Setting the Stage: Towards Principles for Reasonable Image Inferences

Severin Engelmann
engelmas@in.tum.de
Chair of Cyber Trust
Department of Informatics
Technical University of Munich

Jens Grossklags
jens.grossklags@in.tum.de
Chair of Cyber Trust
Department of Informatics
Technical University of Munich

ABSTRACT

User modeling has become an indispensable feature of a plethora of different digital services such as search engines, social media or e-commerce. Indeed, decision procedures of online algorithmic systems apply various methods including machine learning (ML) to generate virtual models of billions of human beings based on large amounts of personal and other data. Recently, there has been a call for a "Right to Reasonable Inferences" for Europe's General Data Protection Regulation (GDPR). Here, we explore a conceptualization of reasonable inference in the context of image analytics that refers to the notion of evidence in theoretical reasoning. The main goal of this paper is to start defining principles for reasonable image inferences, in particular, portraits of individuals. Based on an image analytics case study, we use the notions of first- and second-order inferences to determine the reasonableness of predicted concepts. Finally, we highlight three key challenges for the future of this research space: first, we argue for the potential value of hidden quasi-semantics. Second, we indicate that automatic inferences can create a fundamental trade-off between privacy preservation and "model fit" and, third, we end with the question whether human reasoning can serve as a normative benchmark for reasonable automatic inferences.

CCS CONCEPTS

• **Information systems** → *Recommender systems*; *Personalization*; *Clustering and classification*; • **Security and privacy** → *Social aspects of security and privacy*.

KEYWORDS

Image data, Reasonable inferences, Machine learning

ACM Reference Format:
Severin Engelmann and Jens Grossklags. 2019. Setting the Stage: Towards Principles for Reasonable Image Inferences. In *27th Conference on User Modeling, Adaptation and Personalization Adjunct (UMAP'19 Adjunct), June 9–12, 2019, Larnaca, Cyprus.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3314183.3323846

1 INTRODUCTION

Recently, user modeling techniques have been used to infer aesthetic (e.g., beauty), mental (e.g., beliefs, intentions), emotional (e.g., happiness, depression), and social (e.g., group affiliation) features about individuals based on their personal data as well as their digital footprints. The possibilities of user modeling techniques go far beyond the mere classification of individuals into types of customers: they create virtual models of individuals at an industrial scale based on personal and other data. This data is commonly associated with implicit mental characteristics and social situational factors often unknown to the corresponding individual. Thereby, many big data companies produce billions of virtual models of people to connect a particular informational resource (e.g., an advertising material) to the individual with the most "appropriate" model.

This signifies what we refer to as a hermeneutic shift: parts of the interpretative potential of the person is realized not by the person itself but by the "quasi-semantic power"[1] of textual extraction, image understanding, emotion and speech analysis, location analysis or even inaction interpretation (among others) [4, 25, 34, 49, 50]. Assigning quasi-semantic values to implicit identity claims stands in contrast to The Enlightenment's core idea that humans have the ability to freely and autonomously assign meaning to what they have experienced. From this perspective, user modeling techniques can create tensions with the autonomy of individuals to form a hermeneutic self-concept.

Moreover, the quasi-semantic power of user modeling techniques can lead to consequential discriminatory biases, for example, when credit decisions are based on the collection and analysis of digital footprints unknown to the corresponding individual. The opacity of user modeling processes makes it generally difficult to detect, understand and correct such biases.

Recently, there has been a call for a "Right to Reasonable Inferences" to set legally-binding standards with the purpose to protect individuals against inferences that are privacy-invasive, reputation-damaging, and difficult to verify [45]. Yet, the decisive question is what *reasonable* ought to mean in the context of an automatic inference about a person based on some published media content.

Here, we wish to set the stage for a productive discussion between the computer and social sciences in determining standards for reasonable inferences in image analytics.[2] Based on an image analytics case study using the Clarifai concept prediction prototype[3], we show that inferences about human portraits can be unreasonable when they predict concepts with underlying beliefs that cannot be revised in light of further evidence of the same type. Our claims

UMAP'19 Adjunct, June 9–12, 2019, Larnaca, Cyprus

[1] Since humans are the only semantic engines in nature, see, for example, [11].
[2] Specifically, images that depict human beings.
[3] Available at: https://www.clarifai.com/demo.

are based on an empiricist view of reasonableness[4] that considers a knowledge-object's quality of evidence for a particular inference to qualify as reasonable or unreasonable.

