Contextual Analysis of Social Media: The Promise and Challenge of Eliciting Context in Social Media Posts with Natural Language Processing

Desmond U. Patton
dp2787@columbia.edu
Columbia University

William R. Frey
wf2220@columbia.edu
Columbia University

Kyle A. McGregor
mcgregork@mlhs.org
Lankenau Institute for Medical Research

Fei-Tzin Lee
fl2301@columbia.edu
Columbia University

Kathleen McKeown
kathy@cs.columbia.edu
Columbia University

Emanuel Moss
emoss@gradcenter.cuny.edu
City University of New York

ABSTRACT

While natural language processing affords researchers an opportunity to automatically scan millions of social media posts, there is growing concern that automated computational tools lack the ability to understand context and nuance in human communication and language. This article introduces a critical systematic approach for extracting culture, context and nuance in social media data. The Contextual Analysis of Social Media (CASM) approach considers and critiques the gap between inadequacies in natural language processing tools and differences in geographic, cultural, and age-related variance of social media use and communication. CASM utilizes a team-based approach to analysis of social media data, explicitly informed by community expertise. We use of CASM to analyze Twitter posts from gang-involved youth in Chicago. We designed a set of experiments to evaluate the performance of a support vector machine using CASM hand-labeled posts against a distant model. We found that the CASM-informed hand-labeled data outperforms the baseline distant labels, indicating that the CASM labels capture additional dimensions of information that content-only methods lack. We then question whether this is helpful or harmful for gun violence prevention.

KEYWORDS

qualitative analysis, NLP, ethics, social science

ACM Reference Format:
Desmond U. Patton, William R. Frey, Kyle A. McGregor, Fei-Tzin Lee, Kathleen McKeown, and Emanuel Moss. 2020. Contextual Analysis of Social Media: The Promise and Challenge of Eliciting Context in Social Media Posts with Natural Language Processing. In *Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Societyety (AIES '20), February 7–8, 2020, New York, NY, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3375627.3375841

AIES '20, February 7–8, 2020, New York, NY, USA

ACM ISBN 978-1-4503-7110-0/20/02…$15.00
https://doi.org/10.1145/3375627.3375841

1 INTRODUCTION

Automatic social media content analysis utilizing natural language processing tools has generated discussion as some researchers, communities and policymakers debate the extent to which natural language processing (NLP) can detect cultural influences and nuance in language or correctly decipher the goal or motivations of online speech [5]. Recent research underscores the complexities involved with interpreting text, particularly from communities of color. For example, off the shelf NLP tools incorrectly classified African American text as non-English [3] and have classified African American text incorrectly as hate speech [15]. The dangers and potential harms associated with automated social media analysis can be acute when used for digital surveillance where a lack of context regarding meaning and interpretation of language can have a detrimental impact on communities of color [6, 11, 13].

Decisions made by researchers engaged in NLP analysis require understanding the context of the data and how the algorithmic system will impact and transform behavior and socialization in the world. If an NLP system is not trained to understand context, it is unlikely the system will be able to accurately infer and interpret the meaning of the data [6]. Social media data pose specific challenges related to understanding context and data labeling for algorithmic system development, due to a wide variety of social media platform-specific digital lexicons, syntax, and semantics. This is further complicated by truncated and phonetically spelled text, emojis, and hashtags. There is a dearth of research on contextually driven methodologies for qualitatively analyzing and labelling social media data to use in supervised and semi-supervised machine learning techniques.

The domain-specific nature of social media requires domain expert insights and manual human labeling to accurately interpret and classify context and culturally specific implications of data. Without these insights, it is difficult for researchers to understand the context of social media data, which can lead to low quality annotations and inaccurately labeled training data. Current methods for labeling large amounts of data often rely on crowdsourcing platforms like Amazon Mechanical Turk, which allow researchers to access many annotators to quickly label their data. However, crowdsourced labeling has consistently had quality issues [10]. It is difficult to imagine how a large number of annotators unfamiliar with a domain would be able to label social media data beyond binary classifications where there is a 'right' and 'wrong' answer, to include analysis around meaning, sentiment, and context of digital social behaviors.

