Truman, Or How I Learned to Stop Worrying and Love Simulated Social Media

Lee Taber
ltaber@ucsc.edu

Kevin Weatherwax
wax@ucsc.edu

Alexander Mayben
amayben@ucsc.edu

Sean Fernandes
semafern@ucsc.edu

ABSTRACT
Conducting unbiased empirical research in social media spaces is of great importance. However, it is a highly restrictive space for researchers who are not aligned with, or working from within, the companies who control those spaces. Additionally, even when access to social media networks is available, there are a range of ethical concerns and hurdles around conducting experimental research. Here we present a proof of concept for a system of easily accessible software tools which leverage procedural generation, machine learning, and interactive media to allow researchers– from both technical and non-technical domains– to conduct investigations that involve human personality traits (i.e., Big 5 attributes) in a simulated social media environment. Our past research on personality attributes in social media spaces is both our motivation and exemplar use case, as we have found that affordances and social mores influence individual presentation as measured by personality traits.

INTRODUCTION
Most researchers are not large tech companies; they are largely unable to present participants a fully functioning social media platform with an established user base. Data from large social media platforms (e.g., Facebook, Instagram, Snapchat, etc.) are largely inaccessible to researchers outside of– and sometimes within– their own companies. Unless one already works at a social media company, researchers may find it significantly more difficult to test many of the insights that they might possibly glean from users. Furthermore, faking such a social media would be very difficult using the common practices such as Wizard of Oz (WOZ). Especially if one is attempting to maintain any sort of validity, consistency in posts is very important. In order to further research in this space, collecting empirical data becomes increasingly important for a wide range of research fields.

Personality and social media research is an example of an area where empirical tools would be of great use. Personality is a long-standing and intuitive construct that we use as a way to understand how people behave across different situations. Much research has been collected on how personality is exhibited online, especially in social media [10, 24, 1, 6, 11] However, this research is mostly correlational and is not useful for making causal claims about the relationship between personality and online platforms. Furthermore, little of this research has informed the creation of personas, or character archetypes that tend to portray specific traits. The issue of correlation is significant in the social media research space. Without control over certain aspects of the social media platform, performing empirical research becomes especially challenging. Measuring correlations is about the limit of what most researchers can accomplish with current tools.

In this paper we present our first steps at developing an ecosystem of tools to simulate, manipulate, and record human interactions on social media platforms. We build upon an existing simulated social media framework from DiFranzo's Truman platform [4], but we also utilize machine learning processes so as to generate and validate potential simulated posts. The development of this system was motivated by our own past work investigating personality traits in social media. However, this is also a good exemplar use case to present simulated social media posts, as the core principles and language are widely understood and thus accessible beyond our own discipline.

Personality
The "Big Five" (OCEAN) personality trait taxonomy is a highly validated and well-studied survey [14, 5, 21]. The five personality traits are commonly referred to by the acronym OCEAN: Openness to Experience, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. Openness to Experience is related to the range of a person's interests, intellectual curiosity, and aesthetic sense. Someone who rates high on Openness is likely to have a wide range of interests and enjoy tackling new ideas. Someone who rates low on Openness is likely to focus on a few areas and enjoy routines. Conscientiousness is related to responsibility, organization, and time-keeping. Someone who rates high on Conscientiousness is likely to be punctual and keep spaces around them clean and tidy. Someone who rates low on Conscientiousness is likely to procrastinate, have a messy desk, and find trouble being on time for events. Extroversion is related to social activity, energy level, and assertiveness. Someone who rates high on Extroversion is likely to be the life of the party, energetic in social situations, and not afraid to speak their mind. Someone who rates low on Extroversion is likely to stay at home and not engage frequently in social situations. Agreeableness is related to warmth, trust, and respect. Someone who rates higher on Agreeableness is likely to be a kind, forgiving person who sees the best in everyone. Someone who rates lower on Agreeableness is likely to be a cold, mean person who is distrusting of others. Finally Neuroticism is related to

emotional volatility, anxiety, and depression. Someone who rates high on Neuroticism is likely to have mood swings, be a worrier, and/or have depression. Someone who rates low on Neuroticism is likely to be emotionally stable, and to be seen as a generally calm person.

