Visual Identities in Troll Farms: The Twitter Moderation Research Consortium

Accepted for publication in *Social Media and Society* (preprint version: changes still possible)

Marco Bastos (University College Dublin & City St George's, University of London)

Abstract

The Twitter Moderation Research Consortium is a database of network propaganda and influence operations that includes 115,474 unique Twitter accounts, millions of tweets, and over one terabyte of media removed from the platform between 2017-2022. We probe this database using Google's Vision API and Keras with TensorFlow to test whether foreign influence operations can be identified based on the visual presentation of fake user profiles emphasizing gender, race, camera angle, sensuality, and emotion. Our results show that sensuality is a variable associated with operations that replicate the Kremlin-linked Internet Research Agency campaign, being particularly prevalent in influence operations that targeted communities in North and South America, but also in Indonesia, Turkey, and Pakistan. Our results also show that the visual identities of fake social media profiles are predictive of influence operations given their reliance on selfies, sensual young women, K-pop aesthetics, or alternatively nationalistic iconography overlaid with text to convey ideological positioning.

Data Availability

The TensorFlow and Vision AI models created for this study are available upon request from the corresponding author. The data analyzed in this study are available from the corresponding author on reasonable request and subject to Twitter's Terms and conditions governing the sharing of Twitter data.

Introduction

Network propaganda has proved a formidable challenger to the management of centralized social media platforms following the discovery of the Internet Research Agency's (IRA) influence operation targeting the 2016 US presidential election and the ensuing onslaught of COVID-19 conspiracy theories during the pandemic. The streamlined and cost-effective creation of fake social media profiles, typically employed for coordinated inauthentic behavior or as networks of sockpuppet accounts (Bastos & Mercea, 2019), shaped a playbook for influence operations seeding division ultimately deployed in elections around the world, but also in the cottage industry specialized in crafting desirable online personas for romance scams (Faux, 2023).

Social media platforms sought to curb these operations by implementing community guidelines (Facebook, 2018a, 2018b; Twitter, 2018) to protect the health of the public debate. The platforms also enforced election integrity policies designed to prevent the spread of false or misleading information about elections. These policies cover a range of problematic content, including false, misleading, or unverified information about public consultations, but also the incitement of violence to interfere with civic processes, and coordinated reporting, posting, or sharing of information to manipulate the public conversation. Behavior that abused these norms was deemed in violation of the Terms of Service (ToS), Platform Manipulation Policy, or the Spam Policy, and thus subject to removal from the platforms.

Upon identifying the origin (source attribution) and target of these influence operations, social media platforms label, remove, or reduce the visibility of such content depending on the severity and reach of the violation. These initiatives yield databases of influence operations, with Meta curating the Information Operation (IO) Research Archive and Twitter Trust and Safety team leading the company's efforts to safeguard elections and deal with content that could jeopardize healthy conversations online (Harvey & Roth, 2018). Twitter's Civic Integrity policy (Twitter, 2021) would eventually mature into the Twitter Moderation Research Consortium (TMRC, 2022). Starting in 2017-2018 as a reaction to the IRA operations, it shared data with the academic community, initially under the umbrella of Twitter's

Elections Integrity initiative, which identified and ultimately removed false accounts, Twitterbots, and sockpuppets (Elections Integrity, 2018; Twitter, 2019).

The first release of the data included 2752 accounts the company attributed to the IRA. This list was expanded in early 2018 to include 3814 IRA-linked accounts. The Twitter Moderation Research Consortium (TMRC) continued to be updated over the next years, and the final dataset included 115,474 unique Twitter accounts, millions of individual tweets, and more than one terabyte of media removed from the platform due to breaches of the ToS. It included information about user accounts that posted over 100 million tweets (25 million in TMRC14 and TMRC15 and 34 million in the 2018-2019 releases) linked to 57 influence operations carried out in several countries. From this universe of 115,474 accounts, 40,407 contain information about the user profile and a link to the profile image that could be downloaded at the time the TMRC offered access to the data. This is the database we probe to identify the visual identities of fake Twitter profiles.

We probe this database by training custom machine-learning models based on visual attributes of Twitter profiles regularly explored in state-sponsored social media propaganda, including age, gender, race, camera angle, composition, location, emotion, and sensuality—which are important markers for propagandists seeking to infiltrate social groups. These models rely on Google Vision API, Teachable Machine, and the DeepFace Library to process the totality of fake profile images made available through the TMRC. This allows us to test whether the visual identities of Twitter profiles can be used to identify propaganda campaigns, and more specifically, whether the visual features of such profiles are associated with the source and target of the campaign. In the following, we discuss the particulars of the TMRC database, unpack related work in this area, and advance a framework for the scalable detection of visual propaganda on social media.

Previous work

Influence operations on social media employ an array of digital instruments, including Twitterbots, fake accounts, sockpuppets, trolls, and compensated influencers to disseminate their messaging (Bastos & Mercea, 2018; Benkler et al., 2018). These accounts feature profile images as the initial point of engagement with users, and therefore propagandists carefully and deliberately select images that can elicit trust, evoke emotional response, or endorse ideological positioning (Seo, 2014). IRA profiles in particular employed visual aesthetics and a grammar of self-presentation that strategically catered to the targeted subcultures (Xia et al., 2019). Profile pictures also lend credibility to user's communications (Morris et al., 2012), seize attention, and elicit emotions (Rose et al., 2012). Effective visual communication through profile images is therefore central to attaining social embeddedness and minimizing the labor-intensive costs of infiltration (Freelon et al., 2020). Contemporary propaganda strategists have followed suit by aptly manipulating images as instruments to steer collective sentiments and emotions (Weikmann & Lecheler, 2023).

The affordances of social media compound these issues by emphasizing visual over written communication (Highfield & Leaver, 2016). Previous research identified that social media propaganda seeks to embody relatable, familiar, and attractive faces of ordinary people (Bastos et al., 2023), with a clear gender divide encapsulated by unassuming males lacking overt tropes of hegemonic masculinity like strength and dominance in contrast to females characterized by the tangible exploitation of the female body for maximum erotic impact and objectification (Davis, 2018; Rose et al., 2012). This is consistent with the general sexual objectification found in media culture and music videos, where female artists are more likely to be sexually objectified and display sexually alluring behavior (Aubrey & Frisby, 2011). This is nonetheless magnified in fake social media profiles designed for political propaganda, frauds, and scams using desirable online personas, with leaked manuals of pig butchering scams explicitly determining the creation of female profiles featuring naughty but cute nicknames with photos of attractive young women who appear wealthy and educated (Faux, 2023).

There is also research that explored the personality traits expressed in social media images (Celli et al., 2014; Ferwerda & Tkalcic, 2018), as profile pictures are subject to personality inference from the facial appearance that guides adaptive behavior (Zebrowitz & Montepare, 2008). The potential effects of a politician's facial appearance on trait judgment intersect with gender differences, with voters rewarding female candidates who appear more feminine (Carpinella et al., 2016), and driving inferences of competence, which have been found to be predictive of election results (Todorov et al., 2005). In addition to gender, research has also found that race and ethnicity are significant variables in the visual messaging about immigration that intersect with emotional variables like anxiety (Brader et al., 2008). The detection of such attributes from facial images is informed by early approaches in automated feature extraction that identified a person from a facial image (Kanade, 1977), further expanded to recognize attributes such as gender, race, and age, but also emotional states and expressions (Liu et al., 2015).

Research on visual communication has increasingly incorporated methods from computer vision, with a growing body of scholarship dedicated to reviewing automated methods for image analysis (Pearce et al., 2020; Peng et al., 2024). Unfortunately, off-the-shelf computer vision tools are not designed to analyze media effects such as visual framing nor are they intended to identify visual tropes in social media propaganda. To bridge this gap between computer vision and propaganda studies, we devise an analytical framework based on predefined visual concepts to be measured with customized computer visual algorithms (Peng et al., 2024). This methodological approach draws from scholarship on image composition that portrays individuals as powerful or otherwise ordinary, sensual or otherwise average, and assertive instead of meek (George et al., 2024). Definitions of power and sensuality vary widely among individuals and cultural groups but have been theorized along the following coordinates informing our study: confidence, sensuality, and categorization (Tajfel & Turner, 1986).