The current solutions to assess annotation and label quality as well as eliminate bias solely use computational methods [7, 14, 16]. In this paper, we introduce the Contextual Analysis of Social Media (CASM) approach to underscore the importance of qualitative methodologies for eliciting context when using NLP and other artificial intelligence tools.

CASM provides a methodological process for labeling social media data grounded in contextually driven and domain specific decisions leading to the training of an algorithmic system. It bridges an identified gap between inadequacies in current NLP tools and differences in geographic, cultural, and age-related variance of social media use and communication. CASM utilizes a team-based approach to annotating and qualitatively analyzing social media data, explicitly grounded by community expertise and understanding. This process yields rich qualitative analysis as well as in-depth annotations that easily feed into NLP systems to improve accuracy. However, the focus on context also offers an opportunity to think about the ethical risks of this project that are directly related to what improving accuracy enables the prediction and detection of human behavior. In this paper we engage the stakes of the remaining computational error rate and consequences associated with a well-functioning, automated system of context detection.

2 OUR CURRENT RESEARCH

We are engaged in a mixed-method process which includes a qualitative analysis of Twitter communication using the CASM approach to inform a set of machine learning algorithms that detect and predict loss and aggression in Twitter data. We study the Twitter communication of Gakirah Barnes and users in her network. Gakirah Barnes was a 17-year-old self-identified female member of a gang located on the Southside of Chicago. Gakirah created the Twitter ac-count @TyquanAssassin to memorialize her friend Tyquan Tyler who was killed by a rival gang in 2013. Gakirah posted over 27,000 tweets from December 2011 until her own death on April 11, 2014. She used the account to express events of her daily life, ranging from friendship, love, and happiness to trauma, gang violence, and grief.

Our dataset consists of 5,808 tweets by Gakirah Barnes and her top communicators. The initial dataset included many users who were inactive and users not relevant to the communities and contexts we study (e.g., celebrities not from Chicago), so we used snowball sampling to find 214 additional Twitter users in Chicago with social media connections to and engagements with either Gakirah Barnes or her top communicators. We have adapted the traditional snowball sampling approach for social media data by looking for clues and references from one social media user to find another social media user in the network who may be displaying similar behaviors or gang affiliations as the first user [1]. In total, our dataset consists of 279 users.

We apply the CASM approach on a corpus of social media data from youth in Chicago who live in neighborhoods with high rates of community violence. We describe a set of procedures used to contextualize and unpack meaning in text, images, and emoji. Finally, we compare the effectiveness of context in automatically detecting and predicting expressions of loss and aggression in Twitter data.

2.1 Data Acquisition and Corpus Development

Before implementing the CASM approach, there are several preplanning steps that are necessary. First, as with all research, it is important to clarify the research question(s) and study population(s) to ensure an in-depth, contextual approach meets the specific needs of the study. This clarification also involves unearthing considerations which may be specific to the study populations and the domain. Next, the researcher must identify or create a social media corpus by outlining inclusion criteria. For example, location, self-identified demographics of user(s), keywords, hashtags, and other features may be boundaries to include when creating a social media corpus for the CASM approach. Along with the inclusion criteria, it is important to outline potential harms caused by using specific inclusion criteria. Will the research shine a spotlight on specific users and put them in danger they otherwise may not be subjected to and how will these users' protection be considered and ensured? The identified social media corpus may contain language, community and cultural references, music lyrics, and ideas that are unfamiliar to individuals outside the community context. At this point, it is imperative to identify and consult with domain experts who can provide insights into localized language, events, and context that may impact how the social media corpus is perceived and analyzed [8].