All of the above descriptions of the traits describe so-called "offline" behaviors. However, people can also make predictions about someone's personality from a variety of clues. This can include the spaces where they work and live [9], how they conduct themselves [18], and even their handwriting [12]. Within social media spaces, other researchers have correlated personality traits with certain behaviors on social media. For example, Amichai-Hamburger & Vinizky [1] correlated OCEAN traits with Facebook behaviors and found Extroversion correlated with number of friends, and negatively correlated with the amount of posted personal information. Openness correlated with amount of posted personal information as well as the use of certain features within the user's personal info section, such as emoji. Neuroticism was correlated with likelihood of the user posting a photo on their profile, as well as particularly high and low scorers sharing more basic info (such as selfies and personal information). This was hypothesized to be the result of emotionally stable people being willing to share more because they are secure, while emotionally volatile people do so when seeking self-assurance.

Gosling et al. [10] also found correlations between OCEAN traits and Facebook behaviors. Extroverted people correlated with number of friends, number of friends in local networks, likelihood of maintaining an up-to-date presence, and likelihood of commenting on other pages. Openness correlated with number of friends, shared local networks, and photos of new activities, people and things. Conscientiousness was negatively correlated with time spent on Facebook; those who are low on Conscientiousness tend to procrastinate more, and thus spend more time on social media. People who were high on Agreeableness were more likely to view users' pages more often, including both others and their own. Personality is a useful test case for this tool as there is a great deal of existing research related to its correlations with social media behavior.

Social Media Research

Our research partly builds off of an existing simulated social media platform, named Truman by the original researcher [4]. In DiFranzo et al.'s study, researchers created a simulated social media environment in order to better understand which attributes of the system could potentially influence cyberbullying intervention among college age students. The social media is simulated because the participant is the only real actor on the platform. Each person who signs in sees their own version of the platform. All other posts that participants see are "actors"; scripts that post pictures and comments at pre-specified times. The design of the simulation is also extremely clever; to limit the types of images needed, and to avoid having to show a bunch of photos of people (which would be difficult to simulate for a large number of people), their simulated platform, named "EatSnap.Love", is essentially a food documentation-based social network. People are only allowed to post pictures of food, and anything else would be automatically deleted,

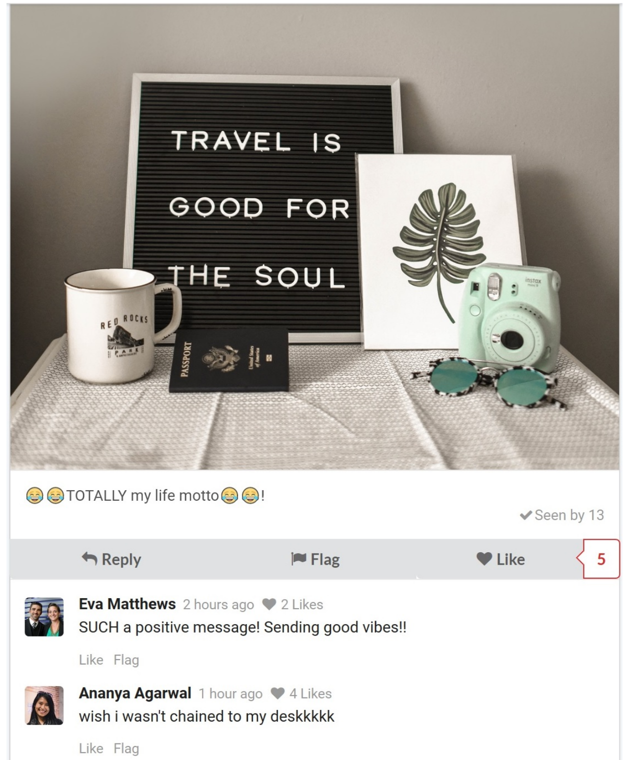

Figure 1. Example of current, human-authored Truman post

ensuring that participants are encouraged not to post anything crazy or risque. Truman also employed two hidden variables within its design communication. The first variable decided on whether the system indicated that other people could see what the user viewed, and the second chose whether to portray the user's audience was large, small, or vacant. These variables aimed to test the influence of social pressure and/or the bystander effect in cases where users may have seen one actor bullying another within a given post. Sadly, this intervention failed, as the vast majority (75%) of participants, regardless of condition, simply didn't do anything about bullying that they witnessed.

