How Personal is Machine Learning Personalization?

Travis Greene
Institute of Service Science
National Tsing Hua University
Hsinchu, Taiwan
travis.greene@iss.nthu.edu.tw

Galit Shmueli
Institute of Service Science
National Tsing Hua University
Hsinchu, Taiwan
galit.shmueli@iss.nthu.edu.tw

December 18, 2019

ABSTRACT

Though used extensively, the concept and process of machine learning (ML) personalization have generally received little attention from academics, practitioners, and the general public. We describe the ML approach as relying on the *metaphor of the person as a feature vector* and contrast this with humanistic views of the person. In light of the recent calls by the IEEE to consider the effects of ML on human well-being, we ask whether ML personalization can be reconciled with these humanistic views of the person, which highlight the importance of moral and social identity. As human behavior increasingly becomes digitized, analyzed, and predicted, to what extent do our subsequent decisions about what to choose, buy, or do, made both by us and others, reflect who we are as persons? This paper first explicates the term *personalization* by considering ML personalization and highlights its relation to humanistic conceptions of the *person*, then proposes several dimensions for evaluating the degree of personalization of ML personalized scores. By doing so, we hope to contribute to current debate on the issues of algorithmic bias, transparency, and fairness in machine learning.

1 Introduction

Personalized products and services are an inescapable fact of modern life. From personalized predictions (Uber ETA, recidivism risk scores, credit scores), to personalized recommendations (Netflix, Spotify, Amazon, Tinder), personalized treatments (medicine and psychiatry), personalized ads, offers, prices, and more, we are inundated with personalized scores[1] based on data about our physical condition, genetic makeup, location, measurable behaviors and interactions. Ideally, personalization reduces information overload and search costs, thereby improving decision making and user experience. The decision maker and scored individual can be the same (self-directed) as in recommender systems, or different (other-directed) as in decision support systems used in law, medicine, business, and finance. In both cases, personalization relies on machine learning (ML).

Nevertheless, ML personalization remains an ambiguous and under-examined concept. A literature search reveals a surprising lack of consensus on its essential characteristics. It is either not explicitly defined, explained circularly, or used in relation to or interchangeably with *customization*, *tailoring*, or *precision marketing*. Cremonesi et al. (2010) say what personalization is *not*: giving "any user a pre-defined, fixed list of items, regardless of his/her preferences." In a paper titled, "What Is Personalization?" Fan and Poole (2006) classify various approaches to personalization. Despite deriving a useful taxonomy of ideal personalization types, the authors overlooked a simple, but fundamental point: that the user is a *person*. Thus, critical questions arise: Does merely assigning a unique score to each user constitute a personalized score? Is a score personalized when it is based (wholly or in part) on the behavior of other persons? Does it matter *who* these "others" are? Does personalization require *personal data*?

These and other questions are critical because lawmakers, journalists, social scientists, managers, and the general public encounter ML personalization everywhere, yet may have a very different concept of what personalization means. These differences can lead to confusion about the origin and accuracy of so-called "personalized" scores. This is especially relevant in light of recent regulations, such as the General Data Protection Regulation's (GDPR) *right to explanation* for cases of automated profiling and its inclusion of a *right to be forgotten*.

---

[1]We refer to recommendations, predictions, and treatments more generally as "scores."

This paper contrasts the process of ML personalization with the humanistic concept of a person and offers some normative criteria for evaluating the level of personalization. A clearer understanding of ML personalization will also contribute to current debates on algorithmic bias, transparency, and fairness in machine learning.

2 The Machine Learning Personalization Perspective: The Person as a Feature Vector

Most commercial "personalization" engines (e.g. recommender systems) rely on a combination of explicitly-given preferences, demographic data, and observed behavior assumed to reflect one's underlying preferences, (i.e., "implicit ratings") such as purchases, repeated use, save/print, delete, reply, mark, glimpse, query (Nichols, 1998). Such data are derived only from what can be observed when using an application or device. This small subset of observed behaviors becomes the basis on which persons are represented in ML. Further, this subset of behavior must be encoded and transformed into a digital (database) representation supporting ML algorithms. Information is lost during the transformation of unstructured data (text, pictures, images, videos) into structured data which can be represented in matrix form. Lastly, the database representation is converted into a *feature vector* in *feature space*.