We proceed as follows. In Section 2, we discuss why image analytics result in epistemic *and* ethical challenges and review related work in Section 2.1. In Section 3, we introduce an empiricist conceptualization of reasonableness that demands that what one is justified in believing is determined exclusively by evidence. We then upload two portraits to the Clarifai web interface image prediction prototype and analyze the reasonableness of the concepts the engine returns (see Section 4). Finally, in Section 5, we consider the potential autonomy-enabling value of hidden quasi-semantics and discuss a fundamental trade-off between privacy and model fit.

2 BACKGROUND

Social media users engage in both explicit[5] and implicit identity claims. Generally, images are among the most prevalent forms of self-presentation techniques on social media. Given their inherent semantic ambiguity, images are considered implicit identity claims. Implicit identity claims are "given off" in various indirect manners. Typical examples of implicit identity claims are showing one's affiliation to certain individuals, social or institutional groups, or expressing preferences and interests in an indirect manner [7, 48]. Indeed, there is evidence that "showing rather than telling" has become the most common self-presentation strategy on social media platforms [21, 43].

Consequently, marketers value images more than other media content. According to Socialbakers, images posted on Instagram[6] create four times more user engagement than other user content on Facebook[7]. Another reason is that image understanding further closes the gap between organic and commercial media content since objects in an image can be classified as products. Overall, there have been significant efforts made in the advancement of image-understanding technologies to model users based on pictorial identity claims in both academia and industry.[8]

When modeling an individual, image-understanding technologies do not simply draw semantics from the content of images but assign, add, and possibly produce their meaning in the first place. Despite their quasiness, user modeling techniques model features of individuals that are likely inaccessible for the individual herself. Thereby, user modeling techniques presumably attempt to transfer what is radically subjective (and therefore difficult if not impossible to falsify) into the realm of objective evaluation. They, therefore, try to explain something that is essentially first-person in third-person terms.

The majority of contemporary philosophical theories on personal identity support the idea that being free in interpreting one's self is a constitutive element of the conceptual boundaries of personal identity [12, 26, 31, 39]. Importantly, a moral status comprising moral rights and duties presupposes autonomy over one's self-concept. In other words, it is *because* individuals can evaluate what they are, shape whatever they wish to be on this basis, that they can be made responsible for what they become [40]. Moral accountability would, therefore, be impossible if individuals did not have the freedom and autonomy to form and negotiate such a hermeneutic self-concept.

Furthermore, empirical studies in psychology have demonstrated that individuals have the ability to attribute meaning to their experiences as a processes of hermeneutic identity formation [24, 36, 37]. Studies by [38] show that individuals interact with other individuals strategically in order to verify their self-concept: self-concept negotiation denotes the verification attempt of a person's self-concept through the interaction with other individuals. Whether individuals perceive user modeling outcomes as a means of technologically-mediated self-verification or self-discontinuity remains to be studied. Yet, hiding a person's quasi-semantic self-concept, i.e. disallowing user modeling techniques to partake in a self-verification process, could have some benefits (see Section 5).

Taken together, an autonomous self-concept emerges when an individual carries out *the psychological work required to attribute meaning to certain experiences*. Image analytics signify a hermeneutic shift because they transform implicit identity claims into explicit declarations of identity. Image analytics are not solely epistemic tools but quasi-semantic engines that potentially interfere with a person's autonomy to freely form a self-concept.

2.1 Related Work

With the rise of search engines in the early 2000s, automatizing the attribution of semantics to images returned high accuracy on object identification [23]. In the context of search tasks, object identification proved to be an efficient strategy.[9] In social media's people-based marketing mere object identification does not suffice for advertisement delivery based on implicit identity claims. Today, learning from content and structure of social network sites as well as correlating aspects about natural persons and groups to online content is a fast-growing research field. In the following, we briefly discuss main trends as they pertain to image data analyses.

