6-2018

From 2,772 segments to five personas: Summarizing a diverse online audience by generating culturally adapted personas

Joni SALMINEN

Sercan SENGUN

Haewoon KWAK
Singapore Management University, hkwak@smu.edu.sg

Bernard J. JANSEN

Jisun AN

See next page for additional authors

f i ® s t m ¤ ñ d @ ¥
PEER-REVIEWED JOURNAL ON THE INTERNET

From 2,772 segments to five personas: Summarizing a diverse online audience by generating culturally adapted personas

by Joni Salminen, Sercan Şengün, Haewoon Kwak, Bernard Jansen, Jisun An, Soon-Gyo Jung, Sarah Vieweg, and D. Fox Harrell

Abstract

Understanding users in the era of social media is challenging, requiring organizations to adopt novel computation-aided approaches. To exemplify such an approach, we retrieved information on millions of interactions with YouTube video content from a major Middle Eastern media outlet, to automatically generate personas that capture how different audience segments interact with thousands of individual content pieces. Then, we used qualitative data to provide additional insights into the automatically generated persona profiles. Our findings provide insights into social media usage in the Middle East and demonstrate the application of a novel methodology that generates culturally adapted personas of social media audiences, summarizing complex social analytics data into human portrayals that are easy to understand by end users in real organizations.

Contents

Introduction

Social media provides an interesting magnifying glass into sociological and behavioral patterns of online audiences. Researchers and organizations can access vast amounts of user data via social media platforms, such as Facebook, Twitter, and YouTube (Kwak, *et al.*, 2010; Abisheva, *et al.*, 2014). However, turning this data into easily understandable insights remains a persistent challenge

for many organizations and individuals (Salminen, Milenković, *et al.*, 2017), especially when dealing with culturally diverse audiences that vary in their topical interests. In addition, cultural differences in social media usage tend to be overlooked (Salminen, Şengün, *et al.*, 2017), ignoring fragmented interests and motivational drivers of an underlying audience. While extant works tend to focus on the Western context of social media users (Joye, 2009), there is a growing need for studying other cultural contexts and regions in the world, as social media usage has rapidly spread globally.

Prior research has established that cultural and societal conditions in the Middle East are distinctively different from those in other parts of the world (Bakhtari, 1995; Karabenick and Moosa, 2005; Zualkernan, *et al.*, 2006), and these overarching conditions are reflected in social media usage, such as photo sharing, use of pseudonyms, and privacy concerns (Harrell, *et al.*, 2017; Ur and Wang, 2013; Millham and Atkin, 2018). The use of social media is therefore culturally embedded (Singh, *et al.*, 2012). Although the impact of culture on social media usage has been explored in prior research, it has not been done for all regions of the world, and there is a great need for regionally focused social media studies (Ngai, *et al.*, 2015).

While some studies have been conducted on the use of social media in the Middle East (Stanger, *et al.*, 2017; Wiest and Eltantawy, 2015), these studies tend to use small datasets instead of large-scale quantitative data. Additionally, none of the extant works use personas to represent Middle Eastern social media audiences, even though personas have a potential to describe social media audiences in an engaging way (Kwak, *et al.*, 2017). To address these shortcomings, this research examines the usage of social media in the Middle East using data-driven personas, namely in the context of YouTube audiences. This context was chosen for three reasons: 1) an opportunity to access a large dataset from a major Middle Eastern media organization; 2) the scarcity of research on social media behavior of Middle Eastern users; and, 3) the high degree of social media adoption in the Middle East. For example, in Qatar, social media penetration is among the highest in the world (Northwestern University in Qatar, 2017), making social media research regionally important and relevant.

Motivated by these considerations, we pose the following questions:

1. How can we automatically generate personas that represent Middle Eastern social media users?

2. What are the benefits of mixed method approaches when developing culturally adapted personas?

To answer the above questions, we applied automatic persona generation (APG) methodology (An, *et al.*, 2017; Jung, *et al.*, 2017) coupled with qualitative data collection. Personas, *i.e.*, fictitious but data-driven user representations (Matthews, *et al.*, 2012), were chosen as an analytical technique due to their abilities to crystallize user statistics into a humanlike format for end users of that information. We demonstrate a mixed-method approach by first analyzing millions of content interactions from the YouTube profile of AJ+ Arabic, a major Middle Eastern media outlet, and then by enriching automatically generated personas with qualitative analysis that includes manually coding 255 public social media profiles from the Middle East and an analysis of interview data. In the next sections of this paper, we review related literature, present methodology and findings, discuss them, and suggest future research avenues.

Literature review

Approach to literature review

In general, social media behavior has been extensively studied. Some scholars have focused on the impact of social media on business practices (Pöyry, *et al.*, 2013), while others have studied the effects of social media on cultural values and norms of users (Stanger, *et al.*, 2017). Here, we focus specifically on the cultural aspects of social media usage in the Middle East. We searched two major academic databases (ScienceDirect and Google Scholar) with relevant key phrases (*e.g.*, "*middle east + social media*", "*culture + social media*") to find prior work on culture's impact on social media behavior, particularly in a Middle Eastern context. Based on manual evaluation, we included only the most relevant articles for a detailed reading, summarized in the following sections.

Use of social media in the Middle East

Impact of norms and a larger societal framework

The context of this study is the Middle East, a region of the world where cultural and religious norms are much valued. In fact, these cultural and religious norms are embedded in nearly all activities ranging from social to political. Although the use of social media has grown in the region, most people tend to consider cultural and religious customs while using social networks. Thus, the use of social media in the Middle East can be understood in a context of Islamic values and practices (Zualkernan, *et al.*, 2006), as well as ethnic and cultural customs.

Stanger, *et al.* (2017) studied how Saudi Arabian youth engaged with social media on a daily basis. They found that although the use of social media was accepted, most users took cultural aspects and norms into consideration. For instance, users paid attention to photos they uploaded and pseudonyms that they selected, as these revealed much about their attitudes toward cultural norms (Stanger, *et al.*, 2017). Abokhodair, *et al.* (2016) and Abokhodair and Vieweg (2016) explored privacy-oriented social media behavior in the Gulf region, describing a regional vision of privacy as a combination of cultural, social, and religious values operating on personal, social, and judicial levels.

Singh (2010) and Wiest and Eltantawy (2015) argued that the use of social media has increased the spread of Western culture into the Arabic world. An example is the continued use of English in social media platforms among youth in the Arab world despite the introduction of Arabic keyboards. Weist and Eltantawy (2015) found that the most preferred language by the social media users, especially Facebook users, in some parts of the Middle East was Arabic; however, about 45 percent preferred to use English. Another study by Haggan (2007) pointed out that most of the text messages sent in Kuwait were typed either in English or English combined with Arabic.

Gender and age differences

Bolton, *et al.* (2013) aimed to understand social media usage of Generation Y, sometimes referred to as digital natives, that is, the first generation that has spent their social lives in a digital environment. They conducted a systematic review and pointed out intra-cohort variations from environmental aspects such as culture, technology, economy, and even politics. The analysis by Al-Jabri, *et al.* (2015) on the use of global social networking sites by Arabs was among the few studies attempting to understand the use of social media by Generation Y outside Western nations, especially in Saudi Arabia. Their findings indicated that social media usage of youth, especially on Twitter, was influenced by freedom of expression, enjoyment, and social interactions (Al-Jabri, *et al.*, 2015).

Stanger, *et al.* (2017) estimated that close to 135 million people in the Gulf region use social media, the largest participation coming from males under the age of 25 years. Thus, this study noted that social media usage in the region is male- and youth-oriented (Stanger, *et al.*, 2017). Women in the Arab world have often been perceived to be culturally submissive to men (Alsharkh, 2012).

However, Arab women have begun to publicly present their opinions and views by posts on social media platforms, such as Facebook, Twitter, and YouTube (Alshark, 2012) as well as utilizing e-mail messages and online blogs to express themselves (Aouragh, 2008).

Need for communication and self-expression

Stanger, *et al.* (2017) found that social media in the Middle Eastern region was primarily used for communication and interaction with others locally and internationally. For instance, Facebook increased the social capital of an individual by shaping what they shared about themselves (Bolton, *et al.*, 2013). Barry and Bouvier (2012) studied the United Arab Emirates (UAE) and Welsh students' use of Facebook, finding that UAE students preferred to use Facebook in collective communications, such as posting a message on Facebook rather than sending it privately to individual recipients, whereas students from Wales used Facebook more significantly to maintain private relationships.