While Truman's original experiment was an unfortunate failure, the underlying systems remain amazingly robust for conducting empirical research. As an early test of this system, one of the authors wanted to understand how social media users evaluate aspects of other users' behaviors to evaluate their personalities. To this end, the author used previous research on personality correlates to social media behaviors [1, 10] in order to generate several personas, each persona attempting to portray one pole of each of the Big 5 dimensions (see Figure 1 for example of human authored posts). Participants in this version were asked to read posts from these and other personas over a period of three days, and at the end guess which "mask" each persona was wearing (visit https://truman.herokuapp.com/ to see for yourself). Personas would use as many attributes of the post format as possible to convey their personality traits. For example, the high-Extroversion persona made more posts, commented frequently on other posts, solicited interaction in

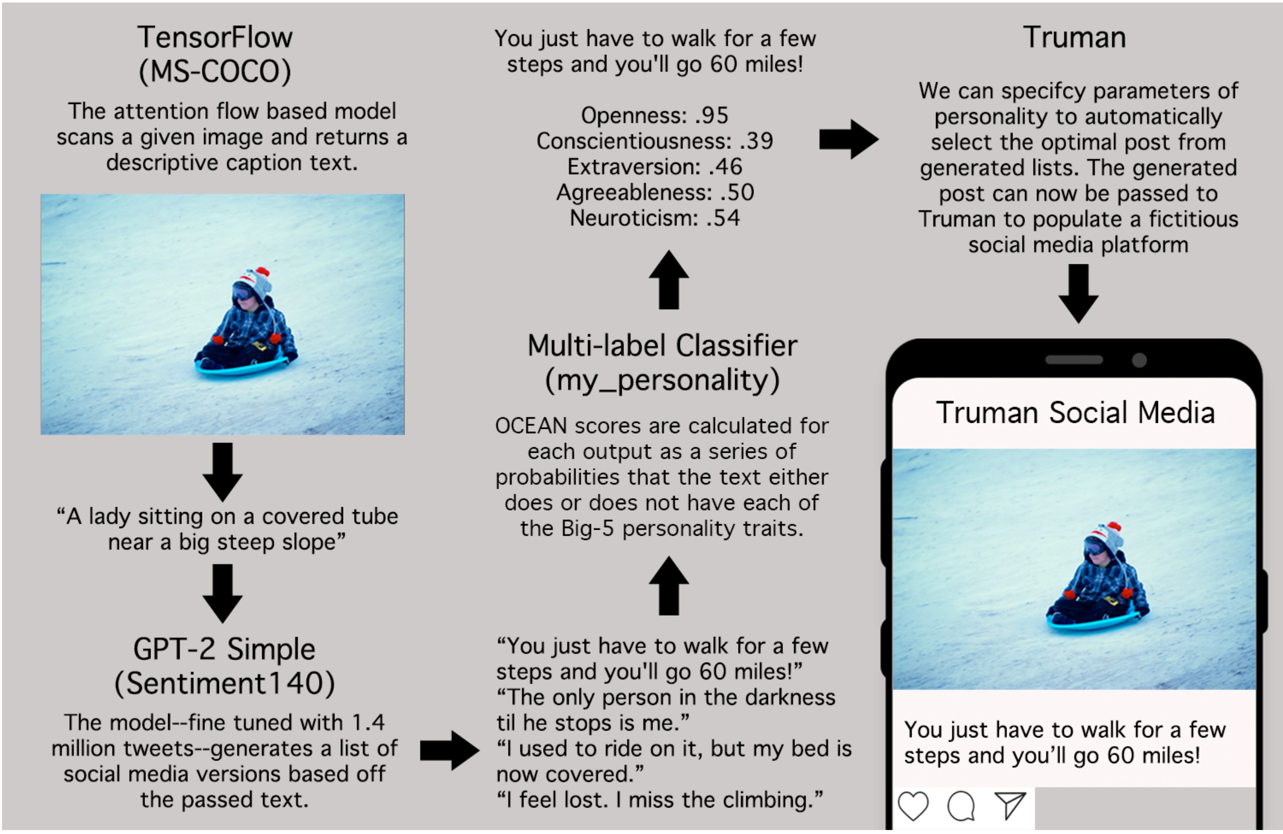

Figure 2. Proof of concept workflow for implementing generative personality type agents in Truman

their posts, had more likes and comments than other personas, and so on. However, this method quickly raised several issues. It was onerous to have one person generate hundreds of posts and comments to make sure that each was exhibiting the correct traits, and it was unclear which aspects people were examining to make their determinations. From these concerns, the author became interested in ways to procedurally generate posts, as well as validate the personality traits being exhibited. This in turn motivated the assemblage and application of this set of tools.

METHOD AND RESULTS

In this section we describe the workflow and assemblage of our toolkit. This serves as a proof of concept (refer to Figure 2 for overview) for procedurally generated, but controllable, stimuli for populating an artificial social media platform with posts measured for personality traits. We will provide a high-level description of each individual piece and how it contributes to the overall proposed system.

Image Captioning

People on the largest social media platforms (e.g., Facebook, Instagram, etc.) [3] generally start conversations with an image and an accompanying piece of text. Normally, the process of manually captioning images to populate platforms like Truman is time consuming and difficult. To address this, we started with an image classifier to parse and describe the content of images. Specifically, we used a machine learning model for captioning visual images, built from a TensorFlow tutorial [25]. This model is