Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-024-03801-z OPEN

Perceived gendered self-representation on Tinder using machine learning

Yan Asadchy [1✉], Andres Karjus [1], Ksenia Mukhina [2] & Maximilian Schich [3]

This paper explores the gendered differences between men and women as perceived through the images on the online dating platform Tinder. While personal images on Instagram, Tumblr, and Facebook have been studied en masse, large-scale studies of the landscape of visual representations on online dating platforms remain rare. We apply a machine learning algorithm to 10,680 profile images collected on Tinder in Estonia to study the perceived gendered differences in self-representation among men and women. Beyond identifying the dominant genres of profile pictures used by men and women, we build a comprehensive map of visual self-representation on the platform. We further expand our findings by analyzing the distribution of the image genres across the profile gallery and identifying the prevalent positions for each genre within the profiles. Lastly, we identify the variability of women's and men's images within each genre. Our approach provides a holistic overview of the culture of visual self-representation on the dating app Tinder and invites scholars to expand the research on gendered differences and stereotypes to include cross-platform and cross-cultural analysis.

[1] School of Humanities, Tallinn University, Tallinn, Estonia. [2] School of Digital Technologies, Tallinn University, Tallinn, Estonia. [3] Baltic Film, Media and Arts School, Tallinn University, Tallinn, Estonia. ✉email: yan.asadchy@tlu.ee

Introduction

Social media platforms and apps, including online dating websites, constitute silos for large amounts of user-generated content, where the inherent bigger picture often remains opaque to individual users and also platform vendors and regulators. The availability of large amounts of data on social media platforms allows researchers to illuminate the broader picture of visual culture. This picture is inherently shaped by the context and features of social media platforms, as well as the community of users and the socio-cultural and geographic circumstances surrounding them. Through text, imagery, and video, users of social media platforms aim to balance representativeness, self-disclosure, and anonymity while ensuring they do not hinder their goals and motivations for using the platform. Research in social media studies, more broadly the arts and humanities, and computational social science has highlighted the cultural significance of the depicted human face, i.e., the centerpiece of most online dating profiles. According to sociologist Anthony Giddens, indeed, so-called "facework commitments" (Giddens and Pierson, 1998) are used by human inhabitants of urban environments (Smith and Duggan, 2013), where unknown encounters are widespread to unconsciously exhibit a range of clues and signals to be perceived as trustworthy. This behavior pattern arguably proliferates to online forums, social media platforms, and marketplaces, where the chance to meet a stranger is even higher. Facial clues are used to communicate trustworthiness, for example, in borrowing campaigns, where photo features such as a smile can predict success (Athey et al., 2022). Such insights are essential for services like online dating apps, where anticipated face-to-face interactions limit opportunities for self-expression (Walther, 2007). Balancing between authenticity and the desired self, online daters must choose their self-representation approaches wisely to create a good impression of themselves while not alienating potential partners by misrepresenting themselves in their photographs (Duguay, 2017).

Additionally, how one's identity is presented depends on several factors, naturally varying across and within social media platforms. There is a large amount of research dedicated to interpersonal communication on social media (Papacharissi, 2010; Whitty and Carr, 2017; Wright and Webb, 2011). Affordances of platforms and limitations of technological devices may enable or limit specific modes of self-representation (Marwick, 2013). While the OkCupid platform, for example, allows users to answer long questionnaires, enabling and emphasizing calculated compatibility, the Tinder platform emphasizes visual self-representation by focusing on profile images to prompt and promote quick judgment based on represented physical aesthetic features. Users are also subjects of "idioms of practice" in different platforms (Gershon, 2019), i.e., emergent conventions of user behavior and platform usage. While it is common practice to post curated selfies on Instagram (Tiidenberg and Gómez Cruz, 2015), this custom may not be a characteristic feature of participation on LinkedIn. The modes of self-representation further depend on the platform's perceived audience and perception of the underlying algorithms. The interaction between users and algorithms in dating apps exemplifies a reciprocal dynamic between dating and sexuality, where users integrate perceived algorithmic logic into their everyday experiences (Pidoux, 2022). Participants of online dating websites make conscious decisions about how to present themselves to potential mates and what information to conceal from unwanted attention (Ellison et al., 2006). This is particularly important for minorities, who want to participate in online dating but may be compelled to cover their face on profile pictures to preserve anonymity (Blackwell et al., 2015). There is a growing body of research across social media studies, digital humanities, and computational social science addressing such issues with a focus on visual media (Leszczynski, 2019; Rose, 2022; Dewdney, 2022). Depending on the research questions, image-driven social media studies and image analysis can be approached from a variety of perspectives, focusing on particular platforms such as Instagram or Flickr, focusing on image genres such as photographs or memes, or focusing on particular visual vernaculars, such as everyday life, war journalism, etc.

Although self-representation on social media is identified as an essential means of social communication in the contemporary world, existing studies are often limited to small scales based on available or chosen methods (Degen and Kleeberg-Niepage, 2023). Such limitation in the method may lead to a limited view and narrowed understanding of relevant phenomena, which are often subject to large amounts and variety of local activity, feeding into the emergence of complex global patterns, i.e., new forms of quality that necessitate cartography and elucidation of the inherent bigger picture. Studies of platforms like Instagram and Tumblr often focus on the "selfie," i.e., a particular genre of photography (Tiidenberg and Whelan, 2017). Selfies alone provide a vast opportunity space for researchers, for example, to study emergent gendered differences, such as the predominance of the left cheek in women's selfies posted on Instagram (Bruno et al., 2015) or differences in facial orientation among heterosexual men and women on Tinder (Sedgewick et al., 2017). Meanwhile, despite their popularity and very distinguishable aesthetic (Leaver et al., 2020), selfies account for only a fraction of images uploaded on these platforms (Caliandro and Graham, 2020; Tifentale and Manovich, 2015). When not analyzed in isolation but taken into account in the context of all other images on a given platform, selfies constitute but a part of a given "style space" (Manovich, 2011) or "meta picture" (Mitchell, 1994), which typically varies from platform to platform, opening novel pathways to establish approaches for critical evaluation and interpretation. For example, such a holistic view may allow us to explore more complex relationships between users through their respective choices and combinations of different photographic styles. At the same time, we must be aware that the practices of users through this lens may not fully capture the users' experience on online dating websites. The "Smart photos" feature, which automatically chooses the best photo for the user (Tinder, 2023a), proximity, online availability, and algorithmic curation (Tong et al., 2016), all create unique individual situations and environments for users to navigate. As such, it is necessary to transcend individual case studies towards a more comprehensive big picture. This is why, in this paper, we explore gendered self-representation, as experienced on Tinder, by collecting a dataset of profile images in a specific geographic region of Estonia over 5 months in 2021.