It is helpful to consider a wide variety of domain expertise, as community members (including young people), sociolinguistics, ethnographers, and other people with specialized knowledge of the various aspects of the social media corpus all may have useful and vastly different knowledge to offer. Social media text is particularly challenging to decode as aspects of performativity, satire, jokes and the like are difficult to identify, define and understand across contexts. As such, it is important to keep track of domain expert insights and how they are represented within each social media post as these insights will be used for identifying and training annotators and their future data analyses.

In our case, we are interested in the role social media plays in gang violence. We identified a user who was mentioned in national news articles that met our inclusion criteria: 1) self-identified gang involvement; 2) frequent engagement on Twitter as evidenced by followers (5,000) and tweets (27,000 in a three-year span). We then consulted with domain experts at violence prevention organizations in the user's area, including the executive director of an organization, violence intervention workers, and formerly gang-involved youth to better understand our social media corpus. The annotation and human labeling of the social media corpus is a tedious and laborious process, one which we cannot expect domain experts to undertake.

We hire and train graduate student annotators to carry out this task. We selected annotators who are current students in a Master of Social Work program. Annotators selected have work experience in 1) adolescent development, 2) criminal justice, and 3) on-the-ground work experience with youth of color. The annotation training includes: 1) a general overview of the domain informed by our domain experts, 2) outlining their role as annotators (e.g., the tasks, purposes, and goals of the analyses), 3) in-depth annotation system tutorial, 4) a week-long deep immersion in the specific social media domain, and 5) annotation practice and feedback (Table 1). Our annotators gain additional insights from domain experts — women, men, and youth of color who have experience with or connections to gangs in Chicago, Illinois. Immersion in the specific social media domain includes a week-long review of twenty Twitter users from our corpus to familiarize themselves with our dataset. Our annotators observe the various ways users curate their online identities through what they share, how they portray themselves, who they engage with, and how frequently they post. For example, the user may post about their relationships, share entertaining videos, or share about their daily activities. An important aspect of this immersion is critical discussions of the ethics surrounding our observations of Twitter users.

After a week-long training, the annotators attend a process meeting with the expert annotator where they share and compare notes on what they observe surrounding each. The process meeting helps new annotators consider contextual features such as offline events, localized language, and emoji usage that may be missing from their initial observations. In week two of training, our annotators are tasked with annotating 100 social media posts. The expert annotator reviews the annotations for any mistakes made by the annotator, such as missing contextual features like images or links in posts, and not utilizing web-based resources when they do not understand domain-specific language and emoji use. The expert annotator provides each annotator with feedback. Finally, they are ready to begin annotating the official social media corpus in our annotation system annotation system [12]. Our annotators first complete the CASM approach individually, then meet weekly as a group to talk through their annotations, ask questions, and reconcile any disagreements.

2.2 Step 1: Baseline Interpretation

Social media data can take forms such as text, emoji, hashtags, memes, images, and videos. The first step in CASM involves collecting baseline data on the annotators' initial impressions and perceptions of the post before seeking any additional contextual information. Annotators are presented with a tweet that has been randomly selected from the corpus. They are then asked to describe in their own words their perception of what is happening in the tweet.

Annotators are trained to acknowledge that their interpretations are inherently informed by bias and their power as annotators. The training involves critically engaging the influence power has on determining the meaning of social media posts. The baseline analysis serves two purposes: First, this initial assessment evaluates what assumptions come up for the annotators. Second, it uncovers how their own positionality affects how they interpret the social media post, which may skew their analysis of the post.

2.3 Step 2: Annotation Process

The annotation process involves focused examination of all biographical and offline information found in the user's text, emojis and hashtags, images, videos, and personal profile page. This systematic process starts with analysis of the original social media post and expands to an analysis of the user's peer network, including any engagements and interactions with the original post.