2.1 Composition

Image composition dedicated to imprinting power and confidence relies on the position of the camera so that the photographed object appears imponent. These compositional choices are broadly defined by whether the subject is shot from above, below, or at eye level to frame the object through high, low, or neutral angle shots (Merkt et al., 2022). Perceptions linked to the quality of the image, but most prominently to camera angles, are influenced by evolutionary cues, social learning, and embodied cognition. Language reflects these perceptions by equating power with upward positions and lack of power with downward trajectories. It also associates vertical angles with dominance or subordination (Meyers-Levy & Peracchio, 1992).

Changes in camera angle thus lead to significant and predictable changes in how the physical and personal characteristics of the photographed object are judged. Low-angle shots often make the object appear taller and stronger and are thus employed to portray power and courage. Eye-level shots, on the other hand, impart a sense of equality, parity, and neutrality. High-angle shots tend to present the photographed subject as weaker or frail and manufacture a sense of vulnerability (Kraft, 1987). Notably, men tend to be depicted from low angles to suggest dominance and power, whereas women are often shown from high angles to suggest fragility or lesser status.

2.2 Sensuality

Broadly conceived, sensuality encompasses the myriad ways humans experience and express sensations of pleasure (Heathwood, 2006, 2007). The more intimate and sexual aspects of sensuality are intricately woven with elements that stimulate multiple senses, often characterized by surfaces, curves, and textures that resonate with femininity, corporeality, and eroticism (Pritchard & Morgan, 2011). This understanding of sensuality drives the depiction of young women in soft advertisements employed by the cosmetic industry (Xie & Zhang, 2013) where sensuality transcends nudity to evoke sexual feelings through sensory stimuli (Ringrow, 2016). While visual elements like parted lips are

commonly associated with sensual tropes, sensuality remains culturally and historically grounded notwithstanding its commodification by the advertising industry to evoke sensory experiences that shape emotional engagement and brand loyalty (Wolf, 2013).

Notions of sensuality vary substantially across cultures and are predicated on cultural norms and belief systems. Western cultures tend to emphasize attributes such as confidence, physical fitness, and the celebration of diverse body types. Conversely, Eastern cultures tend to favor subtlety and slender figures (Starr et al., 2020). In South Asian contexts, notions of sensuality are often intertwined with traditional attire. In Middle Eastern cultures, sensuality is frequently associated with modesty, secrecy, and mystery. African cultures, in contrast, often perceive fuller figures as emblematic of sensuality and a trait of fertility. In some cultures, women with lighter skin tones have been perceived as more feminine and sexually appealing than their darker-skinned counterparts, and Latin American cultures are typically drawn to curves, vivacity, and exuberance (Frost, 1990). In some cultures, including Middle Eastern and South Asian cultures that feature in the TMRC, any form of female nudity in the public space is perceived as sexualized; in other cultures, however, nudity may be featured without overt sexualization.

2.3 Categorization

Another salient dimension in the composition of social media profile images is social categorization through which individuals are grouped based on social information, with gender, race, and age featuring prominently. These categories underpin expected roles, behaviors, and activities assigned to individuals based on societal norms (Butler, 2002; West & Zimmerman, 1987). Gender norms influence the personalization strategies utilized by both authentic and fake social media users, with the visual presentation of gender frequently reinforcing societal norms, particularly in the sexualization and objectification of women (Davis, 2018; Rose et al., 2012). As such, gender is a central aspect in the creation of fake social media profiles and the shaping of digital identities (Muscanell & Guadagno,

2012; Toma & Hancock, 2012), with ‘catfishing’ typifying the centrality of such social constructs in crafting persuasive online personas that adhere to stereotypical gender norms.

The other central category is race, a social construct and stratifier that is central to the persistence of inequalities (Bonilla-Silva, 1997). Race also influences cultural norms and practices through its intersection with intergroup dynamics, intrinsic biases, stereotypes, and prejudices. Consequently, race plays a central role in the shaping of identities, intergroup relations, political engagement, and cultural acceptance (Bonilla-Silva, 1997). Intersecting with other social categories, racial categorizations are wielded to either perpetuate or challenge structures legitimizing power imbalances and inequities (Crenshaw, 2013). This includes the intersectionality of race and gender, further compounding their impact on individual experiences (Smedley & Smedley, 2005). Notably, the complexity of the interaction between gender and race poses significant challenges for machine learning algorithms (Buolamwini & Gebru, 2018).

Objectives

We take stock of the literature reviewed above to test whether the visual tropes exploited by distinct propaganda campaigns can be identified at scale. We leverage TensorFlow’s high-level API Keras and Google Vision API to train custom machine-learning models of social media user profiles. The models are applied to the TMRC database to identify the visual features of state-sponsored social media propaganda, including age, gender, race, angle, quality, sensuality, location, and emotion, along with a range of discrete variables employed to create these profiles. To this end, we test the hypothesis (H1) that the visual parameters of Twitter accounts are predictive of influence operations (unit of analysis: users). We also test an auxiliary hypothesis (H2) that the visual identities of influence operations (IO) on Twitter are predictive of source and targeted countries (unit of analysis: IO campaigns). In the following, we describe the TMRC database and the methods employed in this study.

Data

The TMRC database includes tweets, user accounts, and profile images that have undergone various forms of content moderation measures. Of particular note, the TMRC database not only includes textual content but also information pertaining to profile images that have been subjected to moderation. This repository includes granular data relating to user accounts flagged, removed, or subjected to enforcement measures. Such accounts were categorized according to criteria such as inappropriate content, graphic images, or violations of Twitter's policies and guidelines. The database includes key metrics such as the number of accounts taken down, the number of tweets, languages used by the group of fake accounts, key hashtags, temporal range of account activity, user-reported locations, and technical indicators of location. Taken together, it offers critical insights into the efficacy of content moderation strategies while also foregrounding the challenges that social media platforms must contend with in ensuring user safety and adherence to community standards.

The database includes only networks with significant evidence indicating that state-affiliated entities were knowingly trying to manipulate and distort public conversations. The influence operations taken down by the TMRC include small networks in Bangladesh that engaged in coordinated platform manipulation with a focus on regional political themes and networks in the United Arab Emirates and Egypt that primarily targeted Qatar and Iran while amplifying messaging supportive of the Saudi government. This is in addition to accounts linked to Saudi Arabia's state-run media apparatus that engaged in coordinated efforts to amplify messaging beneficial to the Saudi government. Other small campaigns include the operations of Partido Popular in Spain and a separate network associated with the Catalan independence movement, specifically Esquerra Republicana de Catalunya. It also includes networks in Ecuador tied to the PAIS Alliance political party, which primarily engaged in spreading content about President Moreno's administration.

The TMRC also includes large information operations in Russia, Iran, and Venezuela targeting other countries and/or domestic audiences by leveraging 'spammy' content focused on divisive political

themes, with behavior that mimics the seminal influence operation orchestrated by the IRA. The Iranian cohort posted nearly two million tweets with content that pressed the geostrategic views of the Iranian state. This playbook was also identified in a group of 4248 accounts operating from the United Arab Emirates directed at Qatar and Yemen that employed false personae and tweeted about the Yemeni Civil War and the Houthi Movement. It also includes very large influence operations counting over 200,000 accounts manned by the People's Republic of China (PRC) and dedicated to sowing political discord in Hong Kong and undermining the legitimacy of local protest movements. These accounts were suspended for a range of violations of Twitter's platform manipulation policies, including platform manipulation and spam, coordinated activity, fake accounts, attributed activity, distribution of hacked materials, ban evasion, and what the TMRC referred to as 'violative content.'

The subset of interest to this project includes 115,474 unique Twitter accounts removed from the platform due to breaches in the platform's Terms of Service. These accounts posted over 100 million tweets (25 million in TMRC14 & 15 campaigns alone, in addition to 34 million in the later 2018-2019 data release) linked to 57 influence operations carried out in the Global North and the Global South that mimic, or build upon, the seminal influence operation by the IRA. From this universe of 115K accounts, 40,407 included detailed information about user accounts, with a link to the profile images that could be downloaded at the time the TMRC offered access to this data. As such, a total of 75,067 Twitter accounts were archived with no profile photo or information about the account that could be used to download the profile images (see Appendix 1 for the list of influence operations to which no profile image was made available). Table 1 unpacks this database with a breakdown of campaign targets and source attribution followed by the number of tweets and accounts involved in each influence operation campaign.