attention-based [31]; it learns by focusing on different sections of an image as it decides upon which captions to apply. This allows users to see which parts of the image are "in focus" when each word of a caption is decided upon. It uses InceptionV3 [23] for image classification and feature recognition. The model was trained on the MS-COCO dataset of 221,000 captioned images [16]. Each image in the dataset is segmented into discrete objects, and has five distinct captions, making it ideal for this type of image captioning. Captioning vocabulary was created by tokenizing the content of the MS-COCO captions (i.e., generating and saving text files). After the model was fully trained, we input images from Truman to generate sample captions for later modification. In most cases, these captions accurately identified the main subject of the image, albeit with some confusion about context and background details.

Social Mediafication

Having produced an accurate descriptive caption of a social media image, we then need to adjust the language so that it feels appropriate for a social media post. This adjustment is done by generating a new line of text based off the descriptive caption passed to us above through GPT-2 (Generative Pre-trained Transformer-2). GPT-2 is a generative text algorithm based upon an unsupervised Transformer machine learning model, which generates text based on a pre-trained dataset of several million webpages' worth of content. This text can then

Table 1. Image Caption Tweets

Image	Caption	Tweets
	A cat sitting next to a phone book	"I don't see a cat anymore" "Except she's terrified" "Let me know oh yeah I cant record her making a noise as that sounds like an accident lol" "A cat hollering at a rogue landlord. PICTURE. You might only be responsible for two days, but at least you got it fixed."

be further tuned based on a specific corpus of data, which will specify the output to a specific genre or style. In lay terms; you give the model a chunk of relevant source data and it re-frames its already robust ability to generate cohesive text to produce original content that is thematically/semantically similar to what it was given. GPT-2 was developed by OpenAI and has been leveraged for a wide range of text based natural language projects [20]. However, the original version is notably difficult to use for fine-tuning towards specific implementations. Our team therefore utilized GPT-2-simple, a version of GPT-2 developed by Max Woolf that has been re-translated through code wrapping in Python and TensorFlow [28, 29, 30]. GPT-2-simple takes the GPT-2 algorithm developed by OpenAI, and utilizes these wrappers to streamline the algorithm's tuning and execution.

