicine, reinforcement learning algorithms play an important role in developing personalized or dynamic treatment regimes to aid clinical decision-making processes in healthcare areas related to rehabilitation, medical imaging, diagnosis, dialogue systems, and health management systems (Coronato et al., 2020). Reinforcement learning has been frequently employed in cancer treatment and prevention. Zhao et al. (2009) used it to discover complex associations between sequences of actions and outcomes to discover individualized treatment regimens for cancer. In public health, Figueroa et al. (2021) relied on the trial-and-error methods of reinforcement learning to automatically select personalized social distancing text messages to individuals during the COVID-19 pandemic. Nahum-Shani et al. (2018) described using reinforcement learning to learn just-in-time adaptive interventions with behavioral data collected by mobile devices. In short, adaptive or personalized interventions are valuable because they can reduce negative exposure effects and waste, increase compliance, and enhance the potency of interventions by targeting users (Collins et al., 2004).

While personalized recommendation on commercial platforms is increasingly viewed as analogous to treatment assignment in medical studies or public policy interventions (Schnabel et al., 2016; Yang et al., 2020), this abstraction hides important ethical differences. Commercial firms face relatively few socially and legally enforced safety and external accountability mechanisms; influential economists have even argued that private enterprises have no direct responsibility to promote the social good (Friedman, 2007). Indeed, Facebook and Uber claim they are technology companies, not media publishers or transportation companies, respectively. Their ambiguous legal status is a source of innovation and growth and a reason why "sharing economy" platforms can bypass the ethical duties and social roles traditionally assumed by newspaper publishers or taxi services, for example (Schor, 2016). In reinforcement learning terms, this implies that the specification of their reward functions on their platforms need only reflect their self-interest as profit-maximizing corporations, not the other-regarding aspects derived from their social roles or legal and moral duties, such as promoting a healthy public sphere for debate.

In contrast, the medical, education, and public policy domains are generally more conservative and constrained by various legal rules and ethical ideals. These professionalized domains have specialized education, licensing, and accreditation procedures, are subject to high standards of care, often involve a commitment to public service, and typically rely on a presupposed political or moral vision of what a "good" society or person looks like (see Mittelstadt, 2019). For example, the general goal of medicine is to promote the best possible health of the patient (Anderson & Anderson, 2007), and legal conventions such as *informed consent* are manifestations of Kantian ethical beliefs supporting the need to respect patients' personal autonomy and counter power asymmetries, thus making coercion, manipulation, or deception less likely (O'Neill, 2002). Likewise, the goal of education, in one prominent Western view, is to "emancipate" the minds of students from instrumental rationality and create citizens capable of reflective thinking and democratic participation (Dewey, 1903). Education thus has the normative goal of improving the autonomy of persons, enhancing their self-awareness, reasoning abilities, and ultimately their capacity to govern their own lives (Schaefer et al., 2014). Arguably the reward functions in these domains would by necessity include some measure of user well-being, flourishing, virtue, happiness, or capacity for democratic participation, respectively.

Further, unlike corporate platform-based personalization, public policy interventions are theoretically subject to political accountability and legitimacy mechanisms such as referenda, voting, and rules of transparency that require disclosure in democratic societies. These procedures are political and social analogs of reflective endorsement for social collectives. In medical and health contexts, professional associations, such as the American Medical Association, also function as self-governance mechanisms that impose sanctions on members who violate community ethics codes (Mittelstadt, 2019). Furthermore, tort law is designed to assign legal responsibility for adverse events and permits persons to seek damages for malpractice or negligence. However, no such system of legal responsibility currently exists for corporate platforms or for

individual data scientists employing artificial agents (Asaro, 2007). Finally, privacy laws generally place strict limits on the collection, analysis, and sharing of medical data and health records, thus potentially constraining data collection and the action and state space of reinforcement learning systems relative to purely commercial applications.

In the medical sphere, safety, value alignment, explainability, and causal domain knowledge are paramount, thus making completely autonomous reinforcement learning systems unfeasible or undesirable. Domain knowledge and commonsense reasoning possessed by human medical experts provide sanity checks on spurious correlations—often due to systematic mismeasurement (Mullainathan & Obermeyer, 2017)—that artificial agents may exploit when identifying optimal treatment sequences (Gottesman et al., 2019). Medical devices also undergo rigorous testing before being released to the general public, unlike commercial personalization technologies. The American Food and Drug Administration, for instance, requires AI-based decision support systems to be certified as Software as a Medical Device if used without an accountable physician (Coronato et al., 2020).

Another gap between data scientists in industry and doctors and healthcare professionals involves required training in human subjects research ethics and oaths to do no harm to patients. Medical ethics also distinguishes between therapeutic and non-therapeutic interventions (Brazier & Lobjoit, 2005): therapeutic interventions are expected to benefit the individual patient in question, whereas non-therapeutic interventions may produce no benefit or even a small harm to the patient. The random exploratory actions of platform-based adaptive personalization systems are thus non-therapeutic interventions, analogous to giving digital versions of sham surgeries and sugar pills to subsets of unlucky users.

The ethical standards of academic and industry research differ as well. Since the Declaration of Helsinki, academic data scientists in many countries who collect personal data and interact with human subjects, whether online or offline, have been required to receive Institutional Review Board approval. Their research must also meet accepted standards of confidentiality and data privacy. The Belmont principles guiding biomedical research, for instance, adopt a Kantian "deontological" ethics—similar to the GDPR—and put prima facie duties on scientists in their pursuit of knowledge. Scientific knowledge is considered legitimately acquired when, among other things, it respects the autonomy of research participants (i.e., documenting clear and unambiguous consent) and distributes harms and benefits according to accepted principles of justice (Anderson & Anderson, 2007). For this reason, an industry-academic collaborative study about an emotional contagion that manipulated the News Feeds of over 680,000 Facebook users ignited debates around the ethical standards of platform research (Kramer et al., 2014).

In summary, the optimization-driven, platform-centric consequentialist ethics of reinforcement learning (Card & Smith, 2020) do not clearly or easily map to high-stakes noncommercial contexts. While some may argue that businesses have no social responsibilities beyond profit making, in public policy, education, and medicine at least, greater consensus exists on the moral legitimacy of various intentions and goals; the relationships, rights, and duties of those involved; and the design of public-facing accountability mechanisms when change is needed or when questions of moral and legal responsibility must be decided.

