Structural Ageism in Big Data Approaches

Andrea Rosales[I] & Mireia Fernández-Ardèvol[II]

[I] Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya, Spain,
     arosalescl@uoc.edu
[II] Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya, Spain,
     mfernadezar@uoc.edu

Abstract

Digital systems can track every activity. Their logs are the fundamental raw material of intelligent systems in big data approaches. However, big data approaches mainly use predictions and correlations that often fail in the prediction of minorities or invisibilize collectives, causing discriminatory decisions. While this discrimination has been documented regarding, sex, race and sexual orientation, age has received less attention. A critical review of the academic literature confirms that structural ageism also shapes big data approaches. The article identifies some instances in which ageism is in operation either implicitly or explicitly. Concretely, biased samples and biased tools tend to exclude the habits, interests and values of older people from algorithms and studies, which contributes to reinforcing structural ageism.

Keywords: ageism, discrimination, big data, intelligent systems, literature review

Introduction

In this chapter, we focus on how intelligent systems used in big data approaches for research and innovation purposes are contributing to reinforcing ageism and how to tackle this problem. Our particular interest is in structural ageism, defined as "the way in which society and its institutions sustain ageist attitudes, actions or language in laws, policies, practices or culture" (AGE Platform Europe, 2016: 1). Ageism can be subtle or explicit (Ayalon & Tesch-Römer, 2018) and is even more pervasive than other acknowledged forms of discrimination, such as sexism or racism (Officer & de la Fuente-Núñez, 2018). Stereotypes attached to the use of information and communication technologies (ICTs) can contribute significantly to increasing digital inequality, as they establish implicit social modes of functioning that shape both social behaviours (Greenwald & Banaji, 1995) and digital skills (Lagacé et al., 2015). Discrimination includes practices that "reflect ageist prejudices and deprioritize, disregard or even exclude older people" (AGE Platform Europe, 2016: 1), which also affect the design of and the research on digital media.

The term intelligent systems or artificial intelligence refers to the capacity of any technology to learn from (the continuous flux of) information to obtain analytical capabilities and is often used for automated decision making (EPSC, 2018). Digital systems generate logs of all their activities that, in turn, constitute the inputs for intelligent systems and big data approaches. Digital services, internet service providers and individuals with some computational skills can collect such data. These so-called "big data" (boyd & Crawford, 2012) are used for different purposes, for example tailoring improved individualized digital products that rely on inferences on users' personality traits, age and gender (e.g., Schwartz et al., 2013) or making decisions regarding job positions, mortgages or insurance (O'Neil, 2016). They are also used in research to understand smartphone usage (e.g., Lee et al., 2014) or to identify the profile of users of social media sites (e.g., Bayot & Gon, 2017; Peersman et al., 2011). Although they can take into account all the granularity of the data (Kitchin, 2014), they often result in discriminatory decisions that reproduce (digital) inequality and exclude minorities (Eubanks, 2018; Hajian & Domingo-Ferrer, 2013; O'Neil, 2016). Such discrimination has attracted attention mainly regarding, sex, race, socio-economic background and sexual orientation (Eubanks, 2018; Hajian & Domingo-Ferrer, 2013; O'Neil, 2016).

In this sense, three pieces of evidence inspired this research. First, algorithms often reproduce stereotypes (e.g., Nguyen et al., 2014), as they rely on data sets that emerge from our everyday life. For example, big data approaches can rely on (stereotypical) information that users write on digital media. Second, minorities are poorly represented in big data approaches, because they mainly rely on predictions and correlations (Bonchi et al., 2017) that are acceptable if they suit 80 per cent of the population. However, the predictions are not accurate for the remaining 20 per cent, which tend to be minorities and invisibilized collectives of the population. Third, big data approaches often follow the opaque reasoning that fails to explain their procedures (Rieder & Röhle, 2012). Behind these opaque processes, there can be complex algorithms and human decisions that reproduce judgements, prejudices and ideologies shaped by age stereotypes. Some common prejudices include assuming that everyone would behave in the same way as the developer or the researcher, and consequently controlling the demographics of a sample is not relevant for big data analyses (boyd & Crawford, 2012).

Thus, our research question is: How does ageism shape intelligent systems and what are the implications for big data studies? Thus, the aim of the article is to create awareness of the implicit and explicit mechanisms of ageism inherent in big data approaches and to reflect on ways to tackle ageism. To achieve this aim, we conducted a literature review of research published since 2010. We focused on two particular uses of big data, age prediction and smartphone use analysis. They serve to illustrate how ageism operates by producing biased samples and biased tools by, among others, not controlling the demographics of the sample, involving limitations in the measuring systems and following implicit stereotyped decisions that discriminate against older people.

Theoretical framework

The limitations of big data

Digital systems create logs of all activities. Terabytes of data generated by real users in digital environments help the monitoring, or tracking, of everyday activities, leading to the so-called big data era (boyd & Crawford, 2012). The availability of these massive and complex data sets, or big data, is changing the paradigms of research and innovation in diverse disciplines (Kitchin, 2014), among them computer science, the economy, bio-informatics, sociology and political science (boyd & Crawford, 2012). Big data approaches, which use digital logs to learn from user behaviours, are nowadays pervasive (Schäfer & Van Es, 2017). They can perform systematic analyses of continuous data flows based on statistical instruments (Böhmer et al., 2011; Kosinski et al., 2013). In this sense, big data are used to understand digital practices through statistical models with the intention of predicting the behaviour of the population (e.g., Oulasvirta et al., 2012; Ørmen & Thorhauge, 2015). Algorithms are at the core of intelligent systems and serve different purposes, for example predicting personality, detecting crime or helping developers in tailoring products to users' interests.