For our purposes, we fine-tuned GPT-2-simple on a dataset of social media posts. Specifically, we used the Sentiment140 corpus of approximately 1.6 million Twitter posts gathered by Alec Go, Richa Bhayani, and Lei Huang [8]. Once GPT-2 trained on this corpus (we trained the 124MB model on the corpus for a total of 1,000 runs), the algorithm used our image caption (see 2 1) as a prompt before generating several lines of text. In line with how our corpus is structured, each line of the output represents a separate social media post related to that original caption (i.e., responses). However, since the Twitter data from our corpus is not articulated in narrative chunks, this output is largely nonsensical when taken together. To address this issue, we applied beginning and ending tokens to each tweet in the corpus, which then gave us the ability to stipulate cutoff points. The end result is that the model accepts a descriptive caption, generates a piece of text that it thinks should immediately follow, and then discards the original caption. This leaves a stub of Twitter-style text that is based off of the image description (see Table 1).

OCEAN Classification and Sorting

Finally, after text has been generated, we need a way to then validate any personality traits it appears to be showing. Because we intend to build personas with consistent traits, we need to be able to collect several posts that exhibit the same pattern of traits. In order to validate this text, we run it through a multi-label classifier. Multi-label classifiers categorize text by using multiple non-overlapping binary flags for indicating different attributes. As in Figure 2, each trait is given a probability of the presence or absence of that trait within the text, and are all mutually exclusive from one another. Thus in this model, one's Extroversion has no bearing on their Agreeableness, which in turn delineates the correct empirical approach for evaluating personality traits.

The model we used was built using TensorFlow, trained on the My Personality dataset generated by Kosinski et al. [15]. This dataset is a collection of 10,000 Facebook posts which have been coded for Big Five personality values. This ensures that we can create consistent personality profiles for each post. However, this method also bears some flaws. Ideally, a larger dataset of Facebook posts would result in a more robust model, since the current model often fails to understand text that does not appear in the original dataset. For example, certain instances of netspeak seem to have little impact on the personality rating. It is also difficult to re-interpret the binary presence or absence of personality traits onto the typical 1-5 scale used in personality research; ergo, the probability of a given trait occurring in a given body of text indicates nothing about the perceived strength of said trait. If the probability is in the middle (i.e. around 50%), does that necessarily mean that the person is neutral in regards to that trait? The answer to this is unclear, and may present a useful area for future exploration. This is currently defined as the last stage in our workflow, but we aim to collect the posts, and then move them into Truman as an example of persona creation.

Platform Implementation

As of now, the implementation of our platform is a bit rough. We would ideally have all of the systems working together to a greater degree such that the information is directly passed from one stage to another. Right now, we are passing the information indirectly from one step to another. Another implementation issue that we could address in the short term would be greater validation of automatically generated information. For example, when using GPT-2 to generate a caption, we have to select one based on what makes the most sense for that photo. We could perhaps implement some form of check that makes sure that GPT-2 captions more directly reference an element of the photo (maybe by using the caption generated from the image captioning step). Although this would make GPT-2 generation take a bit longer, we could hopefully avoid having to have eyes on each step of the process to the same degree.

Although this is currently unimplemented, the last step would be to build a collection of social media posts for each pattern of traits (i.e. a persona who is high on Extroversion, and low on everything else), and use those posts for that persona. One way this could be implemented into Truman would be by going through several revisions of GPT-2 generation for a caption, with a caption only "passing" if it is classified within certain limits for each persona. For example, to have a persona as the above example, a caption is only collected if its Extroversion probability is >80% and all other traits are <30%.

One potential issue with this implementation is that there will be little narrative captured between each post. In the current version of Truman, smaller narratives can take place over posts and in comments (for example, two personas get into a fight in the comments, with one declaring that the other is blocked, and not interacting with them anymore). Following DiFranzo's example, clever design might help reduce participant skepticism about the simulated platform's posts. For example, if the social media is framed such that other "users" are posting whatever they are observing at certain specified times, then a participant may be more willing to forgive a lack of cohesion in the posts shown. We can also reduce this by constraining the class of pictures that each persona can use. For example, a persona that is supposed to be in New York could have several photos only from New York, with a relatively consistent tone to the photos. Careful photo selection and clever design can thus help strengthen the illusion of the personas.