INSERT TABLE 1 HERE

Computer vision and object recognition algorithms are optimized for data collected from higher-income households located in the Global North (De Vries et al., 2019). This is a considerable

shortcoming for research on the Global South, which features prominently in the TMRC database. This shortcoming can be partially offset by deploying off-the-shelf tools to mitigate cultural biases in computer vision tools (Peng et al., 2024). This is the case of sensuality, which is subject to cultural and geographic variations in addition to evolving cultural expectations. To this end, we include a diverse set of sensual images sourced from the many countries represented in the TMRC database. The training dataset thus includes images from different cultural settings with a binary category termed 'sensual,' a variable informed by the inferential definitions found in the product management literature (Hofmeester et al., 1996).

Data preprocessing started by removing illustrations lacking identifiable individuals and resulted in a dataset consisting of predominantly neutral angle shots. While high and low-angle shots were evenly distributed, notable disparities emerged in the intersection with race and gender. Neutral angles prevailed across all racial categories, with the highest proportion observed among White individuals (95%). Notably, 9% of images featuring individuals of Indian descent utilized high angles, in sharp contrast to low angles that appeared in only 1% of the images. A similar trend was observed among Latino and Asian groups. Given the overarching prevalence of neutral angles, these differences are compounded by gender, with 80% of images featuring women employing high-angle shots, compared with 50% for men. A significant 68% of sensual images and 75% of amateur photos favored high angles, which were also more likely to convey anger, fear, happiness, and sadness.

Finally, we considered the ethical dilemmas in displaying profile pictures that might have been taken from real people, including of course the many celebrities conspicuously featured in the database, as these disinformation outfits may have misappropriated profile images of real users, even if a sizable share of the data may have been taken from stock photos of celebrities and models. In the end, our concerns were offset by the realization that the images explored in this study have been recontextualized to such an extent that they are detached from potentially existing personas, reflecting

instead the profiles manufactured by disinformation outfits that cannot be promptly associated with the original source of the image.

Methods

This study combines data extraction and automated object recognition (image tagging) for the analysis of state propaganda on social media. It leverages state-of-the-art machine learning platforms for the automatic classification of profile images through optical character recognition (OCR) as well as face, emotion, logo, landmark, color, inappropriate content, and object detection. These features are extracted by deploying customized machine learning algorithms trained with Teachable Machine and DeepFace in addition to the Google Vision API. While Teachable Machine allows users to train custom models, Vision API leverages vast datasets and state-of-the-art computer vision algorithms based on ImageNet (2016), the de facto gold standard for training computer vision algorithms featuring a repository of over 100 million images.

Vision API was launched in February 2016 and is based on the TensorFlow open-source framework. It features optical character recognition (OCR) in addition to face, emotion, logo, landmark, color, inappropriate content, and object detection. Vision API benefits from a comprehensive understanding of diverse image content and has been widely used in media studies (d'Andrea & Mintz, 2019). This vast exposure allows the model to generalize well across a myriad of image types and contexts, leading to higher accuracy and precision in classifications. Google's Teachable Machine, despite leveraging the power of transfer learning through MobileNet, is more constrained due to its training data. Originally designed to discern between 1000 classes, MobileNet offers limited ability to generalize across broader contexts or perform sophisticated fine-tuning.

The classifiers created with Teachable Machine generated a Keras model for downstream analysis. Image composition, including the angle and quality of the photo, was sourced from open image databases, which proved challenging owing to the limited diversity found in open databases. The

difficulty in classifying the photographic angle stems from the relative absence of consistent and clear patterns that can distinguish low, high, and neutral angles. This shortcoming was addressed by using a pre-trained model based on face tilt to classify images into high, low, and neutral, then manually annotating a subset of images to assess the classifier's accuracy, and finally by iteratively refining the model on training data with balanced representation across the relevant categories. Addressing these shortcomings in Teachable Machine is however an imperfect process due to the inherent opacity of the system, whereby performance evaluation relies solely on the output of the model.

Other methodological challenges include training the sensuality classifier. Since sensuality is a subjective concept involving multisensory surfaces, curves, and textures that evoke femininity and corporeality (Pritchard & Morgan, 2011), we relied on visual identifiers such as makeup, skin exposure, and emotion. Similarly, we relied on light exposure, lighting quality, image background, and angle to identify professional and homemade photographs. We also distinguished selfies from other portrait photos by identifying images taken with the camera held at arm's length, as opposed to those taken by using a self-timer, tripod, or a remote trigger. Big close-ups were identified whenever most of the photographic space was occupied by the subject's face from forehead to chin, in contrast to regular close-ups where the face of the subject appears down to the shoulders. Similarly, long shots feature the full body of the subject in contrast to mid-shots portraying the subject from the waist up to include the head and partial torso. The resulting Keras model successfully identified the relevant categories for this study, namely race, gender, emotion, sensuality, number of faces, professional or amateur photo, indoor or outdoor, angle, and framing of the photograph.

INSERT FIGURE 1 HERE

Google Vision API can also process the data to yield similar results, as it provides information about over and under-exposure (underExposedLikelihood) and blur (blurredLikelihood) in the images, which is a proxy for professional and amateur photos. It further identifies 'safe search' and image properties, in addition to face, label, logo, text, and landmark detection. The likelihood ratings provided by the API

are expressed in 6 different values: ‘unknown,’ ‘very unlikely,’ ‘unlikely,’ ‘possible,’ ‘likely,’ and ‘very likely.’ These values were converted to a discrete scale of NA, -10, -5, 0, 5, and 10, with Figure 1 showing the Vision API results for profile images displaying the following emotions: ‘joy,’ ‘anger,’ ‘sorrow,’ and ‘surprise.’ Vision API further estimates the incidence of adult and sensual content with the parameters ‘adult’ and ‘racy,’ which we used to estimate the incidence of images with explicit or implicit sensual overtones, with Figure 2 displaying the incidence of such images in influence operations across the Americas, Middle East, Asia, and Europe.

INSERT FIGURE 2 HERE

In the end, Google Vision API showed a higher accuracy (.87) compared with Teachable Machine (.72). While Teachable Machine had a notably higher recall rate of .80, its precision was significantly lower at .35 compared with Google Vision API’s .65. This suggests that while Teachable Machine was better at correctly identifying sensual images, it also misclassified a larger number of non-sensual images. The F1 Score, which balances precision and recall, was slightly higher for Vision API (.55) than for Teachable Machine (.49). In terms of specificity, Vision API outperformed with a score of .94 against Teachable Machine’s .71, indicating superior ability to correctly identify non-sensual images. The two models are nonetheless broadly in line with respect to key variables used to model the data, with Figure 3 showing the correlation matrix for Vision API and the Keras classifier (Teachable Machine) that supports the results presented in the next section.

INSERT FIGURE 3 HERE

Results

Compositional tropes

We begin by inspecting the distribution of neutral images across 57 influence operations included in the TMRC database. We found the use of neutral images to be higher in TMRC14_AMERICAS_3, a campaign that includes 8920 tweets posted in Spanish where neutral angles account for 97% of the

data. Low angles, on the other hand, prevailed in TMRC14_APAC_1 in Indonesia with a whopping 363,531 tweets, and TMRC14_EUR_5, which targeted Russians and includes 527,152 tweets. These campaigns feature a higher proportion of low-angle shots at 6% and 5%, respectively. In contrast to that, high-angle shots prevail in TMRC14_APAC_3, which targeted Pakistani audiences with 4,418,374 tweets, but also in influence operations targeting Venezuela and Iran. These compositional choices define how fake social media profiles should appear to unwitting users, with the other salient dimension of the photo composition being the shot frame.

Shot frames are clearly segmented by gender across the 57 TMRC influence operations. Women use more selfies (38%) and big close-ups (22%) than long shots (7%). While men also used selfies, the use of mid-shots is considerably higher at 28%. Selfies are also associated with race, as more than half of the images of Blacks and Latinos/Hispanics are selfies and big close-ups. In contrast to that, the use of close-ups and mid-shots is balanced across races averaging 16% and 24%, respectively. Selfies are also associated with the sensuality category, with more than one-third of all sensual images turning out to be selfies. Non-sensual images, on the other hand, make use of mid-shots and long shots at around 26% on average for each type. Similarly, nearly one-third of amateur images are selfies, and professional images rarely feature big close-ups (only 6%), favoring instead long shots or mid-shots (33% and 35%, respectively).

