Large *Human* Language Models: A Need and the Challenges

Anonymous ACL submission

Abstract

As research in human-centered NLP advances, there is a growing recognition of the importance of incorporating human and social factors into NLP models. At the same time, our NLP systems have become heavily reliant on LLMs, most of which do not model authors. To build NLP systems that can truly understand human language, we must better integrate human contexts into LLMs. This brings to the fore a range of design considerations and challenges in terms of what human aspects to capture, how to represent them, and what modeling strategies to pursue. To address these, we advocate for three positions toward creating large *human* language models (LHLMs) using concepts from psychological and behavioral sciences: First, LM training should include the human context. Second, LHLMs should recognize that people are more than their group(s). Third, LHLMs should be able to account for the dynamic and temporally-dependent nature of the human context. We refer to relevant advances and present open challenges that need to be addressed and their possible solutions in realizing these goals.

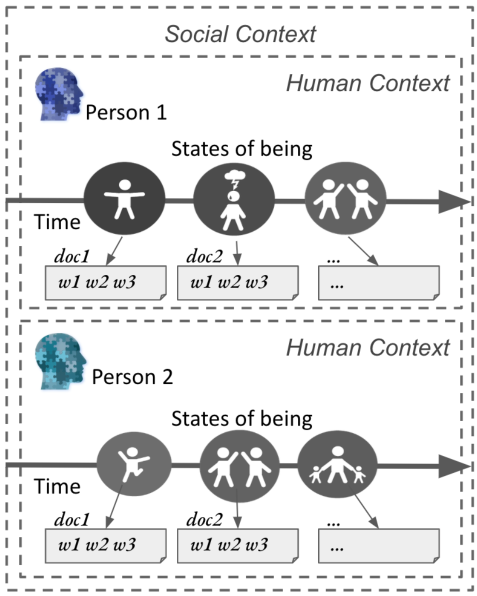

Figure 1: Language expresses the changing human states of being over time. To truly understand human language, language models should have the advantage of the *dynamic human context* along with the context of its neighboring words.

1 Introduction

Language is a fundamental form of *human* expression and communication of thoughts, emotions, and experiences. Learning the meaning of words extends beyond syntax, semantics, and the neighboring words. To truly understand human language, we must look at words in the context of the human generating the language. Figure 1 depicts a view of how our language is moderated by our somewhat stable and changing human states of being over time (Fleeson, 2001; Mehl and Pennebaker, 2003; Heller et al., 2007).

Progress in human-centered NLP research has established the importance of modeling human and social factors, presenting a compelling argument that learning language from linguistic signals alone is not adequate (Hovy, 2018; Bisk et al., 2020; Flek, 2020), and noting that feelings, knowledge and mental states of the speaker and listener referred to as the "Theory of Mind" (Flavell, 2004), along with other social context variables are vital to language understanding (Bisk et al., 2020; Hovy and Yang, 2021). This need is backed by a wealth of empirical evidence demonstrating the benefits of modeling human and social factors (Volkova et al., 2013; Hu et al., 2013; Bamman and Smith, 2015; Lynn et al., 2017; Radfar et al., 2020), and personalized models (Delasalles et al., 2019; Jaech and Ostendorf, 2018; King and Cook, 2020; Welch et al., 2020b).

In parallel, with the advent of Transformers (Vaswani et al., 2017), there have been many advances in language modeling (Devlin et al., 2019; Dai et al., 2019; Liu et al., 2019; Radford et al., 2019) yielding Transformer-based large language

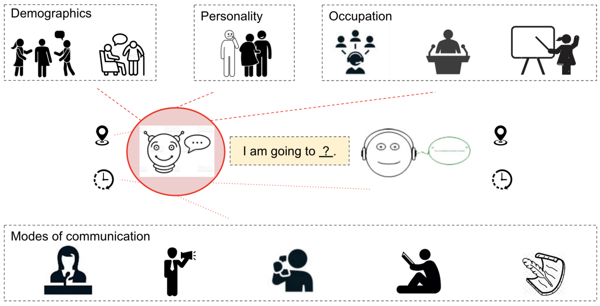

Figure 2: Language is moderated by multiple factors like *who* is speaking to *whom*, *where*, *when*, and other factors like demographics, personality, occupation, modes of communication etc. The author's language is highly dependent on their context, which is referred to as their *human context*.

models (LLMs) as the base of most current NLP systems. LLMs train on a pre-training task and are capable of being applied to a broad set of NLP tasks producing state-of-the-art results. But these language models create word representation only in the context of neighboring words and do not explicitly account for the context of the authors.

Moreover, a person's language can be considered in the rich and complex human context that spans a wide range of aspects.

> *[S]peakers design their utterances to be understood against the common ground they share with their addressees—their common experience, expertise, dialect, and culture.* - Schober and Clark (1989)

Figure 2 illustrates an extensive set of factors that can be considered "human context" which affects how one generates language. A sentence that begins with the phrase "I'm going to...", can be continued in various ways depending on several factors such as (a) *who* is speaking, (b) *where* are they / in what situation and (c) *when* are they speaking, and (d) to *whom* the sentence is addressed including their own time and place. Specific examples of factors include age, personality, occupation, etc., and the forms and modes of communication like public speaking, letter writing, books, phone conversations etc. The speaker's language is, thus, highly dependent on the speaker's states, traits, social and environmental factors (Boyd and Schwartz, 2021), which, collectively, are referred to as the *human context*.

It thus becomes clear that our LLMs can benefit immensely from integrating the human context to truly understand human language but it entails multiple challenges. LLMs can be seen as containing a multitude of personas, and when prompted or primed appropriately can assume a specific one (Patel et al., 2022). Continued scaling may bring benefit from more diversity, such as great ability of GPT-3 and ChatGPT to simulate some forms of human context, especially in generative tasks (Reif et al., 2022). But the models do not input the multi-level structure (documents connected to people) necessary for modeling the richness of human context.

Instead, in this work we call for a more direct and explicit integration of the human contexts when building language models. In particular, we advocate for including the human context directly in language model training, building rich human contexts that account for the fact that people are more than their groups, and the dynamic and temporal-dependent changes to their states of being. In short, we call for building large *human* language models as a step towards better understanding the human language. Furthermore, we discuss open challenges in realizing this vision and their possible solutions.