Gender is a complex concept, and Tinder endorses its users to use all the functionalities to express their true selves (Tinder, 2020). However, when searching for a match on Tinder, the platform prompts its users to show them "men," "women," or "both" in the main view of the app. In Butler's terms, this interface functionality represents "a regulatory regime," forcing users to align with the available options (Butler, 2002). This limitation creates a ground for the emergence of a "prepackaged" understanding of gendered self-representation, which, despite being bound to the interface, invites individuals to creativity in self-expression (Sundén, 2009). Following this platform affordance, in this study, we use machine learning methods to show how gendered visual self-representation is experienced through this limitation—one of many common gender stereotypes programmatically built-in dating apps (Pidoux et al., 2021)

We reconstruct this complex emergent space of gendered visual self-representation to further contextualize and analyze these pictures for their emergent genres of photography (Research question 1) and how they are present within the profiles of men and women (Research question 2). In this study, we use the machine-learning-driven computational method, the "distant viewing" paradigm of

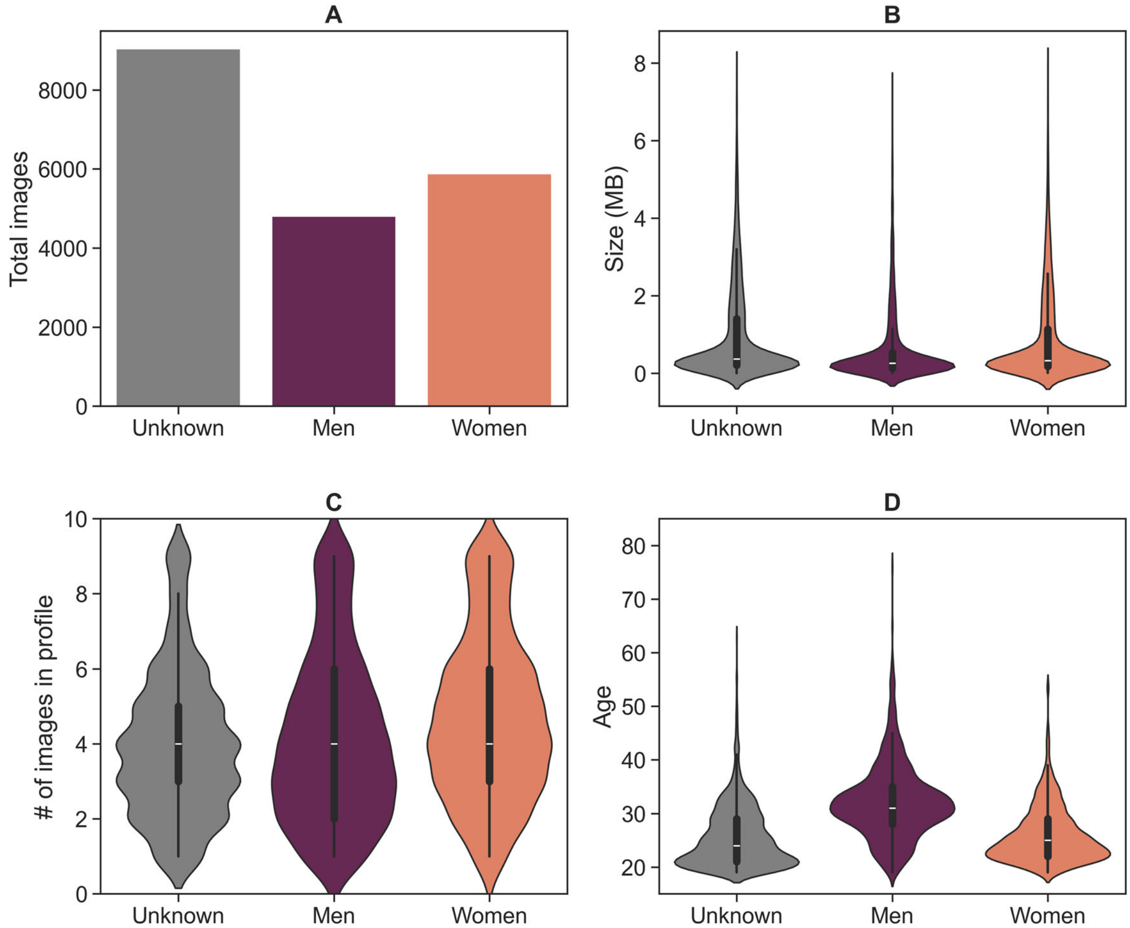

Fig. 1 Summary statistics for men, women, and unknown gender in Tinder in Estonia. A Total number of images in the dataset; **B** image file size distributions as a proxy for image quality are similar across genders; **C** the number of images among the profiles; **D** user age distributions differ, with a median age of 30 for men, 22 for women and 21 for the profiles with no gender label.

Arnold and Tilton (2019), qualitative visual anthropology (Rose, 2022), and privacy-preserving data visualization to identify relevant genres of images used on Tinder in Estonia, to evaluate and to interpret the metapicture of profile images. This approach allows us to highlight the gendered biases, outline the polarized nature of images on Tinder, and identify differences and similarities in prevalent genres of visual self-representation among men and women (research question 3).

First, we describe the data collection protocol, followed by a summary of the collected data. Second, in the data processing chapter, we describe the neural network pipeline and clustering algorithm we use to extract multidimensional representations of images and identify the prevalent genres among them. In the results section, we provide a 2D projection of a multidimensional representation of all images as well as the description of each image genre. We then follow up by providing an overview of how these genres of images are used in men's and women's profiles. We then proceed to the discussions, highlighting the gendered preferences and the compositions of profiles. Finally, in conclusion, we provide general takeaways on visual self-representation on Tinder in Estonia.

Data collection

Visual studies of social media images using digital methods usually cover one platform (Highfield and Leaver, 2016) or multiple (Pearce et al., 2020). While it is known that online dating apps are commonly used together (Couch and Liamputtong, 2008), in this research, we focus only on images from Tinder. Considering the ”liquid nature” of users of online dating platforms, who come and go, we decided to collect images over a relatively long period of 5 Months to create a reasonably comprehensive dataset. We collected 23,499 images of 4660 Tinder users in Estonia between April and August 2021. Using the so-called “sock-puppet” approach, we assembled a corpus of Tinder users in the North-East European country of Estonia, using the Pynder Python package to control the web interface of the Tinder App (Tinder, 2023b). To do this, we created an account on Tinder with settings to see users of all genders and “swipe” left through the profiles of nearby users. When we run out of users nearby, we switch locations, eventually covering the majority of Estonia, including the mainland and major islands.

Following the limitation mentioned above, we refer to gender as described in the interface of the Tinder app and only focus on the profiles that explicitly state ”man” or ”woman.” We removed 9041 images belonging to the 2259 profiles that did not display either of these options from the analysis. The resulting data set consists of 2401 profiles (with 1253 accounts labeled in the profile as ”woman” and 1148 accounts labeled as ”man”), resulting in a total number of 10,680 images across all profiles, of which 6668 images are from “women”, 4012 from “men”. For demonstration purposes and to compare with the gendered profiles, we include individuals who did not make their gender visible under the label ”Unknown” (see Figs. 1 and 2). To protect the anonymity of the users, we did not collect their names, biographies, and locations.