The ethical issues facing reinforcement learning in education are admittedly harder to identify than in public policy, health, and medicine. In the United States at least, education is perhaps seen less and less as a public good—innovation and "marketization" are keywords used to promote the commodification of specialized teaching and research (Olssen & Peters, 2007). While the growth of major education platforms such as Coursera and EdX expand global access to educational materials and make a data-intensive, trial-and-error approach to personalization feasible, basic questions regarding conflicting values and the nature and purposes of education—including their translation into an appropriate reward function—will need to be addressed. To this end, value alignment (van de Poel, 2020) and ethics-based auditing (Mökander et al., 2021) techniques warrant further consideration.

6 Conclusion

Pervasive implicit behavioral data collection combined with reinforcement learning has ushered in a new paradigm of platform-based personalization. In the instrumental pursuit of platform business goals, human persons are now abstracted as environments to control via complex sequences of algorithmic interventions. We identify five emergent features of this new paradigm that raise novel ethical, social, and political issues not easily addressed by current data protection legislation. Nevertheless, by embracing its sociotechnical legacy (Sarker et al., 2019), the IS community is uniquely positioned to lead the way toward ethically aware personalization research. We propose three research directions that involve designing, implementing, and using simulators and sandboxes, as well as fostering collaborations with critical IS, AI ethics, and business ethics researchers in order to tackle the technical, sociotechnical, and ethical problems raised by reinforcement learning-based personalization. These future research directions may also mitigate some of the challenges of

reproducibility, generalizability, and data collection that confront empirical researchers and data scientists when conducting platform-based studies.

IS researchers are now studying how the scale, scope, and speed of reinforcement learning-based "metahuman systems" (Lyytinen et al., 2021) may impact organizations and the future of work. Increasingly adaptive, data-hungry, and automated algorithms—if not carefully designed and monitored—can learn to modify our behavior and attitudes, thereby diminishing our autonomy as self-determining persons and destabilizing our social and political systems. It is imperative to identify and address relevant ethical, social, and political issues before more invasive physiological sources of implicit behavioral data, such as gaze detection, emotion recognition, brain activity, and eye-tracking data, become new learning signals for artificial agents designed to control our behavior at school, work, and other public spaces. When fused with augmented reality and synthetically generated text, audio, and video, personalized reinforcement learning may pose an even greater risk to persons and society. Moreover, reinforcement learning-based personalization technologies will undoubtedly raise new and complex ethical challenges in the higher-stakes domains of medicine, public health, and education.

In closing, we reiterate our plea to consider the ethical implications of advances in personalization technology. Taking the person seriously requires respecting our capacity for autonomy. Toward this normative goal, Dourish and Mainwaring (2012) remind us that before we ask the question of *What might we build tomorrow?* we should first consider our responsibility for what we built yesterday.

References

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2), i-xxxii.

Abdollahpouri, H., Burke, R., & Mobasher, B. (2017). Recommender systems as multistakeholder environments. *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, 347-348.

Abel, D., MacGlashan, J., & Littman, M. L. (2016). Reinforcement learning as a framework for ethical decision making. *Proceedings of the Workshops at the 30th AAAI Conference on Artificial Intelligence*.

Adomavicius, G., & Tuzhilin, A. (2005). Personalization technologies: A process-oriented perspective. *Communications of the ACM*, *48*(10), 83-90.

Aggarwal, C. C. (2016). *Recommender systems*. Springer.

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing*, *91*(1), 34-49.

Aizenberg, E., & van den Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society*, *7*(2), https://doi.org/10.1177/2053951720949566.

Allen, H. J. (2020). Experimental strategies for regulating fintech. *Journal of Law & Innovation*, *3*(1), Article 1.

Anderson, M., & Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. *AI Magazine*, *28*(4), 15.

Armstrong, S., & Mindermann, S. (2018). Occam's razor is insufficient to infer the preferences of irrational agents. Proceedings of the 32nd International Conference on Neural Information Processing SystemsAsaro, P. M. (2007). Robots and responsibility from a legal perspective. *Proceedings of the IEEE*, *4*(14), 20-24.

Avram, M., Micallef, N., Patil, S., & Menczer, F. (2020). *Exposure to social engagement metrics increases vulnerability to misinformation*. Harvard Kennedy School Misinformation Review. https://misinforeview.hks.harvard.edu/article/exposure-to-social-engagement-metrics-increases-vulnerability-to-misinformation/

Badri, S., Prudhvi, R., Lee, K., Lee, D., Tran, T., & Zhang, J. (2016). Uncovering fake likers in online social networks. *Proceedings of the 25th ACM International Conference on Information and Knowledge Management* (pp. 2365-2370).

Bak-Coleman, J. B., Alfano, M., Barfuss, W., Bergstrom, C. T., Centeno, M. A., Couzin, I. D., Donges, J. F., Galesic, M., Gersick, A. S., & Jacquet, J. (2021). Stewardship of global collective behavior. *PNAS*, *118*(27), Article e2025764118.

Balkin, J. (2017). 2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy: The three laws of robotics in the age of big data. *Ohio State Law Journal*, *78* (5), 1217-1241.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462.

Barricelli, B. R., Casiraghi, E., & Fogli, D. (2019). A survey on digital twin: Definitions, characteristics, applications, and design implications. In *IEEE Access 7*, 167653-167671.

Bassen, J., Balaji, B., Schaarschmidt, M., Thille, C., Painter, J., Zimmaro, D., Games, A., Fast, E., & Mitchell, J. C. (2020). Reinforcement learning for the adaptive scheduling of educational activities. *Proceedings of the Conference on Human Factors in Computing Systems*.

Baum, W. M. (2017). *Understanding behaviorism: Behavior, culture, and evolution*. Wiley.

Baumeister, R. E., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). *Ego depletion: Is the active self a limited resource? Journal of Personality and Social Psychology*, *74*(5), 1252–1265.

Bellman, R., & Lee, E. (1984). History and development of dynamic programming. *IEEE Control Systems Magazine*, *4*(4), 24-28.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798-1828.

Berkovsky, S., Freyne, J., & Oinas-Kukkonen, H. (2012). Influencing individually: fusing personalization and persuasion. *ACM Transactions on Interactive Intelligent Systems 2*(2) 1-8.

Bermúdez, J. L. (1995). Nonconceptual content: From perceptual experience to subpersonal

computational states. *Mind & Language*, *10*(4), 333-369.

Berridge, K. C. (2009). Wanting and liking: Observations from the neuroscience and psychology laboratory. *Inquiry*, *52*(4), 378-398.

Bhattacharjee, A., Berger, J., & Menon, G. (2014). When identity marketing backfires: Consumer agency in identity expression. *Journal of Consumer Research*, *41*(2), 294-309.