The use of selfies is comparatively higher in campaigns manned by the IRA and the influence operation of 2019 in Venezuela (33% and 43%, respectively). Similarly, long shots are employed more consistently in the Indonesian campaign (TMRC14_APAC_1). Big close-ups, on the other hand, are prevalent in campaigns that targeted Guatemala, Honduras, and Belize (TMRC14_AMERICAS_1 and TMRC14_AMERICAS_3) where they account for around 44% of the images. These campaigns also feature high levels of sensual undertones, with compositional choices that reinforce traditional gender roles where women are depicted as passive, to be admired or desired, and men are active, the admirers, or the subject that desires (Davis, 2018; Rose et al., 2012).

Twitter profiles targeting regions in the Middle East featured a much higher incidence of illustrations, with 77% of the Twitter profiles targeting the Iranian population (TMRC14_MENA_2) consisting of illustrations and cartoons. The prominence of illustrations suggests an attempt to mask the user identity or to use symbols that resonate with specific ideologies or groups, conceivably to facilitate infiltration. This is in line with propaganda that leverages symbolism and imagery to rally the target population, as many of the illustrations in the TMRC database employ the use of flags and icons overlaid by text to convey ideological talking points. Flags are traditionally used to elicit heightened national identity and ethnic sentiments by driving allegiances to the flag, but they can also be divisive in contexts marked by regional conflicts (Muldoon et al., 2020).

The clenched fist is a recurrent trope in profiles featuring illustrations. It is commonly used to convey unity and solidarity, but it may convey alternative readings when coupled with other symbols (Spierings, 2021). This is particularly the case where the clenched fist appears alongside weaponry to convey retribution and retaliation. The use of captioning, however, far exceeds the boundaries of flags and clenched fists and accounts for a large proportion of images that blur the lines between memes and traditional propaganda tropes, with the grammar of memes driving these compositions. Indeed, the optical character recognition (OCR) applied to the database shows the co-occurrence of images with text (captioning), animals, animations, and drawings, with the variable 'spoof' in Vision API effectively identifying memes in the data. Figure 4a shows the President of Turkey Recep Tayyip Erdoğan and the caption 'Shoulder to shoulder we are always together,' followed by the Russian angel and the caption 'Hello, Khokhly (derogatory Russian term for Ukrainians),' and Hugo Chávez, former President of Venezuela, and the caption 'Two eras. Two warriors. One fight.' Figure 4b shows a similar set of illustrations targeting American users.

INSERT FIGURE 4 HERE

Consistent with previous research (Bastos et al., 2023), sensual images of young women are significantly more likely to be deployed in IRA operations targeting American audiences, but also on

Russian domestic propaganda. This pattern was also observed in influence operations that targeted Venezuelan domestic audiences. Sensual images are also significantly more likely to be deployed in the campaigns TMRC15_APAC_1 and TMRC15_APAC_3, reportedly orchestrated by China, but also in Central America, with the TMRC14_AMERICAS_1, reportedly orchestrated by Russia, and the campaigns TMRC14_AMERICAS_2 and TMRC14_AMERICAS_3 that targeted Guatemalan audiences produced in Guatemala, Mexico, and Russia. They also feature prominently in TMRC14_EUR_5, a Russian influence operation targeting Ukraine. Figure 5 shows a composite of campaigns targeting Turkish, Pakistani, Saudi, Israeli, and Venezuelan audiences. The Keras and Vision API models are relatively consistent with each other with respect to sensuality (adjusted R-square = .38).

INSERT FIGURE 5

The operations manned by the IRA are marked by extensive use of 'happy' and 'joyful' headshots, particularly in the US campaign (joy=.67) compared with domestic operations in Russia (joy=.10). A similar pattern was observed in the operations targeting Venezuelan audiences in this period (joy=.13 and .09), and in the operations in Central America where 'joy' stands at .27 and .32, respectively. The results of the Vision API are consistent with the Keras classifier, though the estimates of the latter are lower likely due to the limited precision of the TensorFlow algorithm for this category. The visual trope of attractive young women is particularly salient in influence operations carried out by China (TMRC15_APAC_3) that targeted the United States in April-October 2022. This cohort included nearly two thousand accounts that posted over 300 thousand messages in English and Chinese, with prominent hashtags including #China, #VladimirPutin, and #AmazingChina. Figure 6 shows a composite of profile photos of this campaign that follows the IRA playbook of exploiting profile photos of attractive young women.

INSERT FIGURE 6 HERE

IRA operations targeting US audiences are more likely to rely on homemade photos compared with other campaigns in the period (r=.64 against .001 to .01 for campaigns targeting Bangladesh, Catalonia, Iran, Russia, and Venezuela). This compositional pattern emerged again in 2022, when a Russian campaign targeting Indonesian audiences resorted primarily to amateurish profile photos (r=.66) compared with a range from .001 to .06 for similar campaigns in 2022 that targeted 13 countries. These differences are also significant for the racial tropes explored by the IRA, with Whites being particularly more prominent in the campaign targeting US audiences in the run-up to the Presidential election of 2016 (r=.12) than in any other campaign that ensued (r<.05).

Another recurrent compositional trope is found in the influence operation carried out by Indonesia between April-November 2021 targeting the local population (TMRC14_APAC_1), with a set of 5914 accounts that posted nearly 400 thousand messages in Indonesian (with prominent hashtags including #papua, #BinmasNokenPolri [Binmas Noken Police], and #IndonesiaNegaraHukum [Indonesian National Law]). The visual identity of these accounts is more likely to feature indoor profile pictures taken from high or low angles in long or mid-shot frames compared with the rest of the influence operations in the past decade. They are also more likely to present professionally made photos of individuals whose face is covered or shown only partially. It features fewer group photos or illustrations, with the lion's share of profiles featuring a single individual usually of East Asian complexion. Much like the IRA operations of 2016, these profile pictures shown in Figure 7 are also more likely to feature unassuming males with K-pop aesthetics. This domestic influence operation strongly resembles the playbook devised by the IRA notwithstanding the different ethnic and cultural groups they targeted.

INSERT FIGURE 7 HERE

Hypothesis testing

We proceed to test hypothesis (H1) positing that the visual parameters of Twitter accounts are predictive of influence operations, with user accounts as the unit of analysis. Our results show that the visual parameters of Twitter accounts are indeed predictive of influence operations. The model based on Vision API was capable of accounting for over one-quarter of the variance in the data when the variable user_reported_location was incorporated into the model. This is to be expected because user_reported_location is a proxy for the target population of the campaign and is therefore predictive of the frames employed in each influence operation. If user_reported_location is removed, the model accounts for less than 10% of the variance in the data, a moderate result given the granular visual information feeding the model.

These models are nonetheless based on the visual dimensions of the data alone, and the model underperforms when numeric and ordinal data from Twitter accounts are included in the model (e.g., the number of tweets, followers, account creation date, language, etc.). In other words, the visual information gleaned from Twitter profiles outperforms the numeric and textual data employed to identify sockpuppet accounts, bots, trolls, and scammers (adjusted R-squared = .2421 compared with .1963, $p < .0001$ for both models). Table 2 summarizes the key findings regarding the visual identities of fake social media profiles in different geographical locations.

INSERT TABLE 2 HERE

The Keras model based on race, composition, angle, quality, indoor/outdoor, number of individuals, and gender also proved moderately powerful by accounting for nearly one-fifth of the variance in the data (adjusted R-squared = .1928, $p < 0.0001$). Visual cues like race and sensuality are particularly effective at identifying influence operations across a range of sources and targets. Similar to the results drawn from the Vision API-based model, incorporating numeric and ordinal data from user profiles such as the number of tweets, followers, and language did not improve the model, with

data extracted from profile images outperforming established metrics for bot detection across the tests performed for this study.

Given the robustness of these results, we accept hypothesis H1 and conclude that the visual dimension of social media accounts offers an effective parameter for identifying influence operations at scale. While visual information outperforms metrics for the identification of bots and trolls, these attributes can be combined in early-warning systems. Indeed, user account information proved useful in modeling campaign targets (as opposed to source attribution). While visual parameters like adult content, incidence of faces expressing joy or sorrow, coded labels, and sensual content continue to outperform typical user account metrics, the combination of these variables renders a model that can account for over one-fourth of the variance in the data (adjusted R-squared = .2765, $p < 0.0001$).

