Review Article (Invited Authors Only)

new media & society
1–20

DOI: 10.1177/14614448211022702
journals.sagepub.com/home/nms
SAGE

Artificial intelligence and mass personalization of communication content—An ethical and literacy perspective

Erik Hermann
IHP—Leibniz-Institut für innovative Mikroelektronik, Germany

Abstract

Artificial intelligence (AI) is (re)shaping communication and contributes to (commercial and informational) need satisfaction by means of mass personalization. However, the substantial personalization and targeting opportunities do not come without ethical challenges. Following an AI-for-social-good perspective, the authors systematically scrutinize the ethical challenges of deploying AI for mass personalization of communication content from a multi-stakeholder perspective. The conceptual analysis reveals interdependencies and tensions between ethical principles, which advocate the need of a basic understanding of AI inputs, functioning, agency, and outcomes. By this form of AI literacy, individuals could be empowered to interact with and treat mass-personalized content in a way that promotes individual and social good while preventing harm.

Keywords

Artificial intelligence, ethics, literacy, mass, personalization

Introduction

Artificial intelligence (AI) is not just a technology but constitutes an encompassing power (re-)shaping daily practices, individual and professional interactions, and environments (Taddeo and Floridi, 2018). Its transformative impact also pertains to how people

Corresponding author:
Erik Hermann, Wireless Systems, IHP—Leibniz-Institut für innovative Mikroelektronik, Im Technologiepark 25, 15236 Frankfurt (Oder), Germany.
Email: hermann@ihp-microelectronics.com

communicate, which content they encounter, and how content is generated and disseminated (Guzman and Lewis, 2020; Hancock et al., 2020; Lewis et al., 2019). Among other things, AI has been and is a powerful force in the personalization of communication content (Sundar, 2020). The sophistication and computational power of AI applications in combination with the availability of big data (e.g. individuals' digital traces) facilitates the unprecedented personalization of communication content and messages at an individual level and likewise on a massive scale to a large audience (Winter et al., 2021). That is, AI enables the mass personalization of communication content (i.e. commercial, editorial/journalistic, and user-generated messages and information). Targeting by tailored persuasive appeals (e.g. Matz et al., 2017), entertainment and commercial (vendor) platforms based on recommender systems (e.g. Milano et al., 2020), news feeds of social network sites (e.g. Bakshy et al., 2015), and automated news production and dissemination including newsbots (e.g. Lewis et al., 2019)—to name but a few—have become part of our daily lives (Kitchin, 2017; Willson, 2017). In spite of the general benefits of personalization such as increased personal relevance and satisfaction of individuals' wants and needs (Sundar and Marathe, 2010), AI-driven mass personalization does not come without ethical concerns, which relate to privacy (e.g. Matz et al., 2019a), agency (e.g. Soffer, 2019; Sundar, 2020), biases and transparency (e.g. Hancock et al., 2020), and filter bubbles and echo chambers (e.g. Levy, 2021)—among other things.

Generally, the mounting pervasiveness of AI systems and application has sparked the debate of ethical principles and values guiding AI development and use (e.g. Cowls et al., 2021; Floridi et al., 2018, 2020; Hagendorff, 2020; Jobin et al., 2019; Mittelstadt, 2019; Morley et al., 2020). To date, the AI ethics landscape is rather fragmented and entails recurring principles (Jobin et al., 2019) that are of high-order, deontological nature (Hagendorff, 2020). Accounting for these principles in practice while taking into account the different stakeholder interests might demand tradeoffs. In light of AI's impact on the individual, economic, and societal level, the AI ethics literature increasingly focuses ethical frameworks of AI for (social) good (Cowls et al., 2021; Floridi et al.,2018, 2020; Taddeo and Floridi, 2018). That approach addresses and attempts to solve the tension between leveraging the benefits and preventing (or at least mitigating) potential harms of AI—to achieve a "dual advantage" for society (Floridi et al., 2018: 694).

To the best of our knowledge, our conceptual analysis is the first study to systematically scrutinize the ethical principles related to AI to AI-driven mass personalization from a multi-stakeholder perspective. Thereby, we provide two important contributions to the AI ethics and communication literature. First, we reveal several interdependent ethical challenges in respect to AI-driven mass personalization. Second, we suggest AI literacy as a mean to leverage these ethical challenges to empower individuals, which eventually benefits society at large.

The remainder of our study is structured as follows. After delineating our methodological approach and illustrating the role and use of AI in mass personalization, we present an overview of the AI ethics literature. Afterwards, we consolidate both perspectives by applying selected ethical principles to AI-powered mass personalization. We conclude our investigation with proposing AI literacy as a potential individual remedy to address the interdependent ethical challenges.

Methodology

We conducted a comprehensive literature search of published papers to identify relevant scholarly work. First, we performed a keyword search of electronic databases (Web of Science, EBSCO, and Google Scholar) using the following keywords: "ethic*," "guidelines," "principles," "framework," (for AI ethics) and "communication," "mass personalization," "personalization," "customization," (for AI in communication) each in combination with "artificial intelligence," "AI," "artificial," "machine learning," "algorithm*," "bots." Second, we examined references of review and seminal articles in both fields (e.g. Guzman and Lewis, 2020; Hancock et al., 2020; Sundar, 2020 and Floridi et al., 2018; Jobin et al., 2019; Mittelstadt, 2019 respectively) and applied an ancestry tree search by screening all papers citing these articles. Third, we performed manual search of journal outlets that turned out to be major sources for journal articles dealing with AI in communication (e.g. *Computers in Human Behavior, Information, Communication & Society, Journal of Computer-Mediated Communication, New Media & Society*) and AI ethics (i.e. *Ethics and Information Technology, Minds and Machines, Nature Machine Intelligence, Science and Engineering Ethics*).

Mass personalization and AI

Personalization refers to the "degree to which receivers perceive a message reflects their distinctiveness as individuals differentiated by their interests, history, relationship network, and so on" (O'Sullivan and Carr, 2018: 1166). As a form of system-initiated personalization, it differs from (user-initiated) customization, where individuals deliberately tailor content by choosing options and/or creating new content (Sundar and Marathe, 2010) and become sources of communicative interactions (i.e. self-as-source; Kang and Sundar, 2016). Mass personalization unifies characteristics of mass communication, that is, technologically mediated content is delivered to large audiences, and interpersonal communication, that is, personalized content reflecting recipients' uniqueness, distinctiveness, or identity (e.g. Kalyanaraman and Sundar, 2006), which is comparable to interpersonal messages by traditional definition (O'Sullivan and Carr, 2018). Although digital media have substantially simplified content personalization, it existed long before the advent of digital media (Sundar and Marathe, 2010). In a seminal essay on personalization of mass media, Beniger (1987) noted,

> The capacity of such mass media for simulating interpersonal communication is limited only by their output technologies, computing power, artificial intelligence; their capacity for personalization is limited only by the size and quality of data sets on the households and individuals to which they are linked. (p. 354)