Binns, R. (2018). Algorithmic accountability and public reason. *Philosophy & Technology*, *31*(4), 543-556.

Bird, S., Barocas, S., Crawford, K., Diaz, F., & Wallach, H. (2016). Exploring or exploiting? Social and ethical implications of autonomous experimentation in AI. *Proceedings of the Workshop on Fairness, Accountability, and Transparency in Machine Learning*.

Bishop, M. (2002). *Computer security: Art and science*. Addison-Wesley.

Bodenhamer, D. J., & Ely, J. W. (2008). *The Bill of Rights in modern America*. Indiana University Press.

Borgeaud, S., Mensch, A., Hoffmann, J., Cai, T., Rutherford, E., Millican, K., Driessche, G. van den, Lespiau, J.-B., Damoc, B., Clark, A., Casas, D. de las, Guy, A., Menick, J., Ring, R., Hennigan, T., Huang, S., Maggiore, L., Jones, C., Cassirer, A., … Sifre, L. (2021). *Improving language models by retrieving from trillions of tokens*. Available at http://arxiv.org/abs/2112.04426.

Bostrom, N. (2012). The superintelligent will: Motivation and instrumental rationality in advanced artificial agents. *Minds and Machines*, *22*(2), 71-85.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Botvinick, M., Wang, J. X., Dabney, W., Miller, K. J., & Kurth-Nelson, Z. (2020). Deep reinforcement learning and its neuroscientific implications. *Neuron*, *107*(4), 603-616.

Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances*, *7*(33).

Brazier, M., & Lobjoit, M. (Eds.) (2005). *Protecting the vulnerable: Autonomy and consent in health care*. Routledge.

Briggs, G. M., & Scheutz, M. (2015). "Sorry, I can't do that": Developing mechanisms to appropriately reject directives in human-robot interactions. *Proceedings of the 2015 AAAI Fall Symposium Series*.

Brusilovsky, P., & Maybury, M. T. (2002). From adaptive hypermedia to the adaptive web. *Communications of the ACM*, *45*(5), 30-33.

Burr, C., Cristianini, N., & Ladyman, J. (2018). An analysis of the interaction between intelligent software agents and human users. *Minds and Machines*, *28*(4), 735-774.

Cai, L., Wu, C., Meimandi, K. J., & Gerber, M. S. (2017). Adaptive mobile behavior change intervention using reinforcement learning. *Proceedings of the 2017 International Conference on Companion Technology*.

Card, D., & Smith, N. A. (2020). On Consequentialism and Fairness. *Frontiers in Artificial Intelligence*, *3*, Article 34.

Chaney, A. J. B., Stewart, B. M., & Engelhardt, B. E. (2018). How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. *Proceedings of the 12th ACM Conference on Recommender Systems* (pp. 224-232).

Chen, M., Beutel, A., Covington, P., Jain, S., Belletti, F., & Chi, E. H. (2019). Top-k off-policy correction for a REINFORCE recommender system. *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining* (pp. 456-464).

Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. *Proceedings of the 31st International Conference on Neural Information Processing Systems*

Christman, J. (1991). Autonomy and personal history. *Canadian Journal of Philosophy*, *21*(1), 1-24.

Churchill, E. F. (2013). Putting the person back into personalization. *Interactions*, *20*(5), 12-15.

Citron, D. K. (2007). Technological due process. *Washington University Law Review*, *85*(6), 1249-1313.

Collins, L. M., Murphy, S. A., & Bierman, K. L. (2004). A conceptual framework for adaptive preventive interventions. *Prevention Science*, *5*(3), 185-196.

Coors, C. (2010). Headwind from Europe: The new position of the German courts on personality rights after the judgment of the European Court of Human Rights. *German Law Journal*, *11*(5), 527-537.

Coronato, A., Naeem, M., de Pietro, G., & Paragliola, G. (2020). Reinforcement learning for intelligent healthcare applications: A survey. *Artificial Intelligence in Medicine*, *109*, Article 101964.

Corral-Acero, J., Margara, F., Marciniak, M., Rodero, C., Loncaric, F., Feng, Y., Gilbert, A., Fernandes, J. F., Bukhari, H. A., Wajdan, A., Martinez, M. V., Santos, M. S., Shamohammdi, M., Luo, H., Westphal, P., Leeson, P., DiAchille, P., Gurev, V., Mayr, M., … Lamata, P. (2020). The "digital twin" to enable the vision of precision cardiology. *European Heart Journal*, *41*(48), 4556-4564.

Costa, L. (2012). Privacy and the precautionary principle. *Computer Law & Security Review*, *28*(1), 14-24.

Couldry, N., & Mejias, U. A. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media*, *20*(4), 336-349.

Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. *Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 191-198).

Cristianini, N., Scantamburlo, T., & Ladyman, J. (2021). The social turn of artificial intelligence. *AI & Society*, *38*, 89-96.

Dan, O., & Loewenstein, Y. (2019). From choice architecture to choice engineering. *Nature Communications*, *10*(1), 1-4.

den Hengst, F., Grua, E. M., el Hassouni, A., & Hoogendoorn, M. (2020). Reinforcement learning for personalization: A systematic literature review. *Data Science*, *3*(2), 107-147.

Dennett, D. C. (2002). *Content and consciousness*. Routledge.

Dewey, J. (1903). The elementary school teacher. *Democracy in Education*, *4*(4), 193-204.

Dezfouli, A., Nock, R., & Dayan, P. (2020). Adversarial vulnerabilities of human decision-making. *PNAS*, *117*(46), 29221-29228.

Dong, J. Q. (2022). Using simulation in information systems research. *Journal of the Association for Information Systems*, *23*(2), 408-417.

Dourish, P., & Mainwaring, S. D. (2012). Ubicomp's Colonial Impulse. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*.

Du, S., & Xie, C. (2021). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. *Journal of Business Research*, *129*, 961-974.

Dutt, R., Deb, A., & Ferrara, E. (2018). "Senator, we sell ads": Analysis of the 2016 Russian Facebook ads campaign. *Communications in Computer and Information Science*, *941*, 151-168.

Ekstrand, M. D., & Willemsen, M. C. (2016). Behaviorism is not enough: Better recommendations through listening to users. *Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 221-224).

Elster, J. (2016). *Sour grapes: Studies in the subversion of rationality*. Cambridge University Press.

European Commission. (2021). *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.* https://digital-strategy.ec.europa.eu/en/library/604proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

European Data Protection Board. (2020). *Guidelines 05/2020 on consent under Regulation 2016/679*. https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf

Evans, C., & Kasirzadeh, A. (2021). *User tampering in reinforcement learning recommender systems*. Avalilable at https://arxiv.org/pdf/2109.04083.