Another potential issue with implementation is that the current model doesn't account for other attributes as mentioned in the introduction (views, comments, likes, posting behavior, etc.). However, these could be set post-hoc for each persona. For example, any personas that score high on Extroversion might automatically receive more likes, and have more auto-generated comments on their posts (again using GPT-2 to generate social media-like posts, except by using the GPT-2 generated post text as the prompt). These sorts of implementation changes would help reduce the amount of necessary human intervention, while still preserving most of the useful aspects of personality research. This would be very useful for social media research, as one could easily tune what the social media looks like through the personas' exhibited behavior. For instance, an study based on this platform could examine how high versus low Agreeableness in personas might influence participant's willingness to troll or bully personas. By keeping other patterns the same, and only flipping a certain switch, one could therefore experimentally test the effect of Agreeableness on perceptions and actions within social media; if everyone on the social media seems to be a jerk, can we prove a new user is more likely to act like one too?

Finally, the current implementation is based upon personalty, as that is the example use case we decided to build for. However, this could easily be tweaked for other potential use cases, depending on the research goals of the study. For example, if one wishes to study if a reader's understanding of bridging social capital is different than their understanding of bonding social capital (as studies of social capital are common in the space of social media research [7, 2, 22]), then all that is needed is to train a different classifier. If one has a corpus of posts that are tagged correctly, then it would be easy to train a model to classify whatever one wishes to study. Other research goals could be achieved through Truman as well. If one is interested in how different affordances shape a user's understanding of the social media (such as ephemerality in Snapchat [26, 17, 19]), then they are able to encode said affordances into Truman. Do people really act worse if the only difference is anonymity? Alter Truman, and provide the exact same posts to each participant. The level of control gained for researchers through the machine learning models and Truman allow for a wide array of studies to be created.

FUTURE WORK

Our short term goal for this project would be to redesign the existing implementation of Truman, and experimentally test it against an equivalent version using the machine learning method as outlined above. We could measure participants on the correct identification of personality attributes, if they chose to post or interact with any agents, as well as self-report data on how easy it was to understand other users, how coherent or believable their posts seemed, and so on. This could also be compared to a baseline using real social media data to see how both methods rank against one another. Further research could then investigate effects of altering certain social variables on the platform. If we manipulate the presence or absence of Agreeableness in a group of social media posts, we could see how this potentially affects participants by measuring log file style data (amount of log-ins, posts, comments, content of comments) as well as participant self report ("On a scale of 1-5, how annoyed were you by the other posters") and qualitative data. This would help present a richer picture of how social mores or social presentation might influence social media behaviors in an empirical study with fairly high ecological validity.

In the long term, making this system available for other researchers would be an important contribution. For instance, many traditional methods and tools for social science research struggle at the intersection of human experience and technology because the breadth of knowledge, experience, and fine-tuned methodologies cannot readily be applied. Supporting personality attributes is itself only a fractional area but it is not hard to imagine that this workflow and set of tools could be re-implemented as a proof of concept to assist in other research domains.

LIMITATIONS

Our method presents a flexible framework for allowing social media researchers to begin to build their own empirical research plans. However, the current implementation has some issues. One potential issue is that the GPT-2 model and the multilabel classifier are trained on slightly different social media, and the output is formatted for a third type of social media. This issues could lead to potential conflicts in the overall effectiveness of the stages. For example, the GPT-2 model is trained on Twitter data, which is heavily text-based and places tight constraints on messages (including a strict character limit) [27, 13]. Facebook is relatively more picture-focused, although the posts the model was trained on don't involve images themselves. Finally, the visual output of Truman is formatted to be similar to Instagram, which is heavily image based, with the caption offering supporting context or a shift of frame. Although GPT-2 is flexible, it could be that the textual nature of Twitter would lead to a different output than if GPT-2 were to be fine-tuned on an equivalent corpus of

Instagram posts. Future research would be advised to verify whether datasets are justifiably equivalent, or, where possible, create their own unique datasets (using human coders for a new classifier dataset, for example).