Prior studies also emphasized the use of social media for political engagement in the Middle East. These studies had a common conclusion, pointing out that the Cultural Revolution in the Arab world was catalyzed by increasingly emerging social media usage (Wiest and Eltantawy, 2015). Mourtada and Salem (2011) highlighted that social media played a critical role in shaping opinions, mobilization, and empowerment of Arabic people, especially youth. Al-Saggaf (2006) explored the potential influence of online technologies in fostering civic engagement in the Middle East. After studying the use of the Internet in Palestine, Aouragh (2008) proposed that social and political agency and activists' tactics among Palestinian citizens had been strengthened by emerging social media. Several researchers have proposed that social media could result in a cultural change in the Arab world (Wasserman, 2011).

A study by Sawyer and Chen (2012) investigated the impact of social media on intercultural adaptation. The study was conducted by directly interviewing international students of different age groups in United States colleges. The results indicated that people intended to use social media to become more incorporated into host cultures as they adapted to that new culture while at the same time maintaining strong connections to their home society's cultural beliefs and norms (Sawyer and Chen, 2012). Harrell, *et al.* (2017) found five themes of social media use in the region using a mixed-methods (computational and qualitative) approach: 1) issues of self-expression; 2) existence or non-existence of Khaleeji features; 3) negotiating social monitoring; 4) forming various levels of social connections; and, 5) contrasting physical and virtual identities.

Summary of earlier work

A common conclusion from prior studies is that social media is seen as an agent of change for individuals to shape their culture and society. These changes could be seen taking place at gender and generational levels. First, women in most of the Gulf nations were gaining opportunities to publicly express their opinions on cultural and religious norms. Social media gave women and men more freedom to interact with the opposite sex. Moreover, through social media, youth in the Middle Eastern countries were being influenced by other cultures, gaining self-confidence to participate in political and social movements such as the Arab Spring, while seeking ways to participate in societal change.

Moreover, there appeared to be a certain perceived conflict between maintaining traditional cultural values and utilizing new forms of self-expression and communication enabled by social media. As the flexibility of social media platforms enables various forms of usage, Middle Eastern audiences seem to have devised their own styles of using social media within cultural boundaries. The literature seems to suggest that individuals can maintain their cultural identities by adopting different usage practices, even when participating in global platforms. However, both culture and social media usage appear to be continuously in flux, strongly suggesting that earlier studies and their findings need to be frequently re-examined.

Table 1: Findings based on literature review.

Theme	Main findings	Example authors
Motives for use	• Freedom of expression, entertainment • Cultural specificities transfer into digital world, including profile images, privacy needs, and use of pseudonyms	Bolton, *et al.* (2013); Ur and Wang (2014); Al-Jabri, *et al.* (2015)
Gender and age differences	• Middle Eastern youth participate in global culture through social media • Western influences, such as language and communicative habits, spread to the region • Gender roles and communication between men and women is shifting	Haggan (2007); Singh (2010); Barry and Bouvier (2012); Wiest and Eltantawy (2015); Stanger, *et al.* (2017)
Social media as change agent	• Empowerment of women is enhanced by social media participation • Political changes, such as Arab Spring, are partly enabled by social media platforms	Al-Saggaf (2006); Mourtada and Salem (2011); Wasserman (2011); Alshark (2012)

Personas as culturally adapted data representations

In the second section of this literature review, we provide a brief overview of persona literature. A persona is an artificial representation of a larger underlying user group (Nielsen and Storgaard Hansen, 2014). Introduced by Cooper (2004), personas are used by software developers, designers, and others to enhance user-driven decision-making. In addition, buyer personas are commonly

considered in marketing to better understand drivers of consumer behavior (Scott, 2007), corresponding to Jenkinson's (1994) idea of going beyond segmentation in that personas communicate about groups using individual-level attributes. Personas can be leveraged to efficiently communicate aggregated user statistics (Jansen, *et al.*, 2017). Furthermore, closely related to personas is the concept of virtual identities, defined as computational surrogates for users in online environments, ranging from social media profiles to avatars (Harrell and Lim, 2017).

The most prominent advantages of personas relate to summarizing multi-dimensional user and customer information into easily understandable formats (Cooper, 2004; Nielsen, 2004). It is customary to use computational techniques to deal with large online datasets (*e.g.*, Beaudouin and Pasquier, 2017). However, compared to traditional data analytics, personas present customer information in a more humane format, therefore enhancing a creation of solutions that are useful in real life (Pruitt and Grudin, 2003).

Traditionally, personas are created via ethnography and/or interviews (Nielsen, 2004). Because these efforts are manual and time-consuming, persona creation tends to take several months and can be expensive. Moreover, manual methods cannot be applied to millions of content interactions. Thus, persona generation via computational methods seems feasible and has been explored by several authors in the past (An, *et al.*, 2017; Chapman, *et al.*, 2008; McGinn and Kotamraju, 2008; Zhang, *et al.*, 2016). Here, we focus on one of the more advanced approaches, called automatic persona generation (APG), introduced in An, *et al.* (2017), Jung, *et al.* (2017), and Kwak, *et al.* (2017). It aims at generating rounded personas from online analytics data. Generally, rounded or complete personas are considered desirable, because they capture key attributes of an audience or user base (Pruitt and Adlin, 2006).

Methodology

Automatic persona generation

Research context and data collection

We collected data from a large Middle Eastern news and media company, Al Jazeera. We particularly focused on one of their media channels, AJ+, which is an online news and media channel owned by the Al Jazeera Media Network. The channel delivers content via YouTube, Facebook, Instagram, and other social media platforms. AJ+ Arabic (AJ+عربي) is the Arabic-language version of the service, that aims at presenting contemporary events and topics from an Arabic perspective. Its goal is to "stimulate dialogue and constructive interaction in the society" (YouTube page of AJ+ Arabic, 2017). Figure 1 shows the channel page.

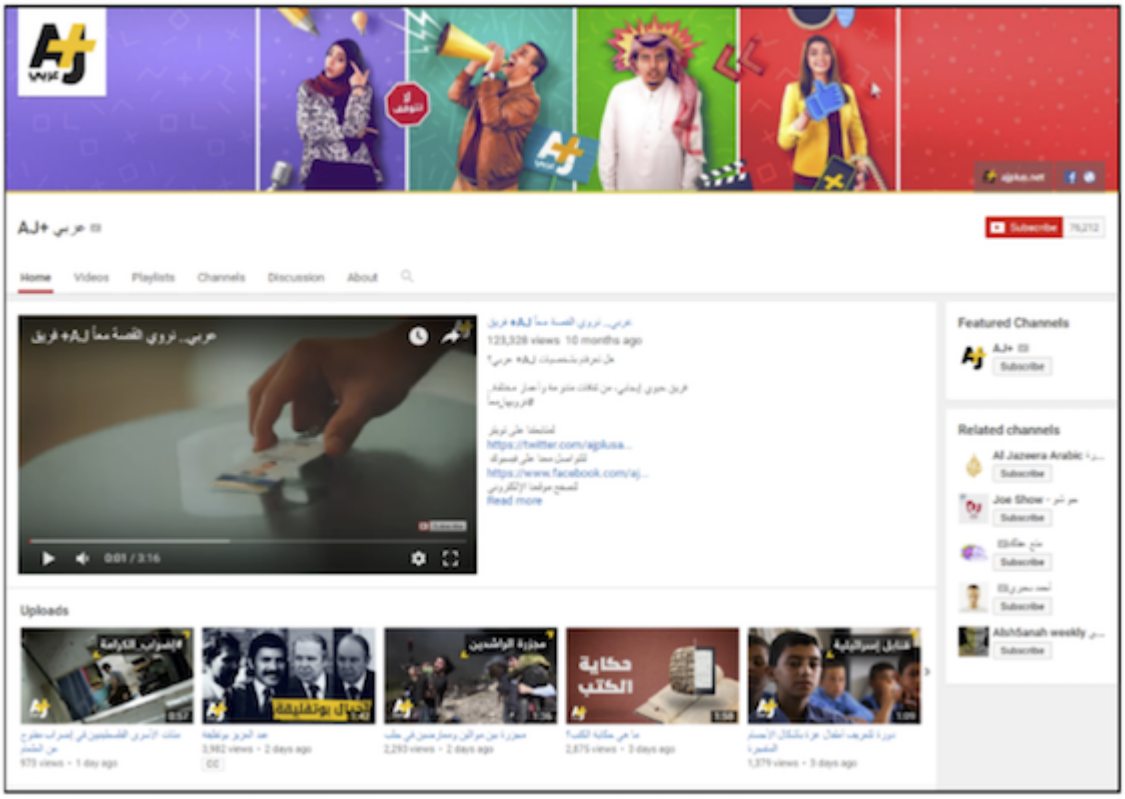


Figure 1: YouTube channel of AJ+ Arabic.