Popularity prediction of image data: Several projects focus on determining the likelihood that certain image postings will achieve high view counts and high positive approval. Using a variety of machine learning approaches the context of a user and posting is taken into consideration to predict the future attention given to a newly posted image (e.g., [15, 27, 46, 47]).

Self-presentation: Various papers explore how (and under what circumstances) individuals strategically manage their social network accounts to aim for more favorable reception by the intended audience (e.g., [32, 41]). In the context of image data, for example, researchers have begun exploring users' management of multiple accounts on Instagram to present themselves to different audiences

[4]The terms "reasonableness" and "rationality" are considered synonymous in this work.

[5]For example, when individuals communicate specific self-relevant information in written form, they usually engage in explicit identity claims: "I am 20 years of age and I like reading biographies of great scientists".

[6]Advertising campaigns on Instagram are run via the Facebook advertising platform including the choice of custom audiences and lookalike audiences: see https://business.instagram.com/advertising/.

[7]https://www.socialbakers.com/blog/instagram-engagement

[8]For example, Amazon: https://aws.amazon.com/de/rekognition/, Microsoft: https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/, Facebook: https://code.fb.com/ai-research/fair-fifth-anniversary/, Google: https://cloud.google.com/vision/.

[9]Object inferences can be semantically ambiguous. For example, while distinct colors and shapes can be mapped to mathematical vectors with relative ease, the same is more difficult with objects containing continuous features [44].

in strategically altered ways. On a “Rinsta” (Real Instagram) account, a curated self is presented to a wider audience; whereas on a “Finsta” (Fake Instagram) account, less perfect material is presented to a hand-selected group of individuals for feedback and banter [21]. Interestingly, research has shown that users perceive their carefully styled images on the Finsta accounts to capture their real self more accurately in comparison to their Rinsta accounts with presumably more “genuine” material [21].

Inferring personality traits and user characteristics from image data: Partly triggered by the Gaydar research study [19] in 2009, significant attention has been given by the research community to finding associations between aspects of user profiles, user relationships, and posts, on the one hand, and traits/characteristics of the user or groups of users, on the other hand. In the context of image data, recent research suggests a relationship between personality traits and style aspects of posted pictures (e.g., hue, brightness and saturation); likewise, the content of pictures can be associated with personality characteristics [8–10].

Previous work also aims to find image characteristics that match specific user groups [17]. Likewise, analyses focus on automatically detecting gender and age from posted image content [16, 33].

Behavioral research has also explored how different personality characteristics (e.g., narcissistic tendencies [20]) impact the perception of image data.

Relationship of mental health and image data: Numerous research projects have focused on uncovering correlations between the usage of social network sites and mental health aspects such as addiction, anxiety, depression or body image (see, for example, a recent review [13]). Similar work can be found that is focused on image data. For example, perusal of attractive pictures of celebrities and peers has been found to be associated with a more negative body image by women [3, 18]. Likewise, uploaded image data can also be revealing of mental health indicators such as related to depression [30].

While there is a plethora of technical research and behavioral studies to understand social network site usage and its impact on users, also in the context of image data, we are unaware of any work that explores principles to develop reasonable standards for image inferences made by automated systems.

3 FIRST STEPS TOWARDS PRINCIPLES FOR REASONABLE IMAGE INFERENCES

3.1 An empiricist view of reasonable inferences

Fundamentally, there are two types of reasoning: practical and theoretical reasoning also sometimes referred to as instrumental and epistemic reasoning, respectively (see for example [35]). Practical reasoning is concerned with the question “What to do?”. Theoretical reasoning asks “What to believe?”. Practical and theoretical reasoning are not mutually exclusive. When choosing a reasonable action for a desirable outcome an individual relies on a theoretically reasonable belief. Thus, practical or instrumental reasoning usually follows theoretical reasoning.