A key idea in ML personalization rests on the metaphor of the *person as a feature vector*. A feature vector is an array of numbers–each representing some descriptive feature or aspect of an object–which can be thought of as the axes of a coordinate system (Kelleher et al., 2015). A 10-dimensional feature vector of a person, for instance, represents a person as an array of 10 numbers, derived from measurements of their observed behavior, and replaces the "person" with a single point in 10-dimensional feature space. Once converted to a point in feature space, the "similarity" of this person to others can be computed by measuring the distance between this point and others in the feature space. Points closer to each other are deemed more similar. Depending on one's modeling and predictive goals, this can indeed be a useful approach. But like any good metaphor, the feature vector approach emphasizes certain similarities at the expense of certain dissimilarities. When we represent a person as a feature vector, we tend to forget the long abstraction process that preceded it. As just one example of a potential problem in this process, Barocas et al. (2017) remark, "[numeric] features incorporate normative and subjective . . . assumptions about measurements, but all features are treated as numerical truth by ML systems." In short, three key characteristics of ML personalization arise from the above process: ML personalization 1) is highly behaviorist in its assumptions, 2) is extremely narrow in predictive scope, and 3) often relies on data of a "community" whose selection is based on predictive goals. Figure 1 illustrates the ML pipeline resulting in the operationalization of a person as a feature vector.
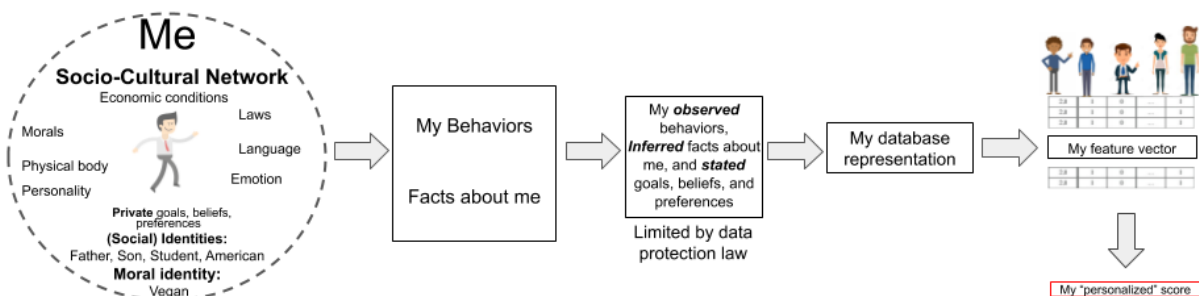

Figure 1: From humanistic concepts of the person to ML operationalization

Three Characteristics of ML Personalization: Behaviorism, Predictive Scope, and Community

First, *ML personalization is strongly behaviorist*. In the behaviorist worldview, "reality" is only what can be observed (Van Otterlo, 2013). Behaviorism sought to eliminate the messy theoretical causal mechanisms of human behavior (beliefs, intentions, goals, etc.) and focus instead on what can be measured and recorded (Skinner, 1965). ML personalization takes this idea one step further by focusing on predictive accuracy rather than on causal factors leading to the observed behavior (Rudin, 2019).

In the representation of the person as a feature vector, "important" attributes are only those that contribute to the algorithm's predictive accuracy. The predictive goal of ML personalization leads to valuing certain aspects of human behavior more than others; for example, behaviors most amenable to measurement tend to be recorded. This leads to a form of selection bias in the representation of the person. Unpredictable behavior is deemed noise, and measured behaviors which do not add to predictive accuracy are deemed redundant. Extraneous features can lead to the "curse of dimensionality" and are removed in order to preserve model parsimony, storage space, and computational efficiency. One example of how ML personalization favors parsimony is the use of unsupervised dimension reduction techniques, such as Singular Value Decomposition (SVD). These techniques rely on the discovery that many of our measured behaviors are simply linear combinations of other behaviors and can thus be discarded with little information loss. We will see how issues of parsimony, scale, and computational efficiency constrain the degree of personalization when we examine Instagram's Explore, which must select relevant content for a focal user from hundreds of millions of other potential users' contributions.

As we discuss in Section 4, the ML approach is at odds with a humanistic view of the person conceived as having an inner and outer self. As persons, our self-conceptions are highly bound up with our reasons for action, and these reasons may stem from our identity (see Section 4). If the reasons we give for our actions are indeed *causes* of our action, as many philosophers believe (see, e.g., Davidson, 1963), then we lose an important aspect of the human experience by focusing solely on prediction. The representation of a person as a feature vector abstracts a "person" away from her socio-cultural network of meanings, identities, and values and represents her as a precisely-defined digital object in order to predict a very specific behavior, limited to a specific application (see Figure 1). Human societies evolve over time and vary in their interpretations of codes of conduct regulating behavior (i.e., morality); yet, due to their abstract generality, mathematical objects and numbers do not. Further, the definitions and properties of mathematical objects remain fixed as a *formal* public system of rules, whereas human morality is an *informal* public system of rules of conduct in which the meaning of behavior is often unclear and evolving (Gert, 1998). This is one possible source of the disconnect in the feature vector representation of the person used in ML personalization.