The generated personas are based on 13,595,605 views on 2,638 videos published between 26 November 2015 and 31 March 2017. The durations of the videos ranges from 15 seconds to 10 minutes. The content in question is short and particularly adapted to social media (*i.e.*, being easy to digest by viewers), focusing on diverse topics, including news, culture, technology, and lifestyles.

Persona generation process

To automatically generate personas from social media data (in this case aggregated video views), our methodology automated the following five steps:

1. Data collection from social media platforms
2. Detection of distinct content interaction patterns by matching demographic groups and content
3. Detection of dominant demographic groups from the set of distinct content interaction patterns
4. Creation of persona prototypes by selecting demographic attributes
5. Enrichment of the prototypes with user attributes (e.g., name, photo).

Due to a recommendation of using only a small number of personas for decision-making (Cooper, 2004), we limited the number of personas shown to end users to 5–15. An example of a resulting persona is shown in Figure 2.

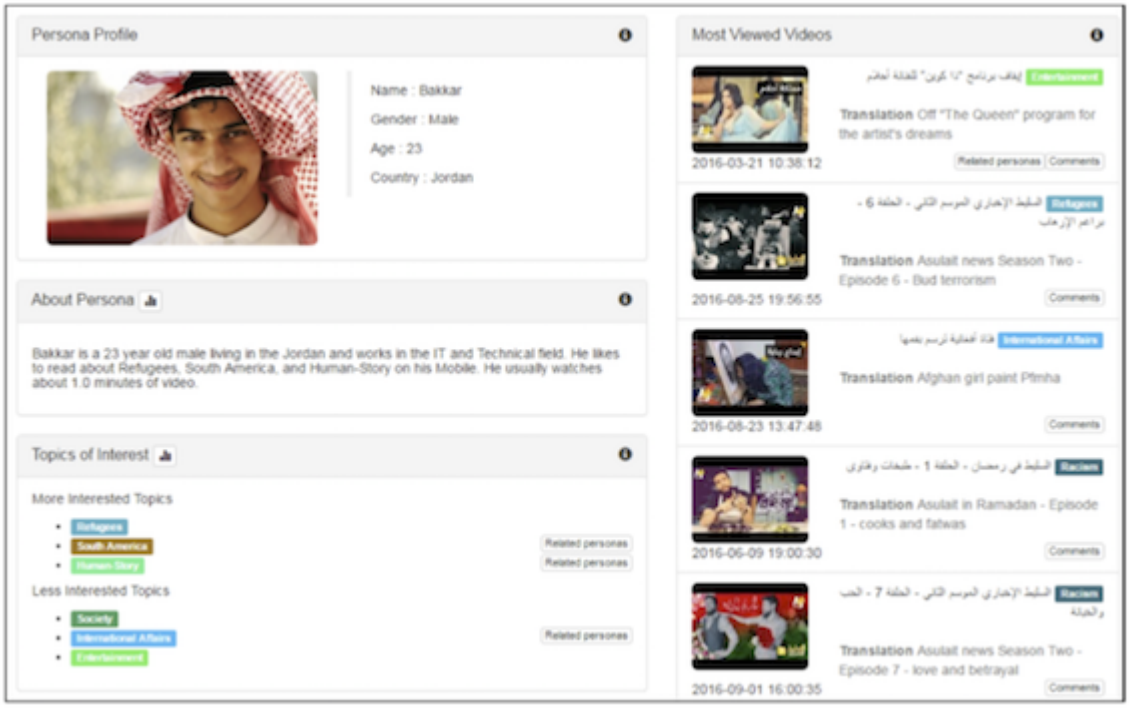


Figure 2: An example of resulting persona from the APG system [1]. The Web site is rendered by using Flask [2], an open source Python Web framework.

The **Persona Profile** section includes the name, gender, age, and country of residence. The name and picture map typical names and pictures to combinations of age/gender/location, obtained from stock photo services and Census statistics. The **About Persona** section contains a summary describing a persona in plain language, obtained by filling in dynamic fields in a static text template. The **Topics of Interest** section shows topics that a given persona has shown the most interest in, based on matrix decomposition, explained below. The **Most Viewed Videos** section lists videos most descriptive to this persona, obtained from YouTube. Viewing behavior is central to personas, as automatically generated personas are based on a combination of behavioral viewing patterns and demographic information (Jansen, *et al.*, 2017).

Computational techniques

As mentioned earlier, we obtained data on user groups of AJ+ Arabic and their behaviors from YouTube. By using this data, we first constructed a large matrix capturing user interactions with video content and then decomposed it to detect latent behavioral patterns. The matrix decomposition is depicted in Figure 3.

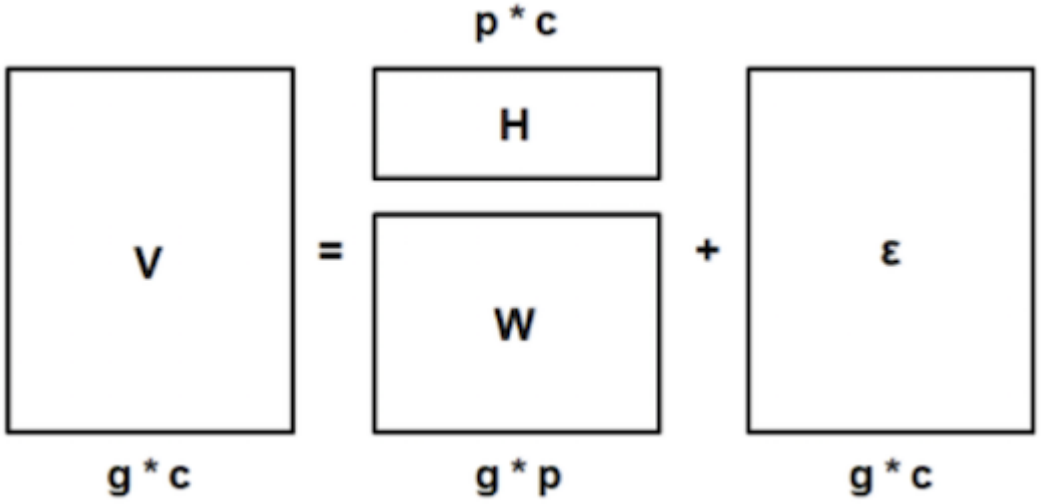


Figure 3: Matrix decomposition for automatic persona generation.

V is a *g* × *c* matrix, where *g* is the number of user groups and *c* is the number of videos. The elements of this matrix are the number of views (the *ViewCount* metric in the YouTube API [3]) of each video by each user group. The user groups are defined by gender, age group, and country so that the maximum number of user groups can be calculated by multiplication (gender × age group × country). Given the available combinations of YouTube, there are 2,772 possible groups. An example is "Male, 25–34, Saudi Arabia".

Given the matrix *V* that contains all user groups' view counts for each video, this matrix is then decomposed into two other matrices: *H* and *W*. Both include the parameter *p*, which is the number of content interaction patterns (*i.e.*, personas) that we want to infer. In other words, *p* represents latent patterns found by the analysis algorithm from the data. This step applies the well-known non-negative matrix factorization (NMF) (Lee and Seung, 1999) whereby we calculate a linear combination of *W* as the basis and *H* as the encoding, ε being the error term. Compared to clustering used in prior works to generate data-driven personas (Aoyama, 2005), NMF can produce multiple behavioral patterns even from a single user group. This is appropriate because different social media behaviors can exist within the same demographic group. In other words, people of the same age and gender may be interested in different content, and these differences are captured by NMF. The use of NMF to generate personas is described in more detail in related work (An, *et al.*, 2017).