CONCLUSION

Overall, this project was successful in accomplishing the project's more broad-stroked goals. We were able to take an image, create a relevant caption, and rate the caption on personality using only machine learning models with minimal human writing. This is a promising start for a set of tools that may be used by current or future researchers within the social media space. Future work along this path would need to clean up, streamline, and make the overall experience more user-friendly for less technical researchers. Furthermore, making the set of tools widely available would be a useful task in itself, as other researchers would be able to conduct empirical research and test long-held ideas toward online socialization. Other future work would be to start implementation of one or more example studies, so as to evaluate how useful the general system is. We aim to continue to build upon these tools in order to provide studies for other researchers, and develop further experiments for our own research goals.

REFERENCES

[1] Yair Amichai-Hamburger and Gideon Vinitzky. 2010. Social network use and personality. *Computers in Human Behavior* 26, 6 (Nov. 2010), 1289–1295. DOI: http://dx.doi.org/10.1016/j.chb.2010.03.018

[2] Moira Burke and Robert E. Kraut. 2014. Growing closer on facebook. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (2014), 4187–4196. DOI: http://dx.doi.org/10.1145/2556288.2557094 ISBN: 9781450324731.

[3] J. Clement. 2020. Global social media ranking 2019. (Feb 2020). https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

[4] Dominic DiFranzo, Samuel Hardman Taylor, Franccesca Kazerooni, Olivia D. Wherry, and Natalya N. Bazarova. 2018. Upstanding by Design: Bystander Intervention in Cyberbullying. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–12. DOI:http://dx.doi.org/10.1145/3173574.3173785

[5] John M. Digman. 1990. Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology* 41, 1 (1990), 417–440. DOI: http://dx.doi.org/10.1146/annurev.ps.41.020190.002221 arXiv: 1011.1669v3 ISBN: 00664308.

[6] Nicole Ellison, Rebecca Heino, and Jennifer Gibbs. 2006. Managing Impressions Online: Self-Presentation Processes in the Online Dating Environment. *Journal of Computer-Mediated Communication* 11, 2 (2006), 415–441. DOI: http://dx.doi.org/10.1111/j.1083-6101.2006.00020.x ISBN: 1083-6101.

[7] Nicole B. Ellison, Charles Steinfield, and Cliff Lampe. 2007. The benefits of facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication* 12, 4 (2007), 1143–1168. DOI: http://dx.doi.org/10.1111/j.1083-6101.2007.00367.x ISBN: 0521832969.

[8] Alec Go, Richa Bhayani, and Lei Huang. 2014. For Academics - Sentiment140. (2014). http://help.sentiment140.com/for-students

[9] Samuel D. Gosling. 2009. *Snoop: What your stuff says about you*. Basic Books, -.

[10] Samuel D. Gosling, Adam A Augustine, Simine Vazire, Nicholas Holtzman, and Sam Gaddis. 2011. Manifestations of Personality in Online Social Networks: Self-Reported Facebook-Related Behaviors and Observable Profile Information. *Cyberpsychology, Behavior, and Social Networking* 14, 9 (Sept. 2011), 483–488. DOI: http://dx.doi.org/10.1089/cyber.2010.0087

[11] Rachel Grieve and Jarrah Watkinson. 2016. The Psychological Benefits of Being Authentic on Facebook. *CyberPsychology, Behavior & Social Networking* 19, 7 (July 2016), 420–425. DOI: http://dx.doi.org/10.1089/cyber.2016.0010

[12] Jacob B. Hirsh and Jordan B. Peterson. 2009. Personality and language use in self-narratives. *Journal of Research in Personality* 43, 3 (2009), 524–527. DOI: http://dx.doi.org/10.1016/j.jrp.2009.01.006 ISBN: 0092-6566.