Second, *ML personalization focuses on predicting a very narrow set of possible behaviors, often limited by the context of the application*. This can make it seem more powerful and accurate than it really is, especially when predictive performance is evaluated. For example, Yeomans et al. (2019) compared the predictions by friends and spouses to those from a simple collaborative filtering (CF) system for predicting a focal user's ratings of jokes. The study found the basic CF system was able to predict more accurately than a friend or spouse. Yet, the experiment included a set of 12 jokes *pre-selected* by the researchers. The much more difficult problem of selecting 12 jokes from a nearly infinite set of possible jokes in all cultures and languages was left to the humans. In essence, the researchers had already personalized a list of jokes to each subject in the study, given their linguistic background, country of origin, and current location. Once narrowed to such a small recommendation space, the algorithm's performance appears quite impressive, but nevertheless hides the fact the hardest task had already been done by humans. A similar argument can be made for personalization on e-commerce sites: by going to a website, a person has already self-selected into a group who would be interested in products offered by the website. Lastly, recidivism prediction in the judicial setting provides another example of the narrow predictive goal of ML personalization. Small changes in the definition of recidivism and its time-horizon– Is it one, two, or three years in the future? Does it include misdemeanors or only felonies?–will drastically affect the algorithm's predicted scores and performance.

Third, *ML personalization uses data not only from the focal user, but also from other users*. This is clear in the case of recommendation systems using social network data, where the interests and preferences of a focal user can be inferred from their direct and indirect connections to other users and groups. But this point is less obvious when when the data are from measured behaviors on an application or device. Though recommendations based on data from other users could lead one to believe that one's personalized scores are not, in fact, "personal," in the sense of stemming from one and the same person, the real issue is whether the "community" chosen reflects one's *chosen* social identity. In other words, personalization can still occur when the personalized scores are based on the behavior of others sharing the same social identity as the focal user. However, this is unlikely to occur in many ML contexts because "nearest neighbor" threshold values are chosen in order to optimize a given error metric, such as RMSE or precision/recall@top-N, not on the basis of any shared social identity (Cremonesi et al., 2010). Recommendations based on the behavior of others who do not share a similar social and moral identity with the focal user are less personalized under our conception of personalization–all things equal. This principle also implies that when dimension reduction techniques are used, such as PCA and SVD (which rely on calculations of total variance across the entire user-item matrix), the selection of a "community" of users making up the user-item matrix is important. Personalized scores using latent factors derived from users who do not share a similar social or moral identity are less personalized in our view. Of course, determining who is truly "similar" to us is a difficult question, but it is one where the methods and concepts of the humanities and social sciences may play an important role.

ML Personalization Case Study: Instagram's Explore

Instagram's personalization system shapes the desires, preferences, and self-conceptions of millions of people every day. By determining which content is displayed to which users, social media personalization systems may play an important role in the formation and maintenance of one's online and offline identity (Helmond, 2010). When scaled to hundreds of millions of users, ML personalization can even have society-wide effects, shifting opinions towards political issues and candidates, as in the 2016 US presidential election. For this reason, we believe it is important to see how ML personalization relates to the "person" and its tacit assumptions regarding observed behavior. We therefore briefly describe Instagram's Explore to illustrate the person as a feature vector approach and the implications discussed in the previous section.

The complicated set of algorithms used for Instagram's Explore recommendations essentially work by creating "account embeddings" in order to infer topical similarity.[2] Roughly put, a feature vector (FV) representation of an Instagram user is derived from the sequence of account IDs a user has visited. Next, the FV is used to determine a "community" of similar accounts. Similarity in this space implies topical similarity, assuming users do not randomly explore different topics, but instead tend to view multiple accounts offering similar content. Once the candidate