However, big data approaches come with questions, not only regarding privacy and ethical protocols (Jensen, 2013; Smith et al., 2012) but also in methodological and epistemological terms (boyd & Crawford, 2012; Schäfer & Van Es, 2017). Digital tracking is meant to provide more objective data (Karikoski & Soikkeli 2013; Möller et al., 2013). As Karikoski and Soikkeli explained, objective data provide a measurement of activities that is exempt from the declarative subjectivities that are common in interviews or surveys (Karikoski & Soikkeli, 2013). Classical sampling techniques can be overcome, because the data volumes are large enough to allow an understanding of trends and correlations at the population level (Ferreira et al., 2012). However, the size of a database does not necessarily grant representativeness at the population level (Ferreira et al., 2012).

Moreover, data are a socio-technical phenomenon: the "crumbs" of a complex system of interrelations between data, systems and society (Letouzé, 2015). Tracked logs respond to the technical processes of the systems, which are not necessarily aligned to the needs of specific analyses and should be understood as a by-product of digital systems. Therefore, an interpretation of the available data is required that introduces subjectivity, which researchers often do not even recognize (boyd & Crawford, 2012). Thus, from the social sciences point of view, critical positions are arising towards the current techno-optimism that accompanies big data, which may suffer from limitations brought by the underlying assumptions, values and biases (boyd & Crawford, 2012).

Ageism and digital media

Age, old age and aging are socially constructed concepts. The concept of a "senior person" often has negative connotations (Garattini & Prendergast, 2015), among others because of the power structures that exclude most older people from decision making (Calasanti & King, 2015). Otherwise, we are living in an aged and aging (Eurostat, 2017) networked society (Castells, 2009), in which an increasing percentage of older people adopt and use digital technologies (Eurostat, 2018). However, older people are

a minority in digital media due to lower levels of adoption (ibid.) and, once adopted, lower levels of use than younger generations (Rosales & Fernández-Ardèvol, 2016a).

Digitally mediated communication has been studied extensively in children and youths, but it has been developed little in the case of the elderly (Mihailidis, 2014). Adolescents and youths have been the main reference generations in ICT studies because they establish trends that later become generalized to the rest of the population (Castells et al., 2006). Otherwise, the older population has systematically been excluded from ICT studies (e.g., Eurostat, 2018; Roca Salvatella, 2016), preventing the development of diverse models of ageing in the digital world. In turn, older people are rarely involved in the design of new technologies, some exceptions include projects developed by Rogers and colleagues (2014) and Righi and colleagues (2018). The use of digital devices and the goals of mediated communication change over time (Ling et al., 2012), because individual values and interests change (Neugarten, 1996). Aging in the broad sense conditions communicative habits and the choice of digital channels (Selwyn et al., 2003). However, when effectively incorporated into empirical analyses, their specificity is often not considered, which makes older people invisible (Hendricks, 2005) by ignoring their particular uses, interests and values.

Older people are often portrayed as less avid users, lacking interest or unable to use digital technologies properly: a thesis that would sustain the statistics because, from the age of 55 years onwards, ownership and use fall below the average (Eurostat, 2018). However, older people constitute a diverse user group (Sawchuk & Crow, 2011). An increasing number of studies have reported that, when older people become users, they can be active users of digital systems (Choudrie & Vyas, 2014; Rosales & Fernández-Ardèvol, 2016a), though it is necessary to distinguish quantity and quality because the centrality of digital media does not depend on the intensity of use. Although older people can (or tend to) be less active users, their use of ICT is relevant and meaningful (Rosales & Fernández-Ardèvol, 2016a).

Discrimination in big data approaches

Big data approaches rely on intelligent systems for decision making by considering individuals' general characteristics, such as where they live, their social network or their credit score, which can lead to race and socio-economic discrimination (O'Neil, 2016). All intelligent systems use a data set for learning the patterns; however, if the learning data set uses collective data and the data are biased against a particular community, the system will reproduce such a stereotype. Instead of evaluating individuals, they are "unfairly treating people on the basis of their belonging to a specific group" (Hajian & Domingo-Ferrer, 2013: 1445). Discrimination also operates by making some collectives invisible (Hendricks, 2005), ignoring their particular practices of use, interests and values in the design of new technologies. Invisibilization reinforces inequality and injustice (Stocchetti, 2018), as poor, marginalized and vulnerable collectives may suffer the negative consequences from weapons of math destruction because data models do not consider their interests (O'Neil, 2016).

Another particular issue creating exclusion is the design of algorithms, based on the beliefs and values of researchers (Uricchio, 2017) that are not supported by empirical evidence. However, while discrimination in big data approaches has attracted attention regarding, sex, race, socio-economic background and sexual orientation (Eubanks, 2018; Hajian & Domingo-Ferrer, 2013; O'Neil, 2016), it has generated less interest regarding age.