[13] David John Hughes, Moss Rowe, Mark Batey, and Andrew Lee. 2012. A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior* 28, 2 (March 2012), 561–569. DOI: http://dx.doi.org/10.1016/j.chb.2011.11.001

[14] Oliver P. John, Laura P. Naumann, and Cristopher J. Soto. 2008. Paradigm shift to the integrative Big Five Trait taxonomy. *Handbook of personality: Theory and research* (2008), 114–158. DOI: http://dx.doi.org/10.1016/S0191-8869(97)81000-8 ISBN: 978-1-59385-836-0.

[15] Michal Kosinski, David Stillwell, Pushmeet Kohli, Yoram Bachrach, and Thore Graepel. 2012. Personality and Website Choice. *WebSci* (2012), 0–3. ISBN: 9781450302678.

[16] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, -, -, 740–755.

[17] Sarah McRoberts, Haiwei Ma, Andrew Hall, and Svetlana Yarosh. 2017. Share First, Save Later: Performance of Self through Snapchat Stories. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. -, -, 6902–6911. DOI: http://dx.doi.org/10.1145/3025453.3025771

[18] Laura P. Naumann, Simine Vazire, Peter J. Rentfrow, and Samuel D. Gosling. 2009. Personality Judgments Based on Physical Appearance. *Personality and Social Psychology Bulletin* 35, 12 (2009), 1661–1671. DOI: http://dx.doi.org/10.1177/0146167209346309 ISBN: 0146-1672.

[19] Lukasz Piwek and Adam Joinson. 2016. "What do they snapchat about?" Patterns of use in time-limited instant messaging service. *Computers in Human Behavior* 54 (2016), 358–367. DOI: http://dx.doi.org/10.1016/j.chb.2015.08.026 Publisher: Elsevier Ltd ISBN: 0747-5632.

[20] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI Blog* 1, 8 (2019), 9.

[21] Christopher J. Soto and Oliver P. John. 2017. The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology* 113, 1 (2017), 117–143. DOI: http://dx.doi.org/10.1037/pspp0000096

[22] Charles Steinfield, Nicole B. Ellison, and Cliff Lampe. 2008. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology* 29, 6 (2008), 434–445. DOI: http://dx.doi.org/10.1016/j.appdev.2008.07.002 arXiv: 1011.1669v3 ISBN: 0193-3973.

[23] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the Inception Architecture for Computer Vision. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. CHI, -.

[24] Lee Taber and Steve Whittaker. 2018. Personality Depends on The Medium: Differences in Self-Perception on Snapchat, Facebook and Offline. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–13. DOI: http://dx.doi.org/10.1145/3173574.3174181

[25] Tensorflow Tutorials. 2020. Image captioning with visual attention. (2020). https://www.tensorflow.org/tutorials/text/image_captioning

[26] J. Mitchell Vaterlaus, Kathryn Barnett, Cesia Roche, and Jimmy A. Young. 2016. "Snapchat is more personal": An exploratory study on Snapchat behaviors and young adult interpersonal relationships. *Computers in Human Behavior* 62 (2016), 594–601. DOI: http://dx.doi.org/10.1016/j.chb.2016.04.029 Publisher: Elsevier Ltd.

[27] Sophie F Waterloo, Susanne E Baumgartner, Jochen Peter, and Patti M Valkenburg. 2018. Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media & Society* 20, 5 (May 2018), 1813–1831. DOI: http://dx.doi.org/10.1177/1461444817707349

[28] Max Woolf. 2019a. How To Make Custom AI-Generated Text With GPT-2. (2019). https://minimaxir.com/2019/09/howto-gpt2/

[29] Max Woolf. 2019b. Python package to easily retrain OpenAI's GPT-2 text-generating model on new texts. (2019). https://github.com/minimaxir/gpt-2-simple

[30] Max Woolf. 2019c. Train a GPT-2 Text-Generating Model w/ GPU For Free. (2019). https://colab.research.google.com/drive/1VLG8e7YSEwypxU-noRNhsv5dW4NfTGce#scrollTo=H7LoMj4GA4n_

[31] Kelvin Xu, Jimmy Lei Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard S. Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *2015 IEEE International Conference on Machine Learning*. -, -, 2048–2057. DOI: http://dx.doi.org/10.1109/ICDMW.2017.51