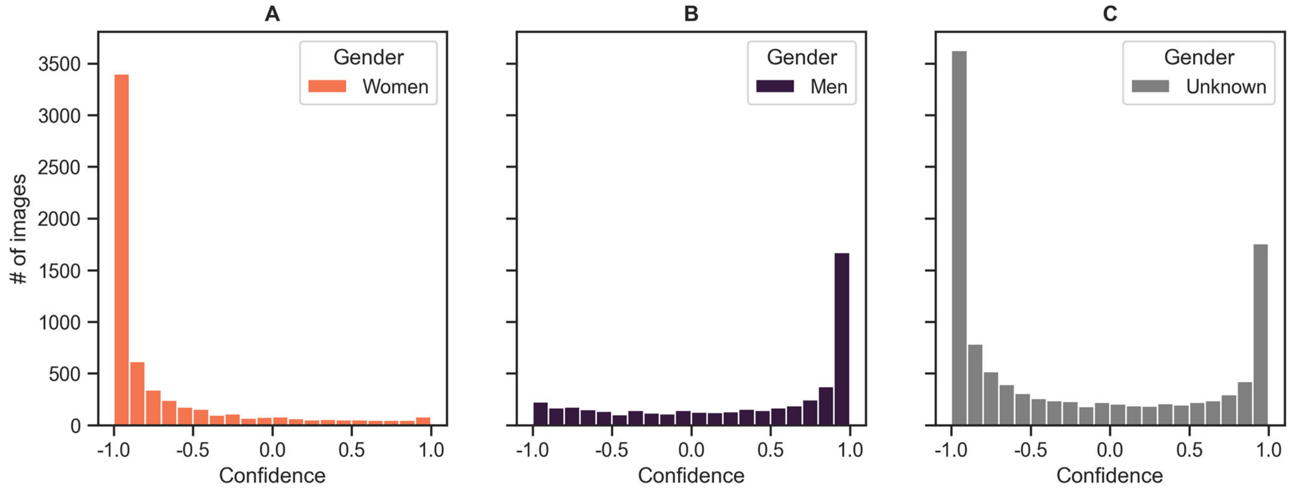

Fig. 2 Confidence of gender identification for profile images using our finetuned VGG16 machine learning model. A for images in women's profiles; **B** in men's profiles; and (**C**) in profiles with unknown gender. In all cases, we find considerable confusion across the spectrum. The distribution for unknown profiles appears to be a combination of unspecified men's and women's profiles.

After the project's end, we deleted the images used in the computational analysis to comply with the ethics committee requirements. To examine differences between groups of men and women, we analyzed the data on an aggregated level.

Data processing

To understand the relationship between the images in the dataset, we embed and represent them as multidimensional vectors (512 dimensions) using the VGG16 machine-learning algorithm from the PyTorch framework (Paszke et al., 2019). VGG16 is a widely used algorithm (Qassim et al., 2017) that achieves 92.7% test accuracy for classification tasks on the ImageNet dataset (Deng et al. 2009). We first resize and center crop all images to fit size 224×224 pixels, which allows us to have a center point that may contain important information. We then used the out-of-the-box VGG16 model trained on ImageNet. We use Adam optimizer with a learning rate of 0.003. We fine-tuned it on a sub-sample of 2000 (20%) random images from the profile pictures to achieve better classification results. Next, using this model with two output nodes, we estimated each image's probability of appearing on women's or men's profiles. We used the cross-entropy loss function, which has a built-in SoftMax activation function. Since we have gender-labeled data, we do not have to rely on biased annotated datasets like ImageNet and carry on these biases (Crawford and Paglen, 2021). The goal here is not to advance these methods but simply to sort the images in our sample on a one-dimensional scale (Fig. 2). For images uploaded by men, the resulting accuracy is 75.3%, and for pictures uploaded by women, the accuracy is 93.5%.

To extend the analysis from this one-dimensional representation and capture multi-dimensional visual similarity of images, we again use the out-of-the-box VGG16 neural network without fine-tuning, only this time, we remove the fully connected layers of the neural network and obtain dense representations of the pictures. The resulting 512-dimensional feature space is then reduced, projected, and visualized as a two-dimensional space using the UMAP algorithm (McInnes et al., 2018). Like any dimension reduction, we acknowledge that UMAP projections are lossy and sensitive to parametrization; nevertheless, we find it a valuable tool for exploring large image datasets. We further use K-means (MacQueen, 1967) on all 512 dimensions to cluster images into arbitrary categories that, upon close reading, can be identified as genres of visual self-representation. While resulting clusters are fuzzy due to the nature of polymorphic visual family resemblance (Karjus et al., 2023; Rosch and Mervis, 1975; Wittgenstein, 1953), separating the large dataset into meaningful categories such as outdoor images, portraits, etc., all together is good enough to identify gendered differences.

To visualize the categories of photography while preserving the anonymity of the users, we further use an out-of-the-box DALL-E 2 (OpenAI, 2023) algorithm created by OpenAI to produce variations of an image (Ramesh et al., 2022) that functions as a privacy-preserving facsimile. Essentially an image-to-text-to-image approach, the algorithm replaces an image encoder in place of a text encoder, allowing for the generation of new variants from an existing image instead of text input. The generated images exhibit a similar semantic meaning and appearance yet vary in composition and detail due to the noise in the encoder. Notably, the source image is entirely encoded before generation, leaving the generator independent of any individual pixel from the original image (cf. call-outs in Fig. 3). This approach allows us to exhibit various images while preserving the users' privacy.

Results

The classification algorithm we used provided us with a one-dimensional representation of the variety of images used by men and women on Tinder. Figure 2 illustrates this tailed distribution of confidences, which goes from −1.0 as the most prototypical image for women to 1.0—the most prototypical image for men. We can speculate that the character of the following distribution of confidences of gendered self-representation results from tedious work of trial and error by men and women online. Previous research on curating profiles suggests that users copy others' behavior online by "grabbing and reusing" (Senft, 2008) and experimenting with various ways of self-representation until they find one that satisfies their goals (Ward, 2017). Images located at the heads of these distributions are visually similar portraits that form a distinct and mainstream way of visual self-representation for men and women, respectively. The confidence distribution for the images in the profiles without explicitly stated gender is similar to the combined distributions of men and women. In other words, our self-labeled sample likely represents the whole dataset.

For women, the images in the head of the distribution are predominantly well-crafted photographs made on a smartphone. The face occupies a large portion of the image, leaving no space

Fig. 3 Visual similarity field and representative genre images. Each data point in the scatter plot corresponds to a downsized 3 × 4 pixel profile image of a Tinder user within our dataset. The representative pictures generated by DALL-E 2 and flanking the plot illustrate the most characteristic photos for each genre. Genres clockwise from the top left corner: *With pets, Framed, Dark Selfies, Mirror Selfies, Bars & Cafes, Urban, Vehicles, Outdoors by the water, Green Fields, Sunglasses, Light Selfies, Glamorous, Intimate selfies.*

for background and other information that would tell about the picture's context. The absence of background, along with the visible traces of editing, such as color filters and skin blur, indicate that women prefer a higher degree of control in constructing the desired impression in the limited environment of the Tinder interface. The long tail of this distribution corresponds to the niche ways of self-representation. It consists of images that are difficult to define as belonging to a specific genre without close reading. It contains photos, pictures of animals, memes, and photographs that challenge gender norms, etc.