Methods

We investigate the particular uses of big data in research since 2010, when articles using large collections of data from digital platforms became common. Thus, we systematically reviewed journal articles published in English between 2010 and 2018 in Scopus and conference proceedings related to big data and human–computer interaction.[1] We reviewed publications that included predicting systems regarding, first, the age of participants in social network sites and, second, smartphone user behaviour based on smartphone logs. For the first topic, the search keywords included age, prediction and social media or social network, which returned 472 valid results. For the second topic, the search keywords were smartphone, logs and user or usage, which produced 279 valid results. Among them, we selected those articles that explicitly reported the methods used to collect and analyse the data or discuss their biases (see Table 1 and Table 2). The total number of revised articles was 33. The analysis of the selected articles focused on identifying common ageist biases, and we concentrated on two, those that were common in most of the selected articles: biased samples and biased tools.

Table 1. *Classification of the reviewed articles on age predictions*

Article	Biased sample	Biased tool
Bayot and Gon (2017)		
Bi et al. (2013)	x	
Culotta et al. (2016)		x
Kosinski et al. (2013)	x	
Liao et al. (2014)		x
Liu and Yang (2012)	x	
Nguyen et al. (2013)		x
Nguyen et al. (2014)	x	
Oktay et al. (2012)	x	
Peersman et al. (2011)		x
Perozzi and Skiena (2015)		x
Rosenthal and McKeown (2011)	x	
Schwartz et al. (2013)		

Table 2. *Classification of the reviewed articles on smartphone usage*

Article	Biased sample	Biased tool
Alvarez-Lozano et al. (2014)		
Böhmer et al. (2011)	x	
De Montjoye et al. (2013)		
Ferdous et al. (2015)	x	
Ferreira et al. (2012)	x	
Ferreira et al. (2014)		
Holz et al. (2015)		x
Ikebe et al. (2012)		
Jones et al. (2015)	x	x
Karikoski and Soikkeli (2013)		
Kiukkonen et al. (2010)		x
Lee et al. (2014)		x
Rahmati et al. (2012)		x
Rosales and Fernández-Ardèvol (2016b)		
Shin et al. (2012)		x
Singh et al. (2013)		
Srinivasan et al. (2014)		x
Wagner et al. (2013)	x	x
Xu et al. (2016)		x
Yan et al. (2012)		x

Age prediction on social media sites

There is a growing interest in the automated prediction of users' age in intelligent systems. Particularly on social media sites, some of the exchanged information could be public (e.g., tweets or Facebook statuses); however, demographic information often stays private. With stakeholders showing a systematic interest in profiling audiences, the prediction of non-public socio-demographic data at the individual level is becoming valuable, as it will "significantly expand the breadth, detail, and interpretability of your measurements" (Popov et al., 2018) .

Methods have investigated a diversity of primary sources of information, such as written texts on social media sites (Bayot & Gon, 2017; Peersman et al., 2011), on Twitter (Liao et al., 2014; Nguyen et al., 2013) or on blogs (Rosenthal & McKeown, 2011). Beyond the contents posted by the user, other sources of information are music listened to (Liu & Yang, 2012) and Facebook likes (Kosinski et al., 2013). Other systems also rely on the name of the users (Oktay et al., 2012), their search queries (Bi et al., 2013) or their behaviour on a selected social media site (Culotta et al., 2016; Perozzi & Skiena, 2015b). However, revised methods often make poor predictions on age for older people (Nguyen et al., 2013). In the following, we describe the identified characteristics that contribute to such a lack of precision.

Biased samples

Predictive systems rely on learning data sets to learn common patterns of use. Intelligent systems compare new users with the patterns of the learning data set to make predictions.

However, in some cases, the characteristics of the learning data set are not provided or not discussed. Thus, it becomes impossible to identify biases, which might entail the exclusion of minorities. Particularly, the learning data set characteristics should be compared with the studied population, as the recruitment can favour the inclusion of some specific profiles. In such cases, predictions would be less accurate for those collectives that are not properly represented in the learning data set.

Three examples illustrate this case. First, Kosinski and colleagues (2013) research on Facebook and Bi and colleagues (2013) study of search queries relied on the same learning data set based on information from 58,000 volunteers. The authors did not explicitly provide the demographics of this learning data set. Thus, it is not possible to evaluate its possible biases. It is likely that older people are a minority in the learning data set, because they are a minority in the digital world, which will shape the predictions of both systems. Second, Oktay and colleagues (2012) predictions relied on the first name of US-based Twitter users. Their learning data set was the US public statistics on naming, at birth, particularly "the frequency of more than 150K different baby names for each year starting from 1881" (Oktay et al., 2012: 2). Migratory movements were not considered in these particular statistics – and some age cohorts might be more affected by immigration from different countries than others, which would shape the learning data set's representativeness. Finally, Liu and Yang (2012) predicted demographics on last. fm based on music. Their learning data set was a public data set of last.fm users (Liu & Yang, 2012); however, the possible biases of the learning data set were not discussed.