After calculating persona prototypes, they are enriched with additional information, including name, photo, and other information noted in Figure 2 and rendered to end users via an online user interface. By using aggregated data in compliance with the social media platforms' terms of service, the APG preserves the privacy of individual users. This is an important property given an increasing overall interest in data privacy. Additionally, since we identify latent content interaction patterns, the resulting personas are based on behavioral data, not only demographic differences. Finally, APG is generalizable to any social media channel that provides programmatic access to aggregated user statistics.

Qualitative analysis

Overview. To further understand the behaviors of automatically generated personas, we revisited two types of qualitative data from previous research about understanding virtual identity uses in the region (Harrell, *et al.*, 2017): 1) manual coding and analysis of public profiles of Instagram users

who had checked into a location in the Middle East in a specific time frame. Eliminating the tourists and expatriates, we were left with 255 public profiles that we were confident that represented Middle Eastern social media users. In addition, 2) we recoded five qualitative interviews with social media users from the Middle East to uncover themes and issues that surfaced while they used various social media services. Overall, the criticism posed toward data-driven personas is that they often remain shallow, not involving in-depth information related to a specific life story (Salminen, Şengün, *et al.*, 2017). However, to realize the full benefits of rounded personas associated with more immersion by end users into the circumstances of portrayed personas (Nielsen, 2004), we turned to additional contextual information to enrich our data-driven persona profiles.

Specifically, there were two reasons for turning to Instagram profiles for qualitative data: 1) YouTube profiles are less personal and self-expressive than those in other popular social networks like Instagram and Facebook, with only behavioral insights coming from user demographics and video viewing history. Analyzing public Instagram profiles in terms of photo content, captions, and tags can provide richer data about life stories and online behaviors of users (Hu, *et al.*, 2014) as well as help capture cultural nuances (Araújo, *et al.*, 2014). Moreover, 2) Instagram is the fastest growing photo sharing app in the Middle Eastern region with 33 percent increase in penetration in 2017 (Northwestern University in Qatar, 2017), thereby making it a rich and relevant source of information about local social media users.

Manual profile analysis. We thereby augmented previously collected profile data with a manual profile coding. The main aim of our manual profile analysis was to uncover nuances of social media usage in the Middle East, specifically how individuals from the region constructed different strategies to negotiate their needs and values within these systems. For this purpose, Instagram was an appropriate source, as it is a popular platform in the region (Northwestern University in Qatar, 2017) and, through the use of public profiles, it offered rich observations into user behavior. Eliminating tourists and expats left us with 255 public profiles that we confidently believed to represent individuals from the region. We analysed the following sections of the chosen user profiles: profile image, text (translated from Arabic), name, and recent posts, coding each into categories.

Interviews. The interviews were conducted with Middle Eastern individuals (five Qatari nationals) utilizing teleconferencing software, lasting between 45 to 90 minutes. Three interviewees were male and two were female with ages ranging between 20 and 30 years old. The interviewee demographics were skewed towards males since regional values of privacy may occasionally refrain females from pursuing personal interactions with strangers. The interviews were conducted in English, a language that the participants and interviewer were comfortable with. The interviews were semi-structured and included open ended questions to uncover attitudes on social media usage considering local cultural norms, sensitivities, and values. The interview questions are described in the Appendix.

Results

Descriptive personas

We first present the automatically generated personas in Table 2, followed by describing qualitative insights with automatically generated personas.

Table 2: Output from the AGP system. The user can choose between 5 and 15 personas; here, the five most

representative personas are chosen.

Persona 1: Omar	
	Name: Omar Gender: Male Age: 29 Country: Saudi Arabia
About Omar: Omar is a 29-year-old male living in Saudi Arabia and works in the Management field. He likes to read about US-affairs, South America, and Human-Story on his Mobile. He usually watches about 1.9 minutes of video.	**People like Omar:** 450,020 people Gender (Male), age (25–34), country (Saudi Arabia), interests (Society, South America, Human Interest Story), and language (Arabic) based potential reach.
Persona 2: Youssef	
	Name: Youssef Gender: Male Age: 26 Country: Morocco
About Youssef: Youssef is a 26-year-old male living in Morocco and works in the Food Preparation and Services field. He likes to read about Israel-Palestine, South America, and Human-Story on his Mobile. He usually watches about 1.3 minutes of video.	**People like Youssef:** 144,620 people Gender (Male), age (25–34), country (Morocco), interests (Israeli-Palestinian Peace Process, South America, Human Interest Story), and language (Arabic) based potential reach.
Persona 3: Bakkar	
	Name: Bakkar Gender: Male Age: 23 Country: Jordan
About Bakkar: Bakkar is a 23-year-old male living in Jordan and works in the IT and Technical field. He likes to read about Refugees, South America, and Human-Story on his Mobile. He usually watches about 1.0 minutes of video.	**People like Bakkar:** 16,040 people Gender (Male), age (18–24), country (Jordan), interests (Refugees International, South America, Human Interest Story), and

	language (Arabic) based potential reach.
Persona 4: Nalkah	
	Name: Nalkah Gender: Female Age: 31 Country: Saudi Arabia
About Nalkah: Nalkah is a 31-year-old female living in Saudi Arabia and works in the Food Preparation and Services field. She likes to read about Technology & Science, South America, and Human-Story on her Mobile. She usually watches about 1.1 minutes of video.	**People like Nalkah:** 3,250 people Gender (Female), age (25–34), country (Saudi Arabia), interests (Science Technology, South America, Human Interest Story), and language (Arabic) based potential reach.
Persona 5: Jahmir	
	Name: Jahmir Gender: Male Age: 28 Country: Iraq
About Jahmir: Jahmir is a 28-year-old male living in Iraq and works in the IT and Technical field. He likes to read about Religion, South America, and Human-Story on his Mobile. He usually watches about 1.2 minutes of video.	**People like Jahmir:** 1,247,020 people Gender (Male), age (25–34), country (Iraq), interests (Religion, South America, Human Interest Story), and language (Arabic) based potential reach.

As can be seen from Table 2, the audience is dominantly young males, from the age group of 25–34. The potential reach expresses the size of each respective persona group on Facebook, as retrieved from the Facebook Marketing API [4] by using targeting criteria shown in Table 2. Figure 4 shows the share of personas from the total video views on the YouTube channel.

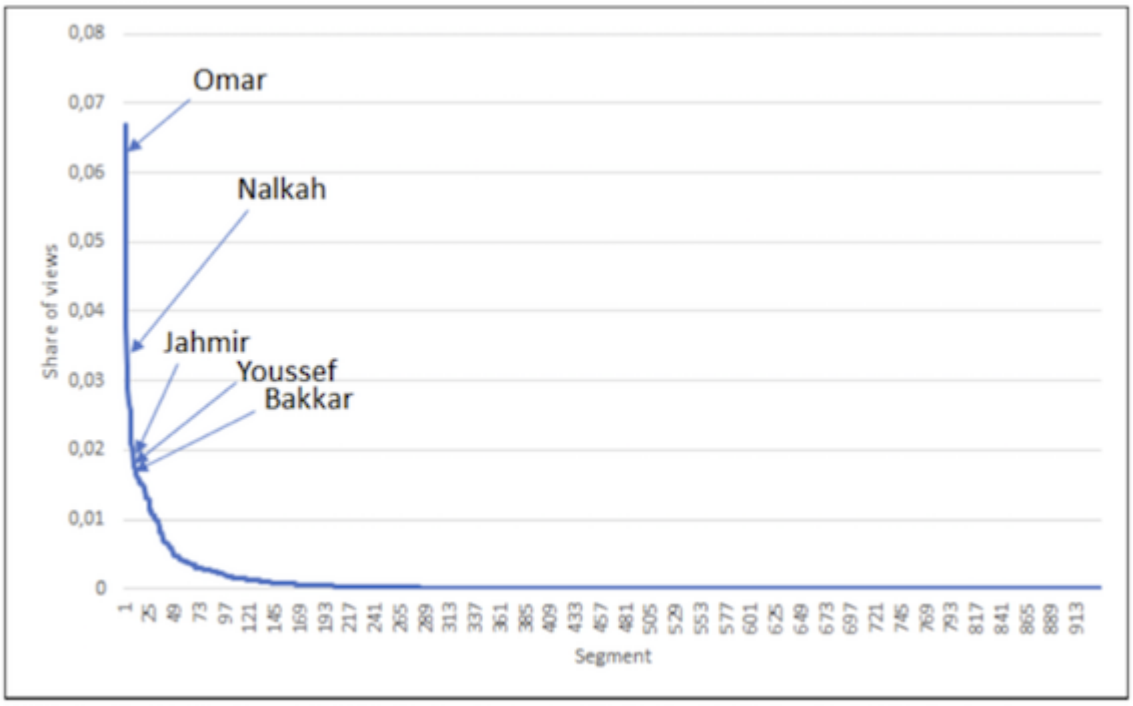


Figure 4: Personas' share of total content views. The videos have views from 937 user groups in 80 countries. The five personas capture a large share of the total views.