The computational power of AI and the availability of massive amounts of (digital and social media) data (e.g. Cappella, 2017; Harari et al., 2016; Matz and Netzer, 2017; Stachl et al., 2020a; Stahl et al., 2021; Winter et al., 2021)—of metrified and tracked individuals (Gerlitz and Helmond, 2013; König et al., 2020)—have relativized some of the limits stressed by Beniger (1987). Thereby, AI allows to personalize content at

unprecedented speed, scale, intensity, and responsiveness. Correspondingly, Kotras (2020) defined mass personalization as "algorithmic processes in which the precise adjustment of prediction to unique individuals involves the computation of massive datasets, compiling the behaviors of very large populations" (p. 2). Thus, mass personalization and algorithms are inextricably intertwined. Put simply, algorithms produce outputs from inputs by means of certain (deterministic) rules or procedures (Hill, 2016). Machine-learning algorithms are methods that utilize data to identify (novel) patterns and underlying rules (Kaplan and Haenlein, 2019). They have the capacity to define or modify decision-making rules autonomously (Mittelstadt et al., 2016). Although machine learning is a focal part of AI, AI is broader due to its ability to perceive data (e.g. language processing) and other human-like capabilities such as moving objects (e.g. robots) or conversation capacities (e.g. chatbots; Kaplan and Haenlein, 2019). The concept of AI is polysemous and not well-defined (Guzman and Lewis, 2020; Stahl et al., 2021). In its simplest sense, AI refers to technologies performing tasks that are associated with some level of human intelligence (Guzman and Lewis, 2020). In more technical terms, AI can be defined as "a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan and Haenlein, 2019: 17).

Algorithms are the integral components of recommender systems (e.g. Lury and Day, 2019; Milano et al.,2020, 2021; Mittelstadt et al., 2016), that are the foundation of and/or intensively used by news intermediaries and aggregators, social media, entertainment, and commercial (vendor) platforms (Bozdag, 2013; Helberger, 2019; Lury and Day, 2019). Recommender systems refer to (algorithmic) functions utilizing information about individual preferences (e.g. products or news items) as inputs to predict how individuals would rate certain items under evaluation (e.g. new items available) and how they would rank a set of items individually or as a bundle (Milano et al., 2020). Inputs from individuals can include any form of reactions (e.g. comments, likes, ratings, reviews) to news, products, or other social, political, cultural, or entertainment stimuli—all of which being indicative of social norms evaluating sociocultural entities. That makes respective recommendations a form of platform-mediated interpersonal communication (Cappella, 2017). Recommender systems can take the form of collaborative filtering, content-based filtering, or hybrid methods (Bozdag, 2013; Lury and Day, 2019; Milano et al., 2020). Collaborative filtering algorithms base their recommendations on target individuals' past behavior, choices, and preferences, and on preferences of other individuals being structurally similar to them (Cappella, 2017; Lury and Day, 2019). These recommendations of future choices based on similar tastes and patterns of past choices can be considered as a surrogate for social influence (Cappella, 2017) or automated word-of-mouth (Bozdag, 2013). However, content-based filtering algorithms make use of discrete characteristics and properties of items to generate recommendations of items with similar characteristics and properties that individuals preferred in the past (Bozdag, 2013; Cappella, 2017; Lury and Day, 2019). Besides, algorithmic content filtering and ranking can personalize, prioritize, and curate content (e.g. search engines, news feeds of social network sites) for individuals (e.g. Bakshy et al., 2015; Bozdag, 2013; Lazer, 2015; Möller et al., 2018; Scharkow et al., 2020; Schwartz and Mahnke, 2021).

Personalization might not only be based on individuals' preferences, interests, demographics, and past behavior, item features and characteristics, or similar tastes of others, but also on psychological factors—the method of psychological targeting (Hirsh et al., 2012; Matz and Netzer, 2017; Matz et al., 2017; Matz et al., 2019b; Stachl et al., 2020a; Winter et al., 2021; Zarouali et al., 2020). AI-powered psychological targeting offers considerable opportunities to tailor (persuasive) appeals to individuals' psychological traits (variability across consumer such as personality traits or values) and states (variability within consumers over time such as mood or emotions) that are computationally predicted from their digital footprints (Matz and Netzer, 2017; Matz et al., 2017). Combining large-scale (digital) data with the computational power of AI allows "an unprecedented understanding of consumers' unique needs as they relate to the situation-specific expressions of more stable motivations and preferences" (Matz and Netzer, 2017: 9). Therefore, AI could be leveraged for both content creation of psychologically tailored appeals and situation-specific and context-aware dissemination of such appeals.

Apart from commercial appeals, the dual role of AI in personalized content production and dissemination also pertains to news and journalistic content (e.g. Bodó, 2019; Bodó et al., 2019; Ford and Hutchinson, 2019; Guzman, 2019; Helberger, 2019; Lewis et al., 2019; Milosavljević and Vobič, 2019; Thurman et al., 2019a, 2019b). While the first generation of news personalization incorporated receiver-initiated customization based on explicitly expressed preferences, the second generation features implicit personalization techniques building on individuals' digital profiles and indirect preference signals (Bodó, 2019; Kunert and Thurman, 2019; Thurman and Schifferes, 2012). In addition, newsbots have developed from rebroadcasters of news content to disseminators of news incorporating chatbot conversation capacities, thereby becoming a third party (person) mediating the sender-receiver relationship (Ford and Hutchinson, 2019; Lokot and Diakopoulos, 2015) or conversational agents (Guzman and Lewis, 2020).

Generally, AI does not only facilitate, mediate, and channel communication (e.g. Hancock et al., 2020), but also functions as a communicator and participant in communicative exchanges itself—a role that has been historically attributed to humans from a communication-theoretical perspective (Gunkel, 2012; Guzman and Lewis, 2020; Schwartz and Mahnke, 2021). Before we shed light on the ethical questions of AI in mass personalization, we provide an overview of the AI ethics literature.

AI ethics

The discourse on moral and ethical implications of AI dates back from 1960 (Samuel, 1960; Wiener, 1960). The increasing pervasiveness and encompassing impact of AI applications and systems have intensified calls for and discussions of accompanying ethical guidelines. That is, "the ethical debate has gone mainstream" (Morley et al., 2020: 2141). Ethical principles related to AI focus on ethical issues in respect to particular features of the technology or the consequences of its use (Stahl et al., 2021). That is, in the tradition of computer and (information) technology ethics (e.g. Brey, 2000, 2012; Moor, 1985, 2005; Royakkers et al., 2018; Wright, 2011), which incorporate recurring principles and themes such as *autonomy, justice, beneficence, non-maleficence, dignity*, and *privacy* (Brey, 2012; Royakkers et al., 2018; Wright, 2011).