---

[2]Example taken from `ai.facebook.com/blog/powered-by-ai-instagrams-explore-recommender-system/`

accounts (the nearest neighbors) have been found, a random sample of 500 pieces of eligible content (videos, stories, pictures, etc.) is taken and a "first-pass" ranking is made to select the 150 "most relevant" pieces of content for a focal user. Then a neural network narrows the 150 to the top 50 most relevant. Notice that at each stage, possible content is narrowed-down and that computational efficiency is a high priority. Finally, a deep neural network is used to predict *specific actions* (narrowed only to those possible on the Instagram app) for 25 pieces of content, such as "like," "save," or "see fewer posts like this." Content is then ranked based on a weighted sum of these very specific behaviors permitted in the app and their associated predicted probabilities. Weights are not determined by users; they are determined by the system engineers and reflect their assumptions about the users' intentions. Finally, a "simple heuristic rule" is used to downrank certain content and diversify the results, reflecting the belief that the person wishes to discover new topics and avoid seeing content from the same person (an assumption about the stability of a user's preferences). After applying the downranking procedure, the content with the highest weighted sum in the "value model" is displayed in decreasing order on the focal user's Explore page.

3   The Person, Personal Data, and the Legal Issues of Personalized Scores

As described and illustrated in Section 2, personalized scores are generated by combining information about your observed behavior with certain assumptions about you and your values to predict a very constrained set of future actions. But *which information* should constitute me as a person, and from *whose viewpoint*? These are difficult philosophical questions. To answer we might first ask, "What is a person?" But this question is problematic because "person" is often defined in such a way as to justify a particular world-view. It can variously mean "human animal," "moral agent," "rational, self-conscious subject", "possessor of particular rights," and "being with a defined personality or character" (Schechtman, 2018). Nevertheless, Western thought has focused on two general aspects to the person, one of which is revealed through the etymology of the word itself.

"Person" comes from the Latin *persona*, which derives from the Ancient Greek word for a type of mask worn by dramatic actors. For the Ancient Greeks, the idea of a person was inherently connected to context and role. One's persona was a specific kind of self-identity that was public, socially defined, and varied depending on context. In time, the outward-facing Greek conception was complemented by a Christian emphasis on self-reflection and awareness, resulting in a focus on the human capacity for rational introspection (Douglas and Ney, 1998). Later thinkers expanded the idea of the person into two main parts: an outer *personnage* of public roles and masks, and an inner conscience, identity, and consciousness. In short, Western thought has generally viewed the person as consisting in two co-existing inner and outer domains.

We note that the word personalization contains the adjective *personal*, implying that for a score to be personalized, it must be based, at least in part, on *personal data*. For example, Liang et al. (2006) write that personalization is "a process of collecting and using *personal* information to uniquely tailor products, content and services to an individual." Yet, researchers in ML have largely overlooked the crucial connection between personalized scores and legal definitions of personal data. A source of complexity is that different legal regimes define personal data differently. For instance, the most influential data protection law at the moment, the EU's GDPR, defines personal data as "any information relating to an identified or identifiable natural person" (Article 4). Such a broad definition means personal data could constitute anything from browser cookies, to location data, to even a combination of non-sensitive measurements, if there are sufficiently few or unique observations to single out individuals. Context determines whether data are personal data and must be processed and protected as such (Greene et al., 2019).

Consent is another legal issue in the generation of personalized scores, especially when *sensitive data* such as race, gender, or political affiliation are involved. The GDPR's reliance on explicit consent has important moral implications. Consent implies choice in how one represents oneself publicly and also which inferences data processors should be able to draw about one's identity (see, e.g., Kosinski et al., 2013). The GDPR's extension of a right to rectification and erasure (the so-called "right to be forgotten") further entrenches the importance of autonomy in deciding the factual basis on which one will be judged, classified, and potentially discriminated against, and the extent to which our identities as persons are reflected in observations of our public behavior. This notion has been called *informational privacy* (Moreau, 2010; Shoemaker, 2010). In short, the existing literature on ML personalization overlooks two key points: first, that it depends on differing legal definitions of personal data; and second, informational privacy concerns are generally not considered in the design of personalized systems.

4   Identity, Self, and Data Regulation

ML researchers and practitioners are typically concerned with "practical identity," while philosophers, social psychologists, and sociologists usually speak of "moral" or "social" identity (De Vries, 2010; Manders-Huits, 2010). A third kind of identity fuses the two kinds of identity and has sparked much writing in the in the humanities and social sciences. This third kind of identity analyzes the difficulties in the formation of a single, unitary digital identity in light of rapidly evolving technology, particularly technology based on the distributed storage and collection of observable behavioral data. These fused digital representations of our practical, moral, and social identities have been variously termed *data citizens*, *data doubles*, *data shadows* and *digital subjects* (Goriunova, 2019).