As automatically generated personas are created by assigning a demographic group with each latent content interaction pattern revealed by the NMF computation, each persona covers a degree of other demographic groups as well. We illustrate the demographic composition of the five personas in Table 3.

Table 3: Five most important demographic subgroups for the personas. Weights represent the importance of the corresponding group. The higher the weight of a certain group, the more important that group is for the corresponding persona.

	Country	Age	Gender	Weight
Omar	SA	25	Male	267.45
	SA	25	Female	132.05
	SA	18	Male	128.58
	IQ	25	Male	87.52
	SA	18	Female	85.50
Youssef	MA	25	Male	44.07
	MA	18	Male	35.54
	MA	35	Male	19.90
	MA	18	Female	12.24
	MA	45	Male	9.91
Bakkar	JO	18	Male	29.45

	PS	18	Male	22.83
	JO	25	Male	21.62
	JO	18	Female	21.41
	IL	18	Female	19.30
Nalkah	SA	25	Female	35.09
	SA	18	Female	18.78
	IQ	25	Female	8.25
	IQ	18	Female	7.14
	SA	35	Female	6.45
Yahmir	IQ	25	Male	38.96
	IQ	18	Male	29.91
	IQ	18	Female	21.61
	IQ	35	Male	17.26
	IQ	25	Female	16.29

Insights from qualitative data analysis

Manual analysis of the 255 Instagram profiles revealed 69 (27 percent) female, 111 (44 percent) male, and 75 (29 percent) unspecified gendered users. Coding usage patterns made it possible to come up with various recurring patterns (see Table 4). Two researchers separately re-coded all previous interview data (Harrell, *et al.*, 2017) from the perspectives of personas and social media usage patterns, constructing a *theme map* relevant to our purpose of enriching data-driven persona profiles. We coded both Instagram profiles and interviews separately using grounded theory principles. Then, we matched the categories to generate a unified understanding of social media behavior for Middle Eastern users. Having generated this understanding, it then became possible to enrich the quantitatively generated personas.

We used *action-implicative discourse analysis* (AIDA) (Tracy, 1995) for coding interviews along with techniques of grounded theory (Glaser and Strauss, 1967). Since AIDA is an analysis method based on grounded theory where participants reflect on how "*they respond to interactional problems and challenges they experience [...] with the aim of working toward what they think would create the best situation*" (Agne, 2012), it was especially useful to our purposes of specifically focusing on local and regional nuances of social media use and further query underlying reasons for these cultural and value-based differences. Results are shown in Tables 4 and 5.

Table 4: Recurring patterns in the manual analysis of social media profiles.

Usage pattern	General behaviors	Regional nuances
Commercial usage	Although we eliminated social media accounts of businesses, there were a good number of personal entrepreneurs (*e.g.*,	In these accounts, there was an emphasis on regional styles, identities, and diversities for local interests, instead of context that can be

	hairdressers, makeup artists, photographers, etc.) merging their commercial activities with personal ones.	deemed as Western. Almost all of the entrepreneurs chose to depict themselves in traditional and regional clothing — a choice also replicated in formal usage.
Family usage	A recurring theme was that accounts dominantly focused on family photos and content, rather than personal representations like selfies or daily snapshots.	Especially for female users who did not want to reveal their own faces, or who did not want to appear in photos, these accounts seemed to focus on an archival usage of children growing up. Apart from children and family depictions, these accounts had very few personal posts.
Formal usage (dominantly male)	The selection of a profile image in traditional or "Westernized" garment was a prominent choice for many male users. This theme appeared frequently in interviews. For male accounts with dominantly formal connections (*e.g.*, colleagues, business owners, elderly family members), appearing with traditional clothing of thobes and ghutras was deemed necessary.	These accounts had very few candid selfies, photos in "Westernized" garments (*e.g.*, jeans, suit & tie, etc.), or overall playful content. Instead, there were photos of regional activities like camping in the desert, falconry, or posing with cars, especially SUVs — all of which can be cited as activities representing social status for men in the region. Formal usage was especially common among middle-aged and older users.
Informal usage (male and female)	We compared informal use to what could be deemed as more "globally standard" use. Users appeared in a variety of clothing, utilized more selfies, and expressed experiences of daily	Although these accounts were mainly informal, they almost always had an optimal amount of content with regional clues, like occasional photos with traditional clothing, locales, or activities (cited earlier at formal usage). Informal usage

	life like food, sports, fun activities, etc.	was dominantly common among younger users, and very rare among middle-aged and older users.
Thematic usage (dominantly female)	These were accounts that dominantly had thematic photos instead of self-depictions. Although they could occasionally belong to male users, we observed them to be mainly female.	Due to privacy concerns, some female users did not want to show their faces. As a result, they were mostly sharing thematic photos like landscapes, motivational or religious quotes, flowers, scenery, close-up staged photos of daily activities without individuals (*e.g.*, food, clothes, etc.), and occasionally close-up details of their faces (mostly eyes, lips, etc.).

Table 5: Themes emerging from interviews.

Theme	Codes	Explanation
Formality; Dominantly Male	• Traditional clothing • Khaleeji features • Family connections • Formal connections • Online racism	The issue of formal vs. informal content was voiced in the interviews. Individuals preferred a more formal look (clothing, facial hair, etc.) and content when interacting with elderly family members and colleagues. While interacting with close friends, however, they preferred to have more informal representations and posts. For example, an interviewee referred to his thawb or thobe; "*I don't think [...] in here [outside of formal*

		account] I'm not gonna wear this." Additionally, they were aware that displaying their ethnic identity with regional clothing might provoke racism online; a male interviewee commented: "*I feel there would be more racism involved if there were more options to display certain things.*"
Monitoring; Male & Female	• Multiple account use • Relationship seeking • Connections with strangers • Privacy • Representation • Self-monitoring • Social monitoring	In parallel to a theme of formality, some interviewees expressed the need to maintain multiple accounts on social media platforms: one for formal interactions, one for more informal and less monitored interactions. In contrast to other regions where this behavior might be perceived to be less common, it seemed more acceptable and normal for the region. Since privacy was a concern, interviewees felt safer in exchanging content with strangers through more anonymized accounts and social networks that allowed for more obscure representations and intricate privacy settings. One female interviewee noted that: "*I feel like my account is my privacy, I only add my friends. I wouldn't want a random*

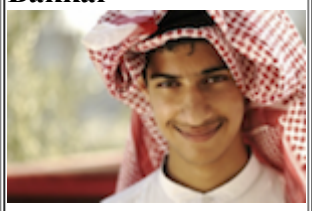
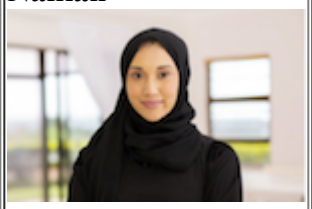
		stranger knowing what I'm watching right now, what I'm playing right now, or what my status is."
Modesty; Dominantly Female	• Participation • Representation • Online dialogue • Connection with strangers	Although the concept of modesty was relevant to both males and females in the region, it especially governed interactions of females online. In the physical world, modesty focuses more on clothing and cultural modes of interaction. In the online world, modesty is manifested in participation in online discussions. To indicate a level of modesty, a female interviewee pointed out that "*usually, a guy should never say his sister's name, in front of his friends.*" While commenting about females talking to strangers an online another male interviewee agreed that "*it's not, ah, modest.*"
Non-permanence; Male & Female	• Anonymity • Escapism • Physical and Online Identity Divide	Despite anonymity being an important concept for users around the globe, it seemed to have special significance in the region where cultural and social expectations and norms orient daily life more intensively. This resulted in individuals seeking anonymous and escapist interactions online, as voiced by one of the male interviewees: "*I*

		feel that the majority of people, if they know they're anonymous then they probably do things they wouldn't do if they weren't."

By integrating the AGPs with findings from qualitative manual coding along with recurring themes from interviews, it became possible to elaborate deeply on the personas and their approach to social media usage, as illustrated in Table 6.

