From Text to Context: Contextualizing Language with Humans, Groups, and Communities for Socially Aware NLP

Adithya V Ganesan[1], **Siddharth Mangalik**[1], **Vasudha Varadarajan**[1], **Nikita Soni**[1], **Swanie Juhng**[1], **João Sedoc**[2], **H. Andrew Schwartz**[1], **Salvatore Giorgi**[3,4] and **Ryan L Boyd**[1]

[1]Stony Brook University [2]New York University [3] University of Pennsylvania
[4] National Institute on Drug Abuse, Intramural Research Program

bit.ly/text2context

1 Description

NLP has conventionally focused on modeling words, phrases, and documents. However, human psychology and behavior underpin the substance of Natural Language. Motivated by the idea that natural language is primarily generated by people, the field has recently witnessed a growth of interdisciplinary empirical work that integrates person-level information. For example, methods have been introduced to model person-level difference in meaning (Welch et al., 2022; Lynn et al., 2017), disentangle group-level biases and dynamics (Hovy and Søgaard, 2015; Shah et al., 2020), and even expose society-level processes reflected in language (Giorgi et al., 2022; Curtis et al., 2018). A demand has emerged for NLP researchers and practitioners to develop a deeper understanding of the individuals, groups, and societies that shape all forms of natural language (Hovy and Yang, 2021).

Natural language is inherently human — neglecting the personal and social aspects of language creates a gap in understanding the function, meaning, and processes that drive natural language (Hovy and Yang, 2021; Flek, 2020). These factors span from individual attributes up to cultural norms of communities. Previous works have demonstrated the importance of contextualizing these social factors along with language in order to better understand the humans behind it (e.g., Volkova et al., 2013; Lukin et al., 2017).

To make NLP systems aware of the linguistic aspects of the multiple levels of human factors, multiple disciplines within the field are beginning to adopt models that consider the hierarchical structure of human influence upon language — specifically, author differences, close-knit group dynamics, and larger societal contexts, as shown in Figure 1. Such influences already permeate texts written by humans; by leveraging established patterns in human thought, emotion, and interpersonal behavior, we enrich our ability to model natural language. Works that integrate the individual author factors, such as age and gender, have found that they can meaningfully improve performance in NLP tasks (Long et al., 2017; Hovy, 2015). Likewise, when studying group dynamics, inclusion of social networks have improved model performance (Yang and Eisenstein, 2017; Farnadi et al., 2018; Mishra et al., 2018; Del Tredici et al., 2019). This effect has also borne out at the community-level, where careful consideration of the socio-demographics of authors improves model outcomes (Curtis et al., 2018; Zamani et al., 2018). Intentional inclusion of the larger contexts that language exists within has become a fundamental component of state-of-the-art modeling techniques.

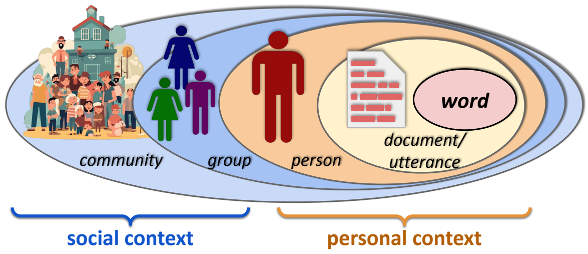

Figure 1: A depiction of the hierarchical structure of how humans influence language. Language found in personal contexts are used to transmit human thought, while also containing direct and latent attributes of the groups they socialize with and cultural aspects of their communities. These levels go beyond traditional view of NLP of seeing language composed of just words, phrases or even documents.

Aimed at the NLP researchers or practitioners who would like to integrate human – individual, group, or societal level factors into their analyses, this tutorial will cover recent techniques and libraries for doing so at each level of analysis. Starting with human-centered techniques that provide benefit to traditional document- or word-level NLP tasks (Garten et al., 2019; Lynn et al., 2017), we undertake a thorough exploration of critical human-

Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 5: Tutorial Abstracts), pages 26–33

level aspects as they pertain to NLP, gradually moving up to higher levels of analysis: individual persons, individual with agent (chat/dialogue), groups of people, and finally communities or societies.

Techniques covered will range from controlling for and correcting biases across demographics, socioeconomic, and other extra-linguistic variables, to leveraging the inherent multi-level structure and placement of language in social contexts. Taken together, participants will acquire techniques for modeling language in human-context that not only offer opportunities for improved accuracies, but also suggest improvements to fairness and social sensibility of NLP in our increasingly digital world.