These principles also characterize ethical frameworks related to AI. In a comprehensive review of 84 documents of principles and guidelines for ethical AI issued by private, public, and research institutions, Jobin et al. (2019) found convergence around the principles *transparency* (referenced in 73 out of 84 documents), *justice and fairness* (68), *non-maleficence* (60), *responsibility* (60), *privacy* (47), *beneficence* (41), and *freedom* and *autonomy* (34). However, no single ethical principle was referenced in all 84 documents. The prevalence of *transparency* could be attributed to the reasoning that it "is not an ethical principle in itself but a proethical condition for enabling or impairing other ethical practices or principles" (Turilli and Floridi, 2009: 105). While frequently referenced principles such as *justice* and *fairness, non-maleficence*, and *privacy* reflect a cautious view on potential risks of AI, the more frequent occurrence of *non-maleficence* as compared to *beneficence* implies the moral obligation to avoid any negative impact of AI and a certain negativity bias (Jobin et al., 2019). The role of *trust* as an AI governance principle is not without opposition and ambiguity, particularly, whether *trust* is a principle in itself or rather an outcome of other foundational principles (e.g. Floridi, 2019; Ryan, 2020; Thiebes et al., 2020). The *solidarity* principle is only featured in 6 out of 84 documents, although it refers to redistributing the benefits of AI to not jeopardize social cohesion (Jobin et al., 2019). In light of persistent and mounting global inequalities, prosperity, and burdens created by AI should be shared, that is, *solidarity* should be considered as a focal ethical principle of AI (Luengo-Oroz, 2019).

The *solidarity* and *beneficence* principles already hint at the need to (equitably) leverage the benefits of AI on a societal level. Ethical frameworks for AI for (social) good (Cowls et al., 2021; Floridi et al.,2018, 2020; Taddeo and Floridi, 2018) are in line with this stance and advocate the following five focal ethical principles: *beneficence, non-maleficence, autonomy, justice*, and *explicability* (Floridi et al., 2018). Figure 1 provides a systematization of the ethical principles identified by Jobin et al. (2019) and Floridi et al. (2018) and how they align.

While the tenet of *beneficence* refers to the promotion of well-being as well as social, environmental, and common good (Jobin et al., 2019; Thiebes et al., 2020), the *non-maleficence* principles caution against the potentially negative aspects of AI. Central to *non-maleficence* are safety, security, privacy, and generally, the prevention of risks and any harm—both accidentally or unintentionally (overuse) and deliberately (misuse) caused (Floridi et al., 2018; Jobin et al., 2019). Across ethical frameworks, *privacy* often constitutes a principle on its own. However, given the emphasis on avoiding infringements of privacy, breaches of data protection, and misuse of data, that is, prevention of harms and risks in respect to personal data and privacy, *privacy* can be subsumed under the *non-maleficence* principle. It is noteworthy that *beneficence* and *non-maleficence* are not opposite ends of a continuum but coexist, although they seem logically equivalent (Floridi et al., 2018). The principle *autonomy* entails self-determination and the power to and whether to decide in an uncoerced way, that is, seeking a balance between human and AI agency and decision-making power (Floridi et al., 2018; Morley et al., 2020). The *justice* principle stresses fairness, avoiding unwanted or unfair biases, and discrimination (Jobin et al., 2019; Morley et al., 2020; Thiebes et al., 2020), as well as sharing benefits and prosperity and fostering solidarity (Floridi et al., 2018). Thus, the *solidarity* principle merges in the *justice* principle.

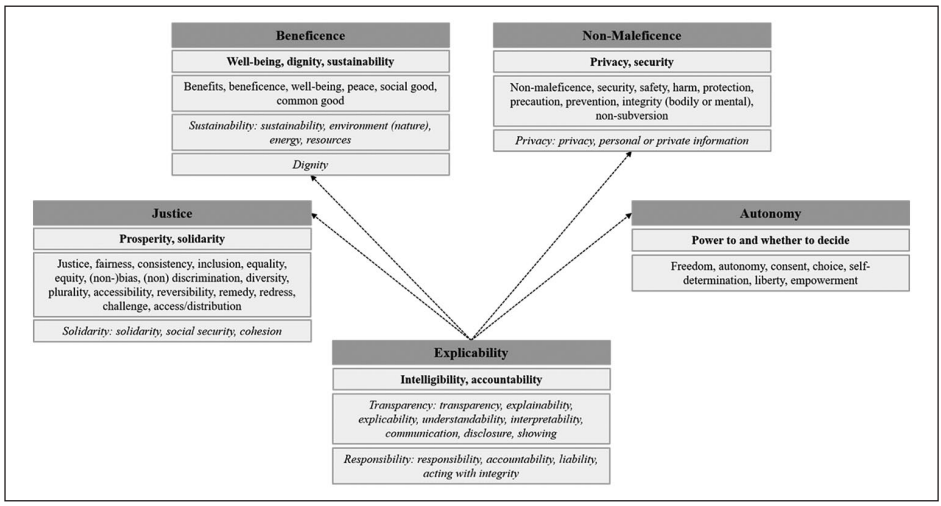

Figure 1. AI ethics map.
Principles in bold are taken from Floridi et al. (2018), while principles in normal font and italics are taken from Jobin et al. (2019). Principles in italics were not subsumed under beneficence, non-maleficence, justice, or explicability, but listed as independent principles by Jobin et al. (2019).

Finally, *explicability* comprises *intelligibility* (i.e. how AI works—the epistemological sense) and *accountability* (i.e. who is responsible for the way AI works—the ethical sense; Floridi et al., 2018). In the literature, different nomenclature and concepts, that is, intelligibility, comprehensibility, interpretability, explainability, and transparency, are used interchangeably and inconsistently (Barredo Arrieta et al., 2020), and are partly misconceived (Rudin, 2019). In a comprehensive review, Barredo Arrieta et al. (2020) identified *intelligibility*, that is, human understanding of a model's function without any need for explaining its internal structure or underlying data processing algorithm, as the most appropriate conceptualization. The narrow relation between *intelligibility* and *accountability* (e.g. Lepri et al., 2018; Martin, 2019; Morley et al., 2020) justifies their subsumption under *explicability*. That is, judgments about *accountability* necessitate a certain understanding of the underlying processes of AI systems and applications (i.e. *intelligibility*; Lepri et al., 2018). Understanding the functionalities (i.e. *intelligibility*) and responsibilities (i.e. *accountability*), in turn, informs evaluations of the other principles by comprehending if and how AI benefits or harms individuals and society (*beneficence* and *non-maleficence*), by drawing conclusions about whether to delegate decisions to AI systems (*autonomy*), and by knowing whom to hold accountable in case of failures or biases (*justice*; Floridi et al., 2018; Thiebes et al., 2020).