Everitt, T., Hutter, M., Kumar, R., & Krakovna, V. (2021). Reward tampering problems and solutions in reinforcement learning: A causal influence diagram perspective. *Synthese*, *198*, 6435-6467.

Eyal, N. (2014). *Hooked: How to build habit-forming products*. Penguin.

Fedus, W., Ramachandran, P., Agarwal, R., Bengio, Y., Larochelle, H., Rowland, M., & Dabney, W. (2020). Revisiting fundamentals of experience replay. *Proceedings of the International Conference on Machine Learning* (pp. 3061-3071).

Figueroa, C. A., Hernandez-Ramos, R., Boone, C. E., Gómez-Pathak, L., Yip, V., Luo, T., Sierra, V., Xu, J., Chakraborty, B., Darrow, S., & Aguilera, A. (2021). A text messaging intervention for coping with social distancing during COVID-19 (StayWell at Home): Protocol for a randomized controlled trial. *JMIR Research Protocols*, *10*(1), Article e23592.

Floridi, L. (2011). The informational nature of personal identity. *Minds and Machines*, *21*(4), 549-566.

Floridi, L. (2016). On human dignity as a foundation for the right to privacy. *Philosophy & Technology*, *29*(4), 307-312.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., & Rossi, F. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689-707.

Fogg, B. J. (2003). *Persuasive technology: Using computers to change what we think and do*. Elsevier.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, *68*(1), 5-20.

Friedman, B., & Hendry, D. (2019). *Value sensitive design: Shaping technology with moral imagination*. MIT Press.

Friedman, M. (2007). The social responsibility of business is to increase its profits. In W.C. Zimmerli, M. Holzinger, K. Richter (Eds.), *Corporate ethics and corporate governance* (pp. 173-178). Springer.

Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines*, *30*(3), 411-437.

Gauci, J., Ghavamzadeh, M., Honglei, L., & Nahmias, R. (2019, October 19). Open-sourcing ReAgent, a modular, end-to-end platform for building reasoning systems. *Meta*. https://ai.meta.com/blog/open-sourcing-reagent-a-platform-for-reasoning-systems/

Gottesman, O., Johansson, F., Komorowski, M., Faisal, A., Sontag, D., Doshi-Velez, F., & Celi, L. A. (2019). Guidelines for reinforcement learning in healthcare. *Nature Medicine 2019 25:1*, *25*(1), 16-18.

Gottlieb, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*(11), 585-593.

Greene, T., Martens, D., & Shmueli, G. (2022). Barriers to academic data science research in the new realm of algorithmic behaviour modification by digital platforms. *Nature Machine Intelligence*, *4*(4), 323-330.

Greene, T., Shmueli, G., Ray, S., & Fell, J. (2019). Adjusting to the GDPR: The impact on data scientists and behavioral researchers. *Big Data*, *7*(3), 140-162.

Grover, V., Lindberg, A., Benbasat, I., & Lyytinen, K. (2020). The perils and promises of big data research in information systems. *Journal of the Association for Information Systems*, *21*(2), 268-291.

Gu, S., Lillicrap, T., Sutskever, I., & Levine, S. (2016). Continuous deep *q*-learning with model-based acceleration. *Proceedings of the International Conference on Machine Learning* (pp. 2829-2838).

Hadfield-Menell, D., Russell, S. J., Abbeel, P., & Dragan, A. (2016). Cooperative inverse reinforcement learning. *Proceedings of the 30th International Conference on Neural Information Processing Systems* (pp. 3909-3917).

Haugen, F. (2021). *Statement of Frances Haugen*: *Whistleblower aid.* U.S. Senate Committee on Commerce, Science, & Transportation. https://www.commerce.senate.gov/services/files/589FC8A558E-824E-4914-BEDB-3A7B1190BD49

Helmond, A. (2015). The platformization of the web: Making web data platform ready. *Social Media+ Society*, *1*(2), 1-11

Helwig, C. C. (2006). The development of personal autonomy throughout cultures. *Cognitive Development*, *21*(4), 458-473.

Hildebrandt, M. (2015). *Smart technologies and the end (s) of law: novel entanglements of law and technology*. Edward Elgar.

Hildebrandt, M. (2022). The issue of proxies and choice architectures. Why EU law matters for recommender systems. *Frontiers in Artificial Intelligence*, *5*, Article 789076.

Hildebrandt, M., & de Vries, K. (2013). *Privacy, due process and the computational turn: The philosophy of law meets the philosophy of technology*. Routledge.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504-507.

Hirschheim, R., & Klein, H. K. (1994). Realizing emancipatory principles in information systems development: The case for ETHICS. *MIS Quarterly*, *18*(1), 83-109.

Ho, M. K., MacGlashan, J., Littman, M. L., & Cushman, F. (2017). Social is special: A

normative framework for teaching with and learning from evaluative feedback. *Cognition*, *167*, 91-106.

Hutsebaut-Buysse, M., Mets, K., & Latré, S. (2022). Hierarchical reinforcement learning: A survey and open research challenges. *Machine Learning and Knowledge Extraction*, *4*(1), 172-221.

Ibarz, J., Tan, J., Finn, C., Kalakrishnan, M., Pastor, P., & Levine, S. (2021). How to train your robot with deep reinforcement learning: Lessons we have learned. *The International Journal of Robotics Research*, *40*(4-5), 698-721.

Ie, E., Hsu, C., Mladenov, M., Jain, V., Narvekar, S., Wang, J., Wu, R., & Boutilier, C. (2019). *Recsim: A configurable simulation platform for recommender systems.* Available at https://arxiv.org/abs/1909.04847.

Jia, K., Kenney, M., Mattila, J., & Seppala, T. (2018). *The application of artificial intelligence at Chinese digital platform giants: Baidu, Alibaba and Tencent* (ETLAReport 81). https://www.etla.fi/wp-content/uploads/ETLA-Raportit-Reports-81.pdf

Kane, G. C., Young, A. G., Majchrzak, A., & Ransbotham, S. (2021). Avoiding an oppressive future of machine learning: A design theory for emancipatory assistants. *MIS Quarterly*, *45*(1), 371-396.

Kant, I. (1948). Groundwork of the metaphysics of morals (H. J. Paton, Trans.). Hutchinson. (Original work published 1785)

Kant, I. (2017). *Kant: The metaphysics of morals*. (M. Gregor, Trans., L. Denis, Ed.). Cambridge University Press. (Original work published 1797).