The distribution of images classified as men follows a similar trend: the head of visually similar images and a long descending tail. Images in the head of the distribution are predominantly snapshots - casual photographs taken at the moment and exhibiting little to no graphic enhancement. In contrast, images in the long tail consist of photographs of various genres.

While selfies for women and snapshots for men carry the gendered difference of self-representation, images on the spectrum highlight the overlapping practices and visual similarities. Such images, including photos of sports activities, photos of and with pets, and pictures taken in bars and restaurants, can be found in the profiles of both women and men. Knowing the genres of each profile image, we can closely look at how consistent men and women are at self-representation and if there is a place for the gendered differences seen through the combination and order of these genres.

Space of visual self-representation

The classifier-based unidimensional prototypicality scale allowed the exploration of typical gendered genres. To gain further insight into genre variation and distributions, we use the UMAP algorithm to reduce the 512-dimensional embedding to a two-dimensional scatter plot and proceed by qualitatively studying the structure of the resulting visual space. On the scatterplot, we observe the dense regions of photos that are visually similar in color and composition. For example, photographs featuring light sky and nature are neatly grouped on the right side of the plot,

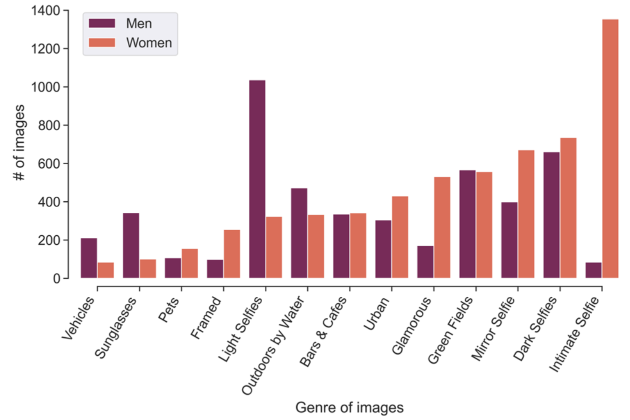

Fig. 4 Different genres of images dominate men's and women's profiles. The following barplot shows counts of men's and women's images in each genre, sorted from lowest to highest number of images in women's profiles.

with images featuring trees, grass, and other greenery above it. While these groups of images are easy to identify, a closer inspection is needed to understand other groups' borders and underlying similarities. We then iteratively performed a K-means clustering with different K-values. We used qualitative observation of the data on each iteration in an attempt to balance between a small number of fuzzy clusters and a large number of similar clusters of potential genres of photography. We ended up with 13 clusters and calculated each cluster's representative images for men and women. Figure 3 shows the two-dimensional representation of all photos, using 3-by-4-pixel versions of the original photographs to illustrate color differences while preserving privacy. Figure 4 shows the distribution of images across the following genres and genders. In the following, we briefly examine each genre cluster.

With pets. This cluster comprises images capturing various interactions with pets, including moments of companionship and standalone portraits of pets. These images are typically light-colored and emphasize the animals. Both men and women are featured alongside their cats or dogs. The photos showcase individuals petting, hugging, feeding, and playing with the pets. 156 (60%) photos with pets are from profiles of women, and 106 (40%) are from men.

Portrait in sunglasses. This collection mainly features images of men and women wearing sunglasses. The images predominantly feature headshots of individuals gazing directly into the camera through the tinted glass. Some photos provide a captivating glance over the frames.

100 (23%) of this genre are found in women's profiles and 342 (77%) in men's.

Dark selfies (Quarter-Length and Snaps). This cluster showcases selfies and snapshots where the subjects often make direct eye contact with the camera. The proportion of the body visible ranges from one-fourth to one-third. The images have a dark color palette. 735 (52%) photos of the Dark Selfies genre are found in women's profiles and 667 (48%) in men's.

Light selfies (Quarter-Length and Snaps). Similarly to the previous category of images, this one showcases a variety of bright and well-lit selfies and snapshots. Individuals in these images tend to make direct eye contact with the viewer. 323 (24%) photos of this genre are found in women's profiles and 1037 (76%) in men's.

Framed self-portraits. This set of images stands out with prominent white or black borders around the photos, which indicates the use of image editing tools to produce the desired result. There are 254 photos (72%) of this genre among women's profiles and 98 (28%) among profiles of men.

Intimate selfie portraits. This group primarily features selfies in which the subjects' faces dominate the frame. The background is dark and hidden, allowing the emphasis to remain on the individual. These images are often headshots or show only a quarter of the body. Notably, this group of images has the most significant imbalance in gender representation, with 94% of photos (1358) among women's profiles and only 85 (6%) among men's.

Mirror selfies. The images in this category capture mirror selfies and candid snapshots, depicting individuals positioned at the center of the frame. Approximately three-fourths of their height is visible, and a distinguishing feature is the proximity of their hand to their head. The challenge for the machine learning lies in distinguishing between the act of taking a mirror selfie and the act of scratching one's head. Women have a total of 673 photos (63%) in this category, while men have 400 (37%).

Bars and cafes. This group showcases portraits of individuals in public venues such as bars, cafes, and restaurants. Compared to selfies and snapshots, these images offer more background context, revealing the complexity of the setting. People are often pictured at tables or chairs, accompanied by food or Beverages. In our sample, women and men have equal numbers of photos in bars and cafes: 342 (50%) and 337 (50%), respectively.

Outdoor by Water. This collection features photographs of individuals engaged in outdoor activities, primarily along coastlines and beaches. The people occupy a smaller portion of the image, directing attention to the natural surroundings—beaches, bodies of water, blue skies, or snowy mountains. Their outfits are tailored to the environment, ranging from swimwear to skiing gear. Three 135 (41%) images of outdoor activities were found in women's profiles and 475 (59%) in men's.

Vehicles. This series of images captures interactions between individuals and various vehicles, including cars, bikes, and yachts. The focus lies on the vehicles, portraying people in natural interactions—driving, repairing, or simply posing beside them. Women have a significantly lower number of images featuring vehicles in their profiles (84 images or 28% of all images), while men have 213 (72%).

Glamorous. This compilation comprises selfies and snapshots with a distinctive feature: extravagant outfits, shirts, dresses, and luxury accessories. This category deviates from the previous ones by emphasizing fashion choices, exquisite patterns, and luxurious items. Women account for three-quarters of all images in this category (531 images or 75%), while men only have 170 (25%).

Green fields. This group presents individuals within natural landscapes, where green grass and towering trees dominate the scene. The subjects stand prominently, offering full-body views against the bright green background. Green fields are equally present on women's and men's profiles: 558 (50%) and 567 (50%).