AI ethics frameworks have in common that they focus high-level ethical principles with little reference to philosophical ethical theories (Stahl et al., 2021). However, a predominantly principled approach is called into question for at least two reasons (e.g. Hagendorff, 2020; Mittelstadt, 2019; Theodorou and Dignum, 2020). First, deontological and normative imperatives and principles (Hagendorff, 2020) lack translation into

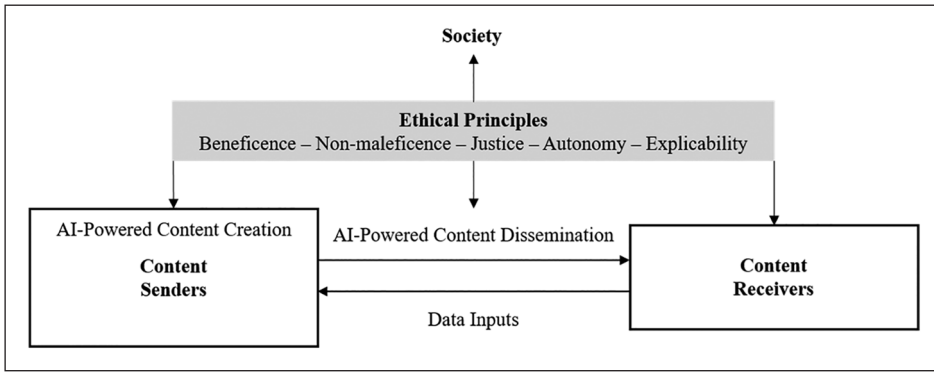

Figure 2. Multi-stakeholder model of ethical principles related to AI-powered mass personalization.

practice through mid-level norms and low-level requirements taking into consideration the legal, technical, and social circumstances (Mittelstadt, 2019). Applied AI ethics could close the gap between principles (*what*) and the practice of *how* to develop ethical AI (Morley et al., 2020). Second, AI is not developed and deployed in isolation, but within the socio-technical system (i.e. people, organizations, their interactions, and processes organizing these interactions) it is operating and unfolding. Therefore, concrete ethical and socio-legal governance and policies are in demand (Cath, 2018; Theodorou and Dignum, 2020).

In the following, we examine ethical principles and controversies of deploying AI for the mass personalization of communication content.

The ethics of AI in mass personalization of communication content

We investigate the ethical implications and concerns of AI-driven mass personalization of communication content from a multi-stakeholder perspective comprising content senders, content receivers, and society at large (see Figure 2). Following communication-theoretical conceptualizations, we refer to senders and receivers. Admittedly, this distinction is far from unequivocal, and transitions can be fluid. For instance, the traditional receivers (i.e. individuals) can now create and share content on their own (i.e. user-generated content). We simplistically define senders as the entities operating and/or economically profiting from AI systems that create or disseminate content, while receivers are the targets the content is directed to. By adopting a multiperspectivity approach, we want to provide a holistic picture of the ethical considerations beyond the individual content receivers. Particularly, phenomena such as filter bubbles, echo chambers, and respective (ideological) polarization that can arise from algorithmic content filtering can have adverse effects for democracy and society at large (e.g. Bozdag and van den Hoeven, 2015; Helberger, 2019). Our multilevel analysis further accounts for the AI-for-social-good perspective stressed by prior AI ethics literature (Cowls et al., 2021; Floridi

et al.,2018, 2020; Taddeo and Floridi, 2018). Correspondingly, our analysis is based on the ethical principles suggested by this stream of research, that is, *beneficence, non-maleficence, autonomy, justice*, and *explicability* (Floridi et al., 2018; Morley et al., 2020).

Beneficence

First and foremost, AI-powered mass personalization of communication content can be harnessed to match individuals' preferences (e.g. Matz and Netzer, 2017), to increase personal relevance and to satisfy—or at least approach—individuals' wants and needs (e.g. Sundar and Marathe, 2010), and to improve the attractiveness and usability of products, services, messages, and content, which, in turn, increases acceptance, usage, satisfaction, and loyalty (e.g. Stachl et al., 2020b). Moreover, mass-personalized content can serve as substitute or shortcut for extensive information search and gathering through information (pre-)filtering and selection (Cappella, 2017) leading to better information and efficiencies (e.g. time savings) on the content receiver side (e.g. Helberger, 2019). These advantages on the content receiver level also benefit the content senders in terms of adoption rates, satisfaction, loyalty and retention, profit, and resource efficiencies.

In general, AI-powered mass personalization is not configured and employed for the sake of it but relates to specific benefits and a clear purpose, which implies justification—one requirement for *beneficence* (Morley et al., 2020). Nevertheless, judgments of benefits, goodness, and hence, the *beneficence* principle can be ambiguous (D'Acquisto, 2020). That is, *beneficence* on the content sender and receiver level does not necessarily imply *beneficence* on the societal level. As mentioned earlier, personalization entails the (pre-)selection of content and recommendations individuals are exposed to, which can eventually result in content receivers' selective exposure to content. Selective exposure and limited content diversity could lead to polarization, that is, strengthening of individuals' original attitude or position (Stroud, 2010), echo chambers, or filter bubbles, that is, individuals are only encountering content from like-minded individuals or content selected by algorithms according to individuals' previous behavior and interactions with the system, respectively (Bakshy et al., 2015; Bozdag and van den Hoeven, 2015).

While some studies found evidence that mass personalization induces selective exposure and polarization (e.g. Dylko et al., 2017; Levy, 2021), other studies challenge the assumption and concerns that personalization algorithms necessarily or solely cause filter bubbles, echo chambers, or polarization (Bakshy et al., 2015; Berman and Katona, 2020; Flaxman et al., 2016; Fletcher and Nielsen, 2018; Haim et al., 2018; Messing and Westwood, 2014; Möller et al., 2018; Nechushtai and Lewis, 2019; Scharkow et al., 2020; Zuiderveen Borgesius et al., 2016). Despite these equivocal findings, constraints on cross-cutting and counter-attitudinal content that undermines balanced content diversity would limit the *beneficence* of AI-driven mass personalization on a societal level. On the content sender and receiver levels, exploitation of existing individual information for tailoring content by AI systems often constitutes the optimal (standard) strategy to maximize individual utility and satisfaction. However, exploitation strategies could also run the risk of choosing prediction accuracy over satisfaction (e.g. Kotkov et al., 2016) and of underrepresenting new or alternative content (i.e. limiting content diversity) as

compared to more explorative strategies (e.g. Milano et al., 2021). Therefore, some scholars emphasize the importance and value of serendipity, that is, recommendations of items that are relevant, novel, and unexpected (and thus, relatively unpopular and significantly different from individuals' profiles; Kotkov et al., 2016), and diversity-sensitive designs to promote content diversity and satisfaction (e.g. Fletcher and Nielsen, 2018; Helberger et al., 2018; Kotkov et al., 2016; Levy, 2021; Reviglio, 2019). Taken together, the unconditional *beneficence* of AI-powered mass personalization can be called into question, which draws the analogy to the *non-maleficence* principle.

Non-maleficence

In the case of limited content diversity resulting from AI-driven mass personalization, judgments of *beneficence* and *non-maleficence* from a societal perspective are related. That is, mass personalization does not necessarily foster social good (*beneficence* principle not met) but can compromise it (*non-maleficence* principle not met). Contrarily, the respective ethical judgments do not coincide or are even inversely related when comparing the content sender and receiver levels. That means, mass personalization could be beneficent (e.g. for content senders) and maleficent (e.g. for content receivers) at the same time (Milano et al., 2021).

Notably, personal privacy, accuracy, data protection, and quality are central to the *non-maleficence* principle (e.g. Floridi et al., 2018; Morley et al., 2020). Privacy risks can arise (1) when data are gathered without informed consent of individuals, (2) after storage when they are leaked or de-anonymized (i.e. data breaches), (3) when AI systems draw inferences from both individual data (directly), or interaction data with others (indirectly; Milano et al., 2020). The large-scale (digital) data feeding AI systems and applications can aggravate privacy and data protection issues (Baruh and Popescu, 2017), as discussed for algorithms (e.g. Mittelstadt et al., 2016), recommender systems (e.g. Milano et al., 2020), psychological targeting (e.g. Matz et al., 2019a), and personalization in general (e.g. Cloarec, 2020).