Kaptein, C. M., Panos, M., de Ruyter, B., & Aarts, E. (2009). Persuasion in ambient intelligence. *Journal of Ambient Intelligence and Humanized Computing*, *1*(1), 43-56.

Karppi, T., & Crawford, K. (2015). Social media, financial algorithms and the hack crash. *Theory, Culture & Society*, *33*(1), 73-92.

Kenton, Z., Everitt, T., Weidinger, L., Gabriel, I., Mikulik, V., & Deepmind, G. I. (2021). *Alignment of language agents*. Available at https://arxiv.org/abs/2103.14659v1.

Kim, Y., Hassan, A., White, R. W., & Zitouni, I. (2014). Modeling dwell time to predict click-level satisfaction. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining* (pp. 193-202).

Kornblith, H. (2010). What reflective endorsement cannot do. *Philosophy and Phenomenological Research*, *80*(1), 1-19.

Korsgaard, C. M. (1989). Personal identity and the unity of agency: A Kantian response to Parfit. *Philosophy & Public Affairs*, *18*(2), 101-132.

Krakovna, V., Uesato, J., Mikulik, V., Rahtz, M., Everitt, T., Kumar, R., Kenton, Z., Leike, J., & Legg, S. (2020, April 21). Specification gaming: the flip side of AI ingenuity. *Deepmind*. https://deepmind.com/blog/article/Specification-gaming-the-flip-side-of-AI-ingenuity

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788-8790.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems* (pp. 1097-1105).

Krueger, D., Maharaj, T., & Leike, J. (2020). Hidden *Incentives for auto-induced distributional shift*. Available at https://arxiv.org/abs/2009.09153.

Kumar, R., Uesato, J., Ngo, R., Everitt, T., Krakovna, V., & Legg, S. (2020). *REALab: An embedded perspective on tampering.* Available at https://arxiv.org/abs/2011.08820.

Langley, P., & Leyshon, A. (2017). Platform capitalism: the intermediation and capitalization of digital economic circulation. *Finance and Society*, *3*(1), 11-31.

le Moing, G., Ponce, J., & Schmid, C. (2021). CCVS: Context-aware controllable video synthesis. *Proceedings of the 34th International Conference on Neural Information Processing Systems*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.

Leidner, D. E., & Tona, O. (2021). The care theory of dignity amid personal data digitalization. *MIS Quarterly*, *45*(1), 343-370.

Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose: The origins and value of the need for control. *Trends in Cognitive Sciences*, *14*(10), 457-463.

Levine, S. (2021). *Understanding the world through action.* Available at https://arxiv.org/abs/2110.12543.

Li, G., Gomez, R., Nakamura, K., & He, B. (2019). Human-centered reinforcement learning: A survey. *IEEE Transactions on Human-Machine Systems*, *49*(4), 337-349.

Lin, L.-J. (1992). Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning*, *8*, 293-321.

Lindström, B., Bellander, M., Schultner, D. T., Chang, A., Tobler, P. N., & Amodio, D. M. (2021). A computational reward learning account of social media engagement. *Nature Communications*, *12*(1), 1-10.

Littman, M. L. (2015). Reinforcement learning improves behaviour from evaluative feedback. *Nature*, *521*(7553), 445-451.

Liu, H., & Abbeel, P. (2021). *Behavior from the void: Unsupervised active pre-training.* Available at https://arxiv.org/abs/2103.04551.

Liu, P., & Chao, W. (2020). Computational advertising: Market and technologies for internet commercial monetization. In *Computational advertising*. CRC Press.

Lyngs, U., Lukoff, K., Slovak, P., Binns, R., Slack, A., Inzlicht, M., van Kleek, M., & Shadbolt, N. (2019). Self-control in cyberspace: Applying dual systems theory to a review of digital self-control tools. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*.

Lynskey, O. (2015). *The foundations of EU data protection law*. Oxford University Press.

Lyytinen, K., & Hirschheim, R. (1988). Information systems as rational discourse: an application of Habermas's theory of communicative action. *Scandinavian Journal of Management*, *4*(1-2), 19-30.

Lyytinen, K., Nickerson, J. v., & King, J. L. (2021). Metahuman systems = humans + machines that learn. *Journal of Information Technology*, *36*(4), 427-445.

Mahmood, T., & Ricci, F. (2007). Learning and adaptivity in interactive recommender systems. *Proceedings of the 9th International Conference on Electronic Commerce* (pp. 75-84).

Manheim, D., & Garrabrant, S. (2018). *Categorizing variants of Goodhart's Law.* Available at https://arxiv.org/abs/1803.04585.

Marcus, G. (2009). *Kluge: The haphazard evolution of the human mind*. Houghton Mifflin Harcourt.

Marhold, K., & Fell, J. (2021). Electronic vaccination certificates: avoiding a repeat of the contact-tracing "format wars." *Nature Medicine*, *27*(5), 738-739.

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, *114*(48), 12714-12719.

McInerney, J., Lacker, B., Hansen, S., Higley, K., Bouchard, H., Gruson, A., & Mehrotra, R. (2018). Explore, exploit, and explain: Personalizing explainable recommendations with bandits. *Proceedings of the 12th ACM Conference on Recommender Systems* (pp 31-39).

Meinert, C. L. (1980). Toward more definitive clinical trials. *Controlled Clinical Trials*, *1*, 249-261.

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., Wood, C. E., Michie, S., Johnston, M., Wood, C. E., Richardson, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., & Cane, J. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, *46*(1), 81-95.

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *AI and Society*, *35*(4), 957-967.

Mill, J. S. (2015). *On liberty, utilitarianism and other essays* (M. Philp & F. Rosen, Eds.). Oxford University Press.

Miller, R. L., Brickman, P., & Bolen, D. (1975). Attribution versus persuasion as a means for modifying behavior. *Journal of Personality and Social Psychology*, *31*(3), 430-441.

Mills, S. (2022). Personalized nudging. *Behavioural Public Policy*, *6*(1), 150-159.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501-507.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., & Ostrovski, G. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529-533.

Mohai, P., Pellow, D., & Roberts, J. T. (2009). Environmental justice. *Annual Review of Environment and Resources*, *34*, 405-430.

Mökander, J., Morley, J., Taddeo, M., & Floridi, L. (2021). Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. *Science and Engineering Ethics*, *27*(4).

Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research*, *26*(4), 323-339.

Mullainathan, S., & Obermeyer, Z. (2017). Does Machine learning automate moral hazard and error? *American Economic Review*, *107*(5), 476-480.

Murthi, B. P. S., & Sarkar, S. (2003). The role of the management sciences in research on personalization. *Management Science*, *49*(10), 1344-1362.