In this work, we assume an empiricist view that considers a knowledge-object’s quality of evidence to decide whether a particular inference qualifies as reasonable or unreasonable. The empiricist view of a reasonable inference considers whether the belief about a proposition is *proportional* to the evidence available. Generally, the empiricist view on being reasonable in the theoretical sense considers the “goodness” or “fitness” of reasons provided that favors the truth of a proposition. While this conceptualization of reasonableness perhaps seems simple or even trivial, empirical research has demonstrated that individuals exhibit many information-processing biases pursuant to this empiricist account of reasonableness [2, 42].[10]

The goal of this work is to start developing principles for *portrait* image inferences that are eligible to be called reasonable. To do this, we need an example output from an image analytics engine. Here, we use the Clarifai web interface image prediction demo, which is based on deep convolutional neural networks (CNNs). We upload two portraits (see Figure 1 and Figure 2) to this image prediction demo and analyze the reasonableness of the concepts the engine returns. Corresponding to the literature reviewed in Section 2.1, we view a single image as a stand-alone knowledge-object whereby a predicted concept (i.e., the predicted outcome) is based only on the content of that single image.

3.2 Case study: Reasonableness and correctness of predicted concepts for two portraits

Reasonable and correct inferences

Consider the two images in Figure 1 and Figure 2. Is the content of these two images eligible to serve as evidence for the inferences made (see “predicted concepts” top right corner on both images)?

Figure 1 displays the face of a woman. The first three predicted concepts “woman”, “portrait”, and “facial expression” cannot be argued against, just like the first five predicted concepts in Figure 2. Here, the given beliefs about these propositions are *proportional* to the evidence available and therefore these inferences can be said to be reasonable. All of these features can be reasonably inferred from the evidence given. Note that we do not evaluate the potential discriminatory or unfair *consequences* of specific labels, rather we are first and foremost interested in their epistemic justification. For example, returning the label “gender” may lead to consequential discrimination independent from whether it is a (epistemically) reasonable inference. Additionally, considering our two portraits, the features “woman”, “portrait” and “facial expression” (Figure 1) and “portrait”, “eye”, “face”, “guy”, “man” (Figure 2) have been classified correctly.[11] Overall, these inferences are – to a large enough degree – reasonable and correct.

Reasonable inferences with incorrect predictions

Other predicted concepts can in principle be reasonable but seem to have been classified incorrectly for the specific portraits given. In Figure 2, for example, the CNNs predict the concept “smile”, which is incorrect since the person depicted does not seem to smile. Note that this would not be an unreasonable inference since a face can potentially bear a smile. Rather, the accuracy of the training set’s classification (i.e., the ground truth) is insufficient in returning an otherwise reasonable inference correctly. In this specific case, the

[10] For example, category mistakes, anchoring, representative bias, ignoring the context, framing effects etc.

[11] For Figure 2, the predicted concepts “hair”, “model”, “skin” seem to be reasonable and correct as well.

(a) Female portrait

PREDICTED CONCEPT	PROBABILITY
woman	0.980
portrait	0.965
facial expression	0.964
fashion	0.930
pretty	0.930
multicultural	0.925
one	0.911
wear	0.908
promotion	0.899
contemporary	0.858
indoors	0.856
friendly	0.847
arrival	0.846
people	0.842
elegant	0.822
intelligence	0.786
dentition	0.742
casual	0.728
business	0.720
charming	0.714

(b) Predicted concepts

Figure 1: Concept results using the Clarifai image prediction demo for a female portrait. The engine returns predictions on gender "woman", ethnicity-related features "multicultural", cognitive skills "intelligence", and presumably aesthetic features "pretty", "elegant", "friendly", "charming" (among others). For copyright purposes, we artistically rendered the original picture. Original picture ©https://thispersondoesnotexist.com/.

prediction seems to be incorrect but only in relation to an otherwise reasonable assumption made when annotating the training set.