What is the relation between personal data, moral identity, and personalized scores? While many of today's ML algorithms generate different scores given different contextual input data, these scores can fail to reflect our moral identities. For example, scores based on categorical features may not reflect one's social or personal identity. Similarly, explanations for personalized scores at the algorithm or model level (e.g., variable importance scores) may fail to reflect the salience of beliefs, motivations, and intentions we have when explaining our behavior to others. An explanation of a personalized score that leaves out these important aspects of one's moral identity surely cannot be said to relate to one *as a person*. Though we cannot control how society sees us, we can choose to (partly) accept or reject society's categorization of us. This latter aspect of human experience remains largely unaccounted for in ML personalization. As Manders-Huits (2010) argues, "The challenge [of] justice to data subjects as moral persons is to take into account the self-informative perspective that is part of 'identity management.'"

We now provide a brief overview of identity theories and humanistic perspectives drawing mainly from social and personality psychology, consumer behavior, sociology, philosophy, economics, information systems, and recent data protection law. We hope exposure to these ideas can influence future thinking on and design of ML systems generating personalized scores. Table 1 summarizes key concepts related to the person from these disciplines.

Identity from a Social Science Perspective

Theories and concepts of self and identity are fundamental topics in psychology, despite being difficult to precisely define. Personality psychologists Larsen and Buss (2009) define personality as a an organized set of relatively enduring psychological traits that guide a person's interactions with her "intrapsychic, physical, and social environments." These traits can be expressed as the"Big Five" dimensions of personality: Neuroticism, Extraversion, Openness to New Experience, Agreeableness, and Conscientiousness, and have been shown to have varying levels of predictive power.[3] Personality psychology has also recently begun to study the formation of *moral identity*. For example, Aquino et al. (2002) show how highly important moral identities–the collection of certain beliefs, attitudes, and behaviors relating to what is right or wrong–can provide a basis for the construction of one's "self-definition." Finally, social psychologists define the self as an interface between the biological processes of the body and the larger social and cultural context. Further, they highlight the power of the cultural context in defining one's identity: "Without society, the self would not exist in full" (Baumeister and Bushman, 2014, p. 74). As the summaries above illustrate, understanding the person requires description at multiple levels. Some researchers therefore provide multi-dimensional theories of the person, focusing on traits, personal concerns, and life stories (McAdams, 1995).

According to identity theorists in sociology and psychology, the self is fluid and occupies multiple social roles (identity theory) or group identities (social identity theory) that coexist and vary over time. The dynamism of the self-concept is reflected in its numerous sub-components or self-representations: the past, present, and future self; the ideal, "ought," actual, possible, and undesired self (Markus and Wurf, 1987). Identity theorists highlight how a unified self-conception arises from the variety of meanings given to various social roles the self occupies (Stryker and Burke, 2000). Common types of social identities relate to one's ethnicity, religion, political affiliation, job, and relationships. People may also identify strongly with their gender, sexual orientation, and various other "stigmatized" identities, such as being homeless, an alcoholic, or overweight (Deux 2001). In short, social psychologists generally agree one's self-concept is one of the most important regulators of one's behavior.

The self-concept has also influenced postmodernist consumer behavior research. Consumers now take for granted their sometimes paradoxical identities, beliefs, and behaviors in everyday life (Fuat Firat et al., 1995). In the postmodern condition one "listens to reggae, watches a Western, eats McDonald's food for lunch and local cuisine for dinner, wears Paris perfume in Tokyo and retro clothes in Hong Kong" (Lyotard, 1984).

Identity, the Self, and Morality from a Philosophical Perspective

How do philosophers understand the link between the identity and the self? For some, the self arises from an identification of structural sameness over time, experienced as a kind of narrative (Ricoeur, 1994). Though individual cells of our bodies are constantly renewed, a person's essential structure nevertheless remains similar enough to be identified as the "same" over time. For others, such as Taylor (1989), social and communal life link the two concepts. Taylor (1989) argues that one's identity is tied to one's moral values and a defining social community. He writes, "I define who I am by defining where I speak from, in the family tree, in social space, in the geography of social statuses and functions, in my intimate relations to the ones I love, and also crucially in the space of moral and spiritual orientation within which my most important defining relations are lived out." Our identities therefore rest on the constant interplay between how others view us and how we view ourselves. These "community" identities further entail various moral obligations to others that influence and motivate our behavior .