Myers M D, & Klein H K. (2011). A set of principles for conducting critical research in information systems. *MIS Quarterly*, *35*(1), 17-36.

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, *52*(6), 446-462.

Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), Article 20180089.

Ng, A. Y., & Russell, S. (2000). Algorithms for inverse reinforcement learning. *Proceedings of the 17th International Conference on Machine Learning.*

Nissenbaum, H. (2004). Privacy as contextual integrity. *Wash. L. Rev.*, *79*, 119.

Nussbaum, M. C. (2001). *Women and human development: The capabilities approach*. Cambridge University Press.

Official Journal of the European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27595 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 (General Data Protection Regulation)*. EUR-Lex. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Olssen, M., & Peters, M. A. (2007). Neoliberalism, higher education and the knowledge economy: From the free market to knowledge capitalism. *Journal of Education Policy*, *20*(3), 313-345.

O'Neill, O. (2002). *Autonomy and trust in bioethics*. Cambridge University Press.

OpenAI. (2022, November 30). ChatGPT: Optimizing language models for dialogue. *OpenAI Blog.* https://openai.com/blog/chatgpt

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., & Ray, A. (2022). *Training language models to follow instructions with human feedback.* Available at https://arxiv.org/abs/2203.02155.

Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.

Pasquale, F. (2015). *The black box society*. Harvard University Press.

Pei, C., Lin, X., Yang, X., Sun, F., Cui, Q., Jiang, P., Ou, W., & Zhang, Y. (2019). Value-aware recommendation based on reinforcement profit maximization. *Proceedings of the World Wide Web Conference* (pp. 3123-3129).

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39-50.

Persons, T. M. (2018). *Artificial intelligence: Emerging opportunities, challenges, and implications for policy and research* (GAO18-644T). U.S. Government Accountability Office. https://www.gao.gov/products/gao-18-644t

Posner, E. A., & Weyl, E. G. (2019). *Radical markets*. Princeton University Press.

Prunkl, C. (2022). Human autonomy in the age of artificial intelligence. *Nature Machine Intelligence*, *4*(2), 99-101.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., & Jackson, M. O. (2019). Machine behaviour. *Nature*, *568*(7753), 477-486.

Rasheed, A., San, O., & Kvamsdal, T. (2020). Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access*, *8*, 21980-22012.

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Raz, J. (1987). Government by consent. *NOMOS*, *29*, 76-95.

Rhoen, M. (2017). Rear view mirror, crystal ball: Predictions for the future of data protection law based on the history of environmental protection law. *Computer Law & Security Review*, *33*(5), 603-617.

Riedmiller, M., Springenberg, J. T., Hafner, R., & Heess, N. (2021). *Collect & infer: A fresh look at data-efficient reinforcement learning*. Available at https://arxiv.org/abs/2108.10273.

Rohde, D., Bonner, S., Dunlop, T., Vasile, F., & Karatzoglou, A. (2018). *RecoGym: A reinforcement learning environment for the problem of product recommendation in online advertising.* Available at https://arxiv.org/abs/1808.00720.

Rollwage, M., & Fleming, S. M. (2021). Confirmation bias is adaptive when coupled with efficient metacognition. *Philosophical Transactions of the Royal Society B*, *376*(1822).

Ropohl, G. (1999). Philosophy of socio-technical systems. *Society for Philosophy and Technology Quarterly Electronic Journal*, *4*(3), 186-194.

Rouvroy, A., & Poullet, Y. (2009). The right to informational self-determination and the value of self-development: Reassessing the importance of privacy for democracy. In S. Gutwirth, Y. Poullet, P. De Hert, C. de Terwangne, & S. Nouwt (Eds.), *Reinventing data protection?* (pp. 45-76). Springer.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206-215.

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Penguin.

Russell, S., Dewey, D., & Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. *AI Magazine*, *36*(4), 105-114.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford.

Ryan, T., Chester, A., Reece, J., & Xenos, S. (2014). The uses and abuses of Facebook: A review of Facebook addiction. *Journal of Behavioral Addictions*, *3*(3), 133-148.

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. *MIS Quarterly*, *43*(3), 695-719.

Schaefer, G. O., Kahane, G., & Savulescu, J. (2014). Autonomy and enhancement. *Neuroethics*, *7*(2), 123-136.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85-117.

Schnabel, T., Swaminathan, A., Singh, A., Chandak, N., & Joachims, T. (2016). Recommendations as treatments: Debiasing learning and evaluation. *International Conference on Machine Learning*, 1670-1679.

Schor, J. (2016). Debating the sharing economy. *Journal of self-governance and management economics*, *4*(3), 7-22.

Schuler, D., & Namioka, A. (Eds.) (1993). *Participatory design: Principles and practices*. CRC Press.

Sell, J. (2008). Introduction to deception debate: *Social Psychology Quarterly*, *71*(3), 213-214.

Sen, A. (1984). *Resources, values and development*. Harvard University Press.

Sharma, P. (2020). *Coronavirus news, markets and AI: The COVID-19 diaries.* Routledge India.

Shi, J. C., Yu, Y., Da, Q., Chen, S. Y., & Zeng, A. X. (2019). Virtual-Taobao: Virtualizing real-world online retail environment for reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence*.

Shmueli, G., & Tafti, A. (2023). How to "improve" prediction using behavior modification. International *Journal of Forecasting*, *39*(2), 541-555.

Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence*, *299*, Article 103535.

Singh, S., Lewis, R. L., & Barto, A. G. (2009). Where do rewards come from? *Proceedings of the Annual Conference of the Cognitive Science Society* (pp. 2601-2606).

Smith, B. (2021). How TikTok reads your mind. *The New York Times*. www.nytimes.com/2021/12/05/business/media/tiktok-algorithm.html

Soman, D., & Yeung, C. (2020). *The behaviourally informed organization*. University of Toronto Press.

Stahl, B. C. (2008). The ethical nature of critical research in information systems. *Information Systems Journal*, *18*(2), 137-163.

Stahl, B. C. (2022). Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence. *International Journal of Information Management*, *62*, Article 102441.

Stahl, B. C., Andreou, A., Brey, P., Hatzakis, T., Kirichenko, A., Macnish, K., Shaelou, S. L., Patel, A., Ryan, M., & Wright, D. (2021). Artificial intelligence for human flourishing-Beyond principles for machine learning. *Journal of Business Research*, *124*, 374-388.

Su, C., Zhou, H., Gong, L., Teng, B., Geng, F., & Hu, Y. (2021). Viewing personalized video clips recommended by TikTok activates default mode network and ventral tegmental area. *NeuroImage*, *237*, 118136.

Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet Policy Review*, *8*(2), https://doi.org/10.14763/2019.2.1410

Sutton, R. (2019, March 13). The bitter lesson. *Incomplete Ideas* http://www.incompleteideas.net/IncIdeas/BitterLesson.html

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.

Tang, P. (2017). Reinforcement mechanism design. *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 5146-5150).

Teschl, M., & Comim, F. (2005). Adaptive preferences and capabilities: Some preliminary conceptual explorations. *Review of Social Economy*, *63*(2), 229-247.

Theocharous, G., Mahadevan, S., Healey, J., & Saad, M. (2019). Personalizing with human cognitive biases. *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization* (pp. 13-17).

Theocharous, G., Thomas, P. S., & Ghavamzadeh, M. (2015). Ad recommendation systems for life-time value optimization. *Proceedings of the 24th International Conference on World Wide Web* (pp. 1305-1310).

Tindell, A. J., Smith, K. S., Berridge, K. C., & Aldridge, J. W. (2009). Dynamic computation of incentive salience: "Wanting" what was never "liked." *The Journal of Neuroscience*, *29*(39), 12220-12228.

Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information Systems Research*, *21*(4), 675-687.

Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., & Abbeel, P. (2017). Domain randomization for transferring deep neural networks from simulation to the real world. *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*.

Touretzky, D. S., & Saksida, L. M. (1997). Operant conditioning in Skinnerbots. *Adaptive Behavior*, *5*(3-4), 219-247.

Urbina, F., Lentzos, F., Invernizzi, C., & Ekins, S. (2022). Dual use of artificial-intelligence-powered drug discovery. *Nature Machine Intelligence*, *4*(3), 189-191.

van de Poel, I. (2020). Embedding values in artificial intelligence (AI) systems. *Minds and Machines*, *30*(3), 385-409.

van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford University Press.

van Doorn, J., Lemon, K. N., Mittal, V., Nass, S., Pick, D., Pirner, P., & Verhoef, P. C. (2010). Customer engagement behavior: Theoretical foundations and research directions. *Journal of Service Research*, *13*(3), 253-266.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 5998-6008).

Wang, K., Zou, Z., Deng, Q., Shang, Y., Zhao, M., Wu, R., Shen, X., Lyu, T., & Fan, C. (2021). *RL4RS: A real-world benchmark for reinforcement learning based recommender system*. Available at https://arxiv.org/abs/2110.11073.

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3-4), 279-292.

Wegner, D. M. (2004). Précis of the illusion of conscious will. *Behavioral and Brain Sciences*, *27*(5), 649-659.

Wertenbroch, K., Schrift, R. Y., Alba, J. W., Barasch, A., Bhattacharjee, A., Giesler, M., Knobe, J., Lehmann, D. R., Matz, S., Nave, G., Parker, J. R., Puntoni, S., Zheng, Y., & Zwebner, Y. (2020). Autonomy in consumer choice. *Marketing Letters*, *31*(4), 429-439.

Wheeler, S. C., & Berger, J. (2007). When the same prime leads to different effects. *Journal of Consumer Research*, *34*(3), 357-368.

Whittlestone, J., Arulkumaran, K., & Crosby, M. (2021). The societal implications of deep

reinforcement learning. *Journal of Artificial Intelligence Research*, *70*, 1003-1030.

Wiener, N. (1988). *The human use of human beings: Cybernetics and society*. Da Capo Press.

Wiesel, T., Pauwels, K., & Arts, J. (2011). Practice prize paper—Marketing's profit impact: Quantifying online and off-line funnel progression. *Marketing Science*, *30*(4), 604-611.

Winick, B. J. (1992). On autonomy: Legal and psychological perspectives. *Villanova. Law Review*, *37*, 1705-1771.

Wolf, S. (1993). *Freedom within reason*. Oxford University Press on Demand.

Wong, K. L. X., & Dobson, A. S. (2019). We're just data: Exploring China's social credit system in relation to digital platform ratings cultures in Westernised democracies. *Global Media and China*, *4*(2), 220-232.

Wright, P. (2002). Marketplace Metacognition and Social Intelligence. *Journal of Consumer Research*, *28*(4), 677-682.

Wu, J., Zhang, Z., Feng, Z., Wang, Z., Yang, Z., Jordan, M. I., & Xu, H. (2022). *Sequential Information design: Markov persuasion process and its efficient reinforcement learning.* Available at https://arxiv.org/abs/2202.10678.

Wu, Y.-H., & Lin, S.-D. (2018). A low-cost ethics shaping approach for designing reinforcement learning agents. *Proceedings of the AAAI Conference on Artificial Intelligence*.

Xin, X., Karatzoglou, A., Arapakis, I., & Jose, J. M. (2020). Self-supervised reinforcement learning for recommender systems. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 931-940).

Yampolskiy, R. v. (2020). Unpredictability of AI: On the impossibility of accurately predicting all actions of a smarter agent. *Journal of Artificial Intelligence and Consciousness*, *7*(1), 109-118.

Yang, J., Eckles, D., Dhillon, P., & Aral, S. (2020). *Targeting for long-term outcomes.* Available at https://arxiv.org/abs/2010.15835.

Yeung, K. (2016). "Hypernudge": Big Data as a mode of regulation by design. *Information, Communication & Society*, *20*(1), 118-136.

Zarsky, T. Z. (2016). Incompatible: The GDPR in the age of big data. *Seton Hall Law Review*, *47*, 995-1018.

Zbontar, J., Jing, L., Misra, I., LeCun, Y., & Deny, S. (2021). Barlow twins: Self-supervised learning via redundancy reduction. *Proceedings of the 38th International Conference on Machine Learning*.

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2017). Deep Learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*, *52*(1), 1-38.

Zhao, Y., Kosorok, M. R., & Zeng, D. (2009). Reinforcement learning design for cancer clinical trials. *Statistics in Medicine*, *28*(26), 3294-3315.

Zhou, S., Dai, X., Chen, H., Zhang, W., Ren, K., Tang, R., & Yu, Y. (2020). Interactive recommender system via knowledge graph-enhanced reinforcement learning. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 179-188).

Zou, L., Song, J., Xia, L., Liu, W., Ding, Z., & Yin, D. (2019). Reinforcement learning to optimize long-term user engagement in recommender systems. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 2810-2818).

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books.

Zuiderveen Borgesius, F. J. (2015). Personal data processing for behavioural targeting: which legal basis? *International Data Privacy Law*, *5*(3), 163-176.