On other occasions, the bias is more explicit and hides the nuances of older age. This is the case when the research defines age groups that are too broad (e.g., those over 40 years old) or establishes upper age limits based on technical restrictions. Research on Twitter conducted in 2012 limited the predictions until 80 years old, because otherwise they "do not seem to contain genuine guesses" (Nguyen et al., 2014: 4). Another study, conducted in 2010, focused on individuals born between 1977 and 1988, because they "use social networking, blogs, and instant messaging more than their elders" (Rosenthal & McKeown, 2011: 769).

Biased tools

By building on homophilic ideas, some developers have assumed that an algorithm that suits some users well should suit all users well, particularly minority users. This is often associated with the limitations of the algorithms in making accurate predictions for some groups. Algorithms often use the average age of a user's network to predict his or her age (e.g., Culotta et al., 2016; Perozzi & Skiena, 2015). However, this equation does not apply to older users, as the average age of their network on social media sites is often lower than their age – because very often older people use social media to follow their children and grandchildren, while most of their friends are not part of their digital network. In line with this, authors such as Culotta and colleagues (2016) recognized that their age predictions worked better for younger cohorts.

Although many projects have argued that they predict the age of participants using text provided by users, the algorithm only manages to distinguish between younger cohorts and adults (Liao et al., 2014; Nguyen et al., 2013; Peersman et al., 2011), partly due to the limitation of the algorithm in making better predictions and partly due to all

the nuances of the language used by adults or the limited use of language among young people.

Smartphone log collection

The operating systems of smartphones allow third parties to develop applications that rely on the by-product information that the device provides. Mobile apps therefore allow the collection of data for different purposes, including the analysis of human behaviour.

Smartphone logs are key input for research aimed at understanding different dimensions of mobile communication, from smartphone use (Böhmer et al., 2011) or general communication behaviours (Wagner et al., 2013) to the way in which given contexts shape them (Karikoski & Soikkeli, 2013). At the intersection of social sciences and computer sciences, mobile logs help to predict selected human behaviour dimensions, including human stress (Ferdous et al., 2015), bipolar disorder states (Alvarez-Lozano et al., 2014) and other personality traits (De Montjoye et al., 2013). In the following, we describe the biases identified in this area.

Biased sample

Most ageist practices in intelligent systems design are related to the data set limitations in terms of the representativeness of the studied population and particularly to recruitment procedures that tend to exclude older people.

Smartphone log studies often offer participants a particular service to engage users and allow them to track their logs through an installed app (for instance Böhmer et al., 2011). In such cases, the mobile application should provide an added value, an appealing reason for the users to install it and contribute voluntarily to the study. However, the interest in the service or the research goal is not necessarily neutral and the resulting sample will not be random but incidental. The (learning) sample of the study would therefore be biased towards those individuals who perceive the advantages of the services linked to the research project. For instance, Securacy allows users to assess the security of their network (Jones et al., 2015) and Appazaar supports users in finding new apps (Böhmer et al., 2011) while collecting data for its studies. However, the users of these two apps might already have been concerned with network security or might be early adopters. The participants could be non-average users, and the "sample may not be representative of what all smartphone owners would do" (Ferreira et al., 2012: 12).

Another recruitment approach relies on snowball sampling procedures. Such is the case of Böhmer and colleagues (2011), who advertised their call on popular blogs and social network sites. These media do not necessarily reach a representative sample of the population. For instance, a study promoted a project on Facebook and Twitter "to reach a diverse population sample" (Jones et al., 2015: 1199); however, it made no reflection on the comparatively high popularity of those media among younger cohorts (Greenwood et al., 2016).

Maybe a more relevant issue is the fact that most revised studies did not present the demographics of their samples, which prevents an accurate evaluation of biases. While some articles did not discuss such a lack of information (Ferdous et al., 2015; Wagner et al., 2013), others showed a clear position. For instance, some preferred not to col-

lect demographic data, such as gender or age, to grant the privacy and security of data (Böhmer et al., 2011). Others, instead, argued that the extra steps necessary to collect personal data – such as the acceptance of informed consent – would reduce the willingness to participate in the research (Jones et al., 2015).

Biased tools

Tracking systems are not universal. They can monitor a limited number of smartphones depending on their operating system (OS) version and model. In addition, the technical condition of the smartphone shapes the efficiency of the tracker. These two dimensions exemplify the non-neutrality of the measurement tool and serve to illustrate how tools shape data collection.

Android and iOS dominate the smartphone operating system market, and trackers that focus on them would allow access to most smartphone users. However, different OS versions are in operation at the same time. Notably, the spectrum of Android OS versions is broader than that of iOS, with several handset models of multiple brands relying on different Android OS versions. Any tracking system should ensure its compatibility with all the OSs, brands and models available on the market. However, most of the revised tracking systems argue that they track Android and/or IOS devices but provide no further detail (Holz et al., 2015; Jones et al., 2015; Lee et al., 2014). Few articles have described the limits of the measurement tool, as Wagner and colleagues (2013) did by explaining that the system tracks 1277 different types of devices. Design limitations in the tracking system might exclude some OS versions, which would result in the exclusion of minority handset models from the sample, such as very new or very old models.