In the case of AI-powered mass personalization, there is a tension and potential trade-off between the scale and scope of data inputs for mass personalization and privacy concerns, that is, the heightened amount of data inputs to achieve predictive validity and accuracy of personalization efforts could interfere with data protection and privacy. Since AI system's inferences and predictions are as accurate and reliable as the underlying data, quality and integrity of data are decisive. Biases, inaccuracies, and errors inherent in data could bias results and lead to false conclusions (e.g. Barredo Arrieta et al., 2020; Hancock et al., 2020; Morley et al., 2020). Furthermore, algorithmic predictions are directed to individuals, although inferences are drawn from populations. That becomes problematic when algorithmic inferences are based on (potentially spurious) correlations found in large datasets, because causality is often not established prior to algorithmic decisions (Mittelstadt et al., 2016). Inferior or biased predictions and recommendations can be particularly adverse for individuals if they depend too much on algorithm-generated recommendations (i.e. algorithm overreliance) that could then harm their well-being (Banker and Khetani, 2019). Whether or not individuals are exposed to and influenced by mass-personalized content is narrowly related to the *autonomy* principle.

Autonomy

Autonomy relates to a meta-autonomy or decide-to-delegate model, that is, "humans should always retain the power to decide which decisions to take" on their own or when to cede decision-making control (Floridi et al., 2018: 698). Human agency (i.e. autonomous decisions) and human oversight are focal requirements of *autonomy* (Morley et al., 2020).

AI in mass personalization act as (secondary) gatekeeper creating, selecting, filtering, and disseminating the content individuals eventually encounter (Just and Latzer, 2017; Singer, 2014; Soffer, 2019). Therefore, content receivers' *autonomy* is concerned to the effect that decisions (i.e. agency) are delegated to AI systems at the information collection stage of the decision-making process, particularly, (pre-)filtering of information and options individuals are exposed to. Because personalization, psychological targeting, and recommender systems can serve as adaptive, structural, or informational nudges (Floridi, 2016; Milano et al., 2020; Sunstein, 2016), individuals' decision-making processes are influenced. That is, these kind of interventions shape individuals' choice sets or information related to choices and eventually preferences and decisions. That can be beneficial due to resource efficiency (e.g. time, cognitive resources) and personally relevant content, but also detrimental in case of overreliance on mass-personalized recommendations (e.g. Banker and Khetani, 2019) or due to manipulated or deceptive content (e.g. Hancock et al., 2020; Milano et al., 2020). Human agency could be fostered by preferring reactive personalization (i.e. obtaining permission before providing personalized content) and overt data gathering over proactive personalization (i.e. automatically providing personalized content) and covert data gathering (Sundar, 2020; Sundar and Marathe, 2010).

On the content sender level, AI systems are granted *autonomy* both at the content creation and dissemination stage (Hancock et al., 2020). Therefore, governance mechanism should be implemented to facilitate human agency and oversight and to keep humans in the loop (e.g. Thiebes et al., 2020), particularly, in ethically or morally salient contexts. Generally, the question of *autonomy* on the content sender and receiver levels should not be framed as a dichotomy between human and AI agency, since humans are either treated as passive victims of AI predictions or they are entirely held accountable for any potential negative effect of (neutral) AI that mediates human inputs (Schwartz and Mahnke, 2021).

Justice

As human judgments can be error-prone, biased, and discriminating, so can AI predictions and inferences (Kleinberg et al., 2020; Rich and Gureckis, 2019). Personalization could "segment a population so that only some segments are worthy of receiving some opportunities or information, re-enforcing existing social (dis)advantages" (Mittelstadt et al., 2016: 9). Such profiling leads to "industrialized social discrimination" that creates "winners" and "losers" being worth or not to receive content (Turow and Couldry, 2018: 417). Accordingly, AI-powered mass personalization could discriminate content receivers on the basis of psychological, economic (e.g. income), and demographic factors,

reinforce gender, age, and racial disparities, prejudices, and stereotypes (e.g. Bol et al., 2020; Datta et al., 2015; Kleinberg et al., 2020), and/or target (psychologically, ideologically, or economically) vulnerable groups (e.g. Matz and Netzer, 2017; Matz et al., 2017). As mentioned earlier, biased outcomes, unfair and unequal treatments, and targeting can be attributed to biases in and skewness of underlying data (Barredo Arrieta et al., 2020; Hancock et al., 2020; Morley et al., 2020). Sources of bias include but are not limited to over- and underrepresentation of demographic groups or sensitive features, consideration of misleading proxy features (Barredo Arrieta et al., 2020), or data sparsity in respect to certain individuals, features, and items (Batmaz et al., 2019; Rich and Gureckis, 2019). In light of these multiple sources of biases, diligence and monitoring along the entire data lifecycle and in respect to AI development (e.g. model specification) and deployment are advisable if not indispensable.

Discrimination is not limited to content receivers but can also affect content senders. In the commercial domain, mass personalization (e.g. recommender systems) can be discriminatory by decreasing sales diversity (i.e. a lack of serendipity) and by increasing market share concentration for popular items (Lee and Hosaganar, 2019). Besides, discrimination can arise from senders' presence versus absence on multisided (e-commerce) platforms deploying AI-driven mass personalization and respective unequal market access (Milano et al., 2021). Apart from commercial content, informational and attitudinal content and its senders can be subject to unbalanced representation caused by the issue of selective exposure, echo chambers, and filter bubbles delineated earlier.

Taken together, content senders and receivers can suffer from biases, discrimination, and amplification of existing inequalities due to AI-driven mass personalization, which can, in turn, diminish social good and well-being, which establishes the connection to the *beneficence* and *non-maleficence* principles.

Explicability

Due to the black-box nature of AI systems (i.e. black-box AI), their opacity and lack of accountability (Ananny and Crawford, 2018; Milano et al., 2020; Mittelstadt et al., 2016; Thiebes et al., 2020; Willson, 2017), *explicability* (i.e. *intelligibility* and *accountability*) features a prominently and controversially debated ethical principle, particularly, when high-stake decisions and sensitive, personal data are involved (e.g. Barredo Arrieta et al., 2020; Rudin, 2019). For content receivers, a basic understanding of how AI functions (i.e. *intelligibility*) and personalizes content might be more effective and satisfying than complicated and methodologically detailed explanations causing information overload, irritation, and frustration (Barredo Arrieta et al., 2020). Moreover, content receivers have a legitimate interest in knowing who to hold accountable (i.e. *accountability*) for adverse, biased, or discriminatory outcomes of personalized recommendations and targeting activities. That becomes particularly important if AI systems and algorithms are considered and conceptualized as value-laden rather than neutral (e.g. Martin, 2019).