Atkins (2010) draws on work by Ricoeur and Korsgaard to explain the connection between morality and identity. She claims that "one's identity is the source of one's moral agency, expressed in one's normative reasons." One's identity constitutes "the condition of the possibility for having a perspective from which to perceive, deliberate, and

---

[3]Mischel et al. (2007) show that knowing an individual's Big Five trait profile allows for only weak prediction of a particular behavior in a particular context

Table 1: Perspectives and Key Concepts on the Person and Personal Identity

Philosophy & Humanities	Psychology, Sociology & Consumer Behavior	Economics & Information Systems	Data Regulation & Professional Codes of Conduct
Moral identity	Personality: Big 5	Identity explains irrational behavior	Right to be forgotten
Community-defined identity	Socially-defined identity	Digital identity	Right to rectification
Pursuit of the good life	Culturally-influenced narratives	Online identity affects behavior	Automated profiling
The Postmodern experience	Emotions and moods		Informational privacy
Narrative identity	Hermeneutic interpretation		Agency over digital identity
Inner and outer-facing self	Meaning of symbols, texts, images	Identity as economic choice	Ethically-aligned design

act ... it is a condition of the possibility of morality." For both Taylor and Atkins, the development of a self-identity provides a moral scaffolding upon which possible actions can be evaluated; our identities are critical for embedding us both in the physical and social worlds, and adding a moral dimension to our actions as persons.

In Schechtman (2018)'s influential narrative self-constitution view, personhood and personal identity are fleshed out in terms of a unified and "unfolding developmental structure," analogous to a complex musical sonata. Echoing the ideas of Ricoeur, she writes that human life "is a structural whole that has, by its very nature, attributes that apply to it as a whole which do not necessarily apply to each individual portion." Through the unity of experience–which takes the form of a narrative–the *same person* can be an infant at one time and later in life suffer from dementia. Her notion of personal identity is similar to McAdams' psychological view in that it explains via a diachronic process how identity can evolve over time.

Identity in Information Systems (IS) and Economics

As the World Wide Web grew in popularity, scholars in the field of IS began to examine its impact not just on the organization, but on the individual. Some were influenced by postmodern arguments and pushed for a reconsideration of such fundamental concepts as space, time, the world, and the self (Introna and Whitley, 1996). Erickson (1996), for example, foreshadowed the rise of social media when he examined the "portrayal management" of online identities afforded by personal webpages. The anonymity of the Internet meant that individuals could selectively represent themselves to others without the complexities of physical contact. Nearly two decades later, the study of online communities and identity formation has entered into the mainstream, leading to novel methods of data collection and analysis. A growing stream of research in IS now draws on the constructs of identity and socio-cultural norms to explain online behavior. Burtch et al. (2016), for example, explored how contributing to the same crowdfunding campaign could foster a common identity in users and influence a campaign's fundraising success. Finally, in the more traditional domain of economics, Akerlof and Kranton (2000) presented one of the first and most influential economic analyses of the role of social and group identity on behavior. Their work illustrates how decisions about identity can be used to explain apparently irrational behavior, the creation of behavioral externalities in others, and preference evolution.

Digital Identity in Data Regulation and Professional Codes of Conduct

As we have tried to show, ML personalization conflicts with the various humanistic views of the self and identity. This conflict is paralleled by the recent passage of new data protection laws. The GDPR and the California Consumer Privacy Act (CCPA) both seek to give citizens more control over their personal data and their digital identities. The GDPR's gives data subjects rights to opt out of "automated profiling" with "legal" or other "significant" effects. The GDPR's *right to be forgotten* can be viewed as a potential legal remedy to the issue of one's practical and moral identity diverging over time. Data subjects have the right to destroy and recreate aspects of their digital identities when they no longer represent them as persons. At the same time, important ML societies, such as the IEEE and ACM, have also voiced concern about the effects of ML on human well-being. For example, recent IEEE guidelines state that autonomous and intelligent systems should give people access and control over their personal data and allow them "agency over their digital identity."[4] We hope to establish a standard basis and vocabulary from which the degree of ML personalization can be assessed by regulators and professional groups.

5   Evaluating ML Personalization

In order to better define and align ML personalization with a humanistic view of the person and current trends in data protection law, we suggest several criteria to guide future discussions. But first, two general comments: 1) the question of personalization is a matter of *degree* and not of *kind*; and 2) personalized scores have objective (outer-facing) and subjective (inner-facing) dimensions. As displayed earlier in Figure 1, the inner aspects of the person are inaccessible through observational methods. This implies that personal, subjective input from the

---

[4]The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, 1st ed, 2019 https://ethicsinaction.ieee.org

person herself is crucial in evaluating the level of personalization. The social sciences and humanities are especially well-positioned to contribute to the interdisciplinary development of such measures. At the same time, legal expertise will also be needed to relate these measures to data subjects' rights under current data protection laws, such as the GDPR.