2 Position 1: LM training should include the human context.

2.1 Motivation

Piantadosi et al. (1988) describe a fallacy in statistical models of the world pertaining to modeling individual observations that are part of a group, as if they are independent, a so-called *ecological fallacy*:

> "Serious errors can result when an investigator makes the seemingly natural assumption that the inference from an ecological analysis must pertain either to individuals within the group or to individuals across groups."

Large Language Models exhibit a form of this ecological fallacy, whereby text sequences coming from a common author are treated as if independent and miss the opportunity to capture dependence (Soni et al., 2022).

Motivated by this need for interpreting language in its human context and inducing inter-dependence between different text sequences from an individual, we posit the need to train our base large language models with the human context. One broad way to frame human context-aware language modeling is as follows:

$$Pr(\mathbf{X}|\mathbf{H}) = \prod_{i=1}^{n} Pr(x_i|X_{1:i-1}, H)$$

This *human language modeling* problem generalizes the regular language modeling problem of predicting the next word conditioned on the previous words in a text sequence X to also condition on a human context H. To train LLMs for this human language modeling problem, we need methods to both represent the human context and to include it in our training objective.

2.2 Past Work

A rich body of prior work sought to include human contexts in NLP models, broadly falling into two categories: ones that are closer to the human language modeling frame, and others that are post hoc adaptations of models with human contexts.

Human context-aware language models. Some work on personalized language models account for human contexts through user embeddings and show improvements in predicting mental health like depression (Wu et al., 2020), and user attributes like demographics (Benton et al., 2016), and occupation (Li et al., 2015). These focus more on creating user representations and less on informing language models with the human context. Others pursue continued training on the language of specific users to build user-specific language models (Wen et al., 2013; King and Cook, 2020) achieving substantial gains in perplexity. While these support the call for integrating human contexts in language modeling, we need to go beyond these user-specific models. Learning and storing separate models for each user presents a scalability challenge, as well as limiting the sharing of knowledge across different users thus limiting generalization.

Delasalles et al. get close to creating a more generalized human language model which improves perplexity by 10 points on New york Times and Semantic Scholar corpus.. They use a dynamic latent representation of the author to capture the human context using an LSTM based architecture but lacks the richness of the human context that can be derived from the author's language. Also, the model parameters seem to depend on the number of authors for the static component of the user representation. While this is better than approaches that create one model per user, the growth in parameters limits scalability and generalization. Soni et al. go further towards modeling the rich dynamic human context from the author's historical language and including it in the continued training of a modified GPT-2 based model. They use social media datasets and show LM improvements with perplexity gains of upto 20 points, and improved downstream task performance on four different tasks including sentiment analysis, stance detection, assessing personality, and estimating age.

Post hoc human contextualized models. Two broad groups of methods use human contexts in a post hoc fashion: Personalized application-focused models, and Debiasing methods using semantic subspaces. Some examples of the first group include methods that create user specific feature vectors (Jaech and Ostendorf, 2018; Seyler et al., 2020) or prefixed static user identifiers (Mireshghallah et al., 2022) or prefixed learned user-specific vectors (Zhong et al., 2021; Li et al., 2021) to the word vectors and show improved accuracies for personalized sentiment analysis, personalized search query auto-completion, or personalized explainable recommendation. Others developed hierarchical modeling using historical text from a user to create personalized models to improve personality detection (Lynn et al., 2020) and stance detection (Matero et al., 2021). In the second group, several studies focused on identifying and eliminating word vector subspaces associated with a particular bias such as gender (Bolukbasi et al., 2016; Wang et al., 2020; Ravfogel et al., 2020) and religion (Liang et al., 2020). The broad evidence for personalization and debiasing indicates performance and fairness benefits of modeling human contexts.

Given the move towards large language models as the basis for NLP, we argue that if the base LLMs can be made human context aware, we can learn better and more fair language representations to begin with.

2.3 Challenges and Possible Solutions

C1: Including human context. Training LLMs for the human language modeling problem raises a wide range of challenges. These include deciding how to capture the human context effectively and how to incorporate it in training.

PS1: Before the advent of LLMs, human-centered NLP mainly infused human context H (e.g. demographic value of an author) into a feature space F either using factor additive approaches (Bamman et al., 2014a; Bamman and Smith, 2015; Kulkarni et al., 2016; Welch et al., 2020a):

$$P = g(F + H)$$

or through user factor adaptation (Lynn et al., 2017; Huang and Paul, 2019):

$$O = g(z(F, H))$$

where g is a model trained to output predictions P, and z represents a form of multiplicative compositional function that is used to adapt the feature space to the human context.

We can extend these "pre-LLM" approaches to LLMs by viewing the hidden states or the contextual word vectors as features. The human context can thus be added directly to the contextual word vectors similar to how position embeddings get added or via composition functions that adapt the contextual word vectors conditioning on the human context. More generally, integrating human context into Transformer based LLMs brings up many challenges in terms of modeling, interaction with downstream applications, and data processing.

C2: Modeling decisions. Architectural decisions include which layers to modify, where do we include the human context, how to alter the self-attention mechanism if needed, which components (query, key, value) should include the human context if needed.

PS2: For example, Soni et al. (2022) modify the language modeling task to include the human context as a user vector, which is derived from author's historical text. The new Transformer-based architecture modifies the self-attention computation by using the user vector in the query representation, and recurrently updates the user vector using the hidden states from a later layer. Other works (Zhong et al., 2021; Mireshghallah et al., 2022), as discussed earlier, simply prefix the user representation to the word embeddings when processing through the Transformer based architectures.

These questions and existing works spur us to explore many other architectural solutions for large human language models, along with suitable pre-training tasks or loss functions that include human contexts.

C3: Model applications. Another key challenge is in effectively applying the pre-trained large human language models on the target downstream tasks and applications.

PS3: For instance, (i) the pre-training task may be built similar to downstream task training i.e., we add a classification or regression head on top of the pre-trained language model and fine-tune for target downstream tasks like a traditional large language model, (ii) the pre-trained model can be trained with downstream task-specific objective i.e., in addition to using the pre-training knowledge, we train the model parameters specific to the target downstream task objective alone, (iii) continue the pre-trained model's training in a multi-task learning setup i.e., we train for the pre-training objective as well as a downstream task-specific objective, or (iv) fine-tune the pre-trained model by exploring changes in the way target downstream task data is processed, for example, limiting the historical language context for downstream task data.

C4: Data processing. Processing human context from user's historical language requires model designs to incorporate solutions to process user-specific data which can be rather long. The runtime and memory complexity of the self-attention mechanism scales quadratically with the sequence length, which often limits their abilities to directly process long input sequences.

PS4: Some approaches to address this limitation include sparsifying the attention mechanism (Beltagy et al., 2020; Kitaev et al., 2020; Qiu et al., 2020; Ye et al., 2019; Roy et al., 2021) or using autoregressive recurrence-based methods (Sukhbaatar et al., 2019; Rae et al., 2019; Dai et al., 2019). Soni et al. used a recurrence mechanism (Dai et al., 2019; Yoshida et al., 2020) over temporally ordered blocks to handle the long historical texts of each user. Another option is to use retrieval-augmentation mechanisms which can expand the model's ability to incorporate information beyond a limited context (Guu et al., 2020).

Other open questions include how much historical language is sufficient to capture the human context, whether adding more language will help build a better human context, and whether we need to process even longer documents in a single pass, among other intriguing considerations.

3 Position 2: LHLMs should recognize that people are more than their group(s).

3.1 Motivation

Human context is not limited to a specific social and demographic group they belong to. Rather it is a mix of the multiple human attribute groups they may belong to and their unique characteristics and idiosyncrasies. Even with their groups, it is not always a binary association, there are varying degrees to which an individual might align with the group traits.

Psychology and Psychopathology have a wealth of literature suggesting that people should not be put in discrete bins but instead should be placed in a dimensional structure by characterizing them as a mixture of continuous factors (McCrae and Costa Jr, 1989; Ruscio and Ruscio, 2000; Widiger and Samuel, 2005). Further, grouping people into discrete bins often uses arbitrary boundaries which may lose the meaningful distinctions in capturing the human context.

Cross-cultural psychology research has noted the distinctions in individualism and collectivism concurring with the predictions from Hofstede's model (Hofstede, 1984; Hofstede and Bond, 1984).

> "people from the collectivist culture produc[ing] significantly more group and fewer idiocentric self-descriptions than did people from the individualist cultures" -Bochner (1994)

These suggest that it is vital to allow for flexible interactions between individualistic and collectivist aspects of the human context.

Moreover, the rich diversity in people cannot be captured effectively by modeling a narrow sample of variation in human factored groups. In behavioural sciences, Henrich et al. (2010) bring to the attention that most of the research in the field is often limited to humans belonging to the WEIRD (Western, Educated, Industrialized, Rich, and Democratic) group. They argue that this narrow group is mostly an outlier as a representative of humanity in cross-cultural research. This provides a corresponding lesson for NLP research. We should not limit ourselves to a narrow spectrum of specific human factors and only modeling outliers in the human context.

Motivated by these ideas from psychology and behavioral sciences, we argue for breadth, depth, and richness in modeling the human context when training large human language models.

3.2 Past Work

A huge body of work in human-centered NLP has shown the importance of modeling human attributes like demographic factors and social context, and latent human variables in natural language processing. These include works that model factors which are either known explicitly from questionnaires, social profiles, or inferred from the user's language, with the aim of grouping people to analyze language variations among different groups.

Wide variety of human factors. There are many types of human factors that can influence a person's language. Cross-cultural differences and demographics like gender (Volkova et al., 2013) and age (Hovy, 2015) have been shown to influence the perceived meaning of words and aid in multiple text classification tasks (Huang and Paul, 2019), and machine translation (Mirkin et al., 2015; Rabinovich et al., 2017). Several studies have also exploited benefits from social relations (Huang et al., 2014; Yang and Eisenstein, 2017; Zeng et al., 2017; Del Tredici et al., 2019) in sentiment analysis (Hu et al., 2013) and toxic language detection (Radfar et al., 2020). Existing literature has shown correlations in language variation with personality (Schwartz et al., 2013), occupation (Preoţiuc-Pietro et al., 2015), and geographical region (Bamman et al., 2014a; Kulkarni et al., 2016; Garimella et al., 2017) illustrating distinctions in style and perspectives among different groups of people.

Intersectionality of human factors. A person's language is mediated not just by an individual factor but by the intersection of many factors. Some works (Bamman and Smith, 2015; Lynn et al., 2019; Huang and Paul, 2019) have explored using multiple human factors together in their studies. Some classification tasks from different domains (Huang and Paul, 2019) have shown greater benefits in a multi-factored approach of combining gender, age, country, and region, while tasks like sarcasm detection (Bamman and Smith, 2015) and stance detection (Lynn et al., 2019) have performed better by specific author features. These empirical studies indicate the need to explore different combinations of human factors for respective downstream tasks and applications.

Continuous representation of human factors. A discrete group often relies on arbitrary boundaries and a person may belong to multiple groups in varying degrees. Thus, using a continuous representation of human factors may allow us to move away from *hard memberships* in arbitrary groups to a more realistic *soft membership* along factor dimensions.Past works have illustrated language differences based on social network clusters with strong gender orientation, treating gender as more than a binary variable (Bamman et al., 2014b), or by continuous adaptation of real-valued human factors like continuous age, gender, and Big Five personality traits (Lynn et al., 2017).

Latent human factors. A person's language has characteristics that go well beyond those of a specific set of groups they may belong to. To capture a broader set of characteristics, some works explored deriving latent factors from a person's language (Wen et al., 2013; Lynn et al., 2017; Kulkarni et al., 2018). Latent linguistic factors have been shown to capture user attributes (Lynn et al., 2017) and differences in thoughts and emotions of people (Kulkarni et al., 2018). Others create latent representations from user posts using bag-of-words (Benton et al., 2016), sparse-encoded BERT contextual embeddings (Wu et al., 2020), and averaged GRU embeddings (Lynn et al., 2020). Another approach focuses on learning embeddings, i.e., a trainable set of parameters, as latent representations of users (Li et al., 2015; Amir et al., 2016; Zeng et al., 2017; Jaech and Ostendorf, 2018; Welch et al., 2020b). These latent user representations and learned embeddings have yielded benefits in multiple downstream tasks and applications.

Modeling the human context in terms of groups that people belong to has pioneered advances in human-centered NLP. However, humans are more than the groups they belong to. To go further, we need a representation that recognizes the variety, intersectionality of the human factors across continuous dimensions, as well as their unique individual characteristics.

3.3 Challenges and Possible Solutions

C1: Modeling data and representational disparities. To capture the rich human context, we need access to datasets that provide relevant information covering users who are representative of the broad and diverse population(Piantadosi et al., 1988; Henrich et al., 2010; Johnson et al., 2022). Spefically the challenges lie in obtaining datasets: (1) that provide access to user identifiers and historical language that allow us to differentiate the human source of the language, associate explicit human attributes such as sociodemographic or personality attributes, (2) that do not amplify representational disparities (Shah et al., 2020) and span multiple domains such as healthcare (Bean et al., 2023), customer service (Adam et al., 2021), and education (Klein and Nabi, 2019).

PS1: There are multiple avenues for addressing the challenges above. First, there are a wide-variety of large scale datasets that contain author Ids as metadata. For example, Amazon reviews, Reddit posts and comments, blogs, books, and news have associated author Ids as metadata, which can be used to train LHLMs. Second, some representational disparities can be addressed by benchmarking and balancing the types of disparities. For example, we can use various text-based human attribute inference methods to detect and balance for attributes such as age, gender, and other demographics (Tadesse et al., 2018; Wang et al., 2019). Similarly, we can address cultural disparities by making use of research efforts to probe(Arora et al., 2023), identify (Gutiérrez et al., 2016; Lin et al., 2018) and benchmark (Yin et al., 2022) cross-cultural differences. Third, we can also use modeling strategies that are better equipped to handle imbalanced and limited data settings. For example, there is large body of work in low-resource settings for problems such as sentiment analysis (Priyadharshini et al., 2021; Muhammad et al., 2023), hate speech detection (Modha et al., 2021), and machine translation (Ranathunga et al., 2023). Other notable examples include strategies for culturally grounding models using transfer learning (Sun et al., 2021; Zhou et al., 2023), and adaptation strategies for modeling societal values (Solaiman and Dennison, 2021).

Additionally, industries with large user bases are a potential source for language data. Investing in community-wide efforts for publishing and evaluating research over proprietary data and improved industry collaborations can provide access to otherwise unavailable data which can also help further research in this area.

C2: Privacy issues. Modeling user's personal characteristics carries the inherent risk of inadvertent privacy leaks as well as the potential for adversarial or malicious use. The challenge of guarding the privacy of the individuals can be broadly categorized into 2 aspects: (1) Privacy and data control of the data subject, and (2) Licensing model usage, policies, and laws to prevent potential misuse like target marketing: As seen in the past with Cambridge Analytica Facebook dataset, a potential misuse of modeling humans is target marketing (Isaak and Hanna, 2018; Bakir, 2020).

PS2: Existing laws in some parts of the world aim to protect the user privacy and security, such as requiring data anonymization and/or asking for consent to share data. The EU general data protection regulation (GDPR) (Lewis et al., 2017), for example, is considered one of the strongest such law. Italy banned the widespread ChatGPT services citing concerns over breaches of EU data protection laws. The US follows the Institutional Review Board (IRB) approvals process to protect human subjects research with most standards rooting from ethical standards involved in medical research (Goodyear et al., 2007; Miracle, 2016) such as protecting the rights of all research subjects or participants in terms of respect, beneficence, justice, right to make informed decisions, and recognition of vulnerable groups. We should be vigilant in preventing such leaks and have strict licensing and policies to safeguard malevolent uses. In fact, human context aware models themselves can be used towards some of these goals such as recognizing target marketing and preventing its spread. An key part here is in continuing to evolve privacy laws and policies as the models evolve and investing in studies that can better inform these decisions.

C3: Model scalability. Targeting human contexts that go beyond group characteristics and include unique individual characteristics increases the scalability requirements on the models. The key challenge is that the model has to simultaneously capture user-specific contexts as well as scale to multiple users without corresponding increases in model parameters or creating a new model itself for each user. Past work on personalized models have been limited by this scalability issue, whereby either models are user-specific or do not scale well. In some, a separate model is created for each user (King and Cook, 2020), while in others a different user identifier is used for each user (Li et al., 2021; Zhong et al., 2021; Mireshghallah et al., 2022).

PS3: Some use a post-hoc fix which handles any new user seen after training by updating the user embeddings with the new user directly during evaluation (Jaech and Ostendorf, 2018). Delasalles et al. (2019) adopt an LSTM-based approach with a dynamic author representation which consists of user-specific static and dynamic components. These approaches that learn user-specific vectors are relatively more scalable than the ones that learn user-specific models. Soni et al. (2022) eliminate this dependence on user-specific vectors using a single Transformer-based model, where a recurrent user states module is trained to use authors' historical language. While this improves scalability, it is still limited in the amount of historical language it can use due to the compute requirements and context-length considerations. These ideas pave the way for further explorations of solutions to this challenge of building scalable large human language models.

4 Position 3: LHLMs should account for the dynamic and temporally-dependent nature of human context.

4.1 Motivation

> *"[People] are embedded within time, that time is fundamentally important to life as it is lived, and that personality processes take place over time."* -Larsen (1989)

A person's static and dynamic human states are intertwined, where static traits influence the likelihood of entering various dynamic states across time (DeYoung, 2015). Correspondingly, a person's language expresses the changing human states and evolving emotions over time (Fleeson, 2001; Mehl and Pennebaker, 2003; Heller et al., 2007). For the human context to be effective, it must not only be able to model the static human traits and attributes but also the more dynamic human states of being.

Temporal rhythms (e.g. diurnal and seasonal) are also known to affect human mood and behavior, which in turn manifests in their language (Golder and Macy, 2011). We need mechanisms that can capture the patterns of regularity or change in human language and human behaviour over time. For example, studies on NLP for mental health also point to the importance of tracking moments of change over time for assessing sucidal risk (Tsakalidis et al., 2022).

Motivated by these ideas of changing human states and the impact of temporal aspects on human behavior and language, we posit the need for a dynamic and temporally-dependent human context.

4.2 Past Work

Studies that explore the dynamic nature of human context fall into two broad categories, those that: (1) dynamically update user representations to capture changing human states, and those that (2) contextualize using temporally ordered texts and other aspects that demonstrate the recurrent changes from seasonality or other cyclic patterns.

Recurrently updated user representations. As discussed earlier, recurrence mechanisms have been used for building user representations (Delasalles et al., 2019; Soni et al., 2022). This use of recurrently updated author representations is motivated by the need to capture author-specific features that do not change with time along with author's altering expression mode, topic evolution, and their changing human states over time. Delasalles et al. learned a dynamic latent vector using an LSTM model for this purpose, while Soni et al. further use the target user's historical texts to recurrently update the user representation. When learnt over temporally ordered language, these enable capturing the changing human states and temporal aspects as exhibited through their language.

Temporal Modeling. The changing human states over time highlights the need to consider the *temporal* aspect of the human context and its expressions in language. Considering temporally ordered texts allows capturing some notion of temporality in an implicit fashion. Matero et al. (2021) introduced a missing message prediction task over a sequence of temporally ordered social media posts of the target user to build a personalized language model that helps in stance detection. Tsakalidis et al. (2022) proposed a shared task to capture drastic and gradual moments of change in an individual's mood based on their language on social media and to identify how this change helps assess suicidal risk (Boinepelli et al., 2022; V Ganesan et al., 2022). Zhou et al. (2020) use other temporal aspects like typical periodicity or cyclical nature, frequency, and duration to induce common sense in language models but over generic newswire texts with no direct relation to the human contexts of the authors.

We propose using recurring patterns or anomalies can better inform our dynamic human context to capture a better representation of a person as a whole. This enriched human context capturing the periodicity or anomalies in human behavior and their language can also help in multiple mental health applications and early detection.

4.3 Challenges and Possible Solutions

C1: Modeling data. To model the dynamic and temporal changes in language, we need time information in our datasets. Assessment over time can be thought as an additional dimension to the dataset, resulting in a three-dimensional dataset (Larsen, 1989) with user information, text, and time. While it may be possible to obtain a reasonable history of a user's language, obtaining adequate samples across all timestamps is difficult.
PS1: Thus, datasets are likely to have larger "gaps" in the time dimension and models may need to learn to fill or otherwise adequately handle these gaps in temporal text sequences (Matero et al., 2021).

C2: Modeling temporal language and temporal aspects. The positional encoding in a temporally ordered sequence can allow a language model to learn some temporal aspects (e.g. before/after relationships). However, more complex recurrent dynamics at different time scales (e.g. diurnal, weekly, and seasonal) may need other mechanisms that allow the model to explicitly consider the time associated with each text. This raises new challenges in encoding such time information into a temporal embedding and in getting models to use this encoded information. Last, pushing models to consider temporal information may also require developing new language modeling objectives.
PS2: Predicting what follows can often be modeled by focusing on the immediate local dependencies (in a Markovian sense). However, to force models to consider different temporal scales we can consider objectives that frame predicting what will be said after a specific temporal interval (e.g. the next day, the same day next week and so on).

5 Conclusion

Building upon the success from two parallels of NLP research: large language models and human-centered NLP, we envision large *human* language models (LHLMs) as the base of future NLP systems. Previous positions taken in human-centered NLP advocate for modeling human and social factors (Hovy, 2018; Bisk et al., 2020; Flek, 2020; Shah et al., 2020; Hovy and Yang, 2021). We go further and call for modeling a richer and dynamic human context in our future large language models. A rich human context captures the personal, social, and situational attributes of the person, and represents both static traits and dynamic human states of being. We put forward three specific positions as steps toward integrating this rich human context in language models to realize the vision of large *human* language models. Our roadmap draws on motivations from multiple disciplines, prior advances in human-centered NLP, and organizes the range of challenges to be met in realizing this vision. We call for our NLP research community to take on the challenge of bringing humans, the originators of language, into our large language models.

Limitations

We elaborate on the three positions we take to create large *human* language models in terms of the need, richness, and dynamic nature of the human context in the main paper. However, the scope of this position is fairly limited, focusing on the details of the human context, only giving social context a brief mention in so far as it's relation to human context. Important social context affecting language include (1) cultural shifts/changes, (2) environmental events like natural disasters, and (3) multi-lingual settings (although most of our discussion is based on the psychological theory that transcends languages). Similarly, we limit our discussion on the need and challenges of the breadth of the domains of the human context. Finally, our discussion of privacy issues is also focused on the human context (refer section 3.3) and thus does not go into required social policies and its effects on language models.

Ethical Considerations

Many of the main points of this paper are in themselves of ethical consideration. We thus use this section to discuss the uncovered considerations. Importantly, while we advocate for large *human* language models and training them with a rich and dynamic human context, we also argue not every use case of LHLMs are of societal benefit. When developing LHLMs to better understand human language and for enabling bias correction and fairness, one should also seek a responsible strategy for the release and use of user-level information which can sometimes be sensitive or private. For such data, user consent and privacy protections are important. Otherwise, such models could be used for targeted content toward training set users without their awareness. While laws in some nations, such as the GDPR, outlaw such use cases, these have not become universal around the world yet.

References

Martin Adam, Michael Wessel, and Alexander Benlian. 2021. Ai-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, 31(2):427–445.

Silvio Amir, Byron C. Wallace, Hao Lyu, Paula Carvalho, and Mário J. Silva. 2016. Modelling context with user embeddings for sarcasm detection in social media. In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, pages 167–177, Berlin, Germany. Association for Computational Linguistics.

Arnav Arora, Lucie-aimée Kaffee, and Isabelle Augenstein. 2023. Probing pre-trained language models for cross-cultural differences in values. In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, pages 114–130, Dubrovnik, Croatia. Association for Computational Linguistics.

Vian Bakir. 2020. Psychological operations in digital political campaigns: Assessing cambridge analytica's psychographic profiling and targeting. *Frontiers in Communication*, 5:67.

David Bamman, Chris Dyer, and Noah A. Smith. 2014a. Distributed Representations of Geographically Situated Language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 828–834, Baltimore, Maryland. Association for Computational Linguistics.

David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. 2014b. Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2):135–160. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/josl.12080.

David Bamman and Noah Smith. 2015. Contextualized Sarcasm Detection on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 9(1):574–577. Number: 1.

Daniel M Bean, Zeljko Kraljevic, Anthony Shek, James Teo, and Richard JB Dobson. 2023. Hospital-wide natural language processing summarising the health data of 1 million patients. *PLOS Digital Health*, 2(5):e0000218.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The Long-Document Transformer. *arXiv:2004.05150 [cs]*. ArXiv: 2004.05150.

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.

Yonatan Bisk, Ari Holtzman, Jesse Thomason, Jacob Andreas, Yoshua Bengio, Joyce Chai, Mirella Lapata, Angeliki Lazaridou, Jonathan May, Aleksandr Nisnevich, Nicolas Pinto, and Joseph Turian. 2020. Experience Grounds Language. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8718–8735, Online. Association for Computational Linguistics.

Stephen Bochner. 1994. Cross-cultural differences in the self concept: A test of hofstede's individualism/collectivism distinction. *Journal of cross-cultural psychology*, 25(2):273–283.

Sravani Boinepelli, Shivansh Subramanian, Abhijeeth Singam, Tathagata Raha, and Vasudeva Varma. 2022. Towards capturing changes in mood and identifying suicidality risk. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, pages 245–250, Seattle, USA. Association for Computational Linguistics.

Tolga Bolukbasi, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam Tauman Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *NIPS*.

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.

Zihang Dai, Zhilin Yang, Yiming Yang, Jaime G Carbonell, Quoc Le, and Ruslan Salakhutdinov. 2019. Transformer-xl: Attentive language models beyond a fixed-length context. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2978–2988.

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.

Edouard Delasalles, Sylvain Lamprier, and Ludovic Denoyer. 2019. Learning Dynamic Author Representations with Temporal Language Models. *2019 IEEE International Conference on Data Mining (ICDM)*, pages 120–129. ArXiv: 1909.04985.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Colin G DeYoung. 2015. Cybernetic big five theory. *Journal of research in personality*, 56:33–58.

John H. Flavell. 2004. Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly*, 50:274 – 290.

William Fleeson. 2001. Toward a structure-and process-integrated view of personality: Traits as density distributions of states. *Journal of personality and social psychology*, 80(6):1011.

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.

Aparna Garimella, Carmen Banea, and Rada Mihalcea. 2017. Demographic-aware word associations. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2285–2295, Copenhagen, Denmark. Association for Computational Linguistics.

Scott A Golder and Michael W Macy. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881.

Michael DE Goodyear, Karmela Krleza-Jeric, and Trudo Lemmens. 2007. The declaration of helsinki.

E.D. Gutiérrez, Ekaterina Shutova, Patricia Lichtenstein, Gerard de Melo, and Luca Gilardi. 2016. Detecting cross-cultural differences using a multilingual topic model. *Transactions of the Association for Computational Linguistics*, 4:47–60.

Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. 2020. Realm: retrieval-augmented language model pre-training. In *Proceedings of the 37th International Conference on Machine Learning*, pages 3929–3938.

Daniel Heller, Jennifer Komar, and Wonkyong Beth Lee. 2007. The dynamics of personality states, goals, and well-being. *Personality and Social Psychology Bulletin*, 33(6):898–910.

Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and brain sciences*, 33(2-3):61–83.

Geert Hofstede. 1984. *Culture's consequences: International differences in work-related values*, volume 5. sage.

Geert Hofstede and Michael H Bond. 1984. Hofstede's culture dimensions: An independent validation using rokeach's value survey. *Journal of cross-cultural psychology*, 15(4):417–433.

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. 2018. The Social and the Neural Network: How to Make Natural Language Processing about People again. In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pages 42–49, New Orleans, Louisiana, USA. 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.

Xia Hu, Lei Tang, Jiliang Tang, and Huan Liu. 2013. Exploiting social relations for sentiment analysis in microblogging. In *Proceedings of the sixth ACM international conference on Web search and data mining - WSDM '13*, page 537, Rome, Italy. ACM Press.

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.

Yu-Yang Huang, Rui Yan, Tsung-Ting Kuo, and Shou-De Lin. 2014. Enriching cold start personalized language model using social network information. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 611–617, Baltimore, Maryland. Association for Computational Linguistics.

Jim Isaak and Mina J. Hanna. 2018. User data privacy: Facebook, cambridge analytica, and privacy protection. *Computer*, 51(8):56–59.

Aaron Jaech and Mari Ostendorf. 2018. Personalized language model for query auto-completion. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 700–705, Melbourne, Australia. Association for Computational Linguistics.

Rebecca L Johnson, Giada Pistilli, Natalia Menédez-González, Leslye Denisse Dias Duran, Enrico Panai, Julija Kalpokiene, and Donald Jay Bertulfo. 2022. The ghost in the machine has an american accent: value conflict in gpt-3. *arXiv preprint arXiv:2203.07785*.

Milton King and Paul Cook. 2020. Evaluating approaches to personalizing language models. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2461–2469, Marseille, France. European Language Resources Association.

Nikita Kitaev, Łukasz Kaiser, and Anselm Levskaya. 2020. Reformer: The efficient transformer. *arXiv preprint arXiv:2001.04451*.

Tassilo Klein and Moin Nabi. 2019. Learning to answer by learning to ask: Getting the best of gpt-2 and bert worlds. *arXiv preprint arXiv:1911.02365*.

Vivek Kulkarni, Margaret L Kern, David Stillwell, Michal Kosinski, Sandra Matz, Lyle Ungar, Steven Skiena, and H Andrew Schwartz. 2018. Latent human traits in the language of social media: An open-vocabulary approach. *PloS one*, 13(11):e0201703.

Vivek Kulkarni, Bryan Perozzi, and Steven Skiena. 2016. Freshman or Fresher? Quantifying the Geographic Variation of Language in Online Social Media. *Proceedings of the International AAAI Conference on Web and Social Media*, 10(1):615–618. Number: 1.

Randy J Larsen. 1989. A process approach to personality psychology: Utilizing time as a facet of data. *Personality psychology: Recent trends and emerging directions*, pages 177–193.

Dave Lewis, Joss Moorkens, and Kaniz Fatema. 2017. Integrating the management of personal data protection and open science with research ethics. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 60–65.

Jiwei Li, Alan Ritter, and Dan Jurafsky. 2015. Learning multi-faceted representations of individuals from heterogeneous evidence using neural networks. *arXiv preprint arXiv:1510.05198*.

Lei Li, Yongfeng Zhang, and Li Chen. 2021. Personalized transformer for explainable recommendation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4947–4957.

Paul Pu Liang, Irene Mengze Li, Emily Zheng, Yao Chong Lim, Ruslan Salakhutdinov, and Louis-Philippe Morency. 2020. Towards debiasing sentence representations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5502–5515.

Bill Yuchen Lin, Frank F. Xu, Kenny Zhu, and Seung-won Hwang. 2018. Mining cross-cultural differences and similarities in social media. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 709–719, Melbourne, Australia. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Veronica Lynn, Niranjan Balasubramanian, and H. Andrew Schwartz. 2020. Hierarchical Modeling for User Personality Prediction: The Role of Message-Level Attention. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5306–5316, Online. Association for Computational Linguistics.

Veronica Lynn, Salvatore Giorgi, Niranjan Balasubramanian, and H. Andrew Schwartz. 2019. Tweet Classification without the Tweet: An Empirical Examination of User versus Document Attributes. In *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science*, pages 18–28, Minneapolis, Minnesota. 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, Nikita Soni, Niranjan Balasubramanian, and H. Andrew Schwartz. 2021. MeLT: Message-level transformer with masked document representations as pre-training for stance detection. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2959–2966, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Robert R McCrae and Paul T Costa Jr. 1989. Reinterpreting the myers-briggs type indicator from the perspective of the five-factor model of personality. *Journal of personality*, 57(1):17–40.

Matthias R Mehl and James W Pennebaker. 2003. The sounds of social life: a psychometric analysis of students' daily social environments and natural conversations. *Journal of personality and social psychology*, 84(4):857.

Vickie A Miracle. 2016. The belmont report: The triple crown of research ethics. *Dimensions of critical care nursing*, 35(4):223–228.

Fatemehsadat Mireshghallah, Vaishnavi Shrivastava, Milad Shokouhi, Taylor Berg-Kirkpatrick, Robert Sim, and Dimitrios Dimitriadis. 2022. Useridentifier: Implicit user representations for simple and effective personalized sentiment analysis. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3449–3456.

Shachar Mirkin, Scott Nowson, Caroline Brun, and Julien Perez. 2015. Motivating Personality-aware Machine Translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1102–1108, Lisbon, Portugal. Association for Computational Linguistics.

Sandip Modha, Thomas Mandl, Gautam Kishore Shahi, Hiren Madhu, Shrey Satapara, Tharindu Ranasinghe, and Marcos Zampieri. 2021. Overview of the hasoc subtrack at fire 2021: Hate speech and offensive content identification in english and indo-aryan languages and conversational hate speech. In *Proceedings of the 13th Annual Meeting of the Forum for Information Retrieval Evaluation*, pages 1–3.

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Abinew Ali Ayele, Nedjma Ousidhoum, David Ifeoluwa Adelani, Seid Muhie Yimam, Ibrahim Sa'id Ahmad, Meriem Beloucif, Saif Mohammad, Sebastian Ruder, et al. 2023. Afrisenti: A twitter sentiment analysis benchmark for african languages. *arXiv preprint arXiv:2302.08956*.

Ajay Patel, Nicholas Andrews, and Chris Callison-Burch. 2022. Low-resource authorship style transfer with in-context learning. *arXiv preprint arXiv:2212.08986*.

Steven Piantadosi, David P Byar, and Sylvan B Green. 1988. The ecological fallacy. *American journal of epidemiology*, 127(5):893–904.

Daniel Preoţiuc-Pietro, Johannes Eichstaedt, Gregory Park, Maarten Sap, Laura Smith, Victoria Tobolsky, H. Andrew Schwartz, and Lyle Ungar. 2015. The role of personality, age, and gender in tweeting about mental illness. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 21–30, Denver, Colorado. Association for Computational Linguistics.

Ruba Priyadharshini, Bharathi Raja Chakravarthi, Sajeetha Thavareesan, Dhivya Chinnappa, Durairaj Thenmozhi, and Rahul Ponnusamy. 2021. Overview of the dravidiancodemix 2021 shared task on sentiment detection in tamil, malayalam, and kannada. In *Proceedings of the 13th Annual Meeting of the Forum for Information Retrieval Evaluation*, pages 4–6.

Jiezhong Qiu, Hao Ma, Omer Levy, Wen-tau Yih, Sinong Wang, and Jie Tang. 2020. Blockwise self-attention for long document understanding. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2555–2565, Online. Association for Computational Linguistics.

Ella Rabinovich, Raj Nath Patel, Shachar Mirkin, Lucia Specia, and Shuly Wintner. 2017. Personalized Machine Translation: Preserving Original Author Traits. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1074–1084, Valencia, Spain. Association for Computational Linguistics.

Bahar Radfar, Karthik Shivaram, and Aron Culotta. 2020. Characterizing Variation in Toxic Language by Social Context. *Proceedings of the International AAAI Conference on Web and Social Media*, 14:959–963.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Jack W Rae, Anna Potapenko, Siddhant M Jayakumar, and Timothy P Lillicrap. 2019. Compressive transformers for long-range sequence modelling. *arXiv preprint arXiv:1911.05507*.

Surangika Ranathunga, En-Shiun Annie Lee, Marjana Prifti Skenduli, Ravi Shekhar, Mehreen Alam, and Rishemjit Kaur. 2023. Neural machine translation for low-resource languages: A survey. *ACM Computing Surveys*, 55(11):1–37.

Shauli Ravfogel, Yanai Elazar, Hila Gonen, Michael Twiton, and Yoav Goldberg. 2020. Null it out: Guarding protected attributes by iterative nullspace projection. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7237–7256.

Emily Reif, Daphne Ippolito, Ann Yuan, Andy Coenen, Chris Callison-Burch, and Jason Wei. 2022. A recipe for arbitrary text style transfer with large language models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 837–848, Dublin, Ireland. Association for Computational Linguistics.

Aurko Roy, Mohammad Saffar, Ashish Vaswani, and David Grangier. 2021. Efficient content-based sparse attention with routing transformers. *Transactions of the Association for Computational Linguistics*, 9:53–68.

John Ruscio and Ayelet Meron Ruscio. 2000. Informing the continuity controversy: a taxometric analysis of depression. *Journal of abnormal psychology*, 109(3):473.

Michael F. Schober and Herbert H. Clark. 1989. Understanding by addressees and overhearers. *Cognitive Psychology*, 21:211–232.

H. Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin E. P. Seligman, and Lyle H. Ungar. 2013. Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLOS ONE*, 8(9):e73791. Publisher: Public Library of Science.

Dominic Seyler, Jiaming Shen, Jinfeng Xiao, Yiren Wang, and ChengXiang Zhai. 2020. Leveraging Personalized Sentiment Lexicons for Sentiment Analysis. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*, pages 109–112, Virtual Event Norway. ACM.

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.

Irene Solaiman and Christy Dennison. 2021. Process for adapting language models to society (palms) with values-targeted datasets. *Advances in Neural Information Processing Systems*, 34:5861–5873.

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.

Sainbayar Sukhbaatar, Edouard Grave, Piotr Bojanowski, and Armand Joulin. 2019. Adaptive attention span in transformers. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 331–335, Florence, Italy. Association for Computational Linguistics.

Jimin Sun, Hwijeen Ahn, Chan Young Park, Yulia Tsvetkov, and David R. Mortensen. 2021. Cross-cultural similarity features for cross-lingual transfer learning of pragmatically motivated tasks. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2403–2414, Online. Association for Computational Linguistics.

Michael M Tadesse, Hongfei Lin, Bo Xu, and Liang Yang. 2018. Personality predictions based on user behavior on the facebook social media platform. *IEEE Access*, 6:61959–61969.

Adam Tsakalidis, Jenny Chim, Iman Munire Bilal, Ayah Zirikly, Dana Atzil-Slonim, Federico Nanni, Philip Resnik, Manas Gaur, Kaushik Roy, Becky Inkster, et al. 2022. Overview of the clpsych 2022 shared task: Capturing moments of change in longitudinal user posts. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, pages 184–198.

Adithya V Ganesan, Vasudha Varadarajan, Juhi Mittal, Shashanka Subrahmanya, Matthew Matero, Nikita Soni, Sharath Chandra Guntuku, Johannes Eichstaedt, and H. Andrew Schwartz. 2022. WWBP-SQT-lite: Multi-level models and difference embeddings for moments of change identification in mental health forums. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, pages 251–258, Seattle, USA. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

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.

Tianlu Wang, Xi Victoria Lin, Nazneen Fatema Rajani, Bryan McCann, Vicente Ordonez, and Caiming Xiong. 2020. Double-hard debias: Tailoring word embeddings for gender bias mitigation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5443–5453.

Zijian Wang, Scott Hale, David Ifeoluwa Adelani, Przemyslaw Grabowicz, Timo Hartman, Fabian Flöck, and David Jurgens. 2019. Demographic inference and representative population estimates from multilingual social media data. In *The world wide web conference*, pages 2056–2067.

Charles Welch, Jonathan K. Kummerfeld, Verónica Pérez-Rosas, and Rada Mihalcea. 2020a. Compositional Demographic Word Embeddings. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4076–4089, Online. Association for Computational Linguistics.

Charles Welch, Jonathan K. Kummerfeld, Verónica Pérez-Rosas, and Rada Mihalcea. 2020b. Exploring the Value of Personalized Word Embeddings. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6856–6862, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Tsung-Hsien Wen, Aaron Heidel, Hung-yi Lee, Yu Tsao, and Lin-Shan Lee. 2013. Recurrent neural network based language model personalization by social network crowdsourcing. In *INTERSPEECH*, pages 2703–2707.

Thomas A Widiger and Douglas B Samuel. 2005. Diagnostic categories or dimensions? a question for the diagnostic and statistical manual of mental disorders–. *Journal of abnormal psychology*, 114(4):494.

Xiaodong Wu, Weizhe Lin, Zhilin Wang, and Elena Rastorgueva. 2020. Author2vec: A framework for generating user embedding. *arXiv preprint arXiv:2003.11627*.

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. Place: Cambridge, MA Publisher: MIT Press.

Zihao Ye, Qipeng Guo, Quan Gan, Xipeng Qiu, and Zheng Zhang. 2019. Bp-transformer: Modelling long-range context via binary partitioning. *arXiv preprint arXiv:1911.04070*.

Da Yin, Hritik Bansal, Masoud Monajatipoor, Liunian Harold Li, and Kai-Wei Chang. 2022. Geomlama: Geo-diverse commonsense probing on multilingual pre-trained language models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 2039–2055.

Davis Yoshida, Allyson Ettinger, and Kevin Gimpel. 2020. Adding Recurrence to Pretrained Transformers for Improved Efficiency and Context Size. *arXiv:2008.07027 [cs]*. ArXiv: 2008.07027.

Ziqian Zeng, Yichun Yin, Yangqiu Song, and Ming Zhang. 2017. Socialized Word Embeddings. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pages 3915–3921, Melbourne, Australia. International Joint Conferences on Artificial Intelligence Organization.

Wanjun Zhong, Duyu Tang, Jiahai Wang, Jian Yin, and Nan Duan. 2021. UserAdapter: Few-Shot User Learning in Sentiment Analysis. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1484–1488, Online. Association for Computational Linguistics.

Ben Zhou, Qiang Ning, Daniel Khashabi, and Dan Roth. 2020. Temporal common sense acquisition with minimal supervision. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7579–7589.

Li Zhou, Laura Cabello, Yong Cao, and Daniel Hershcovich. 2023. Cross-cultural transfer learning for Chinese offensive language detection. In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, pages 8–15, Dubrovnik, Croatia. Association for Computational Linguistics.