We proceed to hypothesis H2, which posits that the visual identities of Twitter profiles are predictive of campaign source (unit of analysis: influence operations) by combining the Keras and Vision API classifiers and aggregating the data at the campaign level. The model includes all variables in the database, but no numeric or categorical variable from individual Twitter accounts such as the number of tweets or followers. The model proved quite capable, with multiple R-squared of .91 and adjusted R-squared of .5351. However, these results are not significant at the 5% confidence interval ($p = .136$), and therefore we reject hypothesis H2 and conclude that aggregate visual data must be disaggregated at lower levels to yield significant insights. There are, however, important correlations that can be identified at the campaign level. The Pearson R for the correlation between sensuality and female, and between sensuality and selfie is .75 and .80, respectively ($p < 0.01$). Other significant correlations exist between each campaign and the prevailing framing, quality of the photo, and other compositional features like the incidence of photos outdoors as shown in Figure 3 and Table 2.

Conclusion

The TMRC database offers comprehensive information about Twitter user accounts used in influence operations. This information was originally provided to the US Senate Committee during the Congressional hearings of 2017 when Twitter, Meta, and Alphabet shared information with the US Senate and Congress. This data, unfortunately, is rarely shared with the academic community. While Twitter has offered researchers access to the TMRC database, Meta is yet to share similar data with the academic community. Since Twitter's acquisition by Elon Musk, however, the TMRC data has been deleted and the information about Twitter's Election Integrity initiative is no longer available on the company's website.

Our results show that sensuality is a central trope in foreign influence operations that replicate the seminal campaign of the IRA targeting American audiences, being particularly prevalent in campaigns targeting Latin Americans, but also Turkish, Spanish, and Israeli audiences. We also found that the visual attributes of social media profiles are predictive of influence operations because these campaigns are segmented into groups with clear visual presentations that emphasize selfies, sensual young women, K-pop aesthetics, or alternatively flags and icons overlaid with text to convey ideological talking points.

The prevalence of selfies and close-ups, especially among women and certain racial groups, is noteworthy. Selfies in contemporary culture are often associated with authenticity and personal expression and propagandists exploit this perception to manufacture relatable and verisimilar social media profiles. Propaganda selfies appear designed to portray attractive individuals and maximize one's social media reach. This explains the much higher incidence of selfies in profile photos marked by heightened sensual undertones in the TMRC. Selfies were also found to mostly employ high angles portraying the subject as submissive and inviting, typically in the profile of women, whose images feature selfies at higher rates than any other type of photo.

Lastly, the high incidence of selfies and big close-ups for Blacks and Latinos/Hispanics resonates with the intimacy often associated with contact cultures. Conversely, the preference for long shots in the TMRC14_APAC_1 campaign, which targeted Twitter users in Indonesia, is in line with the noncontact cultural norms of many Asian societies, where personal space and formality are emphasized. The composition of such fake profiles is therefore attuned to the cultural perceptions and norms of their target audience, a considerable challenge for foreign influence operations seeking to manufacture trust and relatability leading to successful infiltration. Finally, these compositional tropes are no longer restricted to political propaganda, as the cottage industry of fraud and romance scams has promptly adopted this playbook to craft desirable personas, particularly attractive young women who appear wealthy and educated.

Funding Acknowledgments

This material is based upon work supported by the Google Cloud Research Credits program with the award GCP19980904, research grant #270353197 “The Visual Frames of Social Media Propaganda,” and Twitter, Inc. research grant 50069SS “The Brexit Value Space and the Geography of Online Echo Chambers.” The author also acknowledges support from the University College Dublin and OBRSS scheme (grants R21650 and R20825) and the National Council for Scientific and Technological Development (grant 406504/2022-9).

Acknowledgments

The corresponding author is thankful to Noel George, Azhar Sham, and Thanvi Ajith for their important contributions to this project.

References

Aubrey, J. S., & Frisby, C. M. (2011). Sexual objectification in music videos: A content analysis comparing gender and genre. *Mass Communication and Society*, *14*(4), 475-501.

Bastos, M. T., & Mercea, D. (2018). The public accountability of social platforms: lessons from a study on bots and trolls in the Brexit campaign. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. https://doi.org/10.1098/rsta.2018.0003

Bastos, M. T., & Mercea, D. (2019). The Brexit Botnet and User-Generated Hyperpartisan News. *Social Science Computer Review*, *37*(1), 38-54. https://doi.org/10.1177/0894439317734157

Bastos, M. T., Mercea, D., & Goveia, F. (2023). Guy Next Door and Implausibly Attractive Young Women: The Visual Frames of Social Media Propaganda. *New Media & Society*, *25*(8), 2014-2033.

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford University Press.

Bonilla-Silva, E. (1997). Rethinking racism: Toward a structural interpretation. *American Sociological Review*, 465-480.

Brader, T., Valentino, N. A., & Suhay, E. (2008). What triggers public opposition to immigration? Anxiety, group cues, and immigration threat. *American Journal of Political Science*, *52*(4), 959-978.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Conference on fairness, accountability and transparency,

Butler, J. (2002). *Gender trouble*. Routledge.

Carpinella, C. M., Hehman, E., Freeman, J. B., & Johnson, K. L. (2016). The gendered face of partisan politics: Consequences of facial sex typicality for vote choice. *Political Communication*, *33*(1), 21-38.

Celli, F., Bruni, E., & Lepri, B. (2014). Automatic personality and interaction style recognition from facebook profile pictures. Proceedings of the 22nd ACM international conference on Multimedia,

Crenshaw, K. (2013). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. In K. Maschke (Ed.), *Feminist legal theories* (pp. 23-51). Routledge.

Davis, S. E. (2018). Objectification, Sexualization, and Misrepresentation: Social Media and the College Experience. *Social Media + Society*, *4*(3), 2056305118786727. https://doi.org/10.1177/2056305118786727

De Vries, T., Misra, I., Wang, C., & Van der Maaten, L. (2019). Does object recognition work for everyone? Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops,

Elections Integrity. (2018). *Data archive* http://web.archive.org/web/20181019093120/https://about.twitter.com/en_us/values/elections-integrity.html

Community Standards, (2018a). https://www.facebook.com/help/975828035803295

Understanding the Facebook: Community Standards Enforcement Report, (2018b). https://fbnewsroomus.files.wordpress.com/2018/05/understanding_the_community_standards_enforcement_report.pdf

Faux, Z. (2023). *Number Go Up: Inside Crypto's Wild Rise and Staggering Fall*. Penguin Random House.

Ferwerda, B., & Tkalcic, M. (2018). Predicting users' personality from Instagram pictures: using visual and/or content features? Proceedings of the 26th conference on user modeling, adaptation and personalization,

Freelon, D., Bossetta, M., Wells, C., Lukito, J., Xia, Y., & Adams, K. (2020). Black Trolls Matter: Racial and Ideological Asymmetries in Social Media Disinformation. *Social Science Computer Review*, 0894439320914853.

Frost, P. (1990). Fair women, dark men: The forgotten roots of colour prejudice. *History of European ideas*, *12*(5), 669-679.

George, N., Sham, A., Ajith, T., & Bastos, M. T. (2024). Forty Thousand Fake Twitter Profiles: A Computational Framework for the Visual Analysis of Social Media Propaganda. *Social Science Computer Review*, 08944393241269394.

Harvey, D., & Roth, Y. (2018). An update on our elections integrity work. https://web.archive.org/web/20210624004136/https://blog.twitter.com/en_us/topics/company/2018/an-update-on-our-elections-integrity-work

Heathwood, C. (2006). Desire satisfactionism and hedonism. *Philosophical Studies*, *128*, 539-563.

Heathwood, C. (2007). The reduction of sensory pleasure to desire. *Philosophical Studies*, *133*, 23-44.

Highfield, T., & Leaver, T. (2016). Instagrammatics and digital methods: Studying visual social media, from selfies and GIFs to memes and emoji. *Communication Research and Practice*, *2*(1), 47-62.

Hofmeester, G. H., Kemp, J. A. M., & Blankendaal, A. C. M. (1996). *Sensuality in product design: a structured approach* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, British Columbia, Canada.