Table 6: Mixed personas. Combining qualitative analysis results with AGPs enables us to construct likely behavioral pattern narratives for our personas.

AGP	Elaboration with qualitative analysis
Omar 	Omar's social media usage focuses on the formal category. Due to his job he has many colleagues and professional connections in social media. As a result, his representation, posts, and vocalized opinions would be more in-line with social and cultural norms and expectations. He might employ more anonymized accounts where he connects informally with close friends or strangers. Being a GCC citizen he is affected by modesty requirements online and his physical and online self is permeated into each other.
Youssef 	Youssef's social media usage focuses on a combination of family and informal categories as a possible young father. He is interested in social issues and local happenings, and he is vocal about them. He is less affected by the themes of monitoring and modesty. Although he might employ more anonymized accounts, he is more likely to be a member of closed communication groups, where politics and social issues precede family and daily matters.

Bakkar 	Bakkar's social media usage focuses on the informal category. As a young user, although he has some formal connections, many of his other connections are informal like close friends and even strangers. He is less affected by modesty and monitoring, as a result, he is more vocal, his visual representation is more candid and "Westernized." Being savvy in IT he is expected to have an intensive social media presence across many platforms. Some of his accounts may be more anonymized and more disassociated from his physical identity.
Nalkah 	Nalkah's social media usage focuses on the thematic category. Although she shows her face in social media, the majority of her posts stay thematic and modest. She utilizes privacy settings of social media systems to her advantage. Although she is a frequent social media user, she does not vocalize her thoughts so often as in comments and remains an observer. She has a few numbers of public posts if ever. Instead of utilizing multiple identities she prefers closed communication systems (*e.g.*, WhatsApp, Snapchat, etc.) to stay in touch with her friends and family.
Jahmir 	Jahmir's social media usage focuses on a combination of family and formal categories. He is not affected by the concerns of modesty and privacy; however, he still maintains a self-imposed modest and private social media presence. His posts often employ religious and motivational quotes. Although he is interested in global issues, his posts mainly consist of family photos and business content. His physical and online identity are overlapping. He does not maintain multiple accounts; however, he is a member of many

	closed groups in platforms like Facebook and WhatsApp.

The integration of qualitative insights into data-driven personas was done by the researcher that coded interviews and Instagram profiles, and verified by another researcher familiar with the YouTube audience of the source material. The qualitative researcher first familiarized himself with the qualitative material, then with the data-driven personas, and based on this joint understanding, wrote the narrative persona profiles as shown in Table 6. These were discussed together with a quantitative persona researcher to reach a common agreement on the narratives.

Discussion and conclusions

Our research answers two issues described in earlier research. First, we address the need for generating personas from large-scale data, often ignored in persona studies due to small sample sizes (Pruitt and Adlin, 2006; Chapman and Milham, 2006). In our approach, large-scale persona generation was completed quantitatively prior to an enrichment of the captured core personas with qualitative insights. Our qualitative analysis builds upon previous results of Harrell, *et al.* (2017) and expands them into the realm of a Middle Eastern persona context, describing an audience in a data-driven but humane way.

APG can utilize millions of content interactions to produce data-driven personas that are broad in their coverage of an audience. In this research, APG was applied to large-scale data from YouTube. Because the quantitatively generated personas had been enriched with qualitative data from Instagram as well as additional interviews, they were more comprehensive than purely quantitative personas that tend to be limited in their ability to describe in-depth motivations of the persona. In general, we postulate that such *culturally adapted personas* combine the speed and accuracy of computational techniques for processing millions of data points, and the strengths of traditional qualitative analysis in providing deep insights and understanding of the motives driving quantitatively observed behaviors. Moreover, even though the resulting personas were data-driven (*i.e.*, based on statistical analyses and qualitative data), they can be understood by end users without analytical sophistication due to a narrative form of writing (*cf.*, Nielsen, 2004). This approach is more cost-effective than pure manual persona generation, implying it could be used in many different industries to increase understandings of online audiences.

Second, we add to knowledge about Middle Eastern social media users, a segment not widely understood, by associating the following behaviors to data-driven personas: 1) formality; 2) monitoring; 3) modesty; and, 4) non-permanence. We also find an interesting regional divergence between Gulf Cooperation Council (GCC) member countries, so that only Saudi Arabia is represented in main personas, but Qatar, United Arab Emirates (UAE), Kuwait, Bahrain, and Oman are not. The differences likely originate from varying content preferences and population sizes. Regarding prior research, personas confirm the findings of Stanger, *et al.* (2017) that women in the region are using less social media than men, as four out of five of our data-driven personas represented male audiences. Although prior literature mentioned the potential of empowerment of Middle Eastern women in social media, the generated personas suggest that this is still an on-going process, as women are under-represented in the dominating personas.

Additionally, our personas concentrated on young audiences (average age of the personas = 27.6 years), consistent with Stanger, *et al.* (2017) about young male users as the dominant segment among Middle Eastern social media users. Although the sample was limited to one organization, the

active YouTube audience of that organization appears in a total of 226 countries and includes millions of views, lending support to the robustness of our findings.

The generation of culturally adapted personas opens avenues for cross-cultural comparisons of online audiences. Although we do not provide such comparisons in this study, our approach can be seen as a step toward that direction, given the flexibility of the APG method. However, while combining automated and manual procedures in persona generation provides several tangible benefits, it is also crucial to evaluate how useful personas are for end users, and to which extent they improve practical decision-making. Therefore, evaluating the impact of mixed-method personas requires further research.

Finally, more formalized measures of combining qualitative and quantitative analyses are needed. As explained earlier, in our research, this was a collaborative effort between quantitative and qualitative researchers, led by a qualitative researcher. However, more advanced forms of validation are needed, possibly including the use of inter-coder agreement tests to formally measure the alignment of persona perceptions (Krippendorff, 1980). As populations in different social networks may differ by structure (Ruths and Pfeffer, 2014), in future research we also need to verify that samples from various platforms, such as YouTube and Instagram used in this research, are formally commensurable.

About the authors

Joni Salminen is a postdoctoral researcher at the Qatar Computing Research Institute, and the Turku (Finland) School of Economics.
E-mail: joolsa [at] utu [dot] fi

Sercan Şengün is a postdoc associate at the Computer Science & Artificial Intelligence Laboratory at the Massachusetts Institute of Technology in Cambridge, Mass.
E-mail: sengun [at] mit [dot] edu

Haewoon Kwak is a senior scientist at the Qatar Computing Research Institute.
E-mail: hkwak [at] hbku [dot] edu [dot] qa

Bernard Jansen is Principal Scientist in the Social Computing Group at the Qatar Computing Research Institute.
E-mail: bjansen [at] hbku [dot] edu [dot] qa

Jisun An is a scientist at the Qatar Computing Research Institute.
E-mail: jan [at] hbku [dot] edu [dot] qa

Soon-Gyo Jung is a research associate at the Qatar Computing Research Institute.
E-mail: sjung [at] hbku [dot] edu [dot] qa

Sarah Vieweg is a scientist at the Qatar Computing Research Institute.
E-mail: sarahvieweg [at] gmail [dot] com

D. Fox Harrell is founder and director of the Imagination, Computation, and Expression Laboratory (ICE Lab) at the Massachusetts Institute of Technology in Cambridge, Mass.
E-mail: fox [at] csail [dot] mit [dot] edu

Acknowledgments

We would like to thank the journalists and content creators from the Al Jazeera Media Network who have collaborated with us on this project.

Notes

1. The live system has a demo version at https://persona.qcri.org/.

2. See http://flask.pocoo.org/.

3. https://developers.google.com/youtube/.

4.https://developers.facebook.com/docs/marketing-apis.

References

A. Abisheva, V.R.K., Garimella, D. Garcia, and I. Weber, 2014. "Who watches (and shares) what on YouTube? and when? Using Twitter to understand YouTube viewership," *WSDM '14: Proceedings of the Seventh ACM International Conference on Web Search and Data Mining*, pp. 593–602. doi: https://doi.org/10.1145/2556195.2566588, accessed 15 May 2018.

N. Abokhodair and S. Vieweg, 2016. "Privacy & social media in the context of the Arab Gulf, " *DIS '16: Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, pp. 672–683. doi: https://doi.org/10.1145/2901790.2901873, accessed 15 May 2018.