Digital logs are the raw data of what are seen as non-intrusive methods for data collection (Kiukkonen et al., 2010; Xu et al., 2016). Researchers have often argued that tracking systems provide everyday life data in everyday settings with little or no disturbance and require no effort from the participants (Holz et al., 2015; Lee et al., 2014; Shin et al., 2012). However, tracking systems affect the battery, memory and processor of the tracked device (Wagner et al., 2013). In this vein, some studies have expressed concern about the battery drain of monitoring systems (Rahmati et al., 2012; Srinivasan et al., 2014; Yan et al., 2012). Concretely, tracking systems have a considerable impact on smartphones with limited RAM and weak batteries. The capacities of the smartphone, therefore, could act as a limiting technical dimension preventing individuals from participating in a tracking study. However, none of the previous studies have reported the available capacities of the tracked smartphones.

Thus, the capacity of the tracking systems and the technical conditions of the smartphones prevent the inclusion of the oldest smartphones in tracking studies. This affects older people more, as they tend to have the oldest devices, renew them less frequently (Jacobson et al., 2017) and often use second-hand devices inherited from their relatives (Fernández-Ardèvol & Ivan, 2013; Oreglia & Kaye, 2012).

Discussion and conclusion

We focused on two particular challenges regarding the inclusion of the older population in big data approaches – age prediction and smartphone log collection. This allowed a

nuanced discussion of ageism in this arena. However, we analysed two different kinds of intelligent systems that are examples of the limitations of big data approaches when it comes to analysing older people.

On the one hand is the difficulty of making accurate predictions for older people; for example, the accuracy of age predictions based on text increases with age, as younger cohorts are more stereotypical in the use of language than adults (Nguyen et al., 2014). Most systems designed to predict the age of users rely on the social and cultural expressions of individuals. However, predicting the age of an individual through cultural manifestations, such as language, is a challenge. In this particular case, the expression of social identity does not necessarily coincide with the biological age (Bucholtz & Hall, 2005), as individuals adapt their language to the context or audience (Eckert, 1998; Holmes, 2013). On the other hand, there are challenges involved in collecting demographics of the sample in big data approaches to be able to assure the representation of all minorities. To collect smartphone logs, users only have to give their consent to use an app; however, to collect demographics, the system should ask explicit questions, and this would imply a reduction in the sample, which researchers might not want to assume (Jones et al., 2015).

Our study unveils two areas in which ageism is in operation. The biases are not new but are known for shaping the IS of big data studies regarding other areas of discrimination. In the following, we contextualize them with regard to age.

Biased samples. We presented examples of how older people are invisibilized by, for instance, not controlling the capacity of the sample to represent the studied population. Recruitment procedures contribute to such biases, which are reinforced by analytical categories that ignore older individuals (e.g., categories aged 40+). We discussed the extent to which these wide categories tend to exclude older people by relying on learning data sets that are biased towards younger populations (e.g., Kosinski et al., 2013; Liu & Yang, 2012).

Biased tools. Cultural values and beliefs shape the instruments that our societies build (Bijker et al., 1989; Letouzé, 2015). Algorithms are no exception, and the beliefs and values of the developers might shape them. For instance, developers and technical specialists might assume that all users would behave as they do (Uricchio, 2017). We showed that homophilic assumptions are in operation in the case of age predictions, when developers assume that older people should behave similarly to them on social media. Moreover, some challenges to including older people successfully could be related to the limitations of the techniques used. This would be the case of opaque algorithms. They are not able to provide explicit descriptions of either their procedures (Rieder & Röhle, 2012) or the technical limitations of their measuring system (Holz et al., 2015; Jones et al., 2015; Lee et al., 2014), which would indirectly exclude older people as they constitute a minority in the digital world.

These biases misrepresent older people in big data approaches, reducing their predictive efficiency and effectiveness. Thus, structural ageism in big data approaches shows similar patterns to those that discriminate minorities and disempowered collectives (Eubanks, 2018; Hajian & Domingo-Ferrer, 2013; O'Neil, 2016). A general principle for tackling discrimination would be to compensate for the underrepresentation of collectives in less favourable conditions (Pedreschi et al., 2009). It is therefore necessary

to design algorithms that are capable of setting up bias-free training data sets (Bolukbasi et al., 2016). It is also essential to take into account all the granularity of the data to make algorithms more inclusive (Kitchin, 2014) by using statistical models that take into account cultural and social practices in the digital media of wider population segments that are usually invisibilized, in this case empirical evidence on the older population and its specific and diverse media practices. This should lead to a virtuous circle that includes this – usually ignored – segment of the population so that the instruments, in this case intelligent predictive systems, improve the quality of their results and, in turn, contribute to overcoming structural ageism.

Funding

Note

1. Specifically, the Machine Learning, Optimization, and Big Data conference (MOD), (Bayot, R. K. & Gon, T., 2017), and the Big data, socialcom and cybersecurity conference (ASE), (Oktay, H., Firat, A. & Ertem, Z., 2012).

References

AGE Platform Europe. (2016). *AGE Platform Europe position on structural ageism*. Brussels, Belgium.

Alvarez-Lozano, J., Osmani, V., Mayora, O. et al. (2014). Tell me your apps and I will tell you your mood. In *Conference on pervasive technologies related to assistive environments (PETRA'14)* (pp. 1-7). Island of Rhodes: ACM Press.

Ayalon, L. & Tesch-Römer, C. (eds.) (2018). *Contemporary perspectives on ageism*. Cham: Springer Open.