Unreasonable inferences due to non-falsifiability

There seem to be inferences that are unreasonable due to their non-falsifiability. For example, both images contain predicted concepts of aesthetic evaluations or judgments. For a judgment to be an aesthetic judgment it necessarily needs to be subjective, making it the exact opposite of an empirical judgment. More generally, judgments on beauty and ugliness are commonly taken to be core examples of aesthetic judgments. In Figure 1, an example of an aesthetic judgment is "pretty" and in Figure 2 "fine-looking". Other, perhaps more indirect, aesthetic evaluations seem to be "elegant", "friendly", and "charming" (Figure 1) as well as "serious" (Figure 2). Overall, such aesthetic judgments of taste are unreasonable since they cannot be falsified by additional evidence of the same type. For such inferences, additional image evidence cannot *in principle* verify or falsify, in other words, change the proposition.[12]

Similarly to aesthetic inferences, another class of inferences are unreasonable due to their non-falsifiability. These inferences contain category mistakes because they take a physical or anatomical property to be evidence for a mental feature. In Figure 1, the facial proportions of the woman are taken to be evidence for her "intelligence" while the face in Figure 2 is taken to be evidence for the person to be "crazy". Portraits seem to be inadequate evidence for a person's mental capabilities or, generally, their mental characteristics. This inference cannot be made more reasonable by providing more portraits of the two people shown in Figure 1 and Figure 2. In other words, the proposition that the person in Figure 2 is actually crazy does not become more likely the more pictures of that person are analyzed. Again, the prediction for such labels can be correct

[12]There are, however, reasonable physical or anatomical inferences, for example, "freckle" in Figure 2.

(a) Male portrait

PREDICTED CONCEPT	PROBABILITY
portrait	0.993
eye	0.986
face	0.974
guy	0.974
man	0.971
fine-looking	0.968
young	0.961
hair	0.938
boy	0.933
people	0.932
blood	0.930
dark	0.916
freckle	0.909
serious	0.909
model	0.909
crazy	0.909
fashion	0.895
funny	0.893
smile	0.890
skin	0.888

(b) Predicted concepts

Figure 2: Concept results using the Clarifai image prediction demo for a male portrait. The engine returns predictions on gender "man", age "young"/"boy", mental "crazy"/"funny", and presumably aesthetic features "fine-looking", "serious" (among others). For copyright purposes, we artistically rendered the original picture. Original picture ©Bruce Gilden.

but only in relation to the unreasonable assumptions made when annotating the training set.

4 ANALYSIS OF THE CASE STUDY

There is an epistemic difference between descriptively identifying the objects "basketball" and "person" and conclusively inferring "Interest person x = basketball", merely because these objects have been identified. In a similar vein, there is a difference between measuring the physical property "wide space between eyes"and the object "glasses" and inferring some measure of intelligence based on these features. In our case study, we generally judged inferences that could be "directly" read off the portrait as reasonable. Such first-order inferences, as one might want to call them, seem epistemically valid and are henceforth difficult to object morally. They are reasonable independent of the predictive strength of the model.

Unreasonable inferences, on the other hand, seem to be predominantly constructed inferences. In our case study, they included claims about the person that could not be observed or accessed through the evidence given. Such second-order inferences presuppose a selection (and naturally a disregard) of specific first-order inferences that – combined – produce a new proposition. Second-order inferences must not necessarily be unreasonable. Consider, for example, the predicted concept "indoors" for the portrait in Figure 1. Predicting whether a depicted scenery is indoors or outdoors is a second-order inference because a single object is unlikely to produce a definite conclusion. The difference is that this second-order inference is responsive to additional evidence of the same type resulting in belief revision. Thereby, an inference is unreasonable in the case that novel or additional evidence becomes available that defeats the previous justification to believe in a proposition. In case of better evidence one ought to change the previously held belief in light of this new evidence. For example, another image of this scenery could in principle provide what Pollock refers to as "rebutting evidence" [29]. The new image is the same type or source of evidence. But because it is a reasonable second-order inference it is responsive to belief revision, which in this case is equivalent to the principles of Bayesian inference.

This claim does not hold for unreasonable second-order inferences. Bayesian inference (or belief revision) cannot convert an unreasonable second-order inference into a reasonable inference (e.g., predicted concept "intelligence" in Figure 1). Such category mistakes can only be reverted by changing the underlying assumption or by gathering different types of evidence but not by considering more evidence with the same category mistake.