Appendix A: Markov Decision Processes and Reinforcement Learning

For illustrative purposes we consider reinforcement learning within the context of a Markov decision process (MDP) The MDP clarifies the mathematical assumptions of the problem and allows theoretical guarantees about the performance of the reinforcement learning algorithm to be derived. We begin with a system composed of two coupled elements, an agent and an environment, and specify five key components (Mahmood & Ricci, 2007): a set of states *S*, a set of actions *A*, a reward function *R(s,a)* defining rewards received by taking action $a \in A$ in state $s \in S$, and a transition function *T(s'|s, a)* defining the new state $s' \in S$ the agent transitions into when action $a \in A$ is taken in state $s \in S$. A discount factor may also be applied to future rewards. We also set a distribution of initial states from which agent-environment interactions begin. As the artificial agent interacts with its environment of human users, it collects and stores episodes of experience (state-action-reward tuples) at each time step until a terminal state is reached, at which point an entire learning episode, or trajectory, is completed. Figure A1 illustrates the MDP in the context of platform-based personalization.

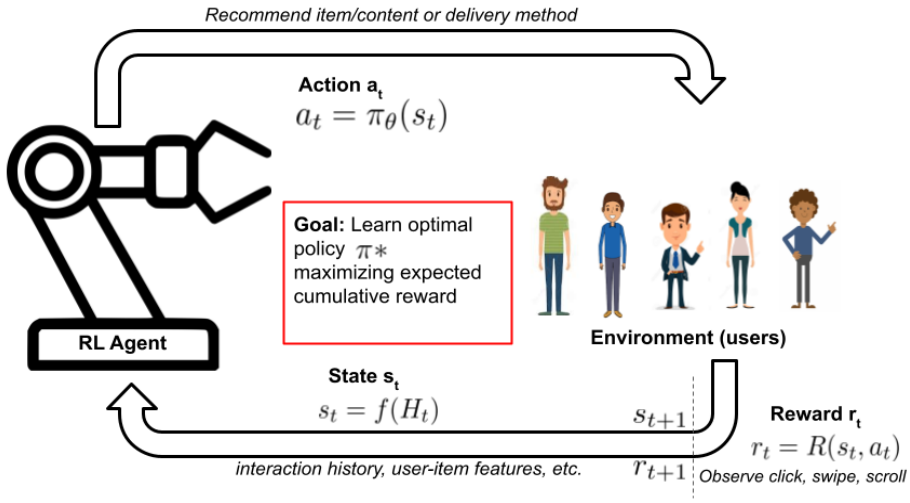

Figure A1. Schematic of Reinforcement Learning-based Personalization

To illustrate the key role of implicit behavioral interaction data in reinforcement learning, consider the *Q*-learning algorithm (Watkins & Dayan, 1992), which estimates a *Q*-value function for state-action pairs. One may think of the *Q*-value function as a predictive function of the expected return of taking an action in a state given the current policy. One interesting aspect of the *Q*-learning process relevant to personalization is that it can be facilitated by reusing previously collected interactions in a replay buffer (Fedus et al., 2020). The replay buffer allows for offline learning and retains previous interaction experience so as to improve sample efficiency and stability during the *Q*-function learning process (Lin, 1992). Once estimated precisely with a sufficient number of interactions, an optimal policy exemplifies the Bellman optimality principle (Bellman & Lee, 1984) by recommending the item with the highest *Q*-value at each time step.

Appendix B

Table B1. Reinforcement Learning Key Terms

Term	Description
Reinforcement learning	The problem of learning (via trial and error) to act in an environment using a reward (feedback) signal
Sequential neural network architectures	Neural network designs repurposed from natural language processing tasks involving sequential prediction and generation of word tokens. Recurrent neural networks (RNNs) and variations, such as long short-term memory (LSTMs) and gated recurrent units (GRUs), can also be used to capture and store sequential or temporal data for predictive tasks. Newer transformer-based architectures generate sequences of tokens, but use a self-attention mechanism instead of recurrence.
Agent/controller	The solution to the problem environment
Environment/controlled system	The problem to be solved by the agent
Multi-armed bandits	A special case of reinforcement learning with immediate rewards and single state environments
Markov decision process	Formalizes the reinforcement learning problem by assuming a set of states and actions and a state transition and reward function. Accounts for long-term value of actions
Markov property	The assumption that given the present state of the environment, the future state is conditionally independent of the past
Policy	A behavioral rule (function) for selecting actions in all possible states of the environment. Can be represented using a neural network, as in deep reinforcement learning
Value iteration methods	Learning a policy indirectly by finding the optimal value function
Policy iteration methods	Gradient based; learning by directly modifying the parameters of the policy function
On-policy learning	The agent learns a behavior policy while interacting with the environment
Off-policy learning	The agent learns a behavior policy from previously stored interactions, possibly collected from a different policy
Model-based reinforcement learning	The agent is provided with a model of the transition dynamics of the environment
Model-free reinforcement learning	The agent learns the transition structure of the environment through direct and repeated interaction

About the Authors

Travis Greene is an assistant professor at Copenhagen Business School's Department of Digitalization. He holds a PhD in business analytics from the Institute of Service Science at National Tsing Hua University in Taiwan. With a background in philosophy (BA/MA) and business (MBA), and research interests in personalization, data science ethics, and AI and data protection law, his research has appeared in journals such as *Nature Machine Intelligence, Journal of the Association for Information Systems, Journal of the Royal Statistical Society*, and *Big Data*. His interdisciplinary work aims to contribute new perspectives, frameworks, and ideas from the humanities into data science, and vice versa.

Galit Shmueli is a Tsing Hua Chair Professor at the Institute of Service Science, College of Technology Management, National Tsing Hua University in Taiwan. Her research focuses on statistical and machine learning methodology with applications in information systems and healthcare, and an emphasis on human behavior. She is the author of multiple books and over 100 peer-reviewed publications. She teaches and designs business analytics courses on machine learning, forecasting, and more. She is also the inaugural editor-in-chief of the *INFORMS Journal on Data Science*, and an IMS Fellow and ISI elected member.

Soumya Ray is a Distinguished Professor of the Institute of Service Science at National Tsing Hua University in Taiwan. Soumya earned his PhD at the University of Wisconsin-Madison. His empirical research investigates user behavior in online information systems, with a special emphasis on community and social networking. His methodological research seeks to apply predictive methods to explanatory modeling. His work has been published or is forthcoming in leading journals such as *Management Science, Information Systems Research, the Journal of Management Information Systems, Information & Management, Decision Sciences*, and more. He teaches courses on service-systems architecture, information security, and computational statistics.