Urban. Similar to the previous group, this collection spotlighted the environment. However, here, the setting is urban, featuring concrete walls, graffiti, and asphalt. The individuals' poses accentuate the urban surroundings, often leaning against walls or

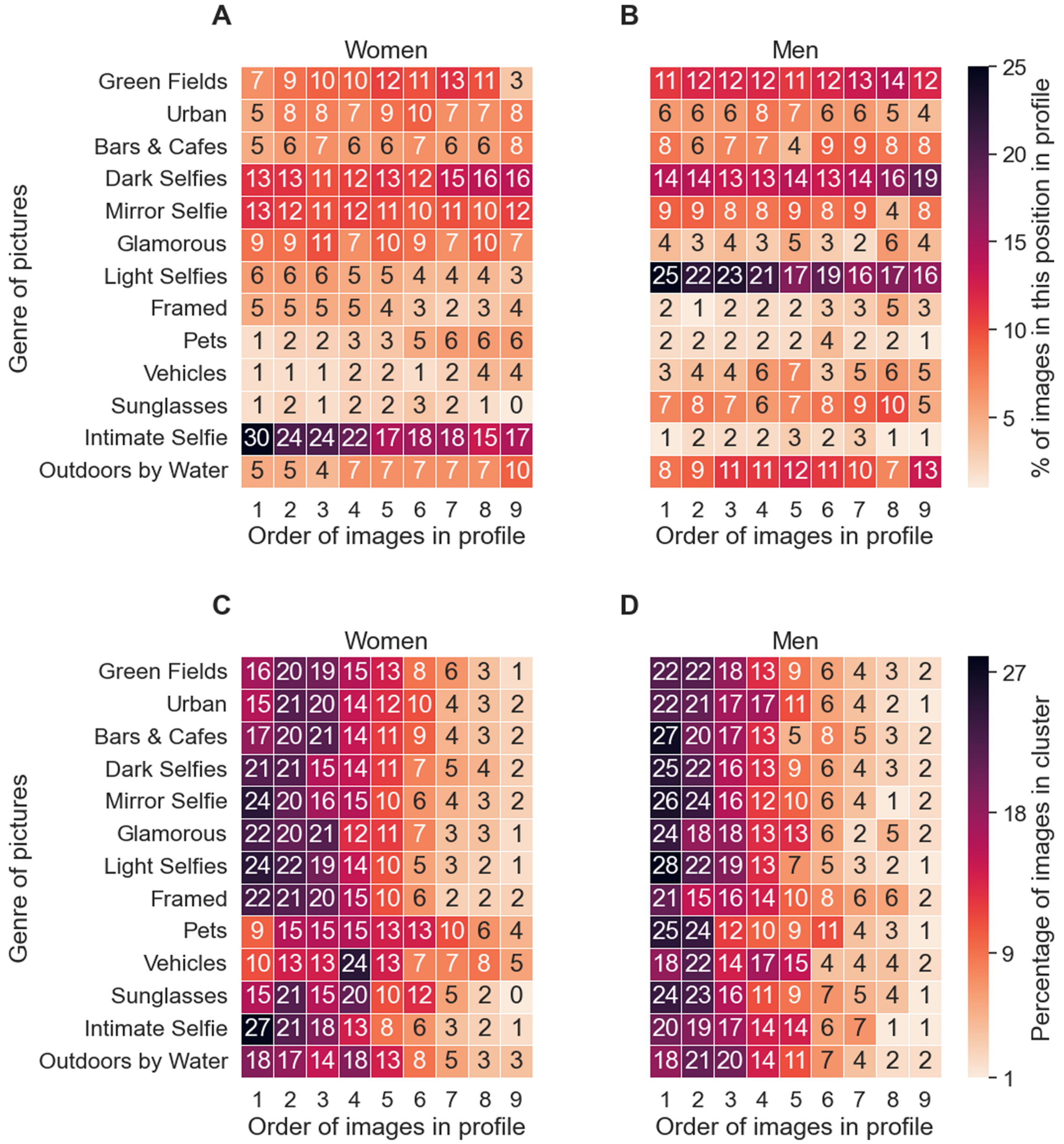

Fig. 5 Genres of pictures are unequally distributed across the profile image galleries of men and women on Tinder. A, **B** Proportion of images of each genre for each position in the profile. The Y-axis is shared for men and women. It is seen how the most popular genre for women is the least popular for men. **C**, **D** Distribution of genres based on the position in the profile gallery sorted by the mean value of proportion for each place for women. The dominant position of each genre reveals the gendered differences as experienced by individuals. Here, we show such gendered differences as men putting images of pets and vehicles at the beginning of their profiles. At the same time, women prefer to show selfies first, followed by pets and vehicles.

immersed in the hustle of city life. 429 images (58%) of women and 305 images (42%) of men in the urban environment were found in our sample.

Gendered differences in self-representation

Tinder allows users to upload up to 9 images to their profile gallery, and profile visitors can go through them individually. To tap into the gendered differences of visual self-representation, we analyze how different images represent each image gallery position and how specific genres are distributed across the profiles. While users can opt-in and use the "Smart Photo" feature to auto-arrange photos (Tinder, 2023a), we compare men's and women's images "as they appear" on the card stack for an individual swiping through the profiles. For each position in the image gallery of men and women, we calculate the number of times an image of each genre appeared in this position. We normalize the values for the total number of pictures in each position. For visual comparison, the resulting distributions are sorted by the most frequent genre for women (see Fig. 5). It is visible that *Intimate Selfies* are the most dominant genre for women, with 30% of the first images of their profiles being *Intimate Selfies*. Looking further at the proportion of genres across other positions in the gallery, we can see that *Intimate Selfies* are consistently more dominant than others. On the one hand, it reflects a large number of images of this genre we identified (calculate the % of all women images) in our dataset, and on the other hand, it shows that women prefer to put *Intimate Selfies* with no regard to the position in the profile. We found a similar distribution in the men's profiles. The genre of *Light Selfies* dominates the shapes,