In selecting topics to cover, we have considered both recency as well as some degree of demonstrated generalization – empirical tests across many domains by the original authors themselves or via replication of the underlying concepts by others. Approximately half of the tools we discuss are developed by others, while those techniques developed by the presenters span multiple labs and even fields of expertise.

In this tutorial, we will detail how emerging techniques tackling this problem confer important advantages across traditional NLP tasks. Since natural language, at its core, is an expression of human cognition and communication (Boyd and Schwartz, 2021), we pay particular attention to methods that draw on theories by researchers in fields as diverse as psychology, sociology, engineering, linguistics and beyond. Our aim is that this tutorial will inspire new researchers to push the boundaries of NLP, such that a new version of this tutorial will be necessary in short order.

2 Type of Tutorial

The tutorial will introduce research that has successfully integrated personal and social factors into traditional NLP as a foundation for **cutting-edge** research in the field. This multidisciplinary work has not been presented at prior *CL tutorials and is timely, given recent excitement in the *CL community for human-aware NLP systems. Unique aspects of this tutorial will include 1) interdisciplinary methods woven together into a coherent framework for human-centered NLP, 2) theory and domain expertise from an interdisciplinary team of presenters, and 3) hands-on demonstrations that facilitate *immediate* uptake and application by attendees[1].

3 Target Audience & Pre-Requisites

Our intended audience for this tutorial is experienced as well as upcoming NLP researchers looking to add human and social contexts to traditional NLP tasks. We expect this tutorial will attract 70-100 attendees.

We expect that attendees will arrive with a practical baseline knowledge of machine learning and computational linguistics. Specifically, we anticipate that our audience will be familiar with Transformer based NLP models, and canonical tasks that the field has been applied to such as: document classification, stance detection, etc.

4 Outline

Introduction (15 minutes)

The 3 hour tutorial will begin with a brief overview of the entire session organized from the individual- to the societal levels of context. We will also introduce the key concepts in behavioral and social science that motivate the techniques that will be discussed in the subsequent sections.

Individual Human Context (40 minutes)

In this session, we will review the methods for producing user representation from language, ranging from simple N gram features to advanced techniques such as Latent Dirichlet Allocation (Schwartz et al., 2013), Word2Vec (Amir et al., 2017; Benton et al., 2016), and Transformer models (Matero et al., 2019; V Ganesan et al., 2021). Importantly, these language-based user representations gain considerable power and effectiveness when integrated with user–level factors (Benton et al., 2016; Huang and Paul, 2019) for analyses. Such user factors include, but are not limited to, personal attributes such as age, gender, personality traits, and past experiences that characterize and differentiate people from one another.

We will showcase different user factor adaptation methods for merging human and social factors with language representations (Yang and Eisenstein, 2017; Lynn et al., 2017). While these methods produce user representations by taking a person's full picture into account, it is also pivotal to preserve the privacy of the individuals. Thus we will also review works (Sawhney et al., 2023;

[1] all materials will be available on bit.ly/text2context

Alawad et al., 2020) demonstrating the successful implementation of human-level NLP systems incorporating differential privacy (Dwork and Roth, 2014) to ensure secure and privacy-preserving NLP practices.

Individuals with Agents (35 minutes)

One way in which NLP systems can see a considerable improvement in their effectiveness/performance is through explicit modeling of the reciprocal influence between the user(s) and the context within which interactions occur. For example, the language that a person generates is determined not only by their accumulated traits, demographics, and psychological characteristics, but also by immediate and distal contextual factors such as the nature of the relationship between communicators, their individual discourse goals, and the broader characteristics of the situation according to psychological theory.

This session will begin by considering the "generator" of language and its mathematical formulation, explicitly beginning with the notion of language emerging in the context of an individual person's collected history of verbal behavior (Soni et al., 2022). Next, we will look at how *individuals* or personas make their way into dialogue and conversational AI systems (Li et al., 2016; Qian et al., 2018), leading to a marked improvement in the modeling of social interactions above and beyond person–level modeling strategies. Finally, we introduce psychology-grounded metrics aimed at assessing conversational AI on an individual level (Giorgi et al., 2023) and how they contrast with the more traditional automatic dialog metrics (Rodríguez-Cantelar et al., 2023).