Bayot, R. K. & Gon, T. (2017). Age and gender classification of tweets using convolutional neural networks. In *Machine learning, optimization, and big data (MOD 2017)* (pp. 337-348). Volterra: Springer.

Bi, B., Shokouhi, M., Kosinski, M. & Graepel, T. (2013). Inferring the demographics of search users: Social data meets search queries. In *Conference on World Wide Web (WWW'13)* (pp. 131-140) Rio de Janeiro: ACM Press.

Bijker, W. E., Hughes, T. P. & Pinch, T. J. (eds.) (1989). *The social construction of technological systems*. London: MIT Press.

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Neural information processing systems (NIPS'16)*. Barcelona. Retrieved from https://www.semanticscholar.org/paper/Man-is-to-Computer-Programmer-as-Woman-is-to-Word-Bolukbasi-Chang/274459c52103f9b7880d0697aa28755ac3366987

Bonchi, F., Hajian, S., Mishra, B. & Ramazzotti, D. (2017). Exposing the probabilistic causal structure of discrimination. *International Journal of Data Science and Analytics*, 3: 1-21.

Boyd, D. & Crawford, K. (2012). Critical questions for big data. *Information and Communication Society*, 15: 662-679.

Bucholtz, M. & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, 7: 585-614.

Böhmer, M., Hecht, B., Schöning J.J., Krüger, A., & Bauer, G. (2011). Falling asleep with Angry Birds, Facebook and Kindle: A large scale study on mobile application usage. In *Human–computer interaction with mobile devices and services (MobileHCI'11)* (pp. 47-56). Stockholm: ACM Press.

Calasanti, T. & King, N. (2015). Intersectionality and age. In J. Twigg & W. Martin (eds.), R*outledge handbook of cultural gerontology* (pp. 193-200). London: Routledge/Taylor and Francis.

Castells, M. (2009). *Communication powe*r. United Kingdom: Oxford University Press.

Castells, M., Fernández-Ardèvol, M., Linchuan Qiu, J. & Sey, A. (2006). *Mobile communication and society: A global perspective*. Cambridge, MA: The MIT Press.

Choudrie, J. & Vyas, A. (2014). Silver surfers adopting and using Facebook? A quantitative study of Hertfordshire, UK applied to organizational and social change. *Technological Forecasting and Social Change*, 89: 293-305.

Culotta, A., Ravi, N. K. & Cutler, J. (2016). Predicting Twitter user demographics using distant supervision from website traffic data. *Journal of Artificial Intelligence Research*, 55: 389-408.

De Montjoye, Y.-A., Quoidbach, J., Robic, F. & Pentland, A. (2013). Predicting personality using novel mobile phone-based metrics. In A. Greenberg, W. Kennedy & N. Bos (eds.), *Social computing, behavioral-cultural modeling and prediction* (pp. 48-55). Heidelberg: Springer

Eckert, P. (1998). Age as a sociolinguistic variable. In F. Coulmas (ed.), *The handbook of sociolinguistics*. Oxford, United Kingdom: Blackwell.

EPSC. (2018). *The age of artificial intelligence – Towards a European strategy for human-centric machines*. Heidelberg: Springer

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police and punish the poor*. New York: St Martin's Press.

Eurostat. (2017). *Population structure and ageing*. Retrieved from http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing [Accessed 2018, March 1].

Eurostat. (2018). Individuals Internet use. Last Internet use in the last 3 months. Table [isoc_ci_ifp_iu]. Retrieved from http://ec.europa.eu/eurostat/web/products-datasets/-/isoc_ci_ifp_iu [Accessed 2018, March 1].

Ferdous, R., Osmani, V. & Mayora, O. (2015). Smartphone app usage as a predictor of perceived stress levels at workplace. In *Proceedings of the 2015 9th international conference on pervasive computing technologies for healthcare (PervasiveHealth'15)* (pp. 225-228). https://doi.org/10.4108/icst.pervasive-health.2015.260192. Istanbul: European Union Digital Library.

Fernández-Ardèvol, M. & Ivan, L. (2013). Older people and mobile communication in two European contexts. *Romanian Journal of Communication and Public Relations*, 15: 83-101.

Ferreira, D., Goncalves, J., Kostakos, V. et al. (2014). Contextual experience sampling of mobile application micro-usage. In *Human–computer interaction with mobile devices & services (MobileHCI'14)* (pp. 91-100). Toronto: ACM Press.

Ferreira, D., Kostakos, V. & Dey, A. K. (2012). Lessons learned from large-scale user studies: Using Android market as a source of data. *International Journal of Mobile Human Computer Interaction*, 4: 28-43.

Garattini, C. & Prendergast, D. (2015). Critical reflections on ageing and technology in the twenty-first century. In D. Prendergast & C. Garattini (eds.), *Aging and the digital life course* (pp. 1-15). New York: Berghahn Books.

Greenwald, A. G. & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102: 4-27.

Greenwood, S., Perrin, A. & Duggan, M. (2016). *Social media update*. Retrieved from http://www.pewinternet.org/2016/11/11/social-media-update-2016/

Hajian, S. & Domingo-Ferrer, J. (2013). A methodology for direct and indirect discrimination prevention in data mining. *IEEE Transactions on Knowledge and Data Engineering*, 25: 1445-1459.