Kanade, T. (1977). *Computer recognition of human faces* (Vol. 47). Birkhäuser Verlag.

Kraft, R. N. (1987). The influence of camera angle on comprehension and retention of pictorial events. *Memory & cognition*, *15*, 291-307.

Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. Proceedings of the IEEE international conference on computer vision,

Merkt, M., Weingärtner, A.-L., & Schwan, S. (2022). Digital images are hard to resist: Teaching viewers about the effects of camera angle does not reduce the camera angle's impact on power judgments. *Acta Psychologica*, *229*, 103687.

Meyers-Levy, J., & Peracchio, L. A. (1992). Getting an angle in advertising: The effect of camera angle on product evaluations. *Journal of Marketing Research*, *29*(4), 454-461.

Morris, M. R., Counts, S., Roseway, A., Hoff, A., & Schwarz, J. (2012). Tweeting is believing?: understanding microblog credibility perceptions. Proceedings of the ACM 2012 conference on computer supported cooperative work,

Muldoon, O. T., Trew, K., & Devine, P. (2020). Flagging difference: Identification and emotional responses to national flags. *Journal of Applied Social Psychology*, *50*(5), 265-275.

Muscanell, N. L., & Guadagno, R. E. (2012). Make new friends or keep the old: Gender and personality differences in social networking use. *Computers in Human Behavior*, *28*(1), 107-112.

Pearce, W., Özkula, S. M., Greene, A. K., Teeling, L., Bansard, J. S., Omena, J. J., & Rabello, E. T. (2020). Visual cross-platform analysis: Digital methods to research social media images. *Information, Communication & Society*, *23*(2), 161-180.

Peng, Y., Lock, I., & Ali Salah, A. (2024). Automated visual analysis for the study of social media effects: Opportunities, approaches, and challenges. *Communication Methods and Measures*, *18*(2), 163-185.

Pritchard, A., & Morgan, N. (2011). Tourist Bodies, Transformation and Sensuality. In P. Bramham & S. Wagg (Eds.), *The New Politics of Leisure and Pleasure* (pp. 153-168). Palgrave Macmillan. https://doi.org/10.1057/9780230299979_10

Ringrow, H. (2016). *The language of cosmetics advertising*. Springer.

Rose, J., Mackey-Kallis, S., Shyles, L., Barry, K., Biagini, D., Hart, C., & Jack, L. (2012). Face it: The impact of gender on social media images. *Communication Quarterly*, *60*(5), 588-607.

Seo, H. (2014). Visual propaganda in the age of social media: An empirical analysis of Twitter images during the 2012 Israeli–Hamas conflict. *Visual Communication Quarterly*, *21*(3), 150-161.

Smedley, A., & Smedley, B. D. (2005). Race as biology is fiction, racism as a social problem is real: Anthropological and historical perspectives on the social construction of race. *American psychologist*, *60*(1), 16–26. https://doi.org/10.1037/0003-066X.60.1.16

Spierings, T. (2021). *The power of symbols in visual propaganda: the meaning behind political logos and flags* Escola Superior de Artes e Design]. Matosinhos, Portugal.

Starr, R. L., Wang, T., & Go, C. (2020). Sexuality vs. sensuality: The multimodal construction of affective stance in Chinese ASMR performances. *Journal of Sociolinguistics*, *24*(4), 492-513.

Tajfel, H., & Turner, J. C. (1986). The Social Identity Theory of Intergroup Behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of Intergroup Relation* (pp. 7-24). Nelson-Hall Publishers.

TMRC. (2022). Overview. https://web.archive.org/web/20220929100519/https://transparency.twitter.com/en/reports/moderation-research.html

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623-1626.

Toma, C. L., & Hancock, J. T. (2012). What lies beneath: The linguistic traces of deception in online dating profiles. *Journal of Communication*, *62*(1), 78-97.

Twitter Privacy Policy, (2018). http://web.archive.org/web/20200331024923/https://twitter.com/en/privacy

Twitter. (2019). *Election integrity policy*. https://web.archive.org/web/20190428071045/https://help.twitter.com/en/rules-and-policies/election-integrity-policy

Twitter. (2021). *Civic integrity*. https://web.archive.org/web/20210130191423/https://about.twitter.com/en/our-priorities/civic-integrity

Weikmann, T., & Lecheler, S. (2023). Visual disinformation in a digital age: A literature synthesis and research agenda. *New Media & Society*, *25*(12), 3696-3713.

West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender & society*, *1*(2), 125-151.

Wolf, N. (2013). *The beauty myth: How images of beauty are used against women*. Random House.

Xia, Y., Lukito, J., Zhang, Y., Wells, C., Kim, S. J., & Tong, C. (2019). Disinformation, performed: self-presentation of a Russian IRA account on Twitter. *Information, Communication & Society*, *22*(11), 1646-1664. https://doi.org/10.1080/1369118X.2019.1621921

Xie, Q., & Zhang, M. (2013). White or tan? A cross-cultural analysis of skin beauty advertisements between China and the United States. *Asian Journal of Communication*, *23*(5), 538-554.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and personality psychology compass*, *2*(3), 1497-1517.

Table 1: Election Integrity and TMRC data releases with number of Twitter users and source attribution

Release	Date	Users	Source	Release	Date	Users	Source
El. Integrity	Oct 2018	3613	Russa (IRA)	**El. Integrity**	Feb 2021	69	Russia (GRU)
El. Integrity	Oct 2018	770	Iran	**El. Integrity**	Feb 2021	238	Iran
El. Integrity	Jan 2019	15	Bangladesh	**El. Integrity**	Feb 2021	35	Armenia
El. Integrity	Jan 2019	2320	Iran	**El. Integrity**	Dec 2021	112	China (Changyu)
El. Integrity	Jan 2019	416	Russia (IRA)	**El. Integrity**	Dec 2021	2048	China (Xinjiang)
El. Integrity	Jan 2019	1960	Venezuela	**El. Integrity**	Dec 2021	276	Mexico
El. Integrity	Mar 2019	4248	United Arab Emirates	**El. Integrity**	Dec 2021	16	Russia; East Africa
El. Integrity	Apr 2019	1019	Ecuador	**El. Integrity**	Dec 2021	50	Russia; North Africa
El. Integrity	Apr 2019	6	Saudi Arabia	**El. Integrity**	Dec 2021	268	Tanzania
El. Integrity	Apr 2019	259	Spain	**El. Integrity**	Dec 2021	277	Venezuela
El. Integrity	Apr 2019	271	UAE; Egypt	**El. Integrity**	Dec 2021	418	Uganda
El. Integrity	Jun 2019	4779	Iran	**TMRC 14 Americas1**	Aug 2022	1379	Russia; USA; Mexico
El. Integrity	Jun 2019	130	Catalonia	**TMRC 14 Americas2**	Aug 2022	249	Guatemala; Mexico
El. Integrity	Jun 2019	4	Russia (IRA)	**TMRC 14 Americas3**	Aug 2022	1780	Russia; Ukraine
El. Integrity	Jun 2019	33	Venezuela	**TMRC 14 Americas4**	Aug 2022	170	USA; UK
El. Integrity	Jul 2019	5241	China	**TMRC 14 APAC 1**	Aug 2022	5914	Indonesia
El. Integrity	Dec 2019	5929	Saudi Arabia	**TMRC 14 APAC 2**	Aug 2022	1198	India; USA
El. Integrity	Mar 2020	71	Ghana; Nigeria	**TMRC 14 APAC 3**	Aug 2022	568	Pakistan
El. Integrity	Apr 2020	2541	Egypt	**TMRC 14 EUR 1**	Aug 2022	1889	Turkey
El. Integrity	Apr 2020	3104	Honduras	**TMRC 14 EUR 2**	Aug 2022	228	Turkey
El. Integrity	Apr 2020	795	Indonesia	**TMRC 14 EUR 3**	Aug 2022	7	Slovenia
El. Integrity	Apr 2020	8558	Serbia	**TMRC 14 EUR 4**	Aug 2022	38	Russia; Ukraine; USA
El. Integrity	Apr 2020	5350	Saudi Arabia; Egypt; United Arab Emirates	**TMRC 14 EUR 5**	Aug 2022	2648	Russia; Philippines; Ukraine; Turkey; India; Bulgaria
El. Integrity	May 2020	23750	China	**TMRC 14 MENA 1**	Aug 2022	504	Saudi Arabia
El. Integrity	May 2020	1152	Russia	**TMRC 14 MENA 2**	Aug 2022	608	UK; Germany; France; Iran; Netherlands; USA
El. Integrity	May 2020	7340	Turkey	**TMRC 14 Africa 1**	Aug 2022	228	Sudan
El. Integrity	Oct 2020	34	Saudi Arabia	**TMRC 15 APAC 1**	Oct 2022	3	China; United States
El. Integrity	Oct 2020	526	Cuba	**TMRC 15 APAC 2**	Oct 2022	22	China; United States
El. Integrity	Oct 2020	5	Russia	**TMRC 15 APAC 3**	Oct 2022	1943	USA; Hong Kong; Singapore; China
El. Integrity	Oct 2020	104	Iran	**TMRC 15 MENA 1**	Oct 2022	37	Iran
El. Integrity	Oct 2020	926	Thailand	**TMRC 15 MENA 2**	Oct 2022	7	Israel; USA; Iran
El. Integrity	Feb 2021	31	Russia (IRA)	**TMRC 15 MENA 3**	Oct 2022	5	Iran