Break (30 minutes)

Groups as Context (35 minutes)

We will go over the methods that place emphasis on treating individuals and groups as interactive entities, with the individual's interactions within a group adding context to documents (Del Tredici et al., 2019; Sawhney et al., 2021; Zamani and Schwartz, 2021). Drawing inspiration from adjacent fields, particularly computational social science, we will show how to analyze the language of user-associated groups (Goldberg et al., 2015), unveil valuable insights into the context of an individual, the evolving dynamics of group language usage over time (Danescu-Niculescu-Mizil et al., 2013), and its influence on individual language patterns (Danescu-Niculescu-Mizil et al., 2011; Ashokkumar and Pennebaker, 2022). By incorporating code demonstrations and references, we will discuss how these methods can enrich multiple traditional NLP tasks.

Communities (40 minutes)

This tutorial session will cover the basics of creating language estimates of spatial communities (e.g., U.S. states or provinces in China). We will cover topics such as aggregation, as in how to move from documents to communities *through* people (Giorgi et al., 2018), selection biases (Giorgi et al., 2022), ecological fallacies (i.e., language patterns at the individual level do not always hold at the community level; Jaidka et al. 2020), and cultural considerations (Havaldar et al., 2023). Participants in this session will be provided with a code notebook to experiment with on their own to examine the gains from proper methods for handling community-level text.

Wrap Up (15 minutes)

We will end the tutorial by briefly summarizing the topics covered across all the sessions, distinguishing the situations for which methods are appropriate, concluding with a perspective on the future of human-centered NLP.

Other than the introduction and wrap-up, the other sessions will have around 70% of the time allocated to talks, followed by interactive sessions with code demonstrations and questions from the audience.

5 Reading List

- User representation through language (Benton et al., 2016; Soni et al., 2022)
- Individual level dialog models (Li et al., 2016)
- Human factor adaptation (Hovy, 2015; Lynn et al., 2017; Soni et al., 2024)
- Groups as Individual Context (Ashokkumar and Pennebaker, 2022; Goldberg et al., 2015)

6 Breadth of Tutorial

Owing to the diverse nature of the sessions and the presenters' backgrounds, about two-thirds of the materials will encompass contemporary research works from other teams, with the other third coming from our works for this tutorial (Schwartz et al., 2013; Soni et al., 2022; Lynn et al., 2017; Giorgi et al., 2022; Jordan et al., 2019).

7 Diversity Considerations

We are an interdisciplinary team composed of computer scientists and psychologist across 3 institutions. We intend to leverage multiple levels of expertise to be accessible to an audience with varied fluency. We have 4 highly-experienced researchers (3 Professors, 1 Data Scientist at NIH) and 5 rising researchers (each with one or more *CL publications). Presenters span multiple demographics, ethnicities, and non-neurotypical backgrounds. This tutorial is aimed at encouraging more human-aware NLP systems through the incorporation of personal, demographic and cultural attributes of the speaker.

8 Tutorial Presenters

Salvatore Giorgi is a senior data scientist for the National Institute of Drug Abuse and the World Well Being Project at University of Pennsylvania. His research focuses on multi-level NLP and bias mitigation. Webpage: https://sjgiorgi.github.io/

João Sedoc is an Assistant Professor in the department of Technology, Operations and Statistics at New York University Stern School of Business. João's research areas are at the intersection of machine learning and natural language processing. His interests include conversational agents, model evaluation, deep learning, and crowdsourcing. Webpage: https://stern.nyu.edu/faculty/bio/joao-sedoc

H. Andrew Schwartz is an Associate Professor at Stony Brook University and Director of the Human Language Analysis Lab. His research focuses on interdisciplinary human-centered NLP, publishing in both computational linguistics and psychological science venues. Webpage: https://www3.cs.stonybrook.edu/~has/

Ryan L. Boyd is a psychologist and computational social scientist. His research uses behavioral science methods to understand how verbal behavior provides clues to how we think, feel, and behave, focusing on domains ranging from personality to society, mental health, human sexuality, and storytelling (e.g., Boyd et al., 2015, 2020). Webpage: https://www.ryanboyd.io

Adithya V Ganesan is a Computer Science PhD student at the Stony Brook University, with research focusing on building NLP systems for Psychological applications. Webpage: https://adithya8.github.io

Siddharth Mangalik is a Computer Science PhD student at Stony Brook University. His research work focuses on methods for examining the language of large-scale communities across time. Webpage: https://smangalik.github.io/