Original Social Media Post. During this phase of the analysis, we examine the randomly selected social media post within our annotation system looking for specific mentions of names, communities, groups, schools, streets, and local institutions that may also be coded and may not be understood by individuals outside the local context. In addition, the annotators identify any words, phrases, emojis, and any other features, specifically identifying words, letters, numbers, punctuation, and abbreviations that may be used as contextually or culturally relevant features.

Utilize Web-Based Resources. We then investigate web-based resources (e.g., emojipedia, Urban Dictionary) to identify other cases in which the contextual or cultural features may be mentioned. This allows our annotators to see the features in various contexts which aids us in deciphering and triangulating meaning. Our annotators iteratively update lists of researched features which they use as a resource for future analysis.

Original Post's Author. Annotators go to the original post on Twitter and study the user who made the post. During this phase of the analysis the annotator examines usernames, reviews any biographical information (e.g., name, birth date, neighborhood, city) and any mentions of their specific location, and reviews photos for clues regarding location, gang affiliations, peer network, and environment. These contextual clues are used to better understand the conditions and factors that may shape a user's communication on social media. Next, the annotators review the last twenty posts from the user to situate their social media engagement. Are there any particular patterns in their posting? Does the post under analysis seem in line or out of place with the ways in which the user has posted previously?

Peer Network. Analysis of a user's peer network seeks to understand who they are connected to and interact with on social media. With any post, we identify anyone who may have been tagged (@) within the post. We then go to that user's page and ask two questions: 1) Who is this person in relation to the original user? 2) Why are they being tagged in the post? When a post is private, we no longer review the user and remove them from the dataset.

Offline Events. Annotators look to see if any offline events are being referenced. During this phase we identify the type of event mentioned, where that event took place, who is being referenced in connection to the event, and if any other user is tagged as being associated with the event. For example, users may reference a party that is happening, a death that has occurred, or remembering a memory that happened offline. Additionally, if we know of an event in a user's neighborhood and the user's posts do not mention it, this also provides valuable information. Offline events often contain contextual features which are specific to a certain domain.

Virality. Annotators review the virality of the post and how far the post is traveling within the network. Our annotators look for features that may be causing a post to have a high potential for virality. For example, we look at who is retweeting or liking a post. Then, they look to confirm any relationship the users have to the author of the original post.

Engagement. Annotators look at the people who reply or comment on the post. Who are the users replying or commenting on the post? What aspects of the post are they replying to or commenting on? Is there content that they are gravitating toward? Can we infer intent in their reply or general comments on the post? Are they attempting to escalate or deescalate the post? For example, we pay attention to whether the commenter is asking a question, questioning the legitimacy or authenticity of the post, or adding additional information.

2.4 Step 3: Interpretation & Contextual Analysis Assessment

After determining the contextual features of the original social media post through community insights, research, and textual, user, and peer network analysis, annotators assess their baseline interpretation. They start by comparing their initial perceptions of the social media post with the contextual features they have uncovered throughout knowledge procurement and contextual analysis. They review where they made assumptions around meaning. Then, annotators explain what they have found in detail, elaborating on the meaning of the original post and the evidence for their determinations. This includes a thorough explanation of the features and meaning they have uncovered throughout the textual and contextual analysis.

For example, if a post contains an image of a person pointing a gun at the camera, an annotator’s initial reaction could be fear leading them to think this user is threatening someone. However, once the annotator goes through the annotation process, they find out that the person is not making a threat. The user is making a joke. Many people would not go through an annotation process to realize that the post is intended to be a joke. However, this first impression of fear could still be a useful interpretation of (possibly inaccurate) ways a user could respond to the tweet.

2.5 Step 4: Labeling

The final task our annotators complete is labeling the social media post. Our annotators go through an iterative process to consider all the contextual features they have unearthed through their analysis to determine the ‘essence’ of the post. The ‘essence’ takes into account the potential intent of the post’s author, while recognizing the various ways the post could be perceived and interpreted by other users on social media. While considering these various potential viewpoints, our annotators apply a label to the post (Table 1).