However, addressing the black-box and opacity issue of AI-driven mass personalization can be challenging for content senders. First, *intelligibility* can interfere with privacy concerns and proprietary boundaries aiming at facilitating exclusivity or competitive advantages (e.g. Ananny and Crawford, 2018; Willson, 2017). Second, *intelligibility* can

be undermined by cognitive (i.e. excessively or insufficiently detailed information, lack of understanding), technical (i.e. methodological and technical complexity), and temporal constraints (i.e. rapid advancements and development cycles; Ananny and Crawford, 2018; Rudin, 2019). Third, transparency and disclosure of non-human identity of AI systems such as (news)bots can compromise their performance and efficiency, which raises the question of how to weigh the benefits and costs of disclosing the non-human AI nature (Ishowo-Oloko et al., 2019). Finally, overall (commercial) content diversity can decline when content receivers are explained why they received certain recommendations (Milano et al., 2020). That is, the item popularity among (similar) other users as explanation could further amplify desirability and popularity of items—a self-reinforcing, popularity-enhancing process could emerge.

AI literacy: a remedy for multiple interdependent ethical challenges?

Our analysis reveals several interdependencies between ethical principles. First and foremost, *explicability* in the form of *intelligibility* can be considered as an enabling principle for the other ethical principles. That is, an entire lack thereof (i.e. black-box AI) impedes individuals' judgments about *beneficence, non-maleficence, justice*, and *autonomy*. Furthermore, *justice* and *autonomy* can determine judgments about *beneficence* and *non-maleficence*, and the latter are related as well—even an inverse relationship between them is possible. In sum, a basic understanding of how AI in the mass personalization context works can be a prerequisite for individuals to assess biases (i.e. *justice*), their decision-making power and agency (i.e. *autonomy*), privacy concerns and underlying data (i.e. *non-maleficence*), and benefits (i.e. *beneficence*). When it comes to subtle and often unconscious techniques of AI-powered mass personalization, content receivers' reflection of everyday practices and interactions with respective content (e.g. Schwartz and Mahnke, 2021) might not suffice. Instead, a form of techno-capital unifying digital media and information literacy (e.g. Choi et al., 2020) might be in demand. We advocate a form of AI literacy to empower consumers in and beyond the mass personalization context. As media literacy is multi-dimensional—typically, cognitive, emotional, aesthetic, and moral dimensions (Potter, 2010)—so AI literacy also requires personal development along various dimensions. We conceptualize AI literacy as individuals' basic understanding of (a) how and which data are gathered; (b) the way data are combined or compared to draw inferences, create, and disseminate content; c) the own capacity to decide, act, and object; (d) AI's susceptibility to biases and selectivity; and (e) AI's potential impact in the aggregate. Increasing awareness and empower individuals to develop AI literacy constitutes a non-trivial and ambitious objective, but it could be an important step to leverage AI for social good. While AI literacy mainly relates to the content receiver level, remedies on the sender and societal level should be contemplated, too.

The existing deontological AI ethics approach might be inappropriate to account for complex interdependencies of ethical principles. Instead, a utilitarian approach weighing benefits and costs across all stakeholders could better account for multiple values, objectives, and utilities at the sender, receiver, and societal levels. Therefore, senders

could equip AI systems with multi-objective utility concepts taking into account ethical principles (e.g. Vamplew et al., 2018). The respective challenging conceptualization and implementation might require embedded ethics approaches proactively integrating ethicists (e.g. McLennan et al., 2020). That could be supplemented by appropriate governance and auditing mechanisms (e.g. Floridi et al., 2018). On the societal level, binding ethical and socio-legal policies (e.g. Stahl et al., 2021; Theodorou and Dignum, 2020) could provide regulatory guidance—the European Union's General Data Protection Regulation is a first step to address the *non-maleficence* (i.e. privacy) principle (Selbst and Powles, 2017).

Conclusion

This article synthesizes the research streams of AI ethics and AI-driven mass personalization. Our conceptual analysis shows that ethical principles related to the AI-for-social-good perspective interdepend and collide, partly, in dependence of the stakeholders concerned. In particular, *beneficence* and *non-maleficence* are not necessarily met, since mass personalization can limit content diversity and/or interfere with privacy. Besides, *explicability* (i.e. *intelligibility* and *accountability*) turns out to be a precursor to the other principles. Therefore, we propose that AI literacy (i.e. a basic understanding of AI inputs, functioning, agency, and outcomes) could empower individuals to judge *beneficence, non-maleficence, autonomy*, and *justice* related to AI-driven mass personalization themselves. AI literacy could further help individuals to retain *autonomy* and agency, since they are better able to identify and assess choice architectures generated through AI-driven mass personalization.

At their core, ethical principles should not be conceived as impediments of actions or (technological) progress; conversely, they should rather enhance the scope of action, autonomy, freedom, and self-responsibility (Hagendorff, 2020). We follow this path and suggest to leverage the ethical principles and respective challenges related to AI-powered mass personalization to conceptualize AI literacy at the receiver level and recommend ethically aligned mechanisms and policies at the sender and societal levels. AI literate and empowered individuals and effective governance mechanisms could promote individual and social good while limiting harm, so that a dual advantage for society could evolve. We hope that some of our thoughts motivate public, private, and/or research institutions to find a balanced set of bottom-up (i.e. AI literacy, embedded ethics approaches) and top-down (i.e. regulations) measures to exploit AI's potential while minimizing its negative externalities.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Erik Hermann https://orcid.org/0000-0003-0895-3562

References

Ananny M and Crawford K (2018) Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. *new media & society* 20(3): 973–989.

Bakshy E, Messing S and Adamic LA (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science* 348(6239): 1130–1132.

Banker S and Khetani S (2019) Algorithm overdependence: how the use of algorithmic recommendation systems can increase risks to consumer well-being. *Journal of Public Policy & Marketing* 38(4): 500–515.

Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, et al. (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58: 82–115.

Baruh L and Popescu M (2017) Big data analytics and the limits of privacy self-management. *new media & society* 19(4): 579–596.

Batmaz Z, Yurekli A, Bilge A, et al. (2019) A review on deep learning for recommender systems: challenges and remedies. *Artificial Intelligence Review* 52(1): 1–37.

Beniger JR (1987) Personalization of mass media and the growth of pseudo-community. *Communication Research* 14(3): 352–371.

Berman R and Katona Z (2020) Curation algorithms and filter bubbles in social networks. *Marketing Science* 39(2): 296–316.

Bodó B (2019) Selling news to audiences—A qualitative inquiry into the emerging logics of algorithmic news personalization in European quality news media. *Digital Journalism* 7(8): 1054–1075.

Bodó B, Helberger N, Eskens S, et al. (2019) Interested in diversity: the role of user attitudes, algorithmic feedback loops, and policy in news personalization. *Digital Journalism* 7(2): 206–229.

Bol N, Strycharz J, Helberger N, et al. (2020) Vulnerability in a tracked society: combining tracking and survey data to understand who gets targeted with what content. *new media & society* 22(11): 1996–2017.

Bozdag E (2013) Bias in algorithmic filtering and personalization. *Ethics and Information Technology* 15(3): 209–227.

Bozdag E and van den Hoeven J (2015) Breaking the filter bubble: democracy and design. *Ethics and Information Technology* 17(4): 249–265.

Brey PAE (2000) Method in computer ethics: towards a multi-level interdisciplinary approach. *Ethics and Information Technology* 2(2): 125–129.