Vasudha Varadarajan is a Computer Science PhD student at Stony Brook University. Her research focuses on using discourse-level NLP for understanding cognitive styles, and also on improving language-based mental health assessments. Webpage: https://vasevarad.github.io

Nikita Soni is a Computer Science PhD student at Stony Brook University. Her research focuses on large language modeling in the additional context of the human behind the language. Webpage: https://www3.cs.stonybrook.edu/~nisoni/

Swanie Juhng is a Computer Science PhD student at Stony Brook University. Her research focuses on developing NLP and ML systems to understand the context of psychological conditions. Webpage: https://swaniejuhng.github.io

9 Ethics Statement

As with most human centered NLP tasks, one must carefully consider issues of privacy and consent, as well as social context and unintended downstream applications. Human level data, which encompasses text as well as non-linguistic data such as self-reports (surveys or health records, for example) and inferred factors (such as language-based estimates of gender or personality), may contain sensitive or identifying information. Thus, care must be taken when collecting, storing, and analyzing data, as well as presenting results (e.g., directly quoting text), in order to not publicize private data or identify individuals. For example, Reddit forums are often self-moderated intimate communities where users may anonymously discuss private and sensitive details related to, among others, mental and physical health, substance use and recovery, and parenting. Identifying personal accounts in such contexts may be especially harmful to individuals (Proferes et al., 2021). Similarly, many studies which use publicly available social media data are classified as not involving human subjects and exempt from Institutional Review Board approval. Thus, the humans behind the social media accounts do not explicitly consent to research studies (Chancellor et al., 2019).

There are also ethical issues around inferring human factors using NLP or machine learning methods. Common tasks such as inferring sociodemographics can suffer from limited representation in data sets (sample biases) or narrow definitions of social constructs (e.g., binary gender). Misclassifications can have unintended downstream consequences which, as more automated systems are deployed in real world situations, are becoming increasingly consequential (Mehrabi et al., 2021). Many algorithms designed to address such issues and remove biases often further marginalize vulnerable groups (Xu et al., 2021).

On the other hand, incorporating human factors may help alleviate biases. For example, when removing selection biases from population-level estimates one must know the socio-demographics of the people within the sample. In the current context, for example, this could mean estimating human factors, such as age and income, at scale across millions of Twitter users. Dialog agents, as another example, can run the risk of mimicking the social and cultural biases in their training data. Thus, forcing diverse ranges of human factors on agents may make them more diverse. Given this range of concerns, addressing ethical issues will be woven into each section of the tutorial.

References

Mohammed M. Alawad, Hong-Jun Yoon, Shang Gao, Brent J. Mumphrey, Xiao-Cheng Wu, Eric B. Durbin, Jong Cheol Jeong, Isaac Hands, David Rust, Linda Coyle, Lynne Penberthy, and Georgia D. Tourassi. 2020. Privacy-preserving deep learning nlp models for cancer registries. *IEEE Transactions on Emerging Topics in Computing*, 9:1219–1230.

Silvio Amir, Glen Coppersmith, Paula Carvalho, Mario J Silva, and Bryon C Wallace. 2017. Quantifying mental health from social media with neural user embeddings. In *Machine Learning for Healthcare Conference*, pages 306–321. PMLR.

Ashwini Ashokkumar and James W Pennebaker. 2022. Tracking group identity through natural language within groups. *PNAS Nexus*, 1(2):pgac022.

Adrian Benton, Raman Arora, and Mark Dredze. 2016. Learning multiview embeddings of Twitter users. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 14–19, Berlin, Germany. Association for Computational Linguistics.

Ryan L. Boyd, Kate G. Blackburn, and James W. Pennebaker. 2020. The narrative arc: Revealing core narrative structures through text analysis. *Science Advances*, 6(32):1–9. Publisher: American Association for the Advancement of Science Section: Research Article.

Ryan L. Boyd and H. Andrew Schwartz. 2021. Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology*, 40(1):21–41. Publisher: SAGE Publications Inc.

Ryan L. Boyd, Steven R. Wilson, James W. Pennebaker, Michal Kosinski, David J. Stillwell, and Rada Mihalcea. 2015. Values in words: Using language to evaluate and understand personal values. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media*, pages 31–40.

Stevie Chancellor, Eric PS Baumer, and Munmun De Choudhury. 2019. Who is the" human" in human-centered machine learning: The case of predicting mental health from social media. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–32.