5 DISCUSSION & CONCLUDING REMARKS

In this discussion paper, we applied an empiricist account of reasoning to determine the reasonableness of predicted concepts in the context of an image analytics case study. This is only one of many possible accounts of reasoning each of which comes with specific trade-offs. Arguably, an empiricist account is autonomy-preserving but limited to first-order inferences about individuals. Regardless of the account of reasonableness, an inference may be reasonable and correct but still be rejected by the individual. Here, one could argue that an inference becomes reasonable only when the data subject agrees with its proposition.

The recent call for a "Right to Reasonable Inferences" proposes a "Right to know about Inferences" and a "Right to rectify Inferences" (among others) [45]. However, hiding the quasi-semantic power of user modeling techniques does have its benefits. By revealing the logic involved in making hermeneutic inferences, the system directly recommends these hermeneutics to the user. It remains to be explored how individuals would perceive information on inferences as given in our two image examples. Revealing at least in part the manner and content of user modeling processes and outcomes enables internalization and conformation to the proposed inferences. Perhaps individuals would welcome such a degree of transparency as a mechanism to "offload" the psychological work necessary to attribute meaning to certain life events. Revealing such inferences to the individual means recognizing their quasi-semantic power in shaping who we are and who we can become – we accept that they have their own narrative capacity. Thus, transparency of user modeling inferences could even exacerbate the polarization effect observed in social media personalization.

Another key challenge is privacy. Image inferences tend to become more reasonable the more personal data is collected and analyzed. This creates a privacy trade-off. The trade-off consists in the observation that a representative model of an individual is possible only at the expense of privacy. For example, ML classifiers must be able to respond to concept drift without "neglecting" the outdated data when learning a model of personal identity [51]. For example, sliding windows of fixed and variable sizes of training data are used to build an updated model [14]. Since both fixed and variable windows are definite in their size, some old data will necessarily be forgotten. What criteria determine which data are to be forgotten and which ones are to be considered in creating an updated representative model of a person? Model fit requires a potentially uninterrupted flow of data possibly resulting in significant privacy challenges [5].

Finally, a key question is whether we should take human reasoning as a benchmark for reasonable automatic inferences. In the empirical literature on human reasoning ...*"the ordinary person is claimed to be prone to serious and systematic error in deductive reasoning, in judging probabilities, in correcting his biases, and in many other activities"* [6]. For example, humans make judgments about cognitive capabilities based on physical properties [1, 28]. Following our image analytics case study, we conclude that inferences about individuals' cognitive and mental features are unreasonable since an image does not provide the kind of evidence needed to justify such claims. This also counts for inferences made about individuals' intentions or goals based on image evidence (see [22]).

Overall, it will remain a pressing ethical challenge to define normative standards of reasonableness that automatic image inferences should comply with.

Acknowledgments: We thank the reviewers for their insightful comments that helped to improve our work. The paper is based on research conducted as part of a Volkswagen Foundation planning grant project.