Label	Examples
Aggression	Threats, Insults, Physical Violence, Taunting
Loss	Memorials, Grief, Incarceration, Death

Table 1: These labels were developed through qualitative analysis and refined through the CASM approach.

2.6 Step 5: Community Validation & Reconciliation

Once the social media post is labeled, the labeled post is reviewed by a domain expert. In this study, we employ two types of domain expertise. First, we consult the expertise of community members from which the social media data derives. These domain experts are Black individuals who have experience with gangs in Chicago, either professionally through intervention work or personally through their own involvement in or affiliations with gangs. Second, we consult the social work researchers on the team who together have over two decades of experience working with youth and are responsible for developing the coding scheme for the study. The post is labeled again by domain experts, who offer insights into their reasoning behind the label they have provided. These labels and insights are used to reconcile the labeling between our annotators and the domain experts to create a final label for the social media post, either aggression, loss, or other. Finally, the labeled social media dataset is sent to the computer science team to use in NLP and machine learning experiments.

3 APPLYING CASM TO NLP

We designed a set of experiments to evaluate if training an NLP classifier on the Twitter data labeled using CASM performs better than training on the same data without a contextual approach. Our experiments utilize a linear-kernel SVM classifier originally described in [2] and used as a strong baseline in [4]. In this method, after basic preprocessing is performed to remove links and tags, unigram, bigram, part-of-speech tag, and emotion features are extracted. Feature selection is performed to prune the feature space. The part-of-speech tagger used in [2] was developed specifically for use on this domain through training on a subset of the corpus labeled with part-of-speech tags. Emotion features are computed using the Dictionary of Affect in Language (DAL). We re-trained on the larger training set in [4] (expanded from our initial work in [2]) to achieve the best performance with the system. We performed grid search to re-tune the loss function, the regularization penalty type and the penalty parameter C. We found that the original settings still performed best even on the new development set.

To examine how well CASM helps in the automatic classification task, we trained the SVM on two separate training sets: one labeled with the qualitative labels only, and one with distant labels that are automatically inferred based on the presence or absence of hand-picked indicator words. The latter method should provide a strong baseline for the performance of distant labeling - while not context-sensitive, it still incorporates the domain expertise of the annotators. We found that the gun emoji (🔫) had highest correlation with aggression and the praying hands emoji (🙏) had highest correlation for loss. For the other label, we randomly sampled from tweets not containing any of the indicators for loss or aggression. We used this method to label previously unlabeled tweets in our much larger dataset. Any tweet with words from the loss indicator set were labeled as loss; tweets with aggression indicators were labeled as aggression. We trained the SVM first with only the qualitative labels and then using only the distantly labeled tweets. We then compared their performance on the labeled test set.

	Loss			Other			Aggression			
Labels	p	r	f	p	r	f	p	r	f	Macro F1
Gold	77.08	56.92	65.49	88.04	95.76	91.74	50	27.59	35.56	64.26
Distant	50.00	48.46	49.22	85.63	84.50	85.06	19.72	24.14	21.71	52.00

Table 2: SVM performance trained on hand-labeled vs distantly-labeled data. The difference between F1 scores is statistically significant with p=0.001.

On the test set, the model trained on hand labels achieved an f-score of 64.26, while the distant model scored 52.00. Details showing precision and recall can be seen in the accompanying table (Table 2). When considering the question of whether it is worthwhile to manually annotate a corpus, it is useful to compare performance of a model using this data to performance of the same model on a dataset created without such investment of effort. In this case, we use as our baseline, performance of our SVM model on a distantly labeled dataset where samples are labeled by the presence or absence of handpicked indicator words for each class - a simple and natural method of gathering data that makes use of our annotators' expertise and does not require them to manually label thousands of tweets. Distant labeling has been used in the past to label sentiment and emotion tweets using hashtags present in the posted tweets [9]. The fact that the same model trained on hand-labeled data substantially outperforms this baseline indicates that these manual annotations are in fact highly useful to our models, and that they do provide additional information that distant labeling does not.