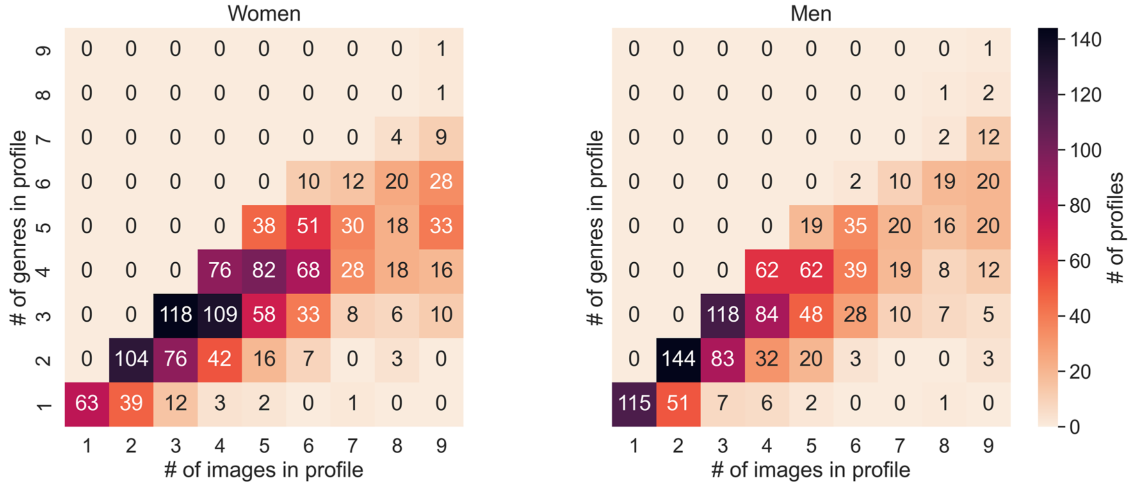

Fig. 6 Women, on average, have more images in their profiles and more unique genres presented than men. Most women's profiles have three images in their profile and three unique genres presented, while most men's profiles have two unique genres across two profile images. The heatmaps show the number of profiles and a relationship between the number of profile images and unique genres within these images. As only one genre is assigned to each image, all values appear on or below the diagonal since the number of genres cannot be larger than the number of images. The number of genres saturates, falling below the diagonal, implying that there are rarely more than 7 out of 13 genres in a profile, even for profiles with 8 or 9 images. This aligns with Miller's cognitive limit of 7+/−2 items in short-term memory (Miller, 1956).

covering 25% of the first images in this category, 22% of the second and 23% of the third profile images. In contrast to the women's most popular genre, *Intimate Selfies* are the least popular among men, with a one- and two-percent chance of all first two profile pictures, respectively. The second most popular genre for women is *Mirror Selfies*, with 13% of first profile images in this category.

We see the sharp difference between the two most popular genres with a 2–3 times less chance that the first image will be a *Mirror Selfie* rather than an *Intimate Selfie*. For men, *Mirror selfies* are the 4th most popular genre, with only an 8–9% chance of appearing anywhere in the profile. For women, *Dark Selfies* are the third most common genre, with chances of appearing anywhere in the profile similar to *Mirror selfies*. Women exhibit the same pattern as men: their eighth and ninth images on the profile are likelier (16%) to be a *Dark Selfie*. *Glamorous* images are the fourth most common genre of photos on women's profiles, accounting for 7% of first images, while for men, this genre of pictures is in seventh place. Photographs featuring green grass and forests can be found in any position of the image gallery of men's and women's profiles alike. *Urban* images and photos of activities *Outdoors by the water* are likely to be seen in the center position of the men's profile or at the very end.

In contrast, *Urban* photographs are more likely to be closer to the end of the profile in women's profiles. Photos in *Bars and cafes* are more likely to be found on the men's profiles than images of any other genre. Every fourth (25%) man in our dataset had a first image of their profile featuring a bar, cafe, restaurant, or other public place, while only 6% of women's profiles had a photo of this category as their first picture. For men, the second most popular genre is *Dark Selfies*, featuring dim lighting and a prevalence of black color, often black and white or desaturated; these images appear anywhere in the profile with an average of 14% chance. The exception is the eighth and ninth photos, with 16 and 19% chances of encountering them. Photographs containing black or white frames around them - a signal of post-processing in other image editing software - are more likely to be seen at the beginning of women's profiles and the end of men's profiles. Images with pets are unlikely to be the first image of the women's profile (1%). However, the likelihood of encountering one increases to 6% towards the end of their profile. Men, in contrast, have *Photos with pets* in any place of their profile with the same probability. *Vehicles* appear on men's profiles more often than women's, with the likelihood increasing towards the middle and end of their image carousel. Women are more likely to have a photograph with a vehicle on the last two positions of their profiles (4%) compared to the first position (1%). *Sunglasses* are predominantly used by men in their profile images. The likelihood of encountering a man wearing sunglasses in their first image is 7%, increasing with the image's position on the profile. Conversely, women are likelier to put an image with sunglasses towards the end of their profile.

We also look at the distribution of the number of unique genres presented in the profiles of men and women. The heatmap in the Fig. 6 illustrates this relationship. Notably, for both men and women, the number of unique genres drops with the increase of profile images. However, if the profile has a maximum number of images (9), they have a more extensive variety in combining unique genres. For both women and men, the variability of genres increases with the number of photos they have in their profile, and the variation is higher for women.

Variation of images

To identify how different the self-representations of men and women are across each genre, we calculate the mean cosine distance between each image for each genre for each gender. While pictures of some genres are similar among men and women (see Fig. 7), other genres (*Pets, Vehicles, Framed*) show more significant variability. In general, medians within groups are similar, showing that men's and women's images do not differ much within each genre. This uniformity of images in our dataset might indicate the existence of unwritten rules of communication and social norms regarding what image to post, which boils down to creating the platform-specific style space.

Discussion

Based on a collected sample of images, we sought to discover gendered differences in visual self-representation on Tinder. The result of our approach is an overview of the visual landscape of

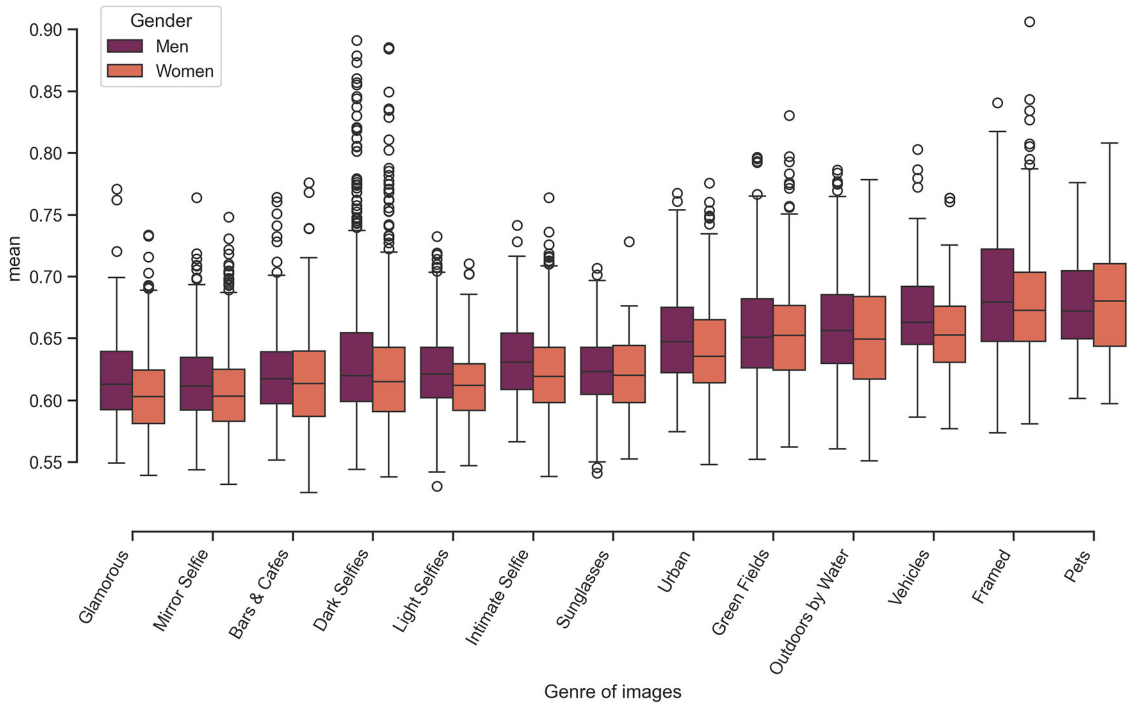

Fig. 7 Box plot of the variability of images within clusters. Sorting by mean variability of cosine similarity reveals the increasing complexity of images across image genres. The higher the mean value, the more uniform self-representation is. Men's images, on average, are slightly more uniform than women's.