Hendricks, J. (2005). Ageism: Looking across the margin in the new millennium. *Generations*, 29: 5-7.

Holmes, J. (2013). *An introduction to sociolinguistics* (4th ed.). New York: Routledge.

Holz, C., Bentley, F., Church, K. & Patel, M. (2015). "I'm just on my phone and they're watching TV": Quantifying mobile device use while watching television. In *Conference on interactive experiences for TV and online video (TVX'15)*. Brussels: ACM Press

Ikebe, Y., Katagiri, M. & Takemura, H. (2012). Friendship prediction using semi-supervised learning of latent features in smartphone usage data. In *Knowledge discovery and information retrieval (KDIR'2012)*. Barcelona: Science and Technology Publications, Lda.

Jacobson, J., Lin, C. Z. & McEwen, R. (2017). Aging with technology: Seniors and mobile connections. *Canadian Journal of Communication*, 42: 331.

Jensen, M. (2013). Challenges of privacy protection in big data analytics. In *BigData'13* (pp. 235-238). doi: 10.1109/BigData.Congress.2013.39

Jones, S. L., Ferreira, D., Hosio, S., Goncalves, J., & Kostakos, V. (2015). Revisitation analysis of smartphone app use. In *Pervasive and ubiquitous computing (UbiComp'15)* (pp. 1197-1208). Osaka: ACM Press

Karikoski, J. & Soikkeli, T. (2013). Contextual usage patterns in smartphone communication services. *Personal and Ubiquitous Computing*, 17: 491-502.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Los Angeles: Sage.

Kiukkonen, N., Blom, J., Dousse, O., Gatica-Perez, D., & Laurila, J. (2010). Towards rich mobile phone datasets: Lausanne data collection campaign. In *Pervasive services (ICPS'10)*. Berlin.

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

Lagacé, M., Charmarkeh, H., Tanguay, J. & Annick, L. (2015). How ageism contributes to the second-level digital divide: The case of Canadian seniors. *Journal of Technologies and Human Usability*, 11: 1-13.

Lee, U., Lee, J., Ko, M. et al. (2014). Hooked on smartphones: An exploratory study on smartphone overuse among college students. In *Human factors in computing systems (CHI'14)* (pp. 2327-2336). Toronto: ACM Press

Letouzé, E. (2015). *Big data and development: General overview primer*. Data-Pop Alliance. Retrieved from http://datapopalliance.org/wp-content/uploads/2015/12/Big-Data-Dev-Overview.pdf

Liao, L., Jiang, J., Ding, Y. et al. (2014). Lifetime lexical variation in social media. In *Artificial intelligence (AAAI'14)* (pp. 1643-1649).

Ling, R., Bertel, T. F. & Sundsøy, P. R. (2012). The socio-demographics of texting: An analysis of traffic data. *New Media & Society*, 14: 281-298.

Liu, J-.Y. & Yang, Y.-H. (2012). Inferring personal traits from music listening history. In *Music information retrieval with user-centered and multimodal strategies (MIRUM '12)* (p. 31).

Mihailidis, P. (2014). A tethered generation: Exploring the role of mobile phones in the daily life of young people. *Mobile Media & Communication*, 2: 58-72.

Möller, A., Kranz, M., Schmid, B., Roalter, L. & Diewald, S. (2013). Investigating self-reporting behavior in long-term studies. In *Human factors in computing systems (CHI'13)* (pp. 2931-2940). Paris: ACM Press.

Neugarten, B. L. (1996). *The meanings of age: Selected papers of Bernice L. Neugarten*. Chicago, IL: University of Chicago Press.

Nguyen, D., Gravel, R., Trieschnigg, D. & Meder, T. (2013). "How old do you think I am?": A study of language and age in Twitter. In *AAAI conference on weblogs and social media* (pp. 439-448). Palo Alto, CA: AAAI Press.

Nguyen, D., Trieschnigg, D., Doğruöz, A. S. et al. (2014). Why gender and age prediction from tweets is hard: Lessons from a crowdsourcing experiment. In *The annual meeting of the EPSRC network on vision & language and the technical meeting of the European network on integrating vision and language: A workshop of the international conference on computational linguistics (COLING 2014)* (pp. 1950-1961). Dublin, Ireland: COLING.

O'Neil, C. (2016). *Weapons of math destruction. How big data increases inequality and threatens democracy*. New York: Broadway Books.

Officer, A. & de la Fuente-Núñez, V. (2018). A global campaign to combat ageism. *Bulletin of the World Health Organization*, 96: 295-296.

Oktay, H., Firat, A. & Ertem, Z. (2012). Demographic breakdown of Twitter users: An analysis based on names. *ASE BIGDATA/SOCIALCOM/CYBERSECURITY*, 1-11.

Oreglia, E. & Kaye, J. "Jofish" (2012). A gift from the city: Mobile phones in rural China. In C*omputer-supported cooperative work and social computing (CSCW'15)* (pp. 137-146). Seattle: ACM Press.

Ørmen, J. & Thorhauge, A. M. (2015). Smartphone log data in a qualitative perspective. *Mobile Media & Communication*, 3: 335-350.