Brenda Curtis, Salvatore Giorgi, Anneke EK Buffone, Lyle H Ungar, Robert D Ashford, Jessie Hemmons, Dan Summers, Casey Hamilton, and H Andrew Schwartz. 2018. Can twitter be used to predict county excessive alcohol consumption rates? *PloS one*, 13(4):e0194290.

Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. 2011. Mark my words! linguistic style accommodation in social media. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, page 745–754, New York, NY, USA. Association for Computing Machinery.

Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, pages 307–318, New York, NY, USA. ACM.

Marco Del Tredici, Diego Marcheggiani, Sabine Schulte im Walde, and Raquel Fernández. 2019. You shall know a user by the company it keeps: Dynamic representations for social media users in NLP. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4707–4717, Hong Kong, China. Association for Computational Linguistics.

Cynthia Dwork and Aaron Roth. 2014. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9:211–407.

Golnoosh Farnadi, Jie Tang, Martine De Cock, and Marie-Francine Moens. 2018. User profiling through deep multimodal fusion. In *Proceedings of the*

Eleventh ACM International Conference on Web Search and Data Mining*, pages 171–179.

Lucie Flek. 2020. Returning the N to NLP: Towards contextually personalized classification models. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7828–7838, Online. Association for Computational Linguistics.

Justin Garten, Brendan Kennedy, Joe Hoover, Kenji Sagae, and Morteza Dehghani. 2019. Incorporating demographic embeddings into language understanding. *Cognitive Science*, 43(1):e12701.

Salvatore Giorgi, Shreya Havaldar, Farhan Ahmed, Zuhaib Akhtar, Shalaka Vaidya, Gary Pan, Lyle H Ungar, H Andrew Schwartz, and Joao Sedoc. 2023. Human-centered metrics for dialog system evaluation. *arXiv preprint arXiv:2305.14757*.

Salvatore Giorgi, Veronica E Lynn, Keshav Gupta, Farhan Ahmed, Sandra Matz, Lyle H Ungar, and H Andrew Schwartz. 2022. Correcting sociodemographic selection biases for population prediction from social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 228–240.

Salvatore Giorgi, Daniel Preoţiuc-Pietro, Anneke Buffone, Daniel Rieman, Lyle Ungar, and H. Andrew Schwartz. 2018. The remarkable benefit of user-level aggregation for lexical-based population-level predictions. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1167–1172, Brussels, Belgium. Association for Computational Linguistics.

Amir Goldberg, Govind Manian, Will Monroe, Christopher Potts, and Sameer B. Srivastava. 2015. Fitting in or standing out? The tradeoffs of structural and cultural embeddedness. *Academy of Management Proceedings*, 2015(1):12263.

Shreya Havaldar, Bhumika Singhal, Sunny Rai, Langchen Liu, Sharath Chandra Guntuku, and Lyle Ungar. 2023. Multilingual language models are not multicultural: A case study in emotion. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 202–214, Toronto, Canada. Association for Computational Linguistics.

Dirk Hovy. 2015. Demographic factors improve classification performance. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 752–762, Beijing, China. Association for Computational Linguistics.

Dirk Hovy and Anders Søgaard. 2015. Tagging performance correlates with author age. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 483–488, Beijing, China. Association for Computational Linguistics.

Dirk Hovy and Diyi Yang. 2021. The importance of modeling social factors of language: Theory and practice. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 588–602, Online. Association for Computational Linguistics.

Xiaolei Huang and Michael J. Paul. 2019. Neural user factor adaptation for text classification: Learning to generalize across author demographics. In *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (*SEM 2019)*, pages 136–146, Minneapolis, Minnesota. Association for Computational Linguistics.

Kokil Jaidka, Salvatore Giorgi, H Andrew Schwartz, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2020. Estimating geographic subjective well-being from twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences*, 117(19):10165–10171.

Kayla N Jordan, Joanna Sterling, James W Pennebaker, and Ryan L Boyd. 2019. Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proceedings of the National Academy of Sciences*, 116(9):3476–3481.

Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill Dolan. 2016. A persona-based neural conversation model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 994–1003, Berlin, Germany. Association for Computational Linguistics.

Yunfei Long, Qin Lu, Rong Xiang, Minglei Li, and Chu-Ren Huang. 2017. Fake news detection through multi-perspective speaker profiles. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 252–256, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Stephanie Lukin, Pranav Anand, Marilyn Walker, and Steve Whittaker. 2017. Argument strength is in the eye of the beholder: Audience effects in persuasion. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 742–753, Valencia, Spain. Association for Computational Linguistics.