Objective dimensions include the extent to which legally-defined personal data are used (*personal data component*), and the percent of other users sharing the same input data, personalized score, or discretized recommendation (*uniqueness component*). The personal data component implies that legally-allowable degrees of personalization may vary by country and legal regime. Regarding the uniqueness component, it is not clear how to weight these three aspects in determining the degree of personalization. For example, numeric measurements may uniquely identify you more so than categorical measurements, but these categories may more closely reflect your personal identity and community. Similarly, it is not obvious how to reconcile dimension reduction techniques with personalization. For example, when highly morally-salient measurements are correlated to other measures, should they be removed?

Another objective dimension worth examining is the accuracy of predicted scores (*accuracy component*). It seems reasonable to assume personalized predictions should be more accurate than non-personalized predictions. But how should accuracy be measured? As we noted above, the very narrow predictive goal of ML personalization can obscure the interpretation of some accuracy metrics. An easy way to improve accuracy is simply to narrow the size of the "recommendation set." However, in some cases, accuracy may not be desired; some persons may prefer to trade off predictive accuracy (in the narrow sense above) for informational privacy. After all, under the GDPR, data subjects have basic rights to informational privacy.

Subjective dimensions include the extent to which personalized scores respect the data subject's right to determine his public self, captured in the notion of informational privacy (*self-determination component*). This subjective component is relevant for debates on fair and transparent ML (FATML), as we can use it to sketch out a notion of a fair decision. Current legal and ML approaches to fairness operate on the level of the "database representation" of the person (i.e., practical identity), completely ignoring the person's moral identity. A fair decision should capture the person's intended public self and associated goals, attitudes, meanings and motivations in that context (*right reasons component*). The relevance of reasons given in algorithmic explanations can be assessed, for example, using relevance feedback techniques already in use in information retrieval. Lastly, another subjective dimension is the moral importance of the personalization context (*moral importance component*). For example, how should the different subjective and objective components be weighted when assessing the personalization of a system that recommends a movie versus a bail amount?

Finally, our proposed list of evaluation components is by no means exhaustive. In fact, some of these components might interact with or even contradict other aspects of personalization. For example, serendipitous recommendations[5] (Ge et al., 2010) can stimulate evolution in one's moral identity, yet might conflict with the *accuracy component*. Allowing users to decide for themselves how they would like to weight these aspects may be one solution.

6   Future Work and Conclusion

Going forward, we believe that as ML personalization becomes increasingly rooted in our daily lives, we should consider new approaches to data collection, such as those used in qualitative social science (e.g., hermeneutics and discourse analysis). By incorporating more diverse forms of personal data into personalized scores, we may be able to reduce the gap between one's identity as a *person* embedded in social and cultural space and as a *feature vector* embedded in feature space. As we pointed out, however, the ML personalization process will never be perfect.

Our presentation of the ML pipeline can also be expanded to consider further ML operations, such as how and which data are used for training algorithms. For example, personalized scores for a given person are computed using models that were trained on data from other persons, typically excluding that person's own data. Chronology of the training set is also important when considering the dynamism of a person's self-concept.

Lastly, these evaluation dimensions are merely an entry point into an interdisciplinary discussion about personalization. We hope to initiate and contribute to such a discussion by describing what the process of ML personalization typically looks like, particularly its defining metaphor of the person as feature vector. We argued that this conception is radically different from that found in the humanities, social sciences, and legal frameworks. As the long-term effects of ML personalization on personal identity, politics, law and society are still uncertain, it is important to critically examine the process of ML personalization. A precise definition of it may not be possible, but highlighting its most salient characteristics may be enough to foster new insights on the issues of algorithmic bias, transparency, and fairness. We end with a quote by the philosopher John Rawls: "Explication is elimination: we start with a concept the expression for which is somehow troublesome; but it serves certain ends that cannot be given up. An explication achieves these ends in other ways that are relatively free of difficulty" (Rawls, 2009, p. 111).

---

[5]Serendipity is used in Instagram's "downranking" procedure, under the assumption that one's moral identity should not remain static and so new topics should be presented in order to provide diverse personalized content.

References

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *Quarterly Journal of Economics*, 115(3):715–753.

Aquino, K., Reed, I., et al. (2002). The self-importance of moral identity. *J of personality and soc psych*, 83(6):1423.

Atkins, K. (2010). *Narrative identity and moral identity*. Taylor & Francis.