Table 2: Summary statistics of the variables in the model

Source	Target	Tweets	Sensual	Sexy	Male	Female	Angle	Quality	Outdoor	Selfie	Closeup	Midshot	Longshot	Pic
Bangladesh	Bangladesh	37486	.10	.00	.38	.00	.25	.50	.63	.13	.13	.00	.00	.63
China	USA	566043	.48	.28	.33	.43	.16	.46	.25	.25	.15	.14	.08	.25
Great Britain	Iran	24311	.13	.02	.21	.02	.09	.79	.16	.05	.04	.07	.04	.77
Guatemala	Guatemala	271873	.38	.22	.37	.32	.13	.45	.26	.15	.08	.15	.10	.29
India	India	340356	.25	.11	.19	.18	.13	.60	.32	.09	.06	.14	.11	.54
Indonesia	Indonesia	363531	.40	.21	.21	.05	-.04	.63	.37	.08	.03	.28	.42	.17
Iran	Iran	14M	.24	.14	.37	.21	.20	.52	.30	.14	.14	.16	.09	.37
Iran	USA	115242	.30	.18	.19	.28	.03	.59	.14	.18	.18	.08	.00	.49
Israel	Israel	24579	.29	.06	.19	.31	.31	.50	.19	.06	.13	.25	.06	.44
Pakistan	Pakistan	4M	.33	.19	.33	.21	.14	.58	.34	.10	.09	.23	.13	.38
Russia	Honduras	43299	.22	.09	.69	.20	.18	.33	.18	.26	.10	.15	.08	.07
Russia	Guatemala	8920	.22	.10	.49	.33	.20	.30	.25	.17	.07	.09	.06	.15
Russia	USA	9M	.47	.31	.53	.31	.15	.45	.26	.33	.13	.23	.10	.13
Russia	Russia	2M	.24	.11	.30	.18	.16	.36	.50	.13	.08	.14	.06	.51
Saudi Arabia	Saudi Arabia	3M	.23	.09	.43	.09	.28	.66	.22	.07	.14	.18	.08	.49
Slovenia	Slovenia	47754	.39	.14	.14	.14	.14	.43	.29	.00	.00	.14	.14	.57
Spain	Spain	13065	.18	.08	.21	.03	-.04	.55	.28	.01	.05	.13	.10	.65
Sudan	Somalia	44326	.18	.06	.28	.02	.02	.64	.22	.06	.06	.13	.07	.66
Turkey	Turkey	13M	.26	.14	.42	.21	.02	.60	.29	.17	.10	.21	.13	.33
USA	Russia	402951	.15	.03	.10	.08	.14	.65	.24	.04	.01	.06	.04	.78
Venezuela	Venezuela	17M	.57	.41	.33	.30	.20	.46	.19	.25	.10	.15	.10	.35

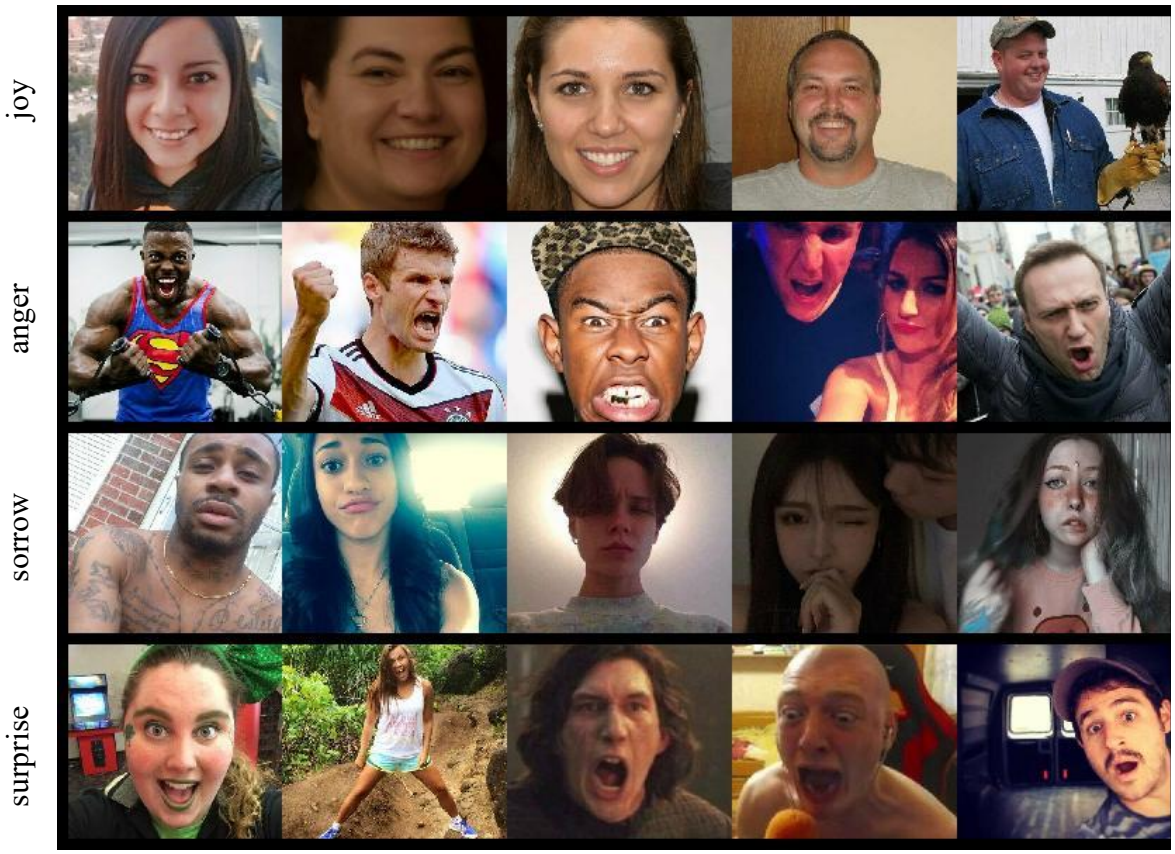

Figure 1: Vision API results for profile images displaying ‘joy,’ ‘anger,’ ‘sorrow,’ and ‘surprise.’