Veronica Lynn, Youngseo Son, Vivek Kulkarni, Niranjan Balasubramanian, and H. Andrew Schwartz. 2017. Human centered NLP with user-factor adaptation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1146–1155, Copenhagen, Denmark. Association for Computational Linguistics.

Matthew Matero, Akash Idnani, Youngseo Son, Salvatore Giorgi, Huy Vu, Mohammad Zamani, Parth Limbachiya, Sharath Chandra Guntuku, and H. Andrew Schwartz. 2019. Suicide risk assessment with multi-level dual-context language and BERT. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 39–44, Minneapolis, Minnesota. Association for Computational Linguistics.

Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6):1–35.

Pushkar Mishra, Marco Del Tredici, Helen Yannakoudakis, and Ekaterina Shutova. 2018. Author profiling for abuse detection. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1088–1098, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. 2021. Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media+ Society*, 7(2):20563051211019004.

Qiao Qian, Minlie Huang, Haizhou Zhao, Jingfang Xu, and Xiaoyan Zhu. 2018. Assigning personality/profile to a chatting machine for coherent conversation generation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI'18, page 4279–4285. AAAI Press.

Mario Rodríguez-Cantelar, Chen Zhang, Chengguang Tang, Ke Shi, Sarik Ghazarian, João Sedoc, Luis Fernando D'Haro, and Alexander Rudnicky. 2023. Overview of robust and multilingual automatic evaluation metrics for open-domain dialogue systems at dstc 11 track 4. *arXiv preprint arXiv:2306.12794*.

Ramit Sawhney, Harshit Joshi, Rajiv Ratn Shah, and Lucie Flek. 2021. Suicide ideation detection via social and temporal user representations using hyperbolic learning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2176–2190, Online. Association for Computational Linguistics.

Ramit Sawhney, Atula Neerkaje, Ivan Habernal, and Lucie Flek. 2023. How much user context do we need? privacy by design in mental health nlp applications. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, pages 766–776.

H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791.

Deven Santosh Shah, H. Andrew Schwartz, and Dirk Hovy. 2020. Predictive biases in natural language processing models: A conceptual framework and overview. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5248–5264, Online. Association for Computational Linguistics.

Nikita Soni, Matthew Matero, Niranjan Balasubramanian, and H. Andrew Schwartz. 2022. Human language modeling. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 622–636, Dublin, Ireland. Association for Computational Linguistics.

Nikita Soni, H Andrew Schwartz, João Sedoc, and Niranjan Balasubramanian. 2024. Large human language models: A need and the challenges. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Mexico City, Mexico. Association for Computational Linguistics.

Adithya V Ganesan, Matthew Matero, Aravind Reddy Ravula, Huy Vu, and H. Andrew Schwartz. 2021. Empirical evaluation of pre-trained transformers for human-level NLP: The role of sample size and dimensionality. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4515–4532, Online. Association for Computational Linguistics.

Svitlana Volkova, Theresa Wilson, and David Yarowsky. 2013. Exploring demographic language variations to improve multilingual sentiment analysis in social media. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1815–1827, Seattle, Washington, USA. Association for Computational Linguistics.

Charles Welch, Chenxi Gu, Jonathan K. Kummerfeld, Veronica Perez-Rosas, and Rada Mihalcea. 2022. Leveraging similar users for personalized language modeling with limited data. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1742–1752, Dublin, Ireland. Association for Computational Linguistics.

Albert Xu, Eshaan Pathak, Eric Wallace, Suchin Gururangan, Maarten Sap, and Dan Klein. 2021. Detoxifying language models risks marginalizing minority voices. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2390–2397, Online. Association for Computational Linguistics.

Yi Yang and Jacob Eisenstein. 2017. Overcoming language variation in sentiment analysis with social attention. *Transactions of the Association for Computational Linguistics*, 5:295–307.

Mohammadzaman Zamani and H Andrew Schwartz. 2021. Contrastive lexical diffusion coefficient: Quantifying the stickiness of the ordinary. In *Proceedings of the Web Conference 2021*, pages 565–574.

Mohammadzaman Zamani, H. Andrew Schwartz, Veronica Lynn, Salvatore Giorgi, and Niranjan Balasubramanian. 2018. Residualized factor adaptation for community social media prediction tasks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3560–3569, Brussels, Belgium. Association for Computational Linguistics.