REFERENCES

[1] Michael Argyle and Robert McHenry. Do spectacles really affect judgements of intelligence? *British Journal of Social and Clinical Psychology*, 10(1):27–29, 1971.
[2] Dan Ariely, George Loewenstein, and Drazen Prelec. "Coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, 118(1):73–106, 2003.
[3] Zoe Brown and Marika Tiggemann. Attractive celebrity and peer images on Instagram: Effect on women's mood and body image. *Body Image*, 19:37–43, 2016.
[4] Buru Chang, Yonggyu Park, Donghyeon Park, Seongsoon Kim, and Jaewoo Kang. Content-aware hierarchical point-of-interest embedding model for successive POI recommendation. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 3301–3307, 2018.
[5] Michela Chessa, Jens Grossklags, and Patrick Loiseau. A game-theoretic study on non-monetary incentives in data analytics projects with privacy implications. In *IEEE Computer Security Foundations Symposium (CSF)*, pages 90–104, 2015.
[6] Jonathan Cohen. Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4(3):317–331, 1981.
[7] Nicole Ellison, Charles Steinfield, and Cliff Lampe. Connection strategies: Social capital implications of Facebook-enabled communication practices. *New Media & Society*, 13(6):873–892, 2011.
[8] Bruce Ferwerda and Marko Tkalcic. Predicting users' personality from Instagram pictures: Using visual and/or content features? In *Conference on User Modeling, Adaptation and Personalization (UMAP)*, pages 157–161. ACM, 2018.
[9] Bruce Ferwerda, Markus Schedl, and Marko Tkalcic. Predicting personality traits with Instagram pictures. In *Workshop on Emotions and Personality in Personalized Systems*, pages 7–10. ACM, 2015.
[10] Bruce Ferwerda, Markus Schedl, and Marko Tkalcic. Using Instagram picture features to predict users' personality. In *International Conference on Multimedia Modeling (MMM)*, pages 850–861, 2016.
[11] Luciano Floridi. Web 2.0 vs. the semantic web: A philosophical assessment. *Episteme*, 6(1):25–37, 2009.
[12] Harry Frankfurt. Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68(1):5–20, 1971.
[13] Rachel Frost and Debra Rickwood. A systematic review of the mental health outcomes associated with Facebook use. *Computers in Human Behavior*, 76:576–600, 2017.
[14] João Gama, Raquel Sebastião, and Pedro Pereira Rodrigues. On evaluating stream learning algorithms. *Machine Learning*, 90(3):317–346, 2013.
[15] Steve Göring, Konstantin Brand, and Alexander Raake. Extended features using machine learning techniques for photo liking prediction. In *International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2018.