Barocas, S., Hardt, M., and Narayanan, A. (2017). Fairness in machine learning. *NIPS Tutorial*.

Baumeister, R. F. and Bushman, B. (2014). *Social psychology and human nature*. Belmont, CA: Cengage Learning.

Burtch, G., Ghose, A., and Wattal, S. (2016). Secret admirers: An empirical examination of information hiding and contribution dynamics in online crowdfunding. *Information Systems Research*, 27(3):478–496.

Cremonesi, P., Koren, Y., and Turrin, R. (2010). Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM.

Davidson, D. (1963). Actions, reasons, and causes. *The journal of philosophy*, 60(23):685–700.

De Vries, K. (2010). Identity, profiling algorithms and a world of ambient intelligence. *Ethics and information technology*, 12(1):71–85.

Douglas, M. and Ney, S. (1998). *Missing persons: A critique of the personhood in the social sciences*, volume 1. Univ of California Press.

Erickson, T. (1996). The world-wide-web as social hypertext. *Communications of the ACM*, 39(1):15–17.

Fan, H. and Poole, M. S. (2006). What is personalization? perspectives on the design and implementation of personalization in information systems. *Journal of Org. Comp. and Elec. Comm.*, 16(3-4):179–202.

Fuat Firat, A., Dholakia, N., and Venkatesh, A. (1995). Marketing in a postmodern world. *European journal of marketing*, 29(1):40–56.

Ge, M., Delgado-Battenfeld, C., and Jannach, D. (2010). Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM.

Gert, B. (1998). *Morality: Its nature and justification*. Oxford University Press on Demand.

Goriunova, O. (2019). The digital subject: People as data as persons. *Theory, Culture & Society*, 36(6):125–145.

Greene, T., Shmueli, G., Ray, S., and Fell, J. (2019). Adjusting to the gdpr: The impact on data scientists and behavioral researchers. *Big data*, 7.

Helmond, A. (2010). Identity 2.0: Constructing identity with cultural software. In *Proceeding of Mini-conference initiative, University of Amsterdam (Amsterdam, ND–Jan 20-22 2010)*. Citeseer.

Introna, L. D. and Whitley, E. A. (1996). Information systems as a social science? the individual perspective. *Information Systems*, 8:16–1996.

Kelleher, J. D., Mac Namee, B., and D'arcy, A. (2015). *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT Press.

Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805.

Larsen, R. and Buss, D. M. (2009). *Personality psychology*. McGraw-Hill Publishing.

Liang, T.-P., Lai, H.-J., and Ku, Y.-C. (2006). Personalized content recommendation and user satisfaction: Theoretical synthesis and empirical findings. *Journal of Management Information Systems*, 23(3):45–70.

Lyotard, J.-F. (1984). *The postmodern condition: A report on knowledge*, volume 10. U of Minnesota Press.

Manders-Huits, N. (2010). Practical versus moral identities in identity management. *Ethics and information technology*, 12(1):43–55.

Markus, H. and Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual review of psychology*, 38(1):299–337.

McAdams, D. P. (1995). What do we know when we know a person? *Journal of personality*, 63(3):365–396.

Mischel, W., Shoda, Y., and Ayduk, O. (2007). *Introduction to personality: Toward an integrative science of the person*. John Wiley & Sons.

Moreau, S. (2010). What is discrimination? *Philosophy & Public Affairs*, 38(2):143–179.

Nichols, D. (1998). Implicit rating and filtering. In *Proc of 5th DELOS Workshop on Filtering and Collab Filtering*.

Rawls, J. (2009). *A theory of justice*. Harvard university press.

Ricoeur, P. (1994). *Oneself as another*. University of Chicago Press.

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.

Schechtman, M. (2018). *The constitution of selves*. Cornell university press.

Shoemaker, D. W. (2010). Self-exposure and exposure of the self: informational privacy and the presentation of identity. *Ethics and Information Technology*, 12(1):3–15.

Skinner, B. F. (1965). *Science and human behavior*. Simon and Schuster.

Stryker, S. and Burke, P. J. (2000). The past, present, and future of an identity theory. *Social Psychology Quarterly*, 63(4):284–297.

Taylor, C. (1989). *Sources of the self: The making of the modern identity*. Harvard University Press.

Van Otterlo, M. (2013). A machine learning view on profiling. *Privacy, Due Process and the Computational Turn-Philosophers of Law Meet Philosophers of Technology. Abingdon: Routledge*.

Yeomans, M., Shah, A., Mullainathan, S., and Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, 32(4):403–414.