Brey PAE (2012) Anticipating ethical issues in emerging IT. *Ethics and Information Technology* 14(4): 267–284.

Cappella JN (2017) Vectors into the future of mass and interpersonal communication research: Big data, social media, and computational social science. *Human Communication Research* 43(4): 545–558.

Cath C (2018) Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A* 376(2133): 20180080.

Choi JR, Straubhaar J, Skouras M, et al. (2020) Techno-capital: theorizing media and information literacy through information technology capabilities. *new media & society*. Epub ahead of print 27 May 2020. DOI: 10.1177/1461444820925800.

Cloarec J (2020) The personalization–privacy paradox in the attention economy. *Technological Forecasting & Social Change* 161: 120299.
Cowls J, Tsamados A, Taddeo M, et al. (2021) A definition, benchmark and database of AI for social good initiatives. *Nature Machine Intelligence* 3(2): 111–115.
D'Acquisto G (2020) On conflicts between ethical and logical principles in artificial intelligence. *AI & Society* 35(4): 895–900.
Datta A, Tschantz MC and Datta A (2015) Automated experiments on ad privacy settings: a tale of opacity, choice, and discrimination. *Proceedings on Privacy Enhancing Technologies* 1: 92–112.
Dylko I, Dolgov I, Hoffman W, et al. (2017) The dark side of technology: an experimental investigation of the influence of customizability technology on online political selective exposure. *Computers in Human Behavior* 73: 181–190.
Flaxman S, Goel S and Rao JM (2016) Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly* 80(S1): 298–320.
Fletcher R and Nielsen RK (2018) Automated serendipity. *Digital Journalism* 6: 8976–8989.
Floridi L (2016) Tolerant paternalism: pro-ethical design as a resolution of the dilemma of toleration. *Science and Engineering Ethics* 22(6): 1669–1688.
Floridi L (2019) Establishing the rules for building trustworthy AI. *Nature Machine Intelligence* 1(6): 261–262.
Floridi L, Cowls J, Beltrametti M, et al. (2018) AI4People—An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines* 28(4): 689–707.
Floridi L, Cowls J, King TC, et al. (2020) How to design AI for social good: seven essential factors. *Science and Engineering Ethics* 26(3): 1771–1796.
Ford H and Hutchinson J (2019) Newsbots that mediate journalist and audience relationships. *Digital Journalism* 7(8): 1013–1031.
Gerlitz C and Helmond A (2013) The like economy: social buttons and the data-intensive web. *new media & society* 15(8): 1348–1365.
Gunkel DJ (2012) Communication and artificial intelligence: opportunities and challenges for the 21st century. *Communication+* 11(1): 1.
Guzman AL (2019) Prioritizing the audience's view of automation in journalism. *Digital Journalism* 7(8): 1185–1190.
Guzman AL and Lewis SC (2020) Artificial intelligence and communication: a human–machine communication research agenda. *new media & society* 22(1): 70–86.
Hagendorff T (2020) The ethics of AI ethics: an evaluation of guidelines. *Minds and Machines* 30(1): 99–120.
Haim M, Graefe A and Brosius HB (2018) Burst of the filter bubble? Effects of personalization on the diversity of Google News. *Digital Journalism* 6(3): 330–343.
Hancock JT, Naaman M and Levy K (2020) AI-mediated communication: definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication* 25(1): 89–100.
Harari GM, Lane ND, Wang R, et al. (2016) Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. *Perspectives on Psychological Science* 11(6): 838–854.
Helberger N (2019) On the democratic role of news recommenders. *Digital Journalism* 7(8): 993–1012.
Helberger N, Karppinen K and D'Acunto L (2018) Exposure diversity as a design principle for recommender systems. *Information, Communication & Society* 21: 2191–2207.
Hill RK (2016) What an algorithm is. *Philosophy & Technology* 29(1): 35–59.

Hirsh JB, Kang SK and Bodenhausen GV (2012) Personalized persuasion: tailoring persuasive appeals to recipients' personality traits. *Psychological Science* 23(6): 578–581.

Ishowo-Oloko F, Bonnefon JF, Soroye Z, et al. (2019) Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation. *Nature Machine Intelligence* 1(11): 517–521.

Jobin A, Ienca M and Vayena E (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1(9): 389–399.

Just N and Latzer M (2017) Governance by algorithms: reality construction by algorithmic selection on the Internet. *Media, Culture & Society* 39(2): 238–258.

Kalyanaraman S and Sundar SS (2006) The psychological appeal of personalized content in web portals: does customization affect attitudes and behavior? *Journal of Communication* 56(1): 110–132.

Kang H and Sundar SS (2016) When self is the source: effects of media customization on message processing. *Media Psychology* 19(4): 561–588.

Kaplan A and Haenlein M (2019) Siri, Siri, in my hand: who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons* 62(1): 15–25.

Kitchin R (2017) Thinking critically about and researching algorithms. *Information, Communication & Society* 20(1): 14–29.

Kleinberg J, Ludwig J, Mullainathan S, et al. (2020) Algorithms as discrimination detectors. *Proceedings of the National Academy of Science* 117(48): 30096–30100.

König R, Uphues S, Vogt V, et al. (2020) The tracked society: interdisciplinary approaches on online tracking. *new media & society* 22(11): 1945–1956.

Kotkov D, Wang S and Veijalainen J (2016) A survey of serendipity in recommender systems. *Knowledge-Based Systems* 111: 180–192.

Kotras B (2020) Mass personalization: predictive marketing algorithms and the reshaping of consumer knowledge. *Big Data & Society* 7(2): 1–14.

Kunert J and Thurman N (2019) The form of content personalization at mainstream, transatlantic news outlets: 2010–2016. *Journalism Practice* 13(7): 759–780.

Lazer D (2015) The rise of the social algorithm. *Science* 348(6239): 1090–1091.

Lee D and Hosaganar K (2019) How do recommender systems affect sales diversity? A cross-category investigation via randomized field experiment. *Information Systems Research* 30(1): 239–259.

Lepri B, Oliver N, Letouzé E, et al. (2018) Fair, transparent, and accountable algorithmic decision-making processes. The premise, the proposed solutions, and the open challenges. *Philosophy & Technology* 31(4): 611–627.

Levy R (2021) Social media, news consumption, and polarization: evidence from a field experiment. *American Economic Review* 111(3): 831–870.

Lewis SC, Guzman AL and Schmidt TR (2019) Automation, journalism, and human–machine communication: rethinking roles and relationships of humans and machines in news. *Digital Journalism* 7(4): 409–427.

Lokot T and Diakopoulos N (2016) News bots: automating news and information dissemination on Twitter. *Digital Journalism* 4: 6682–6699.

Luengo-Oroz M (2019) Solidarity should be a core ethical principle of AI. *Nature Machine Intelligence* 1(11): 494.

Lury C and Day S (2019) Algorithmic personalization as a mode of individuation. *Theory, Culture & Society* 36(2): 17–37.

McLennan S, Fiske A, Celi LA, et al. (2020) An embedded ethics approach for AI development. *Nature Machine Intelligence* 2(9): 488–490.