In addition to our work using support vector machines, we have also developed a classifier using a neural net approach implemented using a Convolutional Neural Network (CNN) [4]. This work was novel in that it used contextual information about the content of past posts by the user of the post the system is currently classifying as well as information about the emotional content of past posts. This recent work also used the same labeled dataset, annotated using CASM. We were able to show that context enabled significant improvement over our baselines and thus, this work is another indication of the importance of context in developing NLP systems for classification.

4 ETHICS

Implementation of the CASM approach requires iterative and ongoing foundational considerations regarding the ethics of interpretation, analysis and sharing of social media data. Our ethics discussion attempts to wrestle with the real-life tensions inherent in using artificial intelligence to study human behavior grounded in violence prevention efforts. Our work sits between two critical issues: 1) Black families wanting their children to be safe and desiring tools that help achieve these ends and 2) digital surveillance and policing enacting and enhancing yet another form of state violence on Black people and communities.

Research involving publicly available social media data has the potential to (in)directly impact study populations in harmful ways. Ethical obligations include clarity of the context and potential vulnerabilities specific to each study population (e.g., heightened police surveillance), adopting various mechanisms to protect the study population (e.g., encrypting and de-identifying the data), and ensuring the research does not amplify vulnerabilities or create further marginalization or harm.

While our research uses publicly available tweets, the users in our dataset (Black youth) face varying levels of marginalization, criminalization, and police surveillance online and offline. We contend with the fact that although our system is arguably "accurate" because we leverage qualitative insights and context, a more accurate system might also indicate harm in this context. The ability to automatically identify aggressive and threatening content from Black youth can also be used as evidence in the criminal justice system, creating an automated pipeline towards furthering e-carceration. Any study that utilizes the CASM approach should be accompanied by a robust set of ethical guidelines that ask the study team to consider: 1) real-world consequences of applying algorithmic tools to complex social problems; 2) measurement for success of NLP outputs (e.g., is "accuracy" an appropriate measure of success?); and 3) the extraction of context for NLP systems and how it is derived, analyzed, and validated.

With these consideration in mind, we implemented various mechanisms to protect our study population from further harm. First, before our annotators are given access to the dataset, they are required to sign the aforementioned EAA. The EAA also includes steps for accountability if one of the practices are not followed. Second, when sharing our work through publications and presentations, we de-identify all social media posts, rendering the text unsearchable, and use images from Flickr: Creative Commons rather than from our dataset to avoid shining a spotlight on our users in the sea of social media posts. Third, we only share our dataset with community partners and other researchers who sign a Memorandum of Understanding (MOU) outlining the intentions and purposes for using the social media dataset. CASM's in-depth interpretation of social media posts requires a dynamic and adaptive understanding of the ethical obligations regarding the safety and protection of social media users. Effective use of CASM requires a critical consideration of context and which tools fit that context.

5 CONCLUSION

The purpose of this article is to present a new method, developed by social work researchers and computer scientists, for confronting bias, leveraging community and domain expertise, and unpacking the promise and challenge of extracting contextual features in social media data. For data scientists, the ability to make sense of and accurately classify social media data is of prime concern [5]. Unsupervised and semi-supervised learning systems may struggle to make meaning of social media posts, which is especially true for data from communities of color [3, 13].

The CASM approach was specifically developed to unearth the nuances and complexities of language within social media posts of Black youth in Chicago; however, throughout development it has become clear that this approach can be utilized to analyze communications in a wide variety of contexts. A possible extension of this approach could be used to develop language identifiers for alt right/hate groups who use non-standard vernacular, syntax, and emoji to communicate across multiple forms of social media. CASM is a group and context agnostic methodology that has wide applicability for use in any culturally specific language enclave where traditional off the shelf language identifiers will underperform due to lack of nuance, context, and culturally specific linguistic expertise.