the profile pictures in a specific context and a deeper, qualitative look at the genres and dynamics of self-representation. Such a combination of close and distant reading opens up an opportunity to look at the culture of visual self-representation from different angles. It allowed us to experience the scale and distribution of genres, something classic approaches with limited samples do not qualify. We used a machine learning algorithm to address the visual similarity of Tinder users' profile images in Estonia. By visualizing the "meta picture" of users' profile images and combining it with such attributes as age and gender, we could identify prevailing genres, such as *Intimate Selfies*, *Dark Selfies, Mirror Selfies, Glamorous pictures, Green Fields, Light Selfies, Urban photos, Outdoors by the water, Bars & cafes, Framed, with Pets, Vehicles*, and *Sunglasses*.

We were aiming to find gendered differences in self-representation practices on Tinder. Our findings suggest that women use more *Intimate Selfies* than men, as they enable a more controlled environment for self-representation through focus on the face and gaze. These findings align with previous research (Qiu et al., 2015), which observed that women are more likely than men to take selfies and prefer more control over their images.

Furthermore, we look at the dominant genres of photographs based on the position in the profile and find that men, compared to women, have more variety of photographs as the first image of the profile. At the same time, women, on average, have more images and use more unique genres to express themselves. In addition, we examined the consistency of image genres across the profiles and the variability of image similarity within genres. We found that genres such as *Intimate Selfies*, *Glamorous*, and *Framed* selfies are dominant solely among women, and *Light* Selfies, *Sunglasses*, and *Vehicles* are mostly popular among men. However, when it comes to similarity within the genres, we do not find a significant difference between images used by men and women, with images *With Pets* and *Bars & Cafes* being an exception where women have a greater variety versus *Framed* for men, respectively. While men and women approach self-representation differently, our research suggests a significant overlap in genre use and distributions within genders.

Limitations

There are several limitations worth mentioning. First, the data was collected from one online dating platform, Tinder, which does not cover other popular services like Bumble and OkCupid. Additionally, the nature of Online Dating platforms makes it hard to reproduce the study results. Usage of online dating sites and apps peaks around middle age (Rosenfeld and Thomas, 2012), and our relatively young sample (with the median age for men at 30 and for women at 22 years; see Fig. 1) not only under-represents the older adults but also makes gendered comparisons imbalanced.

Another limitation comes from the fluid nature of online dating users. Successfully finding a partner or deleting the account due to growing disappointment makes the user base of the dating sites constantly changing, peaking at its most active period during the beginning of the year when individuals make new year resolutions, including finding a new partner. This way, a single short time frame of data collection can capture only part of the online dating culture. Another factor to consider is that we collected data during the COVID-19 pandemic when tourism was restricted. Thus, the population in our sample is smaller yet primarily comprised of locals. Since our data only contains images uploaded by users in Estonia, our findings might differ from the populations of neighboring countries and the rest of the world. The upside of using a sample from a small country (Estonia covers 45k km^2 and has 1.3 M inhabitants). This opens up a possibility for comparative studies, where our approach is used to find cultural differences and the plurality of self-representation tactics between different countries, nations, or parts of the world.

While machine learning and artificial intelligence provide powerful tools for qualitative research, a few limitations arise. By inspecting individual selfies analyzed in this paper, a selfie taken in front of a mirror stands out as a particular sub-genre of the selfie. While the machine learning algorithm did a decent job separating close-up selfies from mirror selfies, which feature prominent objects such as the hand and phone of the person, this particular model did not, for example, differentiate between a bathroom mirror, a mirror in an elevator, or a gym. While they are all mirror selfies, each contains extra contextual information that can be important for correct self-representation.

Conclusions

Images on online dating platforms provide valuable insight into how individuals present themselves in a highly curated environment. Analyzing these images reveals cultural norms, values, and aesthetics that people believe are to make them desirable or acceptable. These practices can reflect and reinforce gender roles and stereotypes. By studying these images, we can examine how traditional gender expectations are challenged in modern dating practices. A large-scale investigation is necessary to understand how identity is constructed and displayed online, as it allows a more holistic view of self-representation practices, often unavailable for small-scale qualitative methods.

Prior research focused on digital ethnographic methods, such as collecting and analyzing a small sample of profile images from dating platforms or conducting interviews with online daters and asking them to craft profiles in a controlled environment. Our study extends research on visual gendered self-representation by examining a relatively large set ($n = 10680$) of profile images on Tinder.

Despite the platform's limitation in providing an ability to select who to see on the card stack, we were able to observe that both men and women use visual clues that do not necessarily conform to this built-in binary disambiguation. This finding challenges traditional notions of attraction and partner selection and suggests that individuals' preferences in visual representation, even in such a dominantly heteronormative platform, are more diverse and complex than simply fitting into limited binary categories, and thus, platforms would be well advised to expand their functionality to account for it. Further understanding of how individuals are looking for a variety of visual cues may make online daters feel more empowered to present themselves authentically rather than conform to mainstream ways. Further research in this direction may lead to a more inclusive understanding of attraction, dating behaviors, and matching. We believe our work will spark more in-depth within-culture and cross-cultural research while informing discussions on gender equality and social dynamics.

Received: 30 December 2023; Accepted: 16 September 2024;
Published online: 27 September 2024

References

Arnold T, Tilton L (2019) Distant viewing: analyzing large visual corpora. Digital Scholarsh Humanit 34(Supplement_1):i3–i16

Athey S, Karlan D, Palikot E, Yuan Y (2022) Smiles in profiles: Improving fairness and efficiency using estimates of user preferences in online marketplaces (No. w30633). National Bureau of Economic Research

Blackwell C, Birnholtz J, Abbott C (2015) Seeing and being seen: co-situation and impression formation using Grindr, a location-aware gay dating app. N Media Soc 17(7):1117–1136

Bruno N, Bertamini M, Protti F (2015) Selfie and the city: a world-wide, large, and ecologically valid database reveals a two-pronged side bias in naïve self-portraits. PLoS ONE 10(4):e0124999

Butler J (2002) Gender trouble. Routledge, New York

Caliandro A, Graham J (2020) Studying Instagram beyond selfies. Soc Media+ Soc 6(2):2056305120924779

Couch D, Liamputtong P (2008) Online dating and mating: the use of the internet to meet sexual partners. Qual Health Res 18(2):268–279