Martin K (2019) Ethical implications and accountability of algorithms. *Journal of Business Ethics* 160(4): 835–850.

Matz SC and Netzer O (2017) Using big data as a window into consumers' psychology. *Current Opinion in Behavioral Sciences* 18: 7–12.

Matz SC, Appel RE and Kosinski M (2019a) Privacy in the age of psychological targeting. *Current Opinion in Psychology* 31: 116–121.

Matz SC, Kosinski M, Nave G, et al. (2017) Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Science* 114(48): 12714–12719.

Matz SC, Segalin C, Stillwell DJ, et al. (2019b) Using computational methods to predict personal image appeal. *Journal of Consumer Psychology* 29(3): 370–390.

Messing S and Westwood SJ (2014) Selective exposure in the age of social media: endorsements trump partisan source affiliation when selecting news online. *Communication Research* 41(8): 1042–1063.

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

Milano S, Taddeo M and Floridi L (2021) Ethical aspects of multi-stakeholder recommendation systems. *The Information Society* 37(1): 35–45.

Milosavljević M and Vobič I (2019) Human still in the loop. *Digital Journalism* 7: 81098–81116.

Mittelstadt BD (2019) Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* 1(11): 501–507.

Mittelstadt BD, Allo P, Taddeo M, et al. (2016) The ethics of algorithms: mapping the debate. *Big Data & Society* 3(2): 1–21.

Möller J, Trilling D, Helberger N, et al. (2018) Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society* 21(7): 959–977.

Moor JH (1985) What is computer ethics? *Metaphilosophy* 16(4): 266–275.

Moor JH (2005) Why we need better ethics for emerging technologies. *Ethics and Information Technology* 7(3): 111–119.

Morley J, Floridi L, Kinsey L, et al. (2020) From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics* 26(4): 2141–2168.

Nechushtai E and Lewis SC (2019) What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior* 90: 298–307.

O'Sullivan PB and Carr CT (2018) Masspersonal communication: a model bridging the mass-interpersonal divide. *new media & society* 20(3): 1161–1180.

Potter WJ (2010) The state of media literacy. *Journal of Broadcasting & Electronic Media* 54: 4675–4696.

Reviglio U (2019) Serendipity as an emerging design principle of the infosphere: challenges and opportunities. *Ethics and Information Technology* 21(2): 151–166.

Rich AS and Gureckis TM (2019) Lessons for artificial intelligence from the study of natural stupidity. *Nature Machine Intelligence* 1(4): 174–180.

Royakkers L, Timmer J, Kool L, et al. (2018) Societal and ethical issues of digitization. *Ethics and Information Technology* 20(2): 127–142.

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.

Ryan M (2020) In AI we trust: ethics, artificial intelligence, and reliability. *Science and Engineering Ethics* 26(5): 2749–2767.

Samuel AL (1960) Some moral and technical consequences of automation—A refutation. *Science* 132(3429): 741–742.

Scharkow M, Mangold F, Stier S, et al. (2020) How social network sites and other online intermediaries increase exposure to news. *Proceedings of the National Academy of Science* 117(6): 2761–2763.

Schwartz SA and Mahnke MS (2021) Facebook use as a communicative relation: exploring the relation between Facebook users and the algorithmic news feed. *Information, Communication & Society* 24(7): 1041–1056.

Selbst AD and Powles J (2017) Meaningful information and the right to explanation. *International Data Privacy Law* 7(4): 233–242.

Singer JB (2014) User-generated visibility: secondary gatekeeping in a shared media space. *new media & society* 16(1): 55–73.

Soffer O (2019) Algorithmic personalization and the two-step flow of communication. *Communication Theory* 2019: qtz008.

Stachl C, Au Q, Schoedel R, et al. (2020a) Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Science* 117(30): 17680–17687.

Stachl C, Pargent F and Hilbert S (2020b) Personality research and assessment in the era of machine learning. *European Journal of Personality* 34(5): 613–631.

Stahl BC, Andreou A, Brey P, et al. (2021) Artificial intelligence for human flourishing—Beyond principles for machine learning. *Journal of Business Research* 124: 374–388.

Stroud NJ (2010) Polarization and partisan selective exposure. *Journal of Communication* 60(3): 556–576.

Sundar SS (2020) Rise of machine agency: a framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication* 25(1): 74–88.

Sundar SS and Marathe SS (2010) Personalization versus customization: the importance of agency, privacy, and power usage. *Human Communication Research* 36(3): 298–322.

Sunstein CR (2016) *The Ethics of Influence: Government in the Age of Behavioral Science*. Cambridge: Cambridge University Press.

Taddeo M and Floridi L (2018) How AI can be a force for good. *Science* 361(6404): 751–752.

Theodorou A and Dignum V (2020) Towards ethical and socio-legal governance in AI. *Nature Machine Intelligence* 2(1): 10–12.

Thiebes S, Lins S and Sunyae A (2020) Trustworthy artificial intelligence. *Electronic Markets*. Epub ahead of print 1 October 2020. DOI: 10.1007/s12525-020-00441-4.

Thurman N and Schifferes S (2012) The future of personalization at news websites. *Journalism Studies* 13(5-6): 775–790.

Thurman N, Lewis SC and Kunert J (2019a) Algorithms, automation, and news. *Digital Journalism* 7(8): 980–992.

Thurman N, Moeller J, Helberger N, et al. (2019b) My Friends, editors, algorithms, and I. *Digital Journalism* 7(4): 447–469.

Turilli M and Floridi L (2009) The ethics of information transparency. *Ethics and Information Technology* 11(2): 105–112.

Turow J and Couldry N (2018) Media as data extraction: towards a new map of a transformed communications field. *Journal of Communication* 68(2): 415–423.

Vamplew P, Dazeley R, Foale C, et al. (2018) Human-aligned artificial intelligence is a multiobjective problem. *Ethics and Information Technology* 20(1): 27–40.

Wiener N (1960) Some moral and technical consequences of automation. *Science* 131(3410): 1355–1358.

Willson M (2017) Algorithms (and the) everyday. *Information, Communication & Society* 20(1): 137–150.

Winter S, Maslowska E and Vos AL (2021) The effects of trait-based personalization in social media advertising. *Computers in Human Behavior* 114: 106525.

Wright D (2011) A framework for the ethical impact assessment of information technology. *Ethics and Information Technology* 13(3): 199–226.

Zarouali B, Dobber T, De Pauw G, et al. (2020) Using a personality-profiling algorithm to investigate political microtargeting: assessing the persuasion effects of personality-tailored ads on social media. *Communication Research*. Epub ahead of print 20 October 2020. DOI: 10.1177/0093650220961965.

Zuiderveen Borgesius FJ, Trilling D, Möller J, et al. (2016) Should we worry about filter bubbles? *Internet Policy Review* 5(1): 1–16.

Author biography

Erik Hermann is post-doctoral researcher at the IHP—Leibniz-Institut für innovative Mikroelektronik. He obtained his PhD from European University Viadrina. His research focuses on the adoption and (ethical) implications of artificial intelligence in marketing, communication, and research and development from both the individual and more holistic perspective.