N. Abokhodair, S. Abbar, S. Vieweg, and Y. Mejova, 2016. "Privacy and Twitter in Qatar: Traditional values in the digital world," *WebSci '16: Proceedings of the 8th ACM Conference on Web Science*, pp. 66–77. doi: https://doi.org/10.1145/2908131.2908146, accessed 15 May 2018.

R.R. Agne, 2012. "Action-implicative discourse analysis," In: W. Donsbach (editor). *International Encyclopedia of Communication*, Malden, Mass.: Blackwell. doi: https://doi.org/10.1002/9781405186407.wbieca009.pub2, accessed 15 May 2018.

I.M. Al-Jabri, M.S. Sohail, and N.O. Ndubisi, 2015. "Understanding the usage of global social networking sites by Arabs through the lens of uses and gratifications theory," *Journal of Service Management*, volume 26, number 4, pp. 662–680. doi: https://doi.org/10.1108/JOSM-01-2015-0037, accessed 15 May 2018.

Y. Al-Saggaf, 2006. "The online public sphere in the Arab world: The war in Iraq on the Al Arabiya Website," *Journal of Computer-Mediated Communication*, volume 12, number 1, pp. 311–334. doi: https://doi.org/10.1111/j.1083-6101.2006.00327.x, accessed 15 May 2018.

Y.N, Alsharkh, 2012. "The social media effect on the families of the Saudi society from the perspective of the youth," Master' thesis, Arizona State University, at http://hdl.handle.net/2286/R.I.14718, accessed 14 January 2018.

J. An, H. Kwak, and B.J. Jansen, 2017 "Personas for content creators via decomposed aggregate audience statistics," *ASONAM '17: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pp. 632–635. doi: https://doi.org/10.1145/3110025.3110072, accessed 15 May 2018.

M. Aouragh, 2008. “Everyday resistance on the Internet: The Palestinian context,” *Journal of Arab & Muslim Media Research*, volume 1, number 2, pp. 109–130.
doi: https://doi.org/10.1386/jammr.1.2.109_1, accessed 15 May 2018.

M. Aoyama, 2005. “Persona-and-scenario based requirements engineering for software embedded in digital consumer products,” *RE ’05: Proceedings of the 13th IEEE International Conference on Requirements Engineering*, pp. 85–94.
doi: https://doi.org/10.1109/RE.2005.50, accessed 15 May 2018.

C.S. Araújo, L.P.D. Corrêa, A.P.C. da Silva, R.O. Prates, and W. Meira, 2014. “It is not just a picture: Revealing some user practices in Instagram,” *Proceedings of 2014 Ninth Latin American Web Congress (LA-WEB)*, pp. 19–23.
doi: https://doi.org/10.1109/LAWeb.2014.12, accessed 15 May 2018.

H. Bakhtari, 1995. “Cultural effects on management style: A comparative study of American and Middle Eastern management styles,” *International Studies of Management & Organization*, volume 25, number 3, pp. 97–118.
doi: https://doi.org/10.1080/00208825.1995.11656662, accessed 15 May 2018.

W. Barry and G. Bouvier, 2012. “Cross-cultural communication: Arab and Welsh students’ use of Facebook,” *Journal of Arab & Muslim Media Research*, volume 4, number 2, pp. 165–184.
doi: https://doi.org/10.1386/jammr.4.2-3.165_1, accessed 15 May 2018.

V. Beaudouin and D. Pasquier, 2017. “Forms of contribution and contributors’ profiles: An automated textual analysis of amateur on line film critics,” *New Media & Society*, volume 19, number 11, pp. 1,810–1,828.
doi: https://doi.org/10.1177/1461444816643504, accessed 15 May 2018.

R.N. Bolton, A. Parasuraman, A. Hoefnagels, N. Migchels, S. Kabadayi, T. Gruber, Y.K. Loureiro, and D. Solnet, 2013. “Understanding Generation Y and their use of social media: A review and research agenda,” *Journal of Service Management*, volume 24, number 3, pp. 245–267.
doi: https://doi.org/10.1108/09564231311326987, accessed 15 May 2018.

C.N. Chapman and R.P. Milham, 2006. “The personas’ new clothes: Methodological and practical arguments against a popular method,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 50, number 5, pp. 634–636.
doi: https://doi.org/10.1177/154193120605000503, accessed 15 May 2018.

C.N. Chapman, E. Love, R.P. Milham, P. ElRif, and J.L. Alford, 2008. “Quantitative evaluation of personas as information,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 52, number 16, pp. 1,107–1,111.
doi: https://doi.org/10.1177/154193120805201602, accessed 15 May 2018.

A. Cooper, 2004. *The inmates are running the asylum*. Indianapolis, Ind.: Sams.

B.G. Glaser and A.L.Strauss, 1967. *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

M. Haggan, 2007. “Text messaging in Kuwait. Is the medium the message?” *Multilingua: Journal of Cross-Cultural and Interlanguage Communication*, volume 26, number 4, pp. 427–449.
doi: https://doi.org/10.1515/MULTI.2007.020, accessed 15 May 2018.

D.F. Harrell and C.-U. Lim, 2017. “Reimagining the avatar dream: Modeling social identity in digital media,” *Communications of the ACM*, volume 60, number 7, pp. 50–61.
doi: https://doi.org/10.1145/3098342, accessed 15 May 2018.

D.F. Harrell, S. Vieweg, H. Kwak, C.-U. Lim, S. Şengün, A. Jahanian, and P. Ortiz, 2017. "Culturally-grounded analysis of everyday creativity in social media: A case study in Qatari context," *C&C '17: Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, pp. 209–221.
doi: https://doi.org/10.1145/3059454.3059456, accessed 15 May 2018.

Y. Hu, L. Manikonda, and S. Kambhampati, 2014. "What we Instagram: A first analysis of Instagram photo content and user types," *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, pp. 595–598, and at https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8118, accessed 15 May 2018.

B.J. Jansen, S.–G. Jung, J. Salminen, J. An, and H. Kwak, 2017. "Viewed by too many or viewed too little: Using information dissemination for audience segmentation," *Proceedings of the Association for Information Science and Technology*, volume 54, number 1, pp. 189–196.
doi: https://doi.org/10.1002/pra2.2017.14505401021, accessed 15 May 2018.

A. Jenkinson, 1994. "Beyond segmentation," *Journal of Targeting, Measurement and Analysis for Marketing*, volume 3, number 1, pp. 60–72.

S. Joye, 2009. "The hierarchy of global suffering: A critical discourse analysis of television news reporting on foreign natural disasters," *Journal of International Communication*, volume 15, number 2, pp. 45–61.
doi: https://doi.org/10.1080/13216597.2009.9674750, accessed 15 May 2018.

S.–G. Jung, J. An, H. Kwak, M. Ahmad, L. Nielsen, and B.J. Jansen, 2017. "Persona generation from aggregated social media data," *CHI EA '17: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1,748–1,755.
doi: https://doi.org/10.1145/3027063.3053120, accessed 15 May 2018.

S.A. Karabenick and S. Moosa, 2005. "Culture and personal epistemology: U.S. and Middle Eastern students' beliefs about scientific knowledge and knowing," *Social Psychology of Education*, volume 8, number 4, pp. 375–393.
doi: https://doi.org/10.1007/s11218-005-1826-3, accessed 15 May 2018.

K. Krippendorff, 1980. *Content analysis: An introduction to Its methodology*. Beverly Hills, Calif. :Sage.

H. Kwak, J. An, and B.J. Jansen, 2017. "Automatic generation of personas using YouTube social media data," *Proceedings of the 50th Hawaii International Conference on System Sciences*, pp. 833–842.

H. Kwak, C. Lee, H. Park, and S. Moon, 2010. "What is Twitter, a social network or a news media?" *WWW ' 10: Proceedings of the 19th International Conference on World Wide Web*, pp. 591–600.
doi: https://doi.org/10.1145/1772690.1772751, accessed 15 May 2018.

D.D. Lee and H.S. Seung, 1999. "Learning the parts of objects by non-negative matrix factorization," *Nature*, volume 401, number 6755 (21 October), pp. 788–791.
doi: https://doi.org/10.1038/44565, accessed 15 May 2018.

T. Matthews, T. Judge, and S. Whittaker, 2012. "How do designers and user experience professionals actually perceive and use personas?" *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1,219–1,228.
doi: https://doi.org/10.1145/2207676.2208573, accessed 15 May 2018.