Crawford K, Paglen T (2021) Excavating AI: the politics of images in machine learning training sets. Ai Soc 36(4):1105–1116

Degen JL, Kleeberg-Niepage A (2023) Profiling the Self in Mobile Online Dating Apps: a Serial Picture Analysis. Human Arenas 6:147–171. https://doi-org.ezproxy.tlu.ee/10.1007/s42087-021-00195-1

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA. 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Dewdney A (2022) The Politics of the networked Image. Dewdney A and Sluis K. The Networked Image in Post-digital Culture. Routledge, London

Duguay S (2017) Dressing up Tinderella: interrogating authenticity claims on the mobile dating app Tinder. Inf Commun Soc 20(3):351–367

Ellison N, Heino R, Gibbs J (2006) Managing impressions online: Self-presentation processes in the online dating environment. J. Comput Mediated Commun 11(2):415–441

Gershon I (2012) The Breakup 2.0: Disconnecting over New Media. Ithaca, NY: Cornell University Press. https://doi.org/10.7591/9780801458637

Giddens A, Pierson C (1998) Conversations with Anthony Giddens: Making sense of modernity. Stanford University Press, Stanford

Highfield T, Leaver T (2016) Instagrammatics and digital methods: studying visual social media, from selfies and GIFs to memes and emoji. Commun Res Pract 2(1):47–62

Karjus A, Solà MC, Ohm T, Ahnert SE, Schich M (2023) Compression ensembles quantify aesthetic complexity and the evolution of visual art. EPJ Data Sci. 12(1):21

Leaver T, Highfield T, Abidin C (2020) Instagram: Visual social media cultures. Polity, Cambridge, UK

Leszczynski A (2019) Digital methods II: Digital-visual methods. Prog Hum Geogr 43(6):1143–1152

Macqueen J (1967) Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. United Kingdom: University of California Press. (Vol. 1, No. 14, pp. 281–297)

Manovich L (2011) Style Space: how to compare image sets and follow their evolution. Software Studies Initiative. Available via https://manovich.net/content/04-projects/073-style-space/70_article_2011.pdf

Marwick AE (2013) Status update: Celebrity, publicity, and branding in the social media age. Yale University Press, London

McInnes L, Healy J, Melville J (2018) Umap: uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426

Miller GA (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol Rev 63(2):81

Mitchell WJ (1994) The reconfigured eye: Visual truth in the post-photographic era. Cambridge, MIT Press

OpenAI (2023) DALL·E 2. https://openai.com/dall-e-2. Accessed 25 Nov 2023

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, … Chintala S (2019) Pytorch: An imperative style, high-performance deep learning library. In Advances in Neural Information Processing Systems 32, pp. 8024–8035

Papacharissi Z (Ed.) (2010) A networked self: Identity, community, and culture on social network sites. Routledge, New York

Pearce W, Özkula SM, Greene AK, Teeling L, Bansard JS, Omena JJ, Rabello ET (2020) Visual cross-platform analysis: digital methods to research social media images. Inf Commun Soc 23(2):161–180

Pidoux J, Kuntz P, Gatica-Perez D (2021) Declarative variables in online dating: A mixed-method analysis of a mimetic-distinctive mechanism. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1):1–32. https://doi.org/10.1145/3449174

Pidoux J (2022) A comparative study of algorithmic–user classification practices in online dating: a human–machine learning process. Porn Stud 10(2):191–209. https://doi.org/10.1080/23268743.2022.2104352

Qassim H, Feinzimer D, Verma A (2017) Residual squeeze vgg16. arXiv preprint arXiv:1705.03004

Qiu L, Lu J, Yang S, Qu W, Zhu T (2015) What does your selfie say about you? Comput Hum Behav 52:443–449

Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M (2022) Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125, 1(2):3

Rosch E, Mervis CB (1975) Family resemblances: Studies in the internal structure of categories. Cognit Psychol 7(4):573–605

Rose G (2022) Visual methodologies: an introduction to researching with visual materials. Visual methodologies, 5th edition, Sage, 1–100

Rosenfeld MJ, Thomas RJ (2012) Searching for a mate: The rise of the Internet as a social intermediary Am Socio Rev 77(4):523–547

Sedgewick JR, Flath ME, Elias LJ (2017) Presenting your best self (ie): the influence of gender on vertical orientation of selfies on Tinder. Front Psychol 8:604

Senft T (2008) Camgirls. Peter Lang Publishing, New York
Smith AW, Duggan M (2013) Online dating relationship. Pew Research Center, Washington, DC
Sundén J (2009) Play as transgression: an ethnographic approach to queer game cultures. In Proceedings of DiGRA 2009 Conference: Breaking New Ground: Innovation in Games, Play, Practice and Theory. London
Tifentale A, Manovich L (2015) Selfiecity: Exploring photography and self-fashioning in social media. In Postdigital aesthetics: Art, computation and design (pp. 109–122). Palgrave Macmillan, London, UK
Tiidenberg K, Gómez Cruz E (2015) Selfies, image and the re-making of the body. Body Soc 21(4):77–102
Tiidenberg K, Whelan A (2017) Sick bunnies and pocket dumps:"Not-selfies" and the genre of self-representation. Pop Commun 15(2):141–153
Tinder (2020) Introducing More Genders on Tinder. https://www.tinderpressroom.com/genders Accessed 24 Dec 2023
Tinder (2023a) Introducing Smart Photos - For The Most Swipeworthy You. https://www.tinderpressroom.com/introducing-smart-photos-for-the-most-swipeworthy-you. Accessed 20 Nov 2023
Tinder (2023b) Tinder | Dating, Make Friends & Meet New People. https://tinder.com/app. Accessed 25 Dec 2023
Tong ST, Hancock JT, Slatcher RB (2016) Online dating system design and relational decision making: Choice, algorithms, and control. Personal Relatsh 23(4):645–662
Walther JB (2007) Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. Comput Hum Behav 23(5):2538–2557
Ward J (2017) What are you doing on Tinder? Impression management on a matchmaking mobile app. Inf Commun Soc 20(11):1644–1659
Wittgenstein L (1953) Philosophical investigations. Philosophische untersuchungen. Macmillan & Co., Oxford
Whitty M, Carr A (2017) Cyberspace romance: The psychology of online relationships. Bloomsbury Publishing, New York
Wright KB, Webb LM (Eds.). (2011) Computer-mediated communication in personal relationships. Peter Lang, New York

Author contributions
Yan Asadchy (YA) is the first and corresponding author, Andres Karjus (AK), Ksenia Mukhina (KM) are contributing authors and Maximilian Schich (MS) is a supervising author. YA, AK, KM, and MS designed the research. YA collected and prepared the data. YA, AK, KM, and MS performed data analysis, visualization, and wrote the manuscript.

Competing interests
Non-financial interests: Author MS serves on the editorial board of the journal

Ethics approval
Decision No. 7 of the Tallinn University Ethics Committee of February 10, 2021

Informed consent
This article does not contain any studies with human participants performed by any of the authors.

Additional information
Correspondence and requests for materials should be addressed to Yan Asadchy.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.