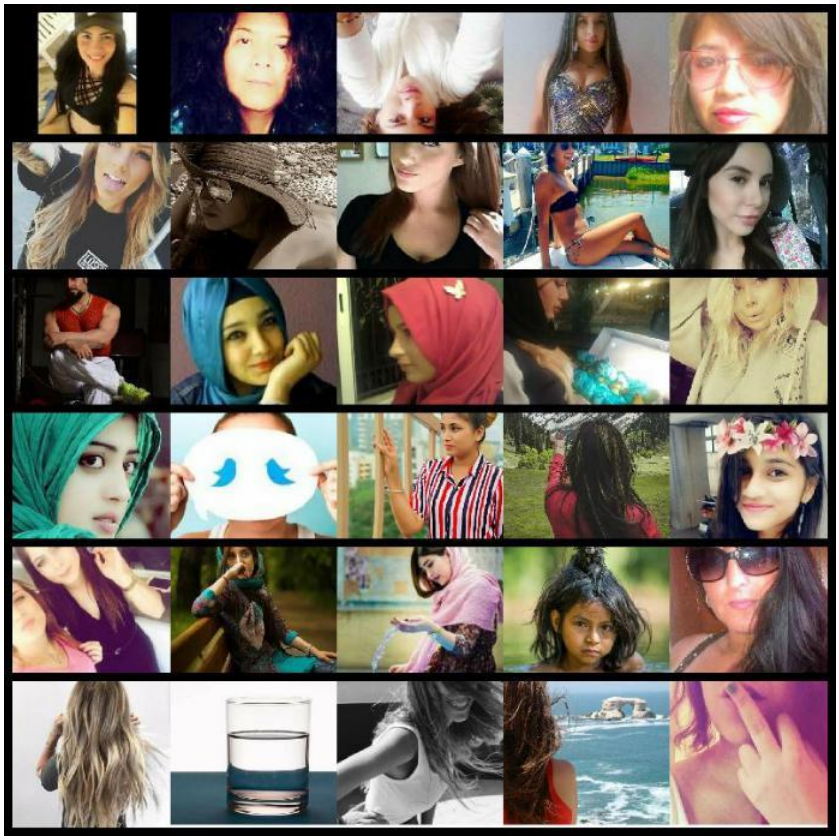

Figure 2: Composite of user profiles used in campaigns targeting Guatemalan, North American, Iranian, Pakistani, and Catalan audiences.

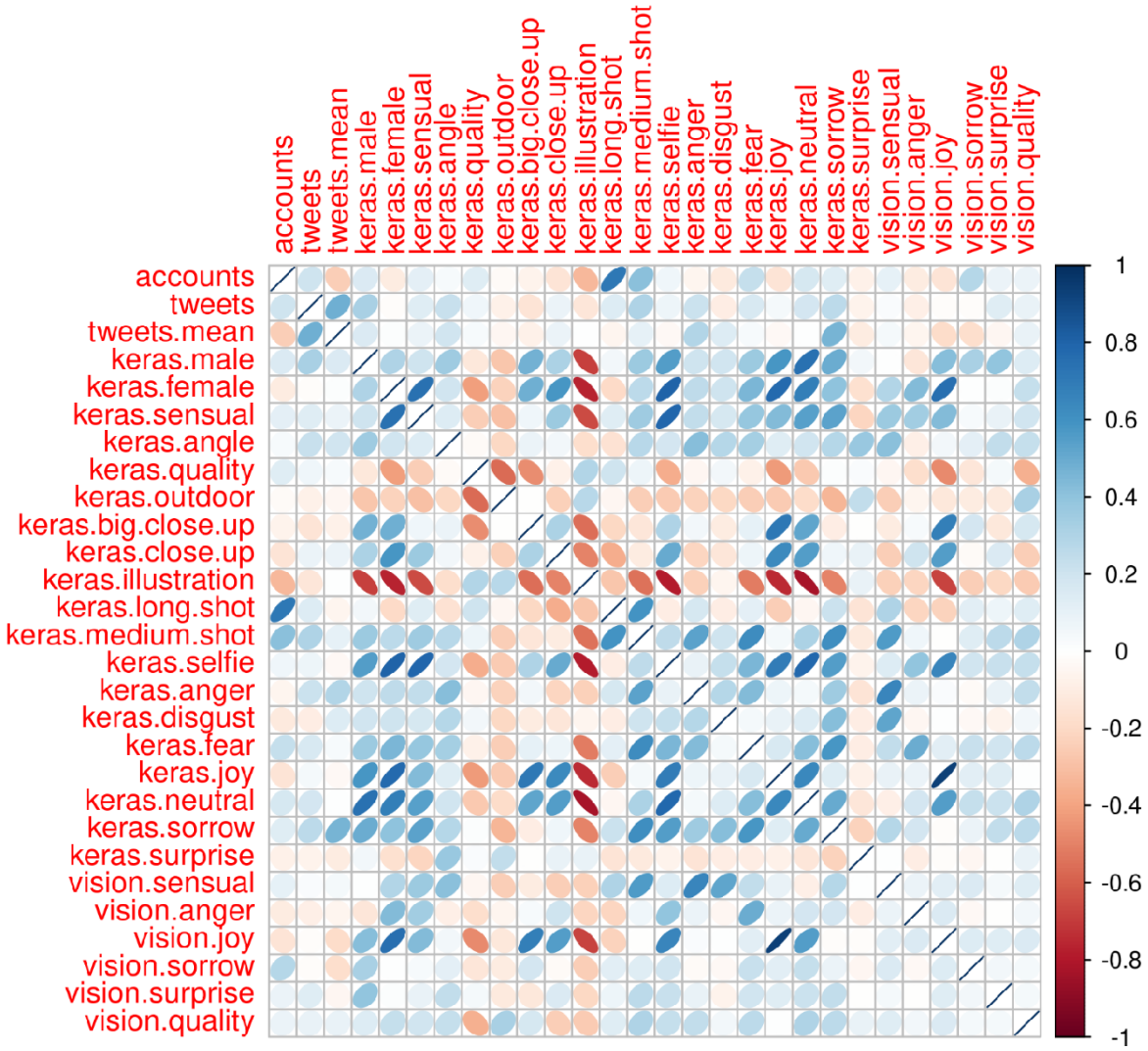

Figure 3: Correlation matrix for Vision API and the Keras classifiers.

Figure 4a: President of Turkey Recep Tayyip Erdoğan and the caption 'Shoulder to shoulder we are always together,' followed by the Russian angel and the caption 'Hello, Khokhly (derogatory Russian term for Ukrainians),' and Hugo Chávez, former President of Venezuela, and the caption 'Two eras. Two warriors. One fight.' Figure 4b: visual tropes targeting American users.

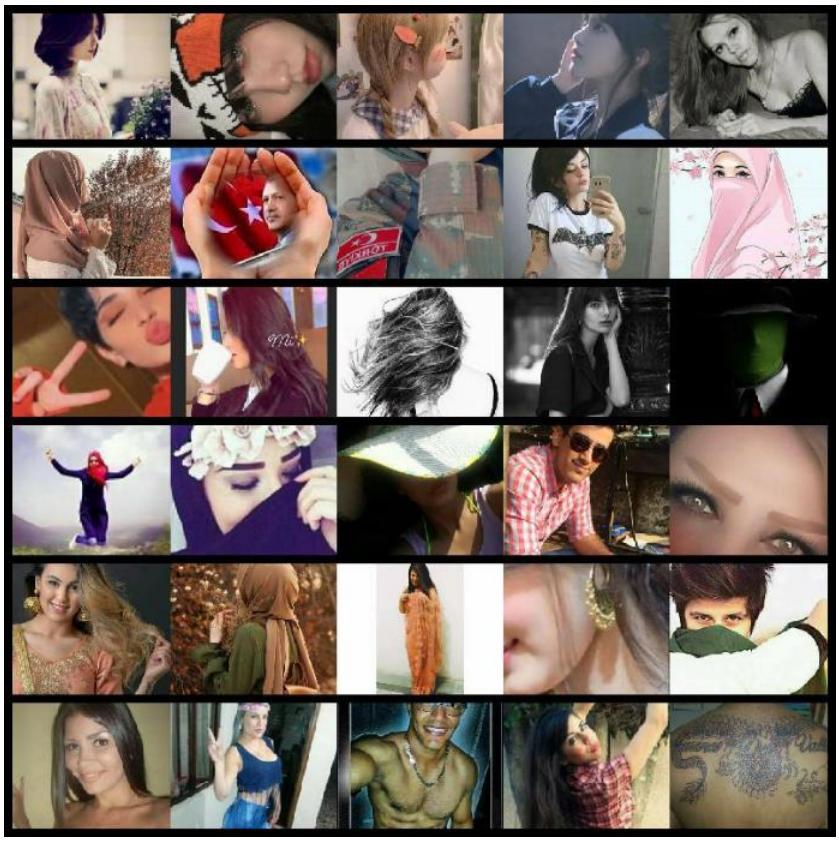

Figure 5: Composite of user profiles used in campaigns targeting Turkish, Pakistani, Saudi, Israeli, and Venezuelan audiences.

Figure 6: Composite of profile photos depicting attractive young women for propaganda.

Figure 7: Composite of profile photos featuring unassuming males with K-pop aesthetics that targeted Indonesian audiences.