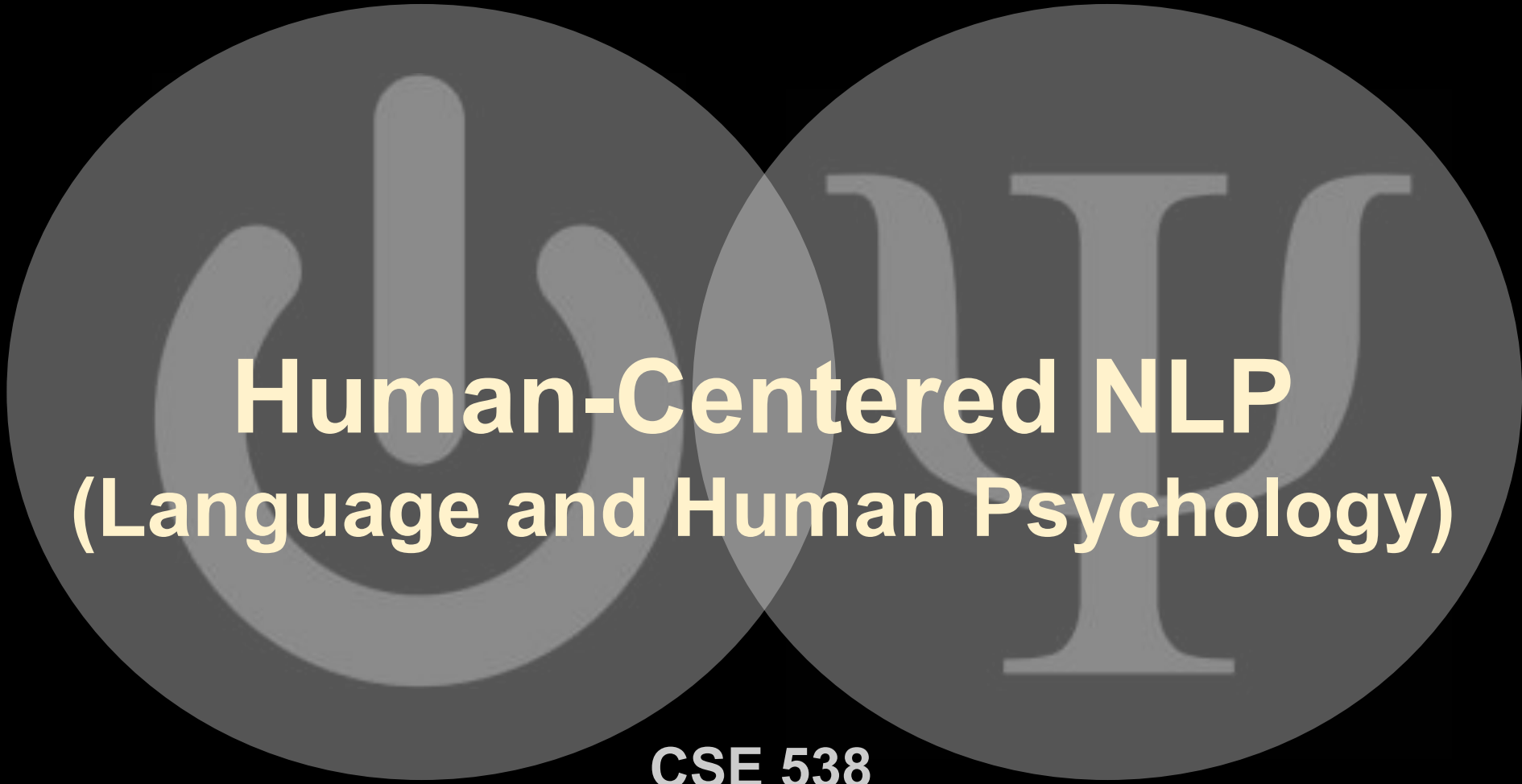
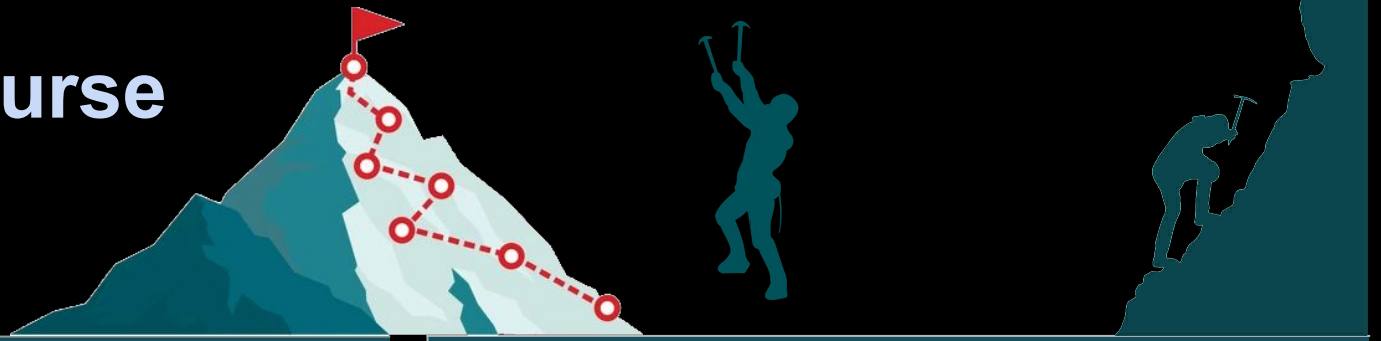


Human-Centered NLP
(Language and Human Psychology)

CSE 538

NLP, The Course



Overall NLP Concept

I. Syntax

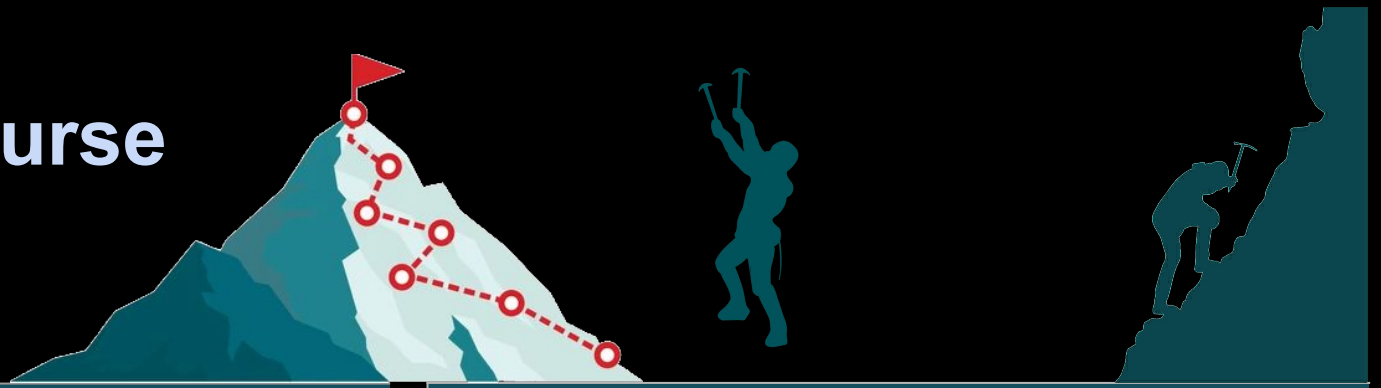
II. Semantics

Overall NLP Concept

III. Language Modeling

IV. Applications

NLP, The Course



Overall NLP Concept

I. Syntax

Introduction to NLP; Tokenization; Words Corpora

One-hot, and Multi-hot encoding.
Parts-of-Speech; Named Entities;

Parsing; Verbal Predicates;Dependency Parsing

II. Semantics

Dependency Parsing; Word Sense Disambiguation

Vector Semantics (Embeddings), Word2vec

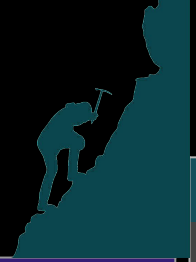
Probabilistic Language Models
Ngram Classifier, Topic Modeling

Overall NLP Concept

III. Language Modeling

IV. Applications

NLP The Course



Overall NLP Concept	Computation or ML
I. Syntax | Classification	
Introduction to NLP; Tokenization; Words Corpora	Regular Expressions; Edit Distance
One-hot, and Multi-hot encoding. Parts-of-Speech; Named Entities;	Maximum Entropy Classifier (LogReg), Gradient Descent,
Parsing; Verbal Predicates;Dependency Parsing	Cross Validation; Regularization Accuracy Metrics; Shift Reduce
II. Semantics | Probabilistic Models	
Dependency Parsing; Word Sense Disambiguation	Term Probabilities; N-d Vectors
Vector Semantics (Embeddings), Word2vec	LDA, Skipgram Model
Probabilistic Language Models Ngram Classifier, Topic Modeling	markov assumption, chain rule, smoothing

Overall NLP Concept	Computation or ML
III. Language Modeling | Transformers	
Ethical Considerations	Model cards, Pred Bias Frmwrk
Masked Language Modeling (autoencoding)	Neural Networks; Backprop Cross-Entropy Loss Self-Attention,
Generative Language Modeling (autoregressive)	Positional encodings The Transformer: Beam Search
Applying LMs	Fine-Tuning, zero-/few-shot, Instruction tuning
IV. Applications | Custom Statistical or Symbolic	
Language and Psychology (advanced sentiment)	Differential Language Analysis; Adaptive Modeling; Human LMing
Speech and Audio Processing, Dialog (chatbots)	Wave Transforms; RNNs
Question Answering, Translation	Multihop Reasoning

NLP The Course

Overall NLP Concept	Computation or ML
I. Syntax | Classification	
Introduction to NLP; Tokenization; Words Corpora	Regular Expressions; Edit Distance
One-hot, and Multi-hot encoding. Parts-of-Speech; Named Entities;	Maximum Entropy Classifier (LogReg), Gradient Descent,
Parsing; Verbal Predicates;Dependency Parsing	Cross Validation; Regularization Accuracy Metrics; Shift Reduce
II. Semantics | Probabilistic Models	
Dependency Parsing; Word Sense Disambiguation	Term Probabilities; N-d Vectors
Vector Semantics (Embeddings), Word2vec	LDA, Skipgram Model
Probabilistic Language Models Ngram Classifier, Topic Modeling	markov assumption, chain rule, smoothing

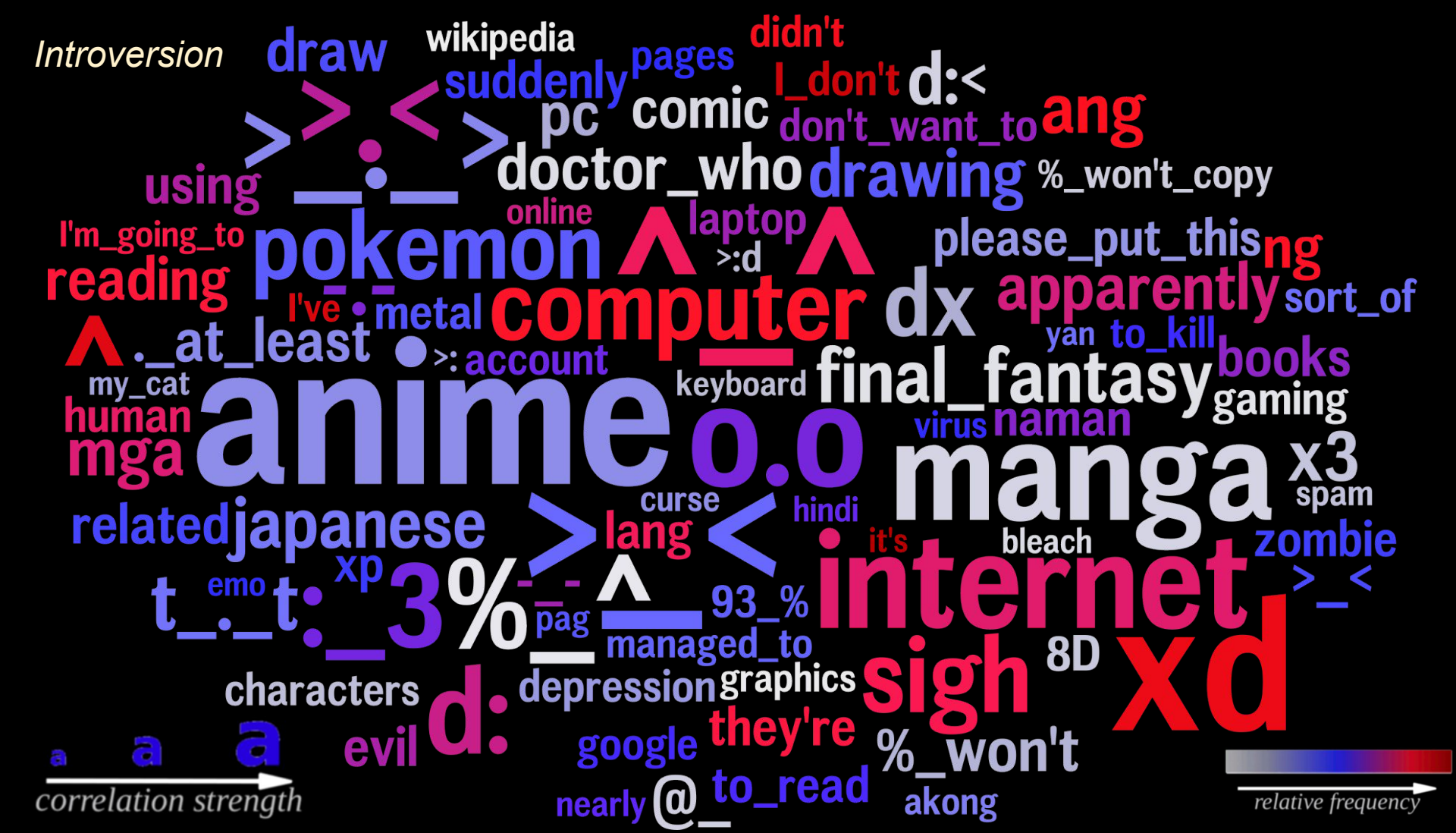
Overall NLP Concept	Computation or ML
III. Language Modeling | Transformers	
Ethical Considerations	Model cards, Pred Bias Frmwrk
Masked Language Modeling (autoencoding)	Neural Networks; Backprop Cross-Entropy Loss Self-Attention,
Generative Language Modeling (autoregressive)	Positional encodings The Transformer: Beam Search
Applying LMs	Fine-Tuning, zero-/few-shot, Instruction tuning
IV. Applications | Custom Statistical or Symbolic	
Language and Psychology (advanced sentiment)	**Differential Language Analysis; Adaptive Modeling; Human LMing**
Speech and Audio Processing, Dialog (chatbots)	Wave Transforms; RNNs
Question Answering, Translation	Multihop Reasoning