J. McGinn and N. Kotamraju, 2008. “Data-driven persona development,” *CHI ’08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1,521–1,524.
doi: https://doi.org/10.1145/1357054.1357292, accessed 15 May 2018.

M.H. Millham and D. Atkin, 2018. “Managing the virtual boundaries: Online social networks, disclosure, and privacy behaviors,” *New Media & Society*, volume 20, number 1, pp. 50–67.
doi: https://doi.org/10.1177/1461444816654465, accessed 15 May 2018.

R. Mourtada and F. Salem, 2011.“Civil movements: The impact of Facebook and Twitter,” *Arab Social Media Report*, volume 1, number 2, pp. 1–30, at https://journalistsresource.org/wp-content/uploads/2011/08/DSG_Arab_Social_Media_Report_No_2.pdf, accessed 15 May 2018.

E.W. Ngai, S.S. Tao, and K.K. Moon, 2015. “Social media research: Theories, constructs, and conceptual frameworks,” *International Journal of Information Management*, volume 35, number 1, pp. 33–44.
doi: https://doi.org/10.1016/j.ijinfomgt.2014.09.004, accessed 15 May 2018.

L. Nielsen, 2004. “Engaging personas and narrative scenarios,” Ph.D. thesis, Department of Informatics, Copenhagen Business School, at http://personas.dk/wp-content/samlet-udgave-til-load.pdf, accessed 15 May 2018.

L. Nielsen and K. Storgaard Hansen, 2014. “Personas is applicable: A study on the use of personas in Denmark,” *CHI ’14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1,665–1,674.
doi: https://doi.org/10.1145/2556288.2557080, accessed 15 May 2018.

Northwestern University in Qatar, 2017. “Media use in the Middle East 2016: A six-nation survey,” at http://www.qatar.northwestern.edu/docs/publications/research-media-use/2016-middle-east-media-use-report.pdf, accessed 27 November 2017.

E. Pöyry, P. Parvinen, and T. Malmivaara, 2013. “Can we get from liking to buying? Behavioral differences in hedonic and utilitarian Facebook usage,” *Electronic Commerce Research and Applications*, volume 12, number 4, pp. 224–235.
doi: https://doi.org/10.1016/j.elerap.2013.01.003, accessed 15 May 2018.

J. Pruitt and T. Adlin, 2006. *The persona lifecycle: Keeping people in mind throughout product design*. Boston: Morgan Kaufmann.

J. Pruitt and J. Grudin, 2003. “Personas: Practice and theory,” *DUX ’03: Proceedings of the 2003 Conference on Designing for User Experiences*, pp. 1–15.
doi: https://doi.org/10.1145/997078.997089, accessed 15 May 2018.

D. Ruths and J. Pfeffer, 2014. “Social media for large studies of behavior,” *Science*, volume 346, number 6213 (28 November), pp. 1,063–1,064.
doi: https://doi.org/10.1126/science.346.6213.1063, accessed 15 May 2018.

J. Salminen, M. Milenković, and B.J. Jansen, 2017 “Problems of data science in organizations: An explorative qualitative analysis of business professionals’ concerns,” *Proceedings of the 17th International Conference on Electronic Business*, pp. 192–201; version at http://iceb.johogo.com/proceedings/2017/ICEB_2017_paper_14-full.pdf, accessed 15 May 2018.

J. Salminen, S. Şengün, H. Kwak, B.J. Jansen, J. An, S.-G. Jung, S. Vieweg, and D.F. Harrell 2017. “Generating cultural personas from social data: A perspective of Middle Eastern users,” *Proceedings of 2017 Fifth International Conference on Future Internet of Things and Cloud*

Workshops (FiCloudW), pp. 120–125.
doi: http://doi.ieeecomputersociety.org/10.1109/FiCloudW.2017.97, accessed 15 May 2018.

R. Sawyer and G.-M. Chen, 2012. "The impact of social media on intercultural adaptation," *Intercultural Communication Studies*, volume 21, number 2, pp. 151–169; version at http://digitalcommons.uri.edu/com_facpubs/15/, accessed 15 May 2018.

D.M. Scott, 2007. *The new rules of marketing and PR*. Hoboken, N.J.: Wiley.

C.L. Singh, 2010. "New media and cultural identity," *China Media Research*, volume 6, number 1, pp. 86–91.

N. Singh, K. Lehnert, and K. Bostick, 2012. "Global social media usage: Insights into reaching consumers worldwide," *Thunderbird International Business Review*, volume 54, number 5, pp. 683–700.
doi: https://doi.org/10.1002/tie.21493, accessed 15 May 2018.

N. Stanger, N. Alnaghaimshi, and E. Pearson, 2017. "How do Saudi youth engage with social media?" *First Monday*, volume 22, number 5, at http://firstmonday.org/article/view/7102/6101, accessed 15 May 2018.
doi: http://dx.doi.org/10.5210/fm.v22i15.7102, accessed 15 May 2018.

K. Tracy, 1995. "Action-implicative discourse analysis," *Journal of Language and Social Psychology*, volume 14, numbers 1–2, pp. 195–215.
doi: https://doi.org/10.1177/0261927X95141011, accessed 15 May 2018.

B. Ur and Y. Wang, 2013. "A cross-cultural framework for protecting user privacy in online social media," *WWW '13 Companion: Proceedings of the 22nd International Conference on World Wide Web*, pp. 755–762.
doi: https://doi.org/10.1145/2487788.2488037, accessed 15 May 2018.

H. Wasserman, 2011. "Mobile phones, popular media, and everyday African democracy: Transmissions and transgressions," *Popular Communication*, volume 9, number 2, pp. 146–158.
doi: https://doi.org/10.1080/15405702.2011.562097, accessed 15 May 2018.

J.B. Wiest and N. Eltantawy, 2015. "Mediatization in the Arab world: A cross-cultural comparison of new media use," *Online Journal of Communication and Media Technologies*, volume 5, number 2, pp. 120–142, and at http://www.ojcmt.net/articles/52/528.pdf, accessed 15 May 2018.

YouTube page of AJ+ Arabic, 2017. "AJ+عربي," at https://www.youtube.com/channel/UC-4KnPMmZzwAzW7SbVATUZQ, accessed 15 May 2018.

X. Zhang, H.-F. Brown, and A. Shankar, 2016. "Data-driven personas: Constructing archetypal users with clickstreams and user telemetry," *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5,350–5,359.
doi: https://doi.org/10.1145/2858036.2858523, accessed 15 May 2018.

I.A. Zualkernan, J. Allert, and G.Z. Qadah, 2006. Learning styles of computer programming students: A Middle Eastern and American comparison," *IEEE Transactions on Education*, volume 49, number 4, pp. 443–450.
doi: https://doi.org/10.1109/TE.2006.882366, accessed 15 May 2018.

Appendix: Interview questions

1. Which social media or digital communication platforms do you use, how often, and for what kind of connections? (*We ask for elaborations on: Facebook, Twitter, Instagram, Snapchat, and WhatsApp, which are all popular in different degrees in the region.*)

2. In deciding whether to use a social media or digital communication platform which factors are important to you? What are the deal-breaker issues why you would NOT use one? (*We ask for elaborations on themes like disclosure, privacy, interactions with strangers, gender, anonymity, monitoring, community, cultural values, and language.*)

3. In any social media or digital communication platform, did you experience any difficulties in building a profile according to your identity? (*We ask for elaborations on race, ethnicity, gender, local values, and norms, etc.*)

4. In any social media or digital communication platform, did you experience any conflict with your cultural values? How about conflicts with other users from diverse cultures? (*We ask for elaborations on themes like disclosure, privacy, interactions with strangers, gender, anonymity, monitoring, community, cultural values, and language.*)

5. In your opinion, what kind of features would make a social media or digital communication platform more regionally and locally usable?

Editorial history

Received 18 March 2018; revised 13 May 2018; accepted 17 May 2018.

From 2,772 segments to five personas: Summarizing a diverse online audience by generating culturally adapted personas
by Joni Salminen, Sercan Şengün, Haewoon Kwak, Bernard Jansen, Jisun An, Soon-Gyo Jung, Sarah Vieweg, and D. Fox Harrell.
First Monday, Volume 23, Number 6 - 4 June 2018
https://firstmonday.org/ojs/index.php/fm/article/download/8415/7416
doi: http://dx.doi.org/10.5210/fm.v23i6.8415