[16] Kyungsik Han, Sanghack Lee, Jin Yea Jang, Yong Jung, and Dongwon Lee. Teens are from mars, adults are from venus: Analyzing and predicting age groups with behavioral characteristics in Instagram. In *Conference on Web Science*, pages 35–44. ACM, 2016.
[17] Kyungsik Han, Yonggeol Jo, Youngseung Jeon, Bogoan Kim, Junho Song, and Sang-Wook Kim. Photos don't have me, but how do you know me? Analyzing and predicting users on Instagram. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, UMAP '18, pages 251–256. ACM, 2018.
[18] Joshua Hendrickse, Laura Arpan, Russell Clayton, and Jessica Ridgway. Instagram and college women's body image: Investigating the roles of appearance-related comparisons and intrasexual competition. *Computers in Human Behavior*, 74:92–100, 2017.
[19] Carter Jernigan and Behram Mistree. Gaydar: Facebook friendships expose sexual orientation. *First Monday*, 14(10), 2009.
[20] Seunga Venus Jin and Aziz Muqaddam. "Narcissism 2.0! Would narcissists follow fellow narcissists on Instagram?" The mediating effects of narcissists personality similarity and envy, and the moderating effects of popularity. *Computers in Human Behavior*, 81:31–41, 2018.
[21] Jin Kang and Lewen Wei. Let me be at my funniest: Instagram users' motivations for using Finsta (aka, fake Instagram). *The Social Science Journal*, 2019.
[22] Owen King. Machine learning and irresponsible inference: Morally assessing the training data for image recognition systems. In Don Berkich and Matteo Vincenzo d'Alfonso, editors, *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence*, pages 265–282. Springer, 2019.
[23] Victor Lavrenko, Raghavan Manmatha, and Jiwoon Jeon. A model for learning the semantics of pictures. In *Advances in Neural Information Processing Systems*, pages 553–560, 2004.
[24] Mark Leary and June Price Tangney. *Handbook of Self and Identity*. Guilford Press, 2011.
[25] Weizhi Ma, Min Zhang, Chenyang Wang, Cheng Luo, Yiqun Liu, and Shaoping Ma. Your tweets reveal what you like: Introducing cross-media content information into multi-domain recommendation. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 3484–3490, 2018.
[26] Alasdair MacIntyre. *After Virtue: A Study in Moral Theology*. University of Notre Dame Press, 1981.
[27] Eric Massip, Shintami Chusnul Hidayati, Wen-Huang Cheng, and Kai-Lung Hua. Exploiting category-specific information for image popularity prediction in social media. In *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 45–46, 2018.
[28] Fhionna Moore, Dimitra Filippou, and David Ian Perrett. Intelligence and attractiveness in the face: Beyond the attractiveness halo effect. *Journal of Evolutionary Psychology*, 9(3):205–217, 2011.
[29] John Pollock and Joseph Cruz. *Contemporary Theories of Knowledge*. Rowman & Littlefield, 1999.
[30] Andrew Reece and Christopher Danforth. Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6(1):15, 2017.
[31] Marya Schechtman. The narrative self. In Shaun Gallagher, editor, *The Oxford Handbook of the Self*. Oxford University Press, 2011.
[32] Gwendolyn Seidman. Expressing the "true self" on Facebook. *Computers in Human Behavior*, 31:367–372, 2014.
[33] Junho Song, Kyungsik Han, Dongwon Lee, and Sang-Wook Kim. "Is a picture really worth a thousand words?": A case study on classifying user attributes on Instagram. *PloS One*, 13(10):e0204938, 2018.
[34] Dusan Sovilj, Scott Sanner, Harold Soh, and Hanze Li. Collaborative filtering with behavioral models. In *Conference on User Modeling, Adaptation and Personalization (UMAP)*, pages 91–99, 2018.
[35] Keith Stanovich. *Decision Making and Rationality in the Modern World (Fundamentals in Cognition)*. Oxford University Press, 2009.
[36] William Swann, Alan Stein-Seroussi, and Brian Giesler. Why people self-verify. *Journal of Personality and Social Psychology*, 62(3):392–401, 1992.
[37] William Swann, Peter Rentfrow, and Jennifer Guinn. Self-verification: The search for coherence. In Mark Leary and June Price Tangney, editors, *Handbook of Self and Identity*, pages 367–383. 2003.
[38] William Swann. Identity negotiation: Where two roads meet. *Journal of Personality and Social Psychology*, 53(6):1038, 1987.
[39] Charles Taylor. Responsibility for self. In Amélie Rorty, editor, *The Identities of Persons*. University of California Press, 1976.
[40] Charles Taylor. *Sources of the Self: The Making of the Modern Identity*. Harvard University Press, 1989.
[41] Leman Pınar Tosun. Motives for Facebook use and expressing "true self" on the Internet. *Computers in Human Behavior*, 28(4):1510–1517, 2012.
[42] Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981.
[43] José Van Dijck. 'You have one identity': Performing the self on Facebook and LinkedIn. *Media, Culture & Society*, 35(2):199–215, 2013.
[44] Jan Van Gemert, Cor Veenman, Arnold Smeulders, and Jan-Mark Geusebroek. Visual word ambiguity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7):1271–1283, 2010.
[45] Sandra Wachter and Brent Mittelstadt. A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review*, 2019.
[46] Bo Wu, Wen-Huang Cheng, Yongdong Zhang, Qiushi Huang, Jintao Li, and Tao Mei. Sequential prediction of social media popularity with deep temporal context networks. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 3062–3068, 2017.
[47] Zhongping Zhang, Tianlang Chen, Zheng Zhou, Jiaxin Li, and Jiebo Luo. How to become Instagram famous: Post popularity prediction with dual-attention. In *IEEE International Conference on Big Data (Big Data)*, pages 2383–2392, 2018.
[48] Shanyang Zhao, Sherri Grasmuck, and Jason Martin. Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior*, 24(5):1816–1836, 2008.
[49] Qian Zhao, Martijn Willemsen, Gediminas Adomavicius, Maxwell Harper, and Joseph Konstan. Interpreting user inaction in recommender systems. In *ACM Conference on Recommender Systems*, pages 40–48, 2018.
[50] Sicheng Zhao, Amir Gholaminejad, Guiguang Ding, Yue Gao, Jungong Han, and Kurt Keutzer. Personalized emotion recognition by personality-aware high-order learning of physiological signals. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 15(1s):Article No. 14, 2019.
[51] Indrė Žliobaitė. Learning under concept drift: An overview. *arXiv preprint arXiv:1010.4784*, 2010.