Oulasvirta, A., Rattenbury, T., Ma, L. & Raita, E. (2012). Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing*, 16: 105-114.

Pedreschi, D., Ruggieri, S. & Turini, F. (2009). Measuring discrimination in socially-sensitive decision records. In *SIAM international conference on data mining* (pp. 581-592). Nevada: Society for Industrial and Applied Mathematics

Peersman, C., Daelemans, W. & Van Vaerenbergh, L. (2011). Predicting age and gender in online social networks. In *International workshop on search and mining user-generated contents (SMUC'11)*, 2011, October 28, Glasgow, Scotland, UK (pp. 37-44). ACM Press.

Perozzi, B. & Skiena, S. (2015a). Exact age prediction in social networks. In *International conference on world wide web* (pp. 91-92). Florence: ACM Press.

Popov, V., Kosinski, M., Stillwell, D. & Kielczewski, B. (2018). Apply magic sauce. Retrieved from https://applymagicsauce.com/research.html [Accessed 2018, January 1].

Rahmati, A., Tossell, C., Shepard, C., Kortum, P. & Zhong, L. (2012). Exploring iPhone usage. In *Human–computer interact with mobile devices and services (MobileHCI'11)*. San Francisco: ACM Press

Rieder, B. & Röhle, T. (2012). Digital methods: Five challenges. In D. M. Berry (ed.), *Understanding digital humanities*. London: Palgrave Macmillan.

Righi, V., Sayago, S., Rosales, A. et al. (2018). Co-designing with a community of older learners for over 10 years by moving user-driven participation from the margin to the centre. *CoDesign*, 14: 32-44.

Roca Salvatella. (2016). *La brecha digital en la ciudad de Barcelona*. Barcelona, Spain.

Rogers, Y., Paay, J., Brereton, M., Vaisutis, K., Marsden, G. & Vetere, F. (2014). Never too old: Engaging retired people inventing the future with MaKey. In *Human factors in computing systems* (CHI'14) (pp. 3913-3922). Toronto: ACM Press

Rosales, A. & Fernández-Ardèvol, M. (2016a). Beyond WhatsApp: Older people and smartphones. *Romanian Journal of Communication and Public Relations*, 18: 27-47.

Rosales, A. & Fernández-Ardèvol, M. (2016b). Smartphones, apps and older people's interests: From a generational perspective. In *Human–computer interaction with mobile devices and services (MobileHCI'16)* (pp. 491-503). Florence: ACM Press.

Rosenthal, S. & McKeown, K. (2011). Age prediction in blogs: A study of style, content, and online behavior in pre-and post-social media generations. In *Meeting of the Association for Computational Linguistics: Human language technologies* (pp. 763-772). Portland: Association for Computational Linguistics

Sawchuk, K. & Crow, B. (2011). Into the grey zone: Seniors, cell phones and milieus that matter. *WI: Journal of Mobile Media*, 5.

Schäfer, M. T. & Van Es, K. (2017). *The datafied society: Studying culture through data*. Amsterdam: Amsterdam University Press.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L. et al. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS One*, 8: e73791.

Selwyn, N., Gorard, S., Furlong, J. & Madden, L. (2003). Older adults' use of information and communications technology in everyday life. *Ageing and Society*, 23: 561–582.

Shin, C., Hong, J.-H. & Dey, A. K. (2012). Understanding and prediction of mobile application usage for smart phones. In *Pervasive and ubiquitous computing (UbiComp'12)* (p. 173). Pittsburgh: ACM Press

Singh, V. K., Freeman, L., Lepri, B. & Pentland, A. (2013). Predicting spending behavior using socio-mobile features. In *Social computing* (pp. 174-179). Washington: IEEE Computer Society Press

Smith, M., Szongott, C., Henne, B., Voigt, G. Von (2012). Big data privacy issues in public social media. *Digital Ecosystems Technologies (DEST'12)*. Campione d'Italia: IEEE Computer Society Press

Srinivasan, V., Moghaddam, S., Mukherji, A. et al. (2014). MobileMiner: Mining your frequent patterns on your phone. In *Joint conference on pervasive and ubiquitous computing (UbiComp'14)* (pp. 389-400). Seattle: ACM Press.

Stocchetti, M. (2018). Invisibility, inequality and the dialectics of the real in the digital age. *Interaçoes*, 34: 23-46.

Uricchio, W. (2017). Data, culture and the ambivalence of algorithms. In M. T. Schäfer & K. Van Es (eds.), *The datafied society: Studying culture through data* (pp. 125-137). Amsterdam, Amsterdam University Press

Wagner, D. T., Rice, A. & Beresford, A. R. (2013). Device analyzer: Understanding smartphone usage. In *International conference on mobile and ubiquitous systems* (pp. 1-12). Tokyo: Springer

Xu, R., Frey, R. M., Fleisch, E. & Ilic, A. (2016). Understanding the impact of personality traits on mobile app adoption – Insights from a large-scale field study. *Computers in Human Behavior*, 62: 244-256.

Yan, T., Chu, D., Ganesan, D., Kansal, A. & Liu, J. (2012). Fast app launching for mobile devices using predictive user context. In *Mobile systems, applications, and services (MobiSys'12)* (pp. 113-126). Low Wood Bay: ACM Press