REFERENCES

[1] Rowland Atkinson and John Flint. 2001. Accessing hidden and hard-to-reach populations: Snowball research strategies. *Social research update* 33, 1 (2001), 1–4.

[2] Terra Blevins, Robert Kwiatkowski, Jamie MacBeth, Kathleen McKeown, Desmond Patton, and Owen Rambow. 2016. Automatically Processing Tweets from Gang-Involved Youth: Towards Detecting Loss and Aggression. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. The COLING 2016 Organizing Committee, Osaka, Japan, 2196–2206. https://www.aclweb.org/anthology/C16-1207

[3] Su Lin Blodgett and Brendan O'Connor. 2017. Racial Disparity in Natural Language Processing: A Case Study of Social Media African-American English. *CoRR* abs/1707.00061 (2017). arXiv:1707.00061 http://arxiv.org/abs/1707.00061

[4] Serina Chang, Ruiqi Zhong, Ethan Adams, Fei-Tzin Lee, Siddharth Varia, Desmond Patton, William Frey, Chris Kedzie, and Kathy McKeown. 2018. Detecting Gang-Involved Escalation on Social Media Using Context. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium, 46–56. https://doi.org/10.18653/v1/D18-1005

[5] Natasha Duarte, Emma Llanso, and Anna Loup. 2018. Mixed Messages? The Limits of Automated Social Media Content Analysis.. In *FAT*. 106.

[6] Madeleine Clare Elish and danah boyd. 2018. Situating methods in the magic of Big Data and AI. *Communication monographs* 85, 1 (2018), 57–80.

[7] Benjamin Fish, Jeremy Kun, and Ádám D Lelkes. 2016. A confidence-based approach for balancing fairness and accuracy. In *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, 144–152.

[8] William R Frey, Desmond U Patton, Michael B Gaskell, and Kyle A McGregor. [n. d.]. Artificial intelligence and inclusion: Formerly gang-involved youth as domain experts for analyzing unstructured twitter data. *Social Science Computer Review* ([n. d.]), 0894439318788314.

[9] Sharath Chandra Guntuku, Anneke Buffone, Kokil Jaidka, Johannes C Eichstaedt, and Lyle H Ungar. 2019. Understanding and measuring psychological stress using social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 214–225.

[10] Robert Kern, Christian Zirpins, and Sudhir Agarwal. 2008. Managing quality of human-based eservices. In *International Conference on Service-Oriented Computing*. Springer, 304–309.

[11] Alexandra Mateescu, Douglas Brunton, Alex Rosenblat, Desmond Patton, Zachary Gold, and danah boyd. 2015. Social Media Surveillance and Law Enforcement. In *Data and Civil Rights*. 2015–2027.

[12] Desmond Patton, Philipp Blandfort, William Frey, Rossano Schifanella, and Kyle McGregor. forthcoming. VATAS: An Open-Source Web Platform for Visual and Textual Analysis of Social Media. *Journal of the Society for Social Work and Research* (forthcoming).

[13] Desmond Upton Patton, Douglas-Wade Brunton, Andrea Dixon, Reuben Jonathan Miller, Patrick Leonard, and Rose Hackman. 2017. Stop and frisk online: theorizing everyday racism in digital policing in the use of social media for identification of criminal conduct and associations. *Social Media+ Society* 3, 3 (2017), 2056305117733344.

[14] Vikas C Raykar and Shipeng Yu. 2012. Eliminating spammers and ranking annotators for crowdsourced labeling tasks. *Journal of Machine Learning Research* 13, Feb (2012), 491–518.

[15] Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. 2019. The Risk of Racial Bias in Hate Speech Detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 1668–1678. https://doi.org/10.18653/v1/P19-1163

[16] Noah Youngs, Dennis Shasha, and Richard Bonneau. 2015. Positive-unlabeled Learning in the Face of Labeling Bias. In *International Conference on Data Mining Workshop (ICDMW)*. 639–645.