Natural
Language
Processing


Psychological
& Health
Sciences

Extraversion

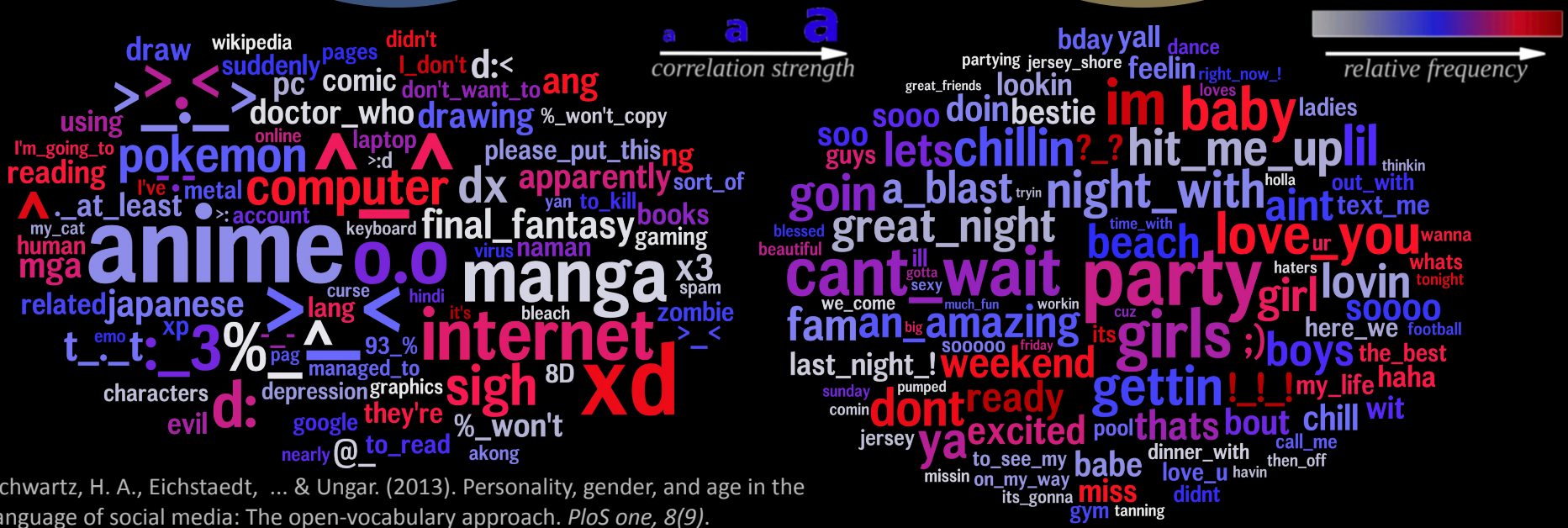
Schwartz, H. A., Eichstaedt, ... & Ungar. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one, 8(9)*.



Schwartz, H. A., Eichstaedt, ... & Ungar. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one, 8(9)*.

Natural Language Processing

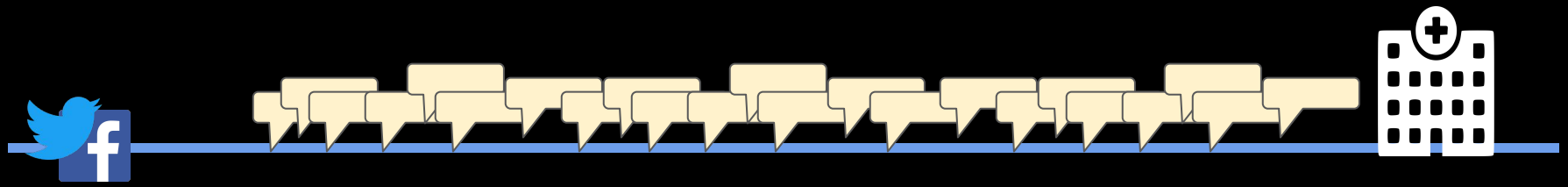
Psychological & Health Sciences



Schwartz, H. A., Eichstaedt, ... & Ungar. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one, 8(9)*.

Natural Language Processing

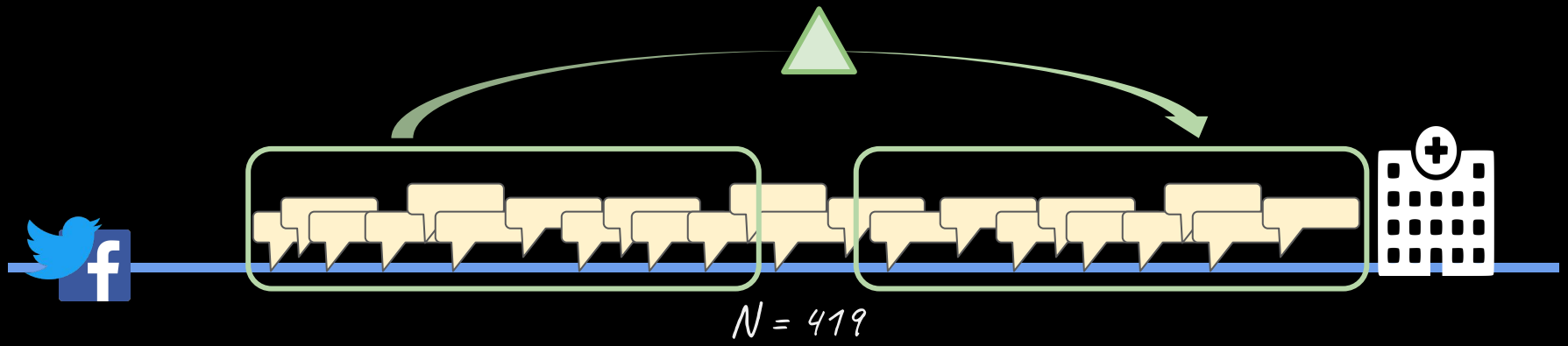
Psychological & Health Sciences



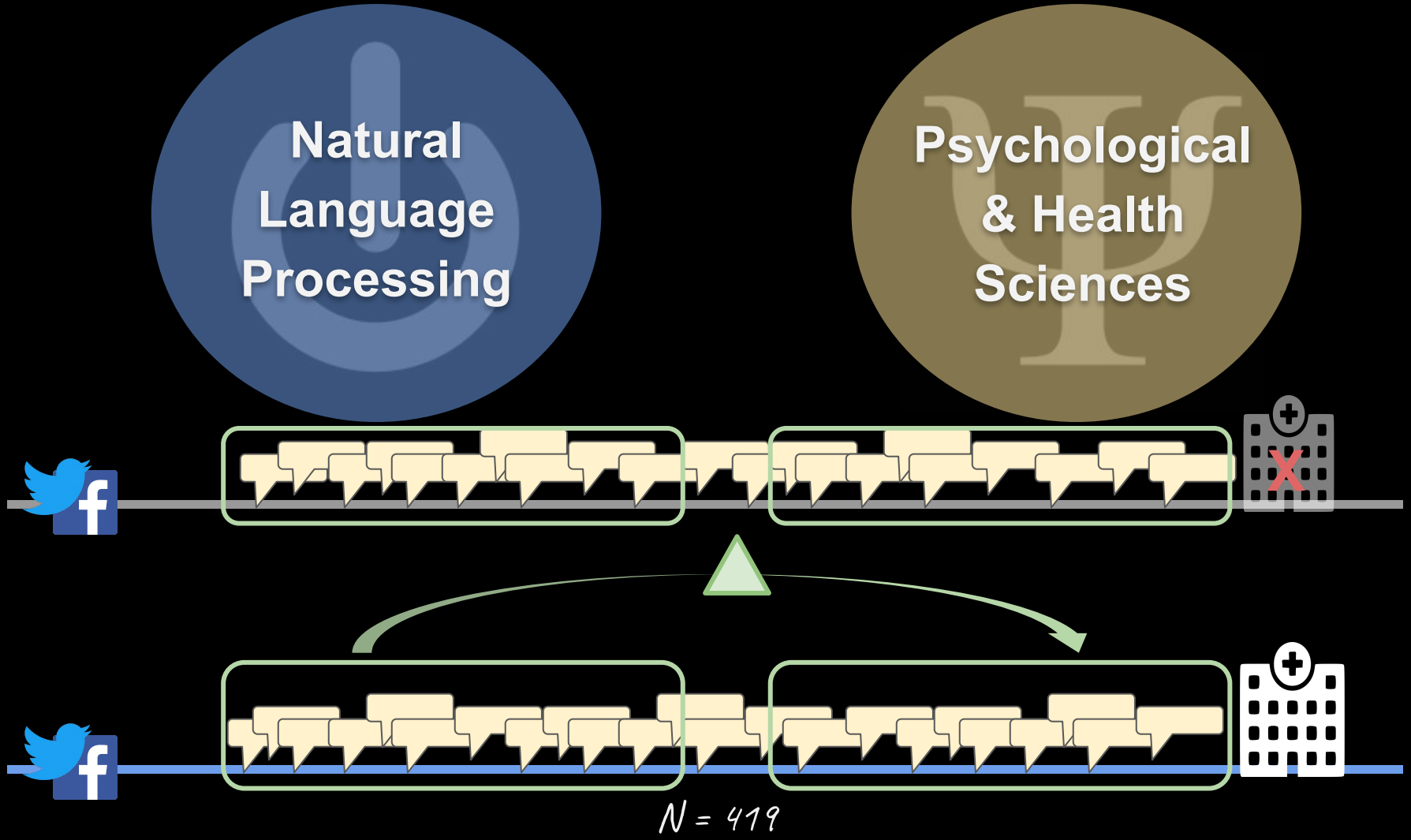
Guntuku, S. C., Schwartz, H. A., Kashyap, A., Gaulton, J. S., Stokes, D. C., Asch, D. A., ... & Merchant, R. M. (2020). Variability in Language used on Social Media prior to Hospital Visits. *Nature - Scientific Reports*, *10(1),* 1-9.

Natural Language Processing

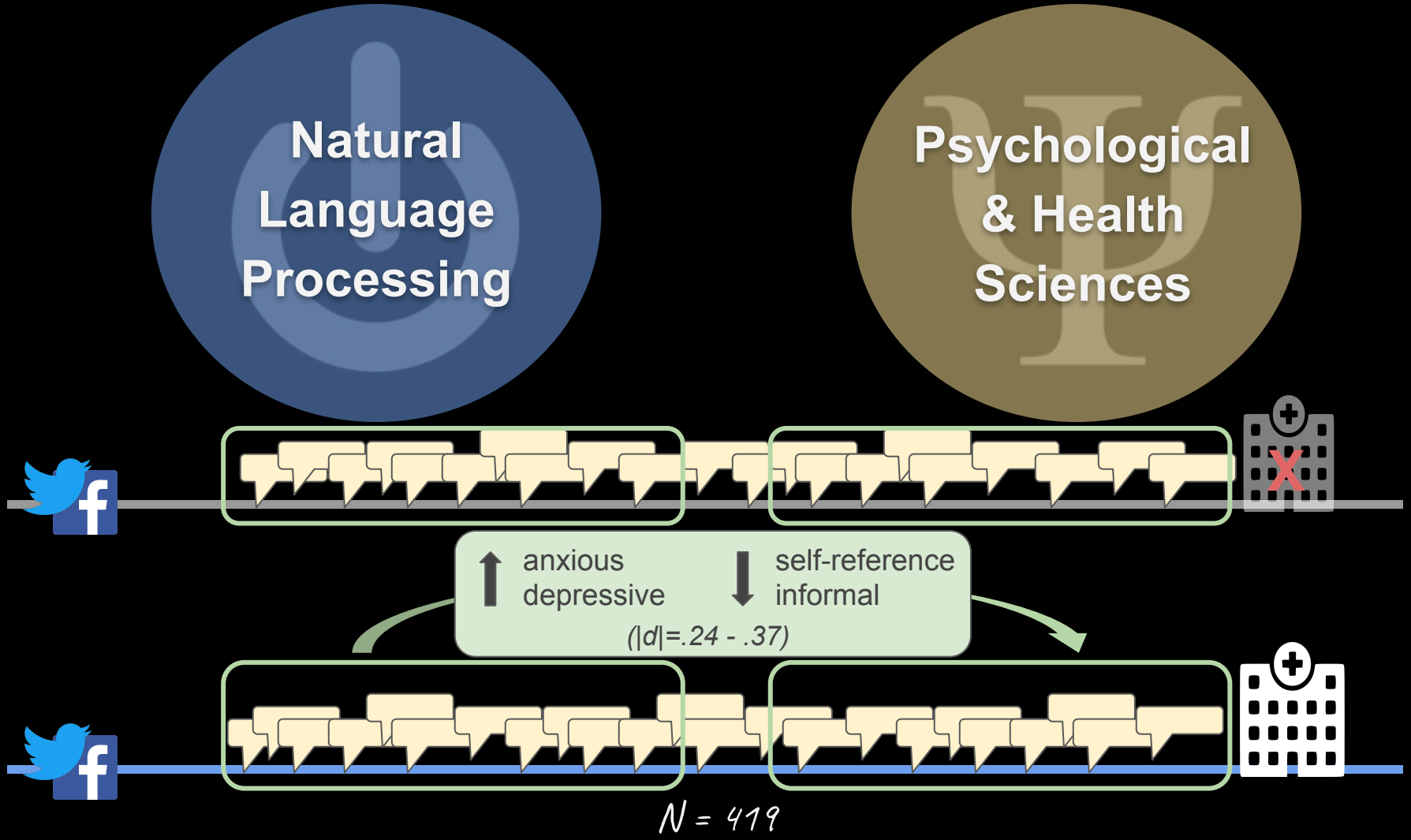
Psychological & Health Sciences



Guntuku, S. C., Schwartz, H. A., Kashyap, A., Gaulton, J. S., Stokes, D. C., Asch, D. A., ... & Merchant, R. M. (2020). Variability in Language used on Social Media prior to Hospital Visits. *Nature - Scientific Reports*, *10(1),* 1-9.



Guntuku, S. C., Schwartz, H. A., Kashyap, A., Gaulton, J. S., Stokes, D. C., Asch, D. A., ... & Merchant, R. M. (2020). Variability in Language used on Social Media prior to Hospital Visits. *Nature - Scientific Reports*, *10(1),* 1-9.



Guntuku, S. C., Schwartz, H. A., Kashyap, A., Gaulton, J. S., Stokes, D. C., Asch, D. A., ... & Merchant, R. M. (2020). Variability in Language used on Social Media prior to Hospital Visits. *Nature - Scientific Reports*, *10(1)*, 1-9.


Natural
Language
Processing


Psychological
& Health
Sciences

Overly Simplified Problem-Statement:

Natural language is written by

Overly Simplified Problem-Statement:

Natural language is written by **people.**

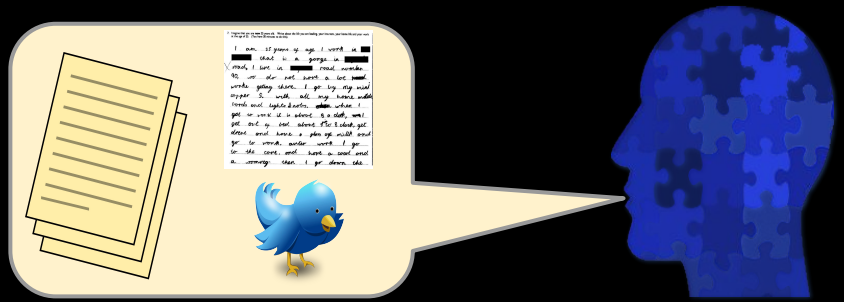
Overly Simplified Problem-Statement:

Natural language is written by **people.**

Problem

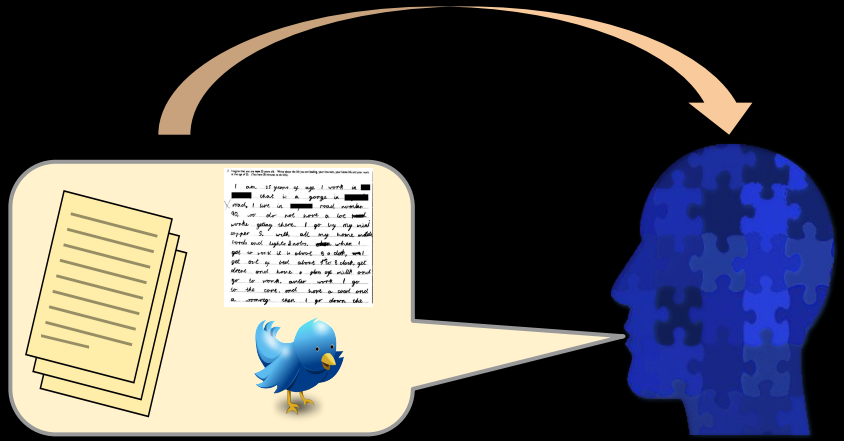
Natural language is written by **people.**

Natural language is generated by people.



People have different beliefs, backgrounds, styles, vocabularies, preferences, knowledge, personalities, …

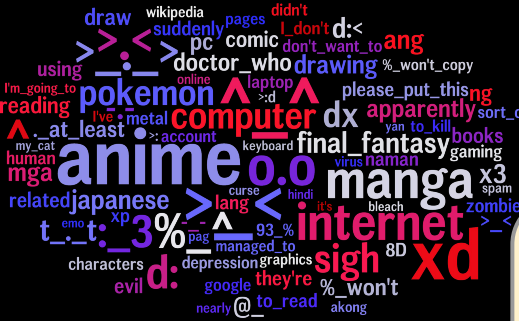
Natural language is generated by people.

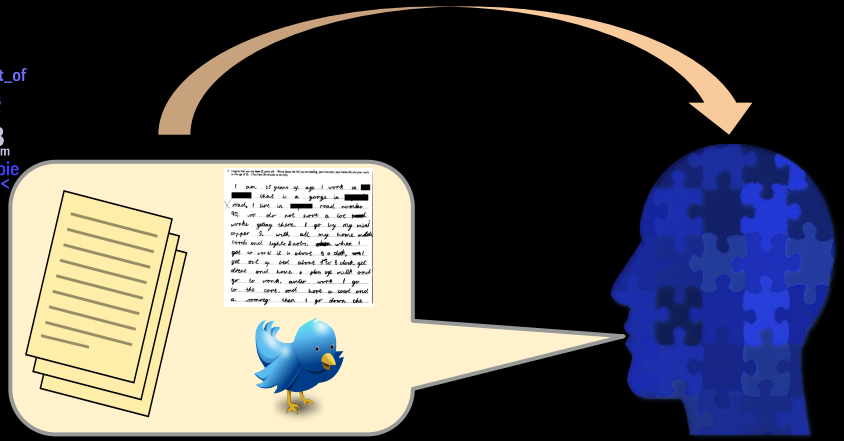


People have different beliefs, backgrounds, styles, vocabularies, preferences, knowledge, personalities, …,

and our language reflects these differences.

Natural language is generated by people.

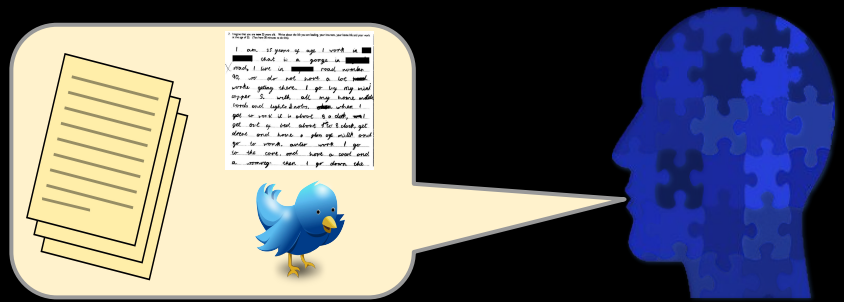




People have different beliefs, backgrounds, styles, vocabularies, preferences, knowledge, personalities, …,

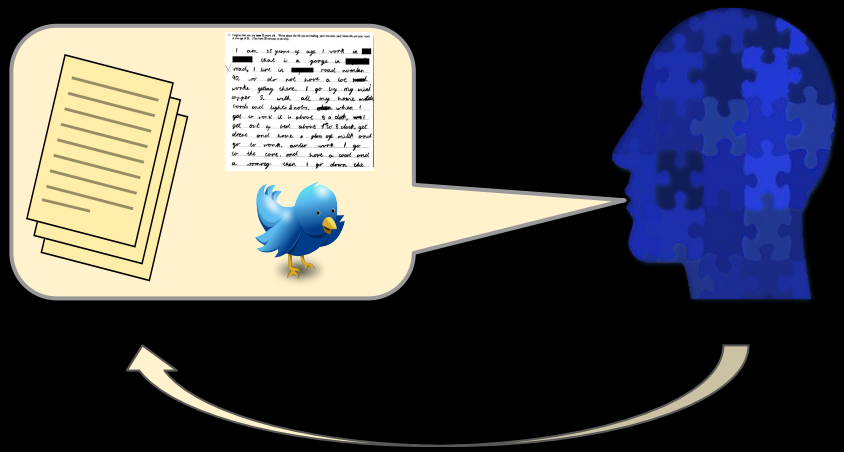
and our language reflects these differences.

Human Centered NLP:



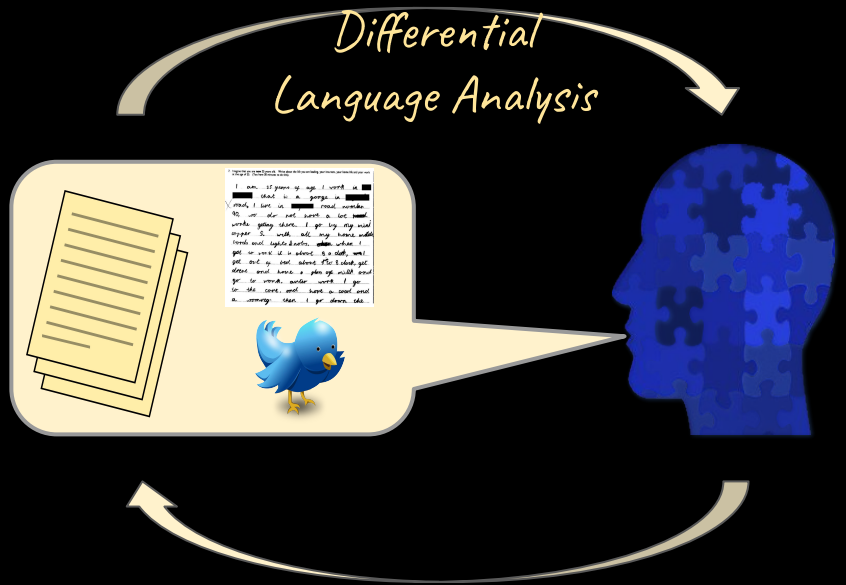
Human Centered NLP:

1. Model language as a human process



Human Centered NLP:

1. Model language as a human process
2. Use language to better understand humans.



Human-Centered NLP – We will cover:

1. Differential Language Analysis
2. Human Factor Adaptation
3. Human Language Modeling

Differential Language Analysis

Input:

Linguistic features

Human or community attribute

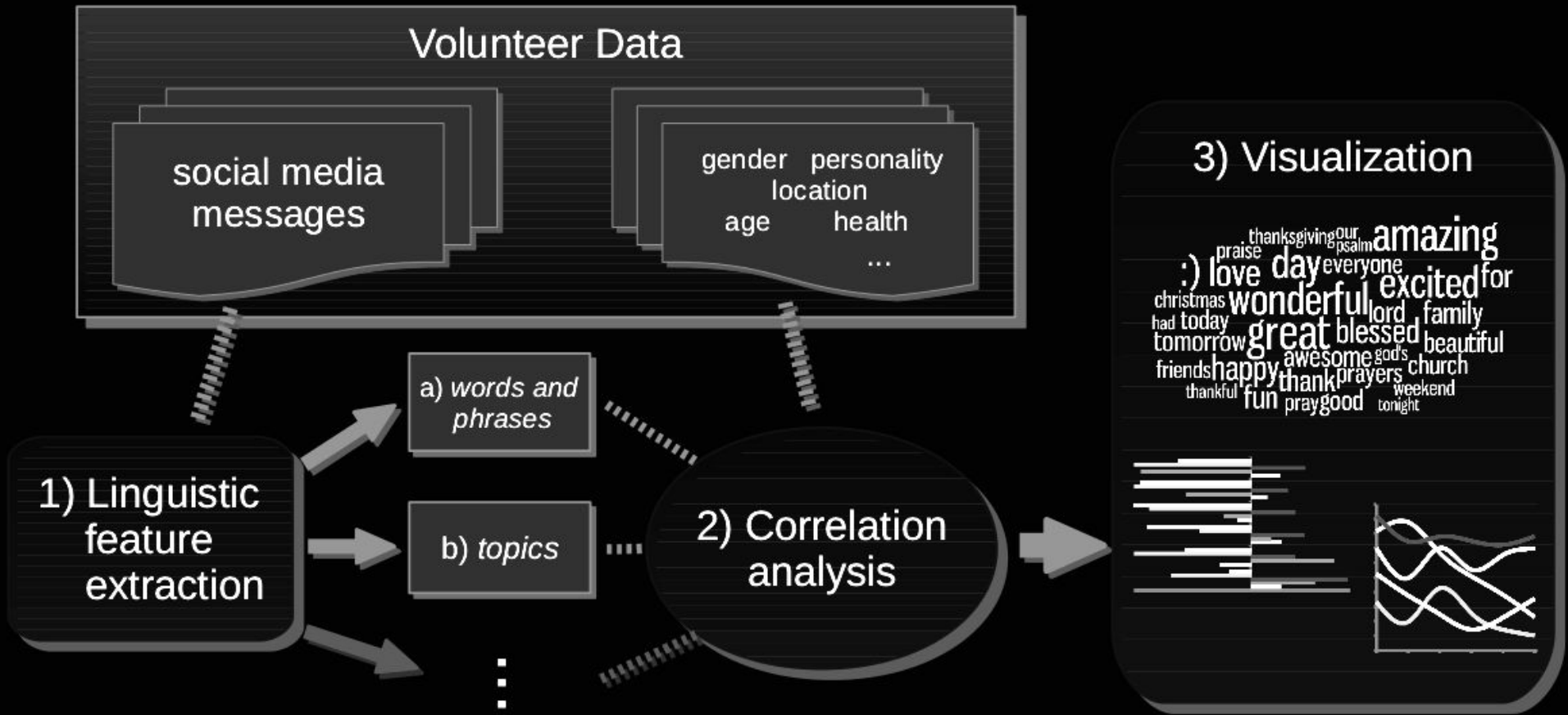
Output:

Features distinguishing attribute

Goal: Data-driven insights about an attribute

E.g. Words distinguishing communities with increases in real estate prices.

Differential Language Analysis



Differential Language Analysis

Methods of Correlation Analysis:

- Pearson Product-Moment Correlation
  Limitation: Doesn’t handle controls

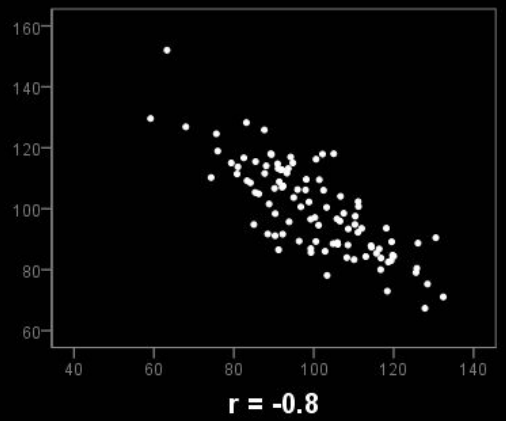
$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Differential Language Analysis

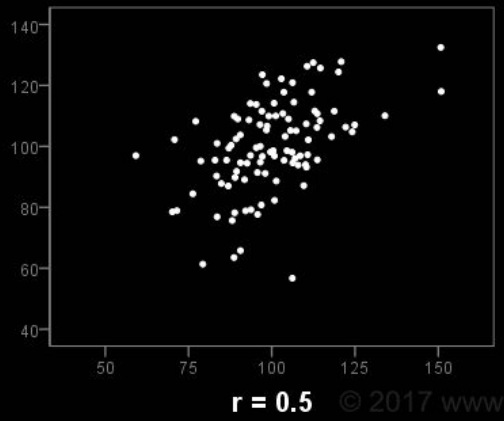
Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

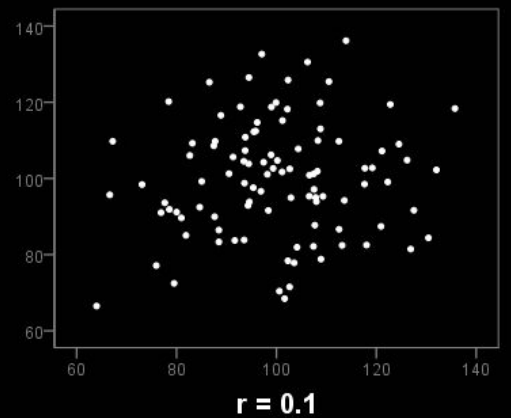
- Pearson Product-Moment Correlation
  Limitation: Doesn't handle controls



r = -0.8



r = 0.5



r = 0.1

Differential Language Analysis

Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation
  Limitation: Doesn't handle controls

- Standardized Multivariate Linear Regression
  Fit the model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$$

Differential Language Analysis

Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation
  Limitation: Doesn't handle controls

- **Standardized** Multivariate Linear Regression
  Fit the model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$$

  Adjust all variables to have "mean center" and "unit variance":

Differential Language Analysis

Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation
  Limitation: Doesn't handle controls

- **Standardized** Multivariate Linear Regression
  Fit the model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$$

Adjust all variables to have "mean center" and "unit variance":

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean
σ = Standard Deviation

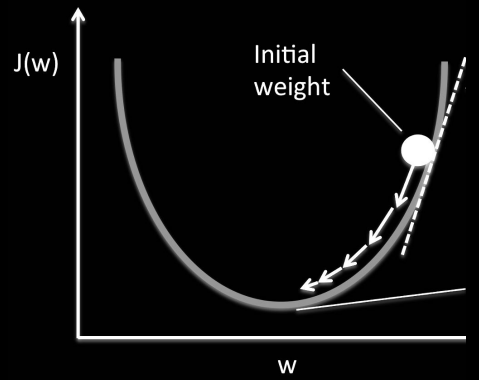
Differential Language Analysis

Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation
  Limitation: Doesn't handle controls

- Standardized Multivariate Linear Regression
  Fit the model:
  $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$$
  Option 1: Gradient Descent:
  $J = \sum (y - \hat{y})^2$ -- "Sum of Squares" Error



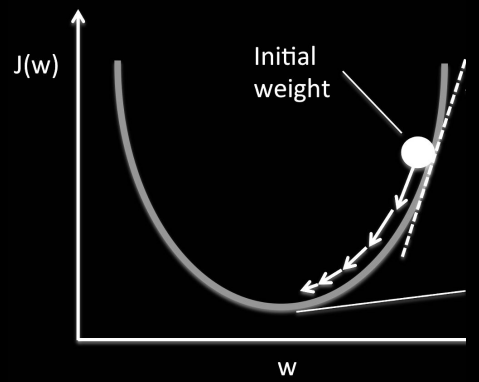
Differential Language Analysis

Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation
  Limitation: Doesn't handle controls

- Standardized Multivariate Linear Regression
  Fit the model:
  $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$$
  Option 1: Gradient Descent:
  $J = \sum (y - \hat{y})^2$ -- "Sum of Squares" Error
  Option 2: Matrix model: $Y = X\beta + \epsilon$



Differential Language Analysis

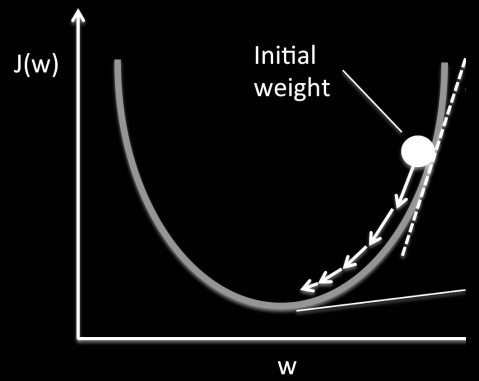
Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation
  Limitation: Doesn't handle controls

- Standardized Multivariate Linear Regression
  Fit the model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$
  Option 1: Gradient Descent:
  $J = \sum (y - \hat{y})^2$ -- "Sum of Squares" Error
  Option 2: Matrix model: $Y = X\beta + \epsilon$
  Matrix Computation Solution:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$



Differential Language Analysis

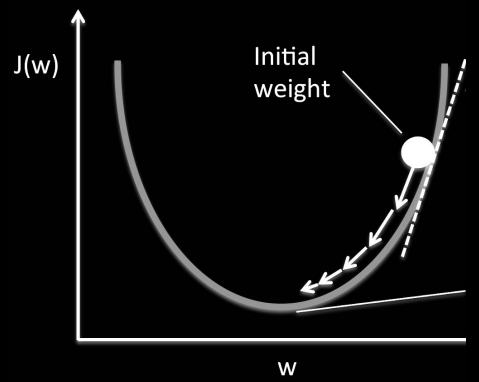
Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation
  Limitation: Doesn't handle controls

- Standardized Multivariate Linear Regression
  Fit the model:
  $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$$
  Option 1: Gradient Descent:
  $J = \sum (y - \hat{y})^2$ -- "Sum of Squares" Error
  Option 2: Matrix model: $Y = X\beta + \epsilon$
  Matrix Computation Solution:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$



Differential Language Analysis

Methods of “Correlation” Analysis for binary outcomes:

- Logistic Regression over Standardized variables
- Odds Ratio

$$\frac{\frac{countA(\text{"horrible"})}{NA}}{1-\frac{countA(\text{"horrible"})}{NA}}$$

$$\frac{\frac{countB(\text{"horrible"})}{NB}}{1-\frac{countB(\text{"horrible"})}{NB}}$$

Differential Language Analysis

Methods of “Correlation” Analysis for binary outcomes:

- Logistic Regression over Standardized variables
- Odds Ratio

$$\frac{\frac{\frac{countA(\text{"horrible"})}{NA}}{1-\frac{countA(\text{"horrible"})}{NA}}}{\frac{\frac{countB(\text{"horrible"})}{NB}}{1-\frac{countB(\text{"horrible"})}{NB}}} \propto log\left(\frac{\frac{countA(\text{"horrible"})}{NA}}{1-\frac{countA(\text{"horrible"})}{NA}}\right) - log\left(\frac{\frac{countB(\text{"horrible"})}{NB}}{1-\frac{countB(\text{"horrible"})}{NB}}\right)$$

Differential Language Analysis

Methods of “Correlation” Analysis for binary outcomes:

- Logistic Regression over Standardized variables
- Odds Ratio

$$\frac{\frac{countA("horrible")}{NA}}{1-\frac{countA("horrible")}{NA}}$$

$$\frac{\frac{countB("horrible")}{NB}}{1-\frac{countB("horrible")}{NB}}$$

$$\propto log\left(\frac{\frac{countA("horrible")}{NA}}{1-\frac{countA("horrible")}{NA}}\right) - log\left(\frac{\frac{countB("horrible")}{NB}}{1-\frac{countB("horrible")}{NB}}\right)$$

$$= log\left(\frac{countA("horrible")}{NA-countA("horrible")}\right) - log\left(\frac{countB("horrible")}{NB-countB("horrible")}\right)$$

Differential Language Analysis

$$log\left(\frac{countA(\text{“}horrible\text{”})}{NA-countA(\text{“}horrible\text{”})}\right) - log\left(\frac{countB(\text{“}horrible\text{”})}{NB-countB(\text{“}horrible\text{”})}\right)$$

- Odds Ratio using Informative Dirichlet Prior

$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i + \alpha_w}{n^i + \alpha_0 - (f_w^i + \alpha_w)}\right) - \log\left(\frac{f_w^j + \alpha_w}{n^j + \alpha_0 - (f_w^j + \alpha_w)}\right) \qquad (20.9)$$

(Monroe et al., 2010; Jurafsky, 2017)

Differential Language Analysis

$$log\left(\frac{countA("horrible")}{NA-countA("horrible")}\right) - log\left(\frac{countB("horrible")}{NB-countB("horrible")}\right)$$

- Odds Ratio using **Informative Dirichlet Prior**

$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i + \alpha_w}{n^i + \alpha_0 - (f_w^i + \alpha_w)}\right) - \log\left(\frac{f_w^j + \alpha_w}{n^j + \alpha_0 - (f_w^j + \alpha_w)}\right) \quad (20.9)$$

(where n^i is the size of corpus i, n^j is the size of corpus j, f_w^i is the count of word w in corpus i, f_w^j is the count of word w in corpus j, α_0 is the size of the background corpus, and α_w is the count of word w in the background corpus.)

(Monroe et al., 2010; Jurafsky, 2017)

Differential Language Analysis

$$log\left(\frac{countA("horrible")}{NA-countA("horrible")}\right) - log\left(\frac{countB("horrible")}{NB-countB("horrible")}\right)$$

- Odds Ratio using **Informative Dirichlet Prior**

$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i + \alpha_w}{n^i + \alpha_0 - (f_w^i + \alpha_w)}\right) \quad \log\left(\frac{f_w^j + \alpha_w}{n^j + \alpha_0 - (f_w^j + \alpha_w)}\right) \qquad (20.9)$$

(where n^i is the size of corpus i, n^j is the s pus j, f_w^i is the count of word w in corpus i, f_w^j is the count of word w in is the size of the background corpus, and α_w is the count of word w in corpus.)

Bayesian term for "smoothing": accounts for uncertainty as a function of event frequency (i.e. words observed less) by integrating "prior" beliefs mathematically.

(Mu

Differential Language Analysis

$$log\left(\frac{countA("horrible")}{NA-countA("horrible")}\right) - log\left(\frac{countB("horrible")}{NB-countB("horrible")}\right)$$

- Odds Ratio using **Informative Dirichlet Prior**

$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i + \alpha_w}{n^i + \alpha_0 - (f_w^i + \alpha_w)}\right) \quad \log\left(\frac{f_w^j + \alpha_w}{n^j + \alpha_0 - (f_w^j + \alpha_w)}\right) \qquad (20.9)$$

(where n^i is the size of corpus i, n^j is the s… …pus j, f_w^i is the count of word w in corpus i, f_w^j is the count of word w in … is the size of the background corpus, and α_w is the count of word w in … corpus.)

Bayesian term for "smoothing": accounts for uncertainty as a function of event frequency (i.e. words observed less) by integrating "prior" beliefs mathematically.
"Informative": the prior is based on past evidence. Here, the total frequency of the word.

(Mu…

Differential Language Analysis

$$log\left(\frac{countA(\text{"horrible"})}{NA-countA(\text{"horrible"})}\right) - log\left(\frac{countB(\text{"horrible"})}{NB-countB(\text{"horrible"})}\right)$$

- Odds Ratio using Informative Dirichlet Prior

$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i + \alpha_w}{n^i + \alpha_0 - (f_w^i + \alpha_w)}\right) - \log\left(\frac{f_w^j + \alpha_w}{n^j + \alpha_0 - (f_w^j + \alpha_w)}\right) \quad (20.9)$$

(where n^i is the size of corpus i, n^j is the size of corpus j, f_w^i is the count of word w in corpus i, f_w^j is the count of word w in corpus j, α_0 is the size of the background corpus, and α_w is the count of word w in the background corpus.)

(Monroe et al., 2010; Jurafsky, 2017)

Differential Language Analysis

$$log\left(\frac{countA("horrible")}{NA-countA("horrible")}\right) - log\left(\frac{countB("horrible")}{NB-countB("horrible")}\right)$$

- Odds Ratio using Informative Dirichlet Prior

$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i + \alpha_w}{n^i + \alpha_0 - (f_w^i + \alpha_w)}\right) - \log\left(\frac{f_w^j + \alpha_w}{n^j + \alpha_0 - (f_w^j + \alpha_w)}\right) \qquad (20.9)$$

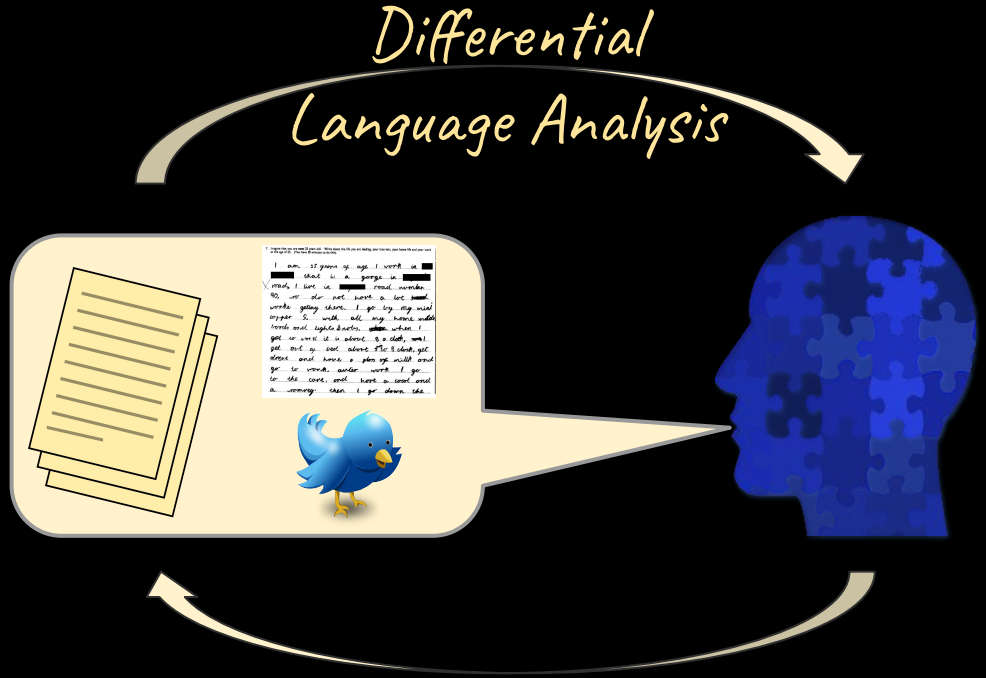
(where n^i is the size of corpus i, n^j is the size of corpus j, f_w^i is the count of word w in corpus i, f_w^j is the count of word w in corpus j, α_0 is the size of the background corpus, and α_w is the count of word w in the background corpus.)

Final score is standardized (z-scored): $\frac{\hat{\delta}_w^{(i-j)}}{\sqrt{\sigma^2\left(\hat{\delta}_w^{(i-j)}\right)}}$, where $\sigma^2\left(\hat{\delta}_w^{(i-j)}\right) \approx \frac{1}{f_w^i + \alpha_w} + \frac{1}{f_w^j + \alpha_w}$

(Monroe et al., 2010; Jurafsky, 2017)

DLATK

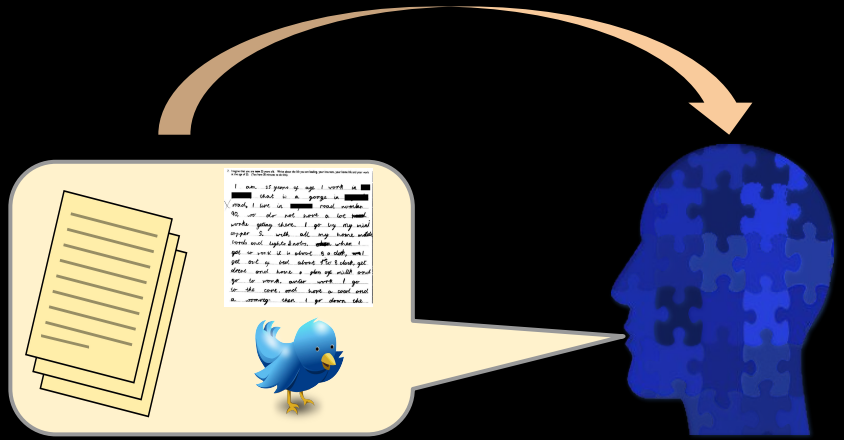
Python Library, CLI, and Colab for DLA



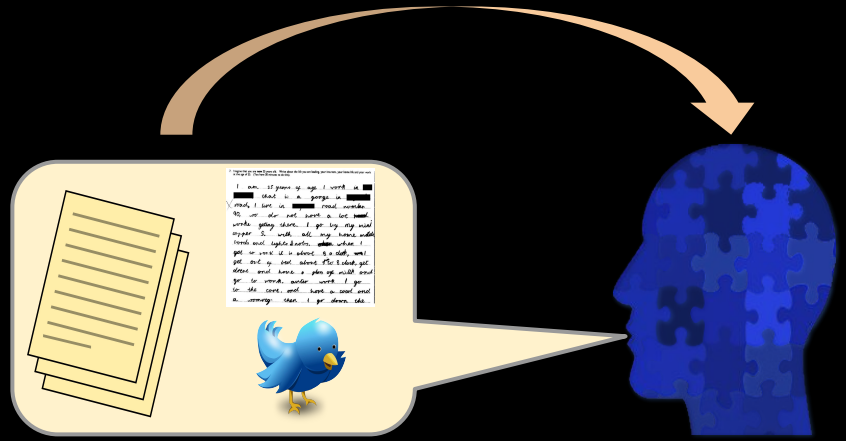
https://dlatk.github.io/

Getting Started in Colab

Natural language is generated by people.



Natural language is generated by people.



Shannon, 1948

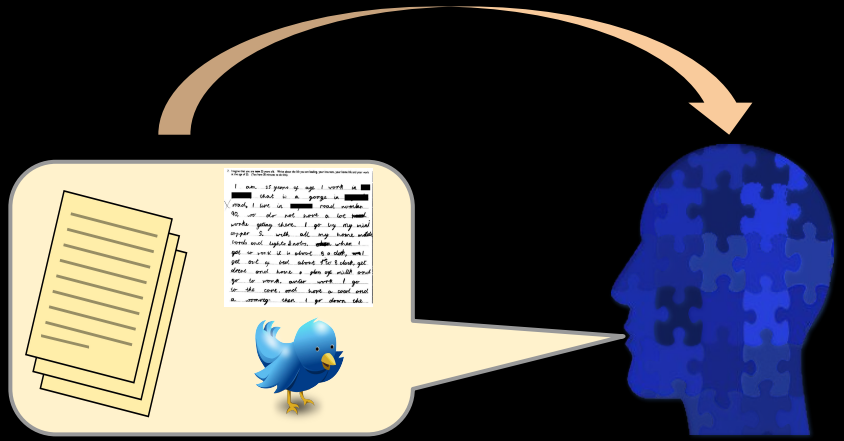
Mosteller & Wallace 1963

Clark & Schober, 1992

Mairesse, Walker, et al., 2007

Hovy & Soogaard, 2015

Natural language is generated by people.



"The common misconception is that language has got to do with words and what they mean. It does not. It has to do with people and what they mean."

Shannon, 1948

Mosteller & Wallace 1963

Clark & Schober, 1992

Mairesse, Walker, et al., 2007

Hovy & Soogaard, 2015

Human-Centered NLP – We will cover:

1. Differential Language Analysis
2. Human Factor Adaptation
3. Human Language Modeling

Human-Centered NLP – We will cover:

1. Differential Language Analysis
2. **Human Factor Adaptation**
3. Human Language Modeling

Approaches to Human Factor Inclusion

1. **Bias Mitigation:** Optimize so as not to pick up on unwanted relationships.

   (e.g. image captioner label pictures of men in kitchen as women)

Approaches to Human Factor Inclusion

1. **Bias Mitigation:** Optimize so as not to pick up on unwanted relationships.

   (e.g. image captioner label pictures of men in kitchen as women)

2. **Additive:** Include direct effect of human factor on outcome.

   (e.g. age and distinguishing PTSD from Depression; covariate in regression)

Approaches to Human Factor Inclusion

1. **Bias Mitigation:** Optimize so as not to pick up on unwanted relationships.

   (e.g. image captioner label pictures of men in kitchen as women)

2. **Additive:** Include direct effect of human factor on outcome.

   (e.g. age and distinguishing PTSD from Depression)

3. **Adaptive:** Allow meaning of language to change depending on human context. (also called “compositional”)

   (e.g. “sick” said from a young individual versus old individual)

Approaches to Human Factor Inclusion

1. **Bias Mitigation:** Opti so as not to pick up on
   (e.g. image captioner label pictures of men in kitchen as women)

What are human “factors”?

2. **Additive:** Include direct effect of human factor on outcome.
   (e.g. age and distinguishing PTSD from Depression)

3. **Adaptive:** Allow meaning of language to change depending on human context. (also called “compositional”)
   (e.g. “sick” said from a young individual versus old individual)

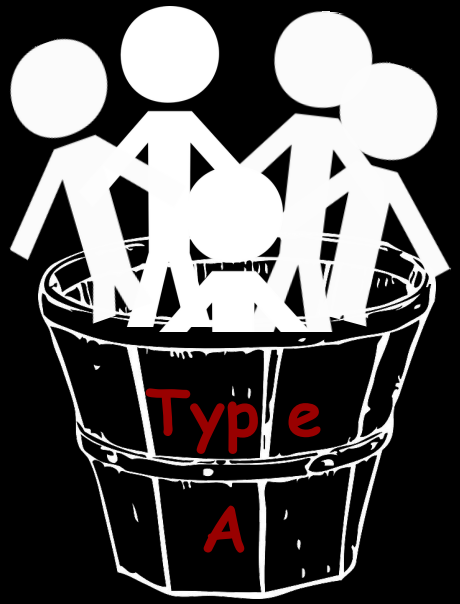
Human Factors

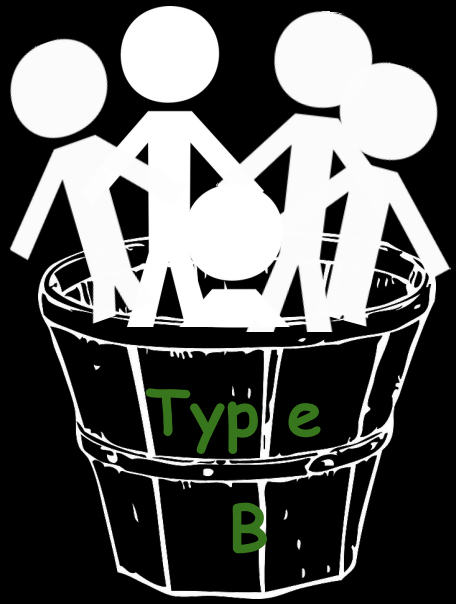
--- Any attribute, represented as a continuous or discrete variable, of the humans generating the natural language.

E.g.

- Gender
- Age
- Personality
- Ethnicity
- Socio-economic status

Human Factors

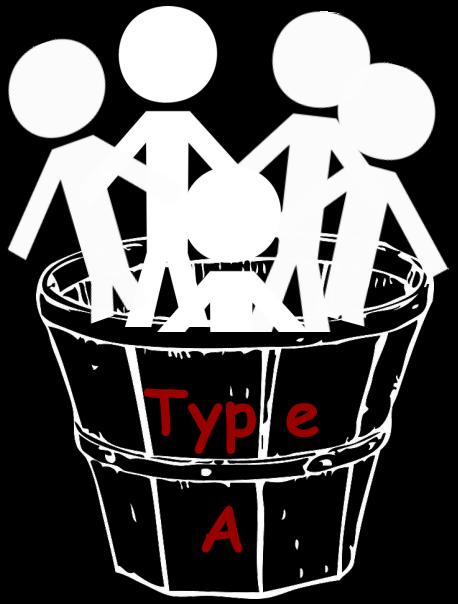




typically requires putting people into discrete bins

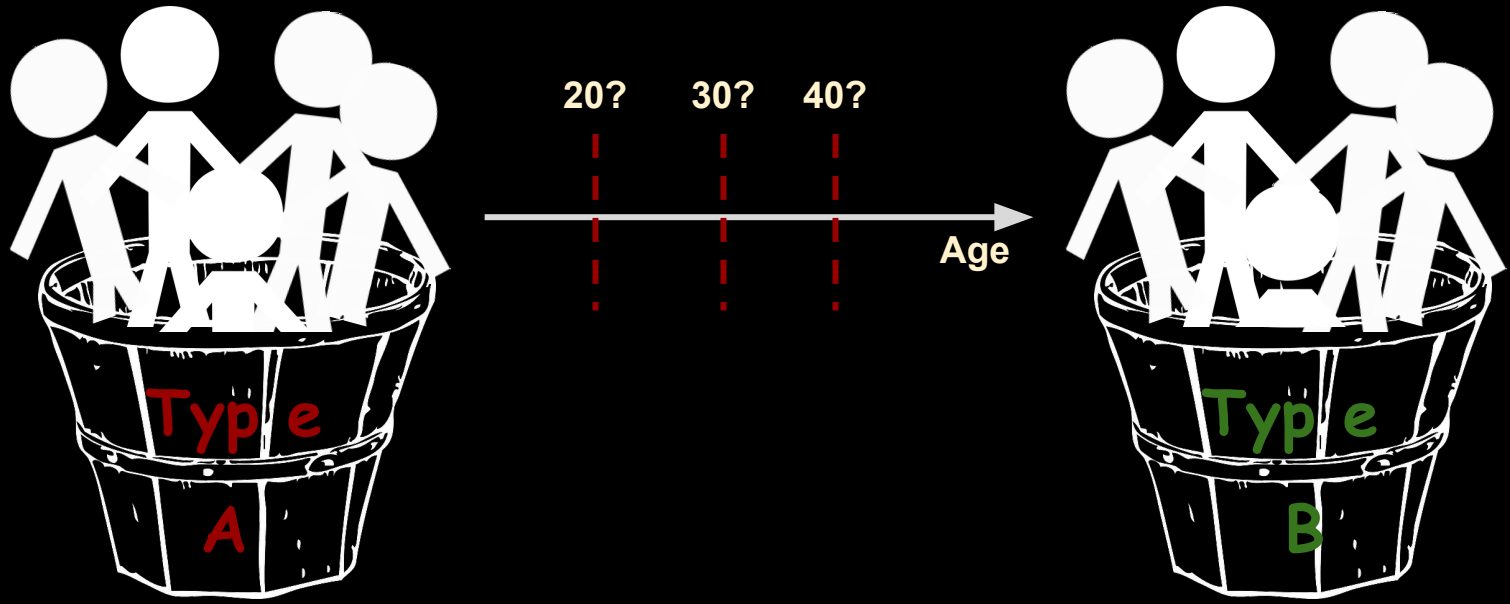
"most latent variables of interest to psychiatrists and personality and clinical psychologists are dimensional [continuous]"

(Haslam et al., 2012)



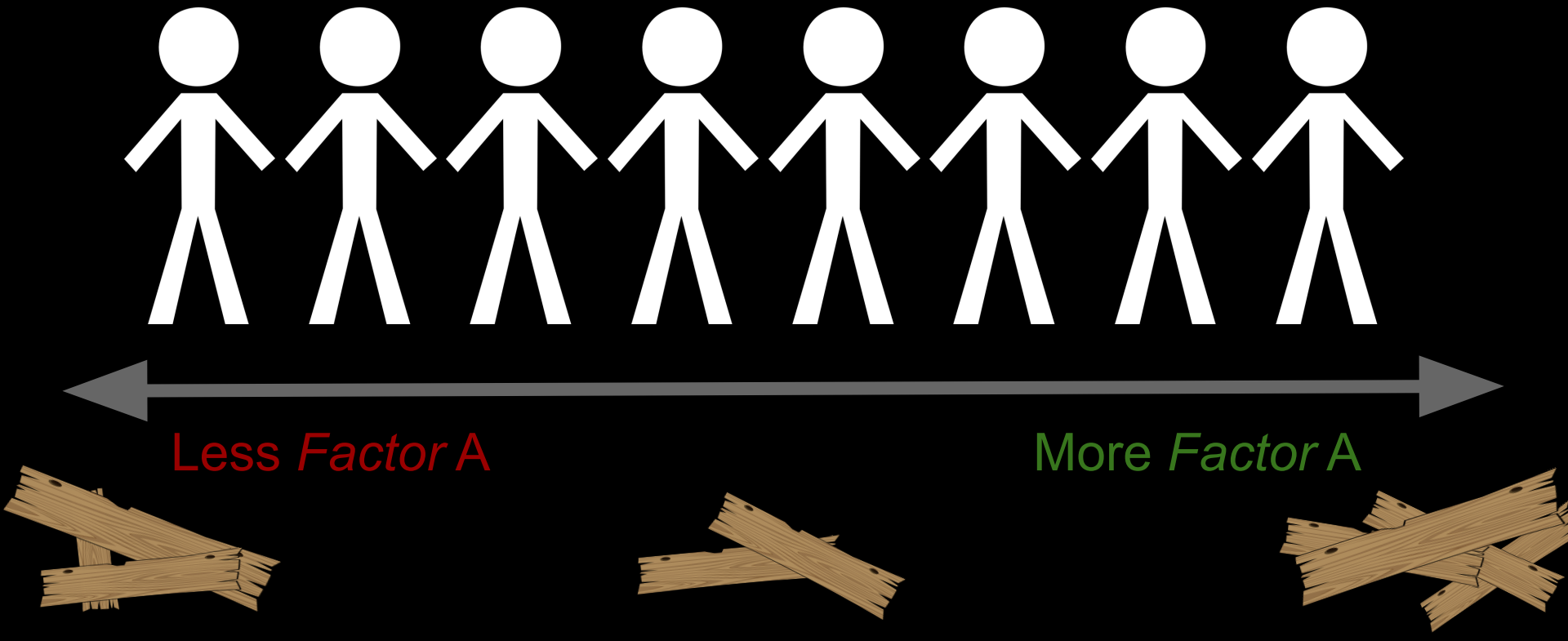
"most latent variables of interest to psychiatrists and personality and clinical psychologists are dimensional [continuous]"

(Haslam et al., 2012)



"most latent variables of interest to psychiatrists and personality and clinical psychologists are dimensional [continuous]"

(Haslam et al., 2012)



Adaptation Approach: Domain Adaptation

Features for: source target

$$\Phi^s(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{x}, \boldsymbol{0} \rangle, \quad \Phi^t(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{0}, \boldsymbol{x} \rangle$$

Frustratingly Easy Domain Adaptation

Hal Daumé III
School of Computing
University of Utah
Salt Lake City, Utah 84112
me@hal3.name

Abstract

We describe an approach to domain adaptation that is appropriate exactly in the case

supervised case. The fully supervised case models the following scenario. We have access to a large, annotated corpus of data fr

Adaptation Approach: Domain Adaptation

Features for: source target

$$\Phi^s(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{x}, \boldsymbol{0} \rangle, \quad \Phi^t(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{0}, \boldsymbol{x} \rangle$$

```
newX = []
for all x in source_x:
  newX.append(x + x + [0]*len(x))
for all x in target_x:
  newX.append(x + [0]*len(x), x)
```

Frustratingly Easy Domain Adaptation

Hal Daumé III
School of Computing
University of Utah
Salt Lake City, Utah 84112
me@hal3.name

Abstract

We describe an approach to domain adaptation that is appropriate exactly in the case

supervised case. The fully supervised case models the following scenario. We have access to a large, annotated corpus of data fr

Adaptation Approach: Domain Adaptation

Features for: source target

$$\Phi^s(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{x}, \boldsymbol{0} \rangle, \quad \Phi^t(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{0}, \boldsymbol{x} \rangle$$

```
newX = []
for all x in source_x:
  newX.append(x + x + [0]*len(x))
for all x in target_x
  newX.append(x + [0]*len(x), x)

newY = source_y + target_y

model = model.train(newX,newY)
```

Adaptation Approach: Factor Adaptation

Human Centered NLP with User-Factor Adaptation

Veronica E. Lynn, Youngseo Son, Vivek Kulkarni
Niranjan Balasubramanian and **H. Andrew Schwartz**
Stony Brook University
Stony Brook, NY
{velynn, yson, vvkulkarni, niranjan, has}@cs.stonybrook.edu

Abstract

We pose the general task of *user-factor adaptation* — adapting supervised learning models to real-valued user factors inferred from a background of their la

and Costa Jr., 1989; Ruscio and Ruscio, 2000; Widiger and Samuel, 2005).

Here, we ask how one can adapt NLP models to real-valued human *factors* – continuous valued attributes that capture fine-grained differences be-

Residualized Factor Adaptation for Community Social Media Prediction Tasks

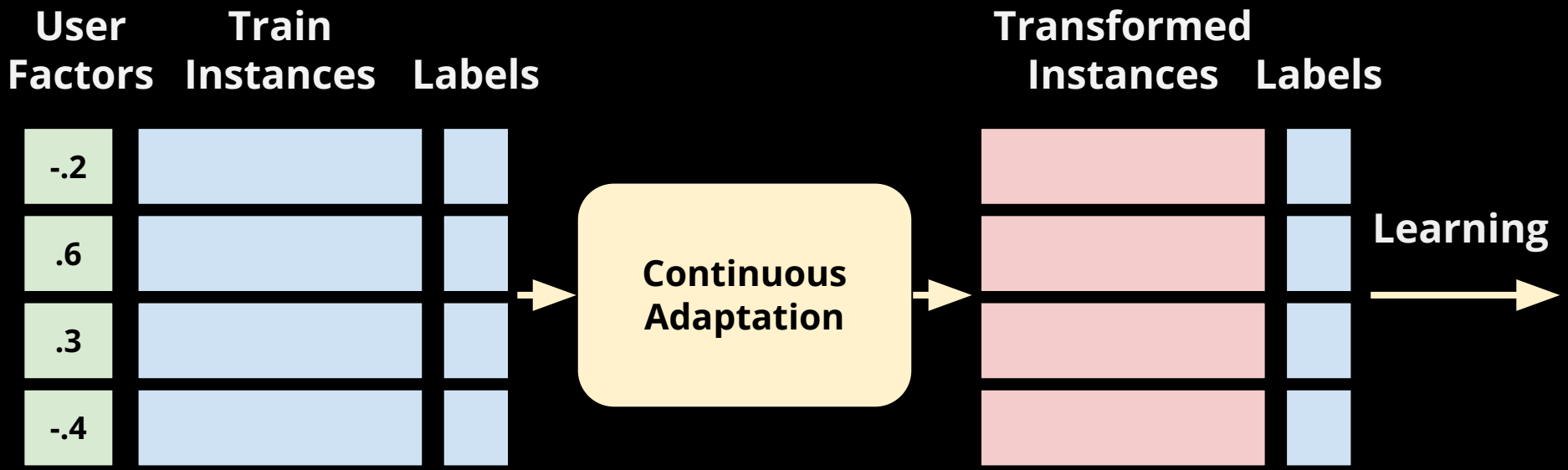
Mohammadzaman Zamani,[1] **H. Andrew Schwartz,**[1] **Veronica E. Lynn,**[1]
Salvatore Giorgi,[2] and **Niranjan Balasubramanian**[1]
[1] Computer Science Department, Stony Brook University
[2]Department of Psychology, University of Pennsylvania
mzamani@cs.stonybrook.edu

Abstract

Predictive models over social media language

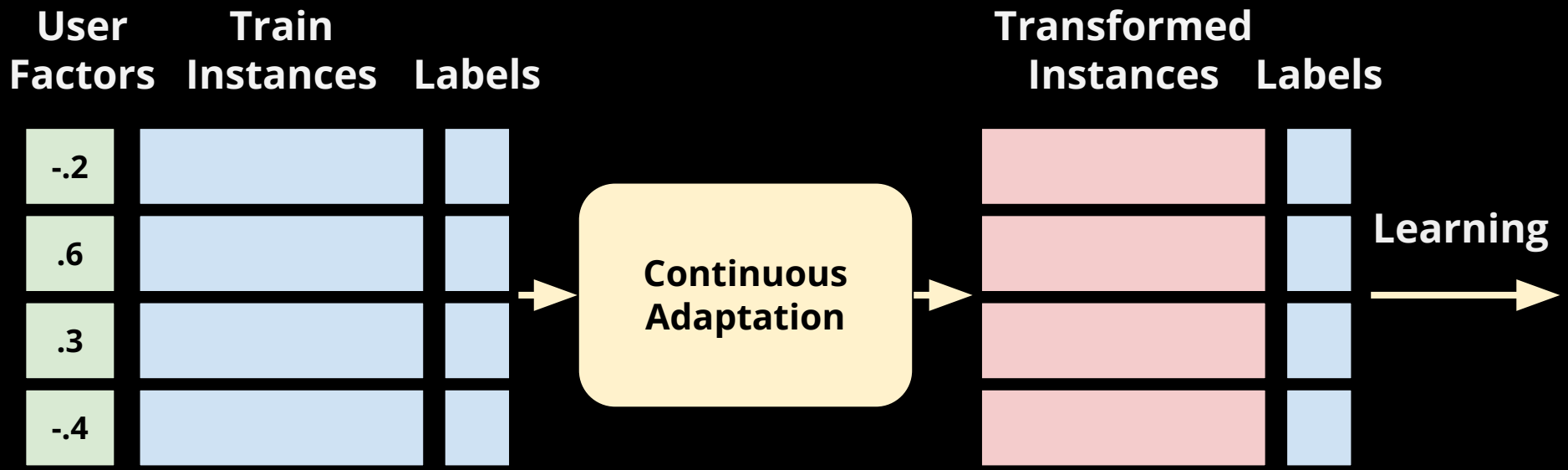
linked to socio-demographic factors (age, gender, race, education, income levels) with many social scientific studies supporting their predictive

Our Method: Continuous Adaptation



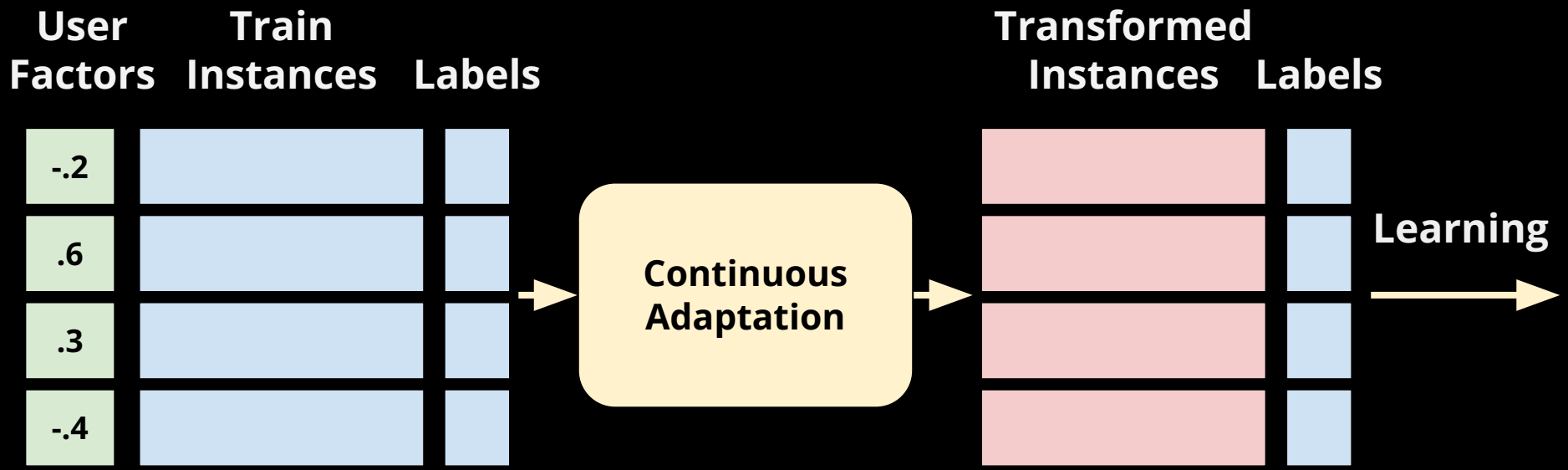
(Lynn et al., 2017)

Our Method: Continuous Adaptation



(Lynn et al., 2017)

Our Method: Continuous Adaptation



Gender Score	Features		Original	Gender Copy
-.2	X	→	X	*compose*(-.2, X)

(Lynn et al., 2017)

User Factor Adaptation: Handling multiple factors

Replicate features for each factor:

A compositional function c combines d user factor scores $f_{u,d}$ with original feature values $\mathbf{x}$:

$$\Phi(\mathbf{x}, u) = \langle \mathbf{x}, c(f_{u,1}, \mathbf{x}), c(f_{u,2}, \mathbf{x}), \cdots, c(f_{u,d}, \mathbf{x}) \rangle$$

(Lynn et al., 2017)

User Factor Adaptation: Handling multiple factors

Replicate features for each factor:

A compositional function c combines d user factor scores $f_{u,d}$ with original feature values $\mathbf{x}$:

$$\Phi(\mathbf{x}, u) = \langle \mathbf{x}, c(f_{u,1}, \mathbf{x}), c(f_{u,2}, \mathbf{x}), \cdots, c(f_{u,d}, \mathbf{x}) \rangle$$

User	Factor Classes	Augmented Instance $\Phi(\mathbf{x}, u)$
User 1	F_1	$\langle \mathbf{x}, \mathbf{x}, \mathbf{0}, \mathbf{0}, \cdots, 0 \rangle$
User 2	F_2	$\langle \mathbf{x}, \mathbf{0}, \mathbf{x}, \mathbf{0}, \cdots, 0 \rangle$
User 3	F_1, F_3	$\langle \mathbf{x}, \mathbf{x}, \mathbf{0}, \mathbf{x}, \cdots, 0 \rangle$
User 4	F_k	$\langle \mathbf{x}, \mathbf{0}, \mathbf{0}, \cdots, 0, \mathrm{x} \rangle$

Table 1: Discrete Factor Adaptation: Augmentations of an original instance vector $\mathbf{x}$ under different factor class mappings. With k domains the augmented feature vector is of length $n(k+1)$.

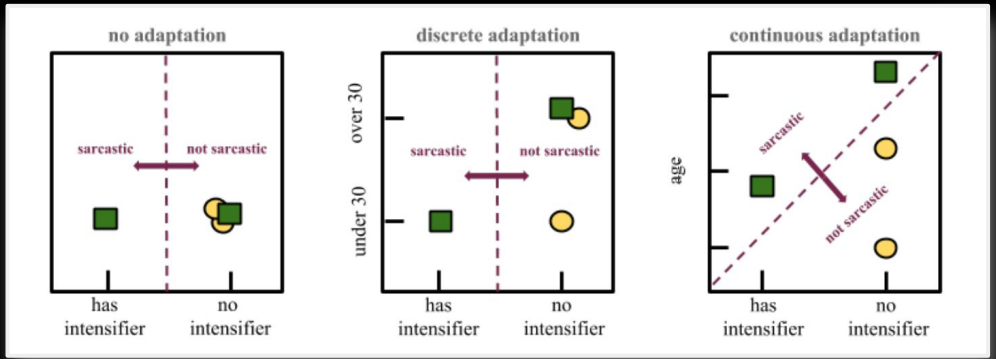
(Lynn et al., 2017)

User Factor Adaptation: Handling multiple factors

Replicate features for each factor:

A compositional function c combines d user factor scores $f_{u,d}$ with original feature values $\mathbf{x}$:

$$\Phi(\mathbf{x}, u) = \langle \mathbf{x}, c(f_{u,1}, \mathbf{x}), c(f_{u,2}, \mathbf{x}), \cdots, c(f_{u,d}, \mathbf{x}) \rangle$$



User	Factor Classes	Augmented Instance $\Phi(\mathbf{x}, u)$
User 1	F_1	$\langle \mathbf{x}, \mathbf{x}, \mathbf{0}, \mathbf{0}, \cdots, 0 \rangle$
User 2	F_2	$\langle \mathbf{x}, \mathbf{0}, \mathbf{x}, \mathbf{0}, \cdots, 0 \rangle$
User 3	F_1, F_3	$\langle \mathbf{x}, \mathbf{x}, \mathbf{0}, \mathbf{x}, \cdots, 0 \rangle$
User 4	F_k	$\langle \mathbf{x}, \mathbf{0}, \mathbf{0}, \cdots, 0, \mathrm{x} \rangle$

Table 1: Discrete Factor Adaptation: Augmentations of an original instance vector $\mathbf{x}$ under different factor class mappings. With k domains the augmented feature vector is of length $n(k+1)$.

(Lynn et al., 2017)

Main Results

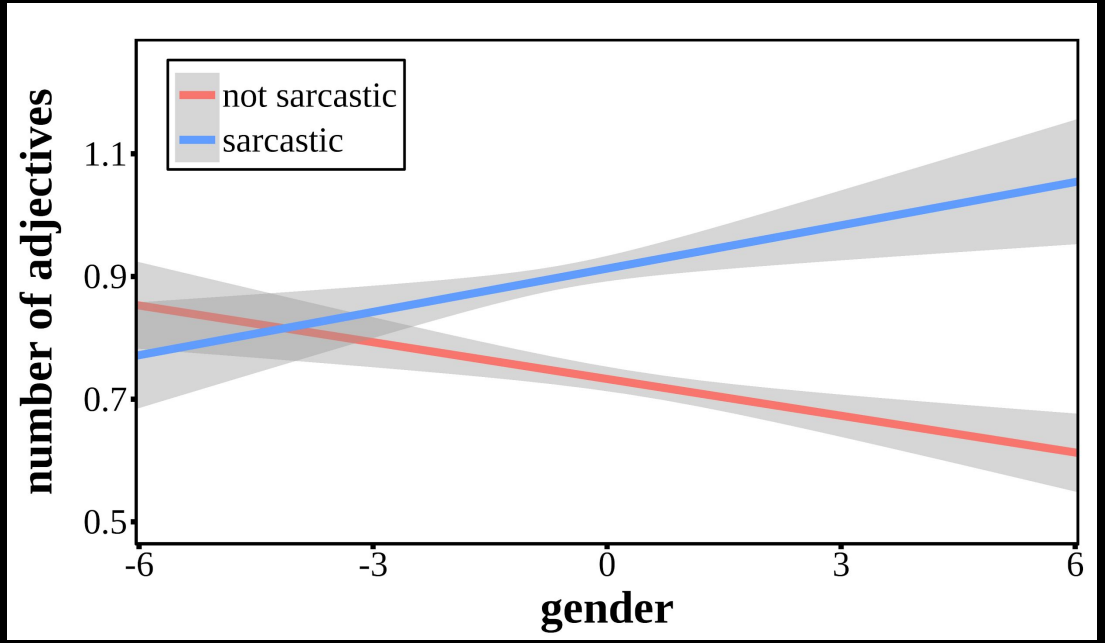
Adaptation improves over unadapted baselines (Lynn et al., 2017)

Task	Metric	No Adaptation	Gender	Personality	Latent (User Embed)
Stance	F1	64.9	**65.1 (+0.2)**	**66.3 (+1.4)**	**67.9 (+3.0)**
Sarcasm	F1	73.9	**75.1 (+1.2)**	**75.6 (+1.7)**	**77.3 (+3.4)**
Sentiment	Acc.	60.6	**61.0 (+0.4)**	**61.2 (+0.6)**	**60.7 (+0.1)**
PP-Attach	Acc.	71.0	70.7 (-0.3)	70.2 (-0.8)	70.8 (-0.2)
POS	Acc.	91.7	**91.9 (+0.2)**	91.2 (-0.5)	90.9 (-0.8)

Example: How Adaptation Helps

Women
more adjectives→sarcasm

Men
more adjectives→no sarcasm



more “male”

more “female”

Problem

User factors are not always available.

Solution: User Factor Inference

past tweets

→ **inferred factors**

Known

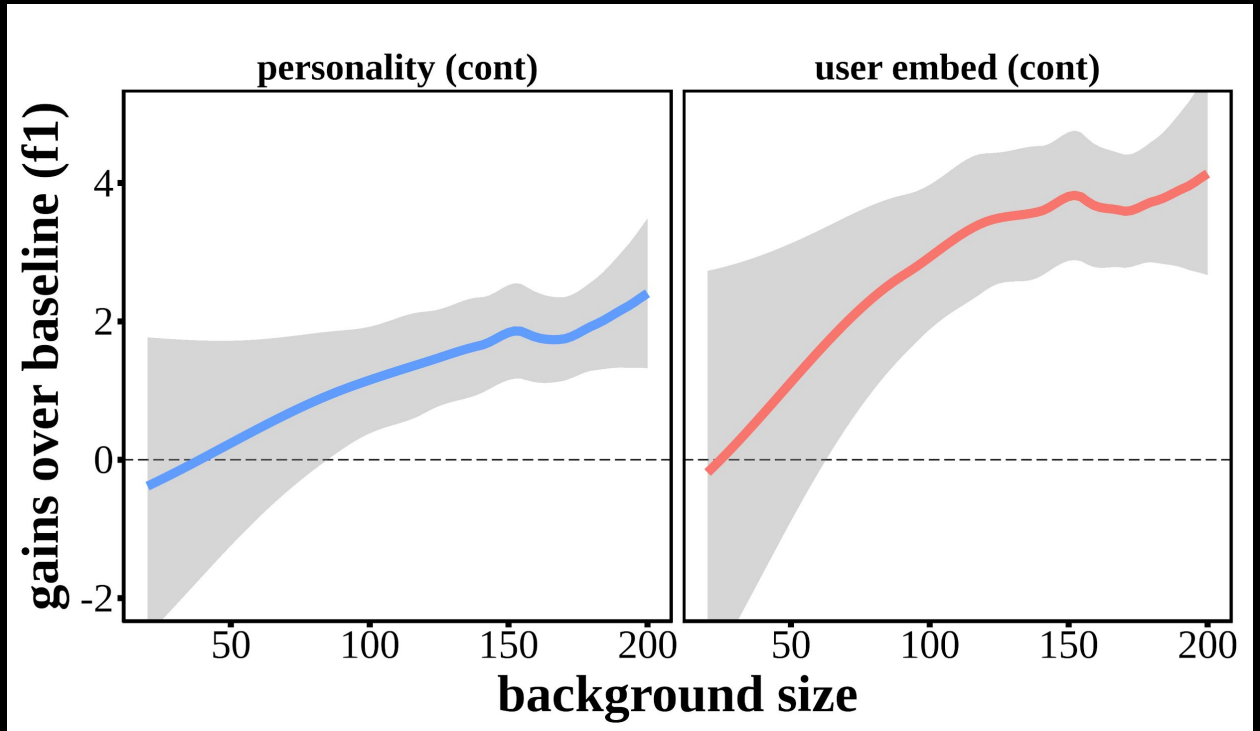
Age (Sap et al. 2014)
Gender (Sap et al. 2014)
Personality (Park et al. 2015)

Latent

User Embeddings
(Kulkarni et al. 2017)
Word2Vec
TF-IDF

Background Size

Using more background tweets to infer factors produces larger gains



Approaches to Human Factor Inclusion

1. **Bias Mitigation:** Optimize so as not to pick up on unwanted relationships.

   (e.g. image captioner label pictures of men in kitchen as women)

2. **Additive:** Include direct effect of human factor on outcome.

   (e.g. age and distinguishing PTSD from Depression)

3. **Adaptive:** Allow meaning if language to change depending on human context. (also called “compositional”)

   (e.g. “sick” said from a young individual versus old individual)

Approaches to Human Factor Inclusion

1. **Bias Mitigation:** Optimize so as not to pick up on unwanted relationships.

   (e.g. image captioner label pictures of men in kitchen as women)

2. **Additive:** Include direct effect of human factor on outcome.

   (e.g. age and distinguishing PTSD from Depression)

3. **Adaptive:** Allow meaning if language to change depending on human context. (also called “compositional”)

   (e.g. “sick” said from a young individual versus old individual)

Human-Centered NLP – We will cover:

1. Differential Language Analysis
2. **Human Factor Adaptation**
3. Human Language Modeling

Human-Centered NLP – We will cover:

1. Differential Language Analysis
2. Human Factor Adaptation
3. **Human Language Modeling**

Language Modeling

probability of a token sequence

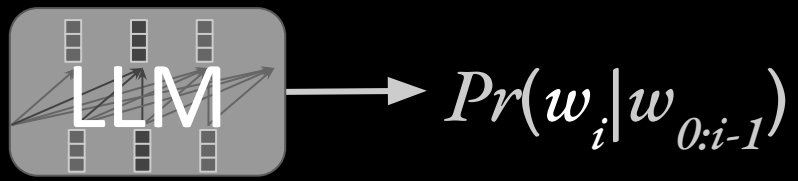
$$Pr(\mathbf{W}) = \prod_{i=1}^{n} Pr(w_i | w_{1:i-1})$$

Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

Language Modeling

Language is generated by _____

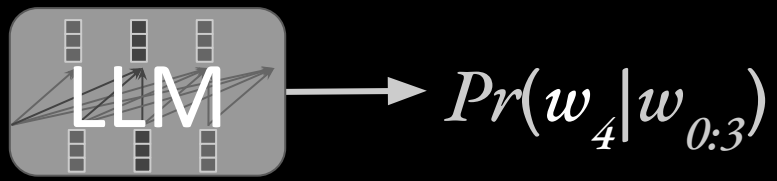
w_0 w_1 w_2 w_3 w_4

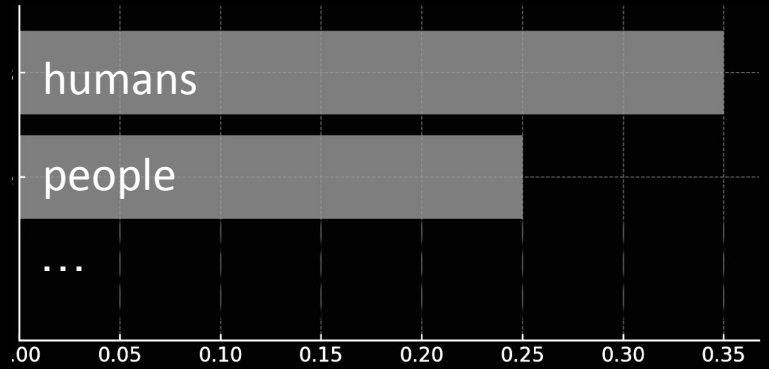


Language Modeling

Language is generated by _____

w_0 w_1 w_2 w_3 w_4





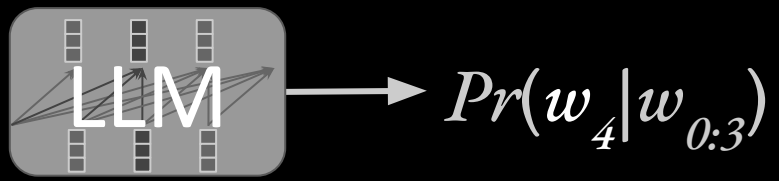
next word probability

(GPT4o)

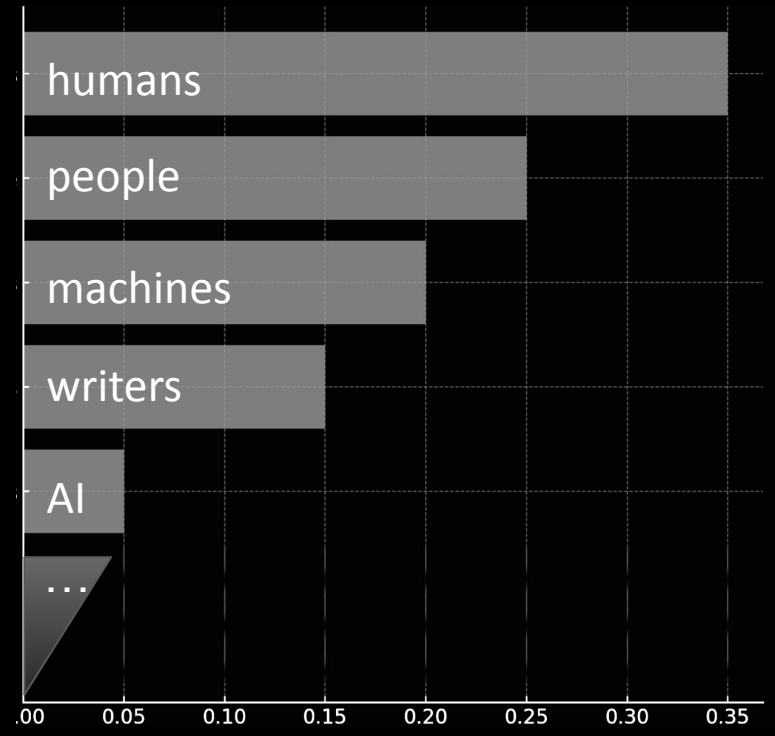
Language Modeling

Language is generated by _____

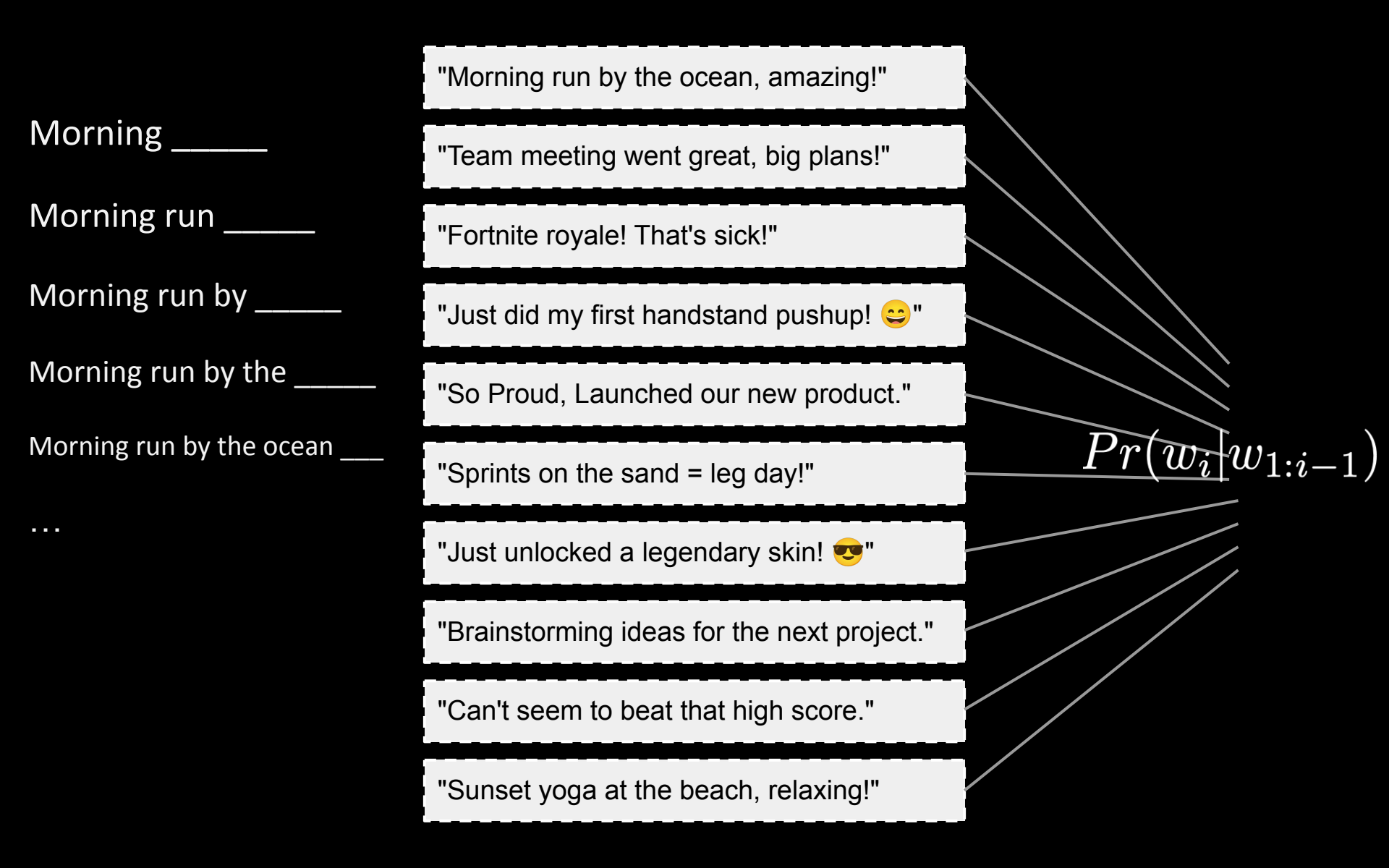
w_0 w_1 w_2 w_3 w_4



(~1 million words)



(GPT4o)


Morning _____
Morning run _____
Morning run by _____
Morning run by the _____
Morning run by the ocean ___
…
"Morning run by the ocean, amazing!"
"Team meeting went great, big plans!"
"Fortnite royale! That's sick!"
"Just did my first handstand pushup! 😄"
"So Proud, Launched our new product."
"Sprints on the sand = leg day!"
"Just unlocked a legendary skin! 😎"
"Brainstorming ideas for the next project."
"Can't seem to beat that high score."
"Sunset yoga at the beach, relaxing!"
$Pr(w_i|w_{1:i-1})$

"Morning run by the ocean, amazing!"

"Team meeting went great, big plans!"

"Fortnite royale! That's sick!"

"Just did my first handstand pushup! 😄"

"So Proud, Launched our new product."

"Sprints on the sand = leg day!"

"Just unlocked a legendary skin! 😎"

"Brainstorming ideas for the next project."

"Can't seem to beat that high score."

"Sunset yoga at the beach, relaxing!"

$$Pr(w_i | w_{1:i-1})$$

"Morning run by the ocean, amazing!"

"Team meeting went great, big plans!"

"Fortnite royale! That's sick!"

"Just did my first handstand pushup! 😄"

"So Proud, Launched our new product."

"Sprints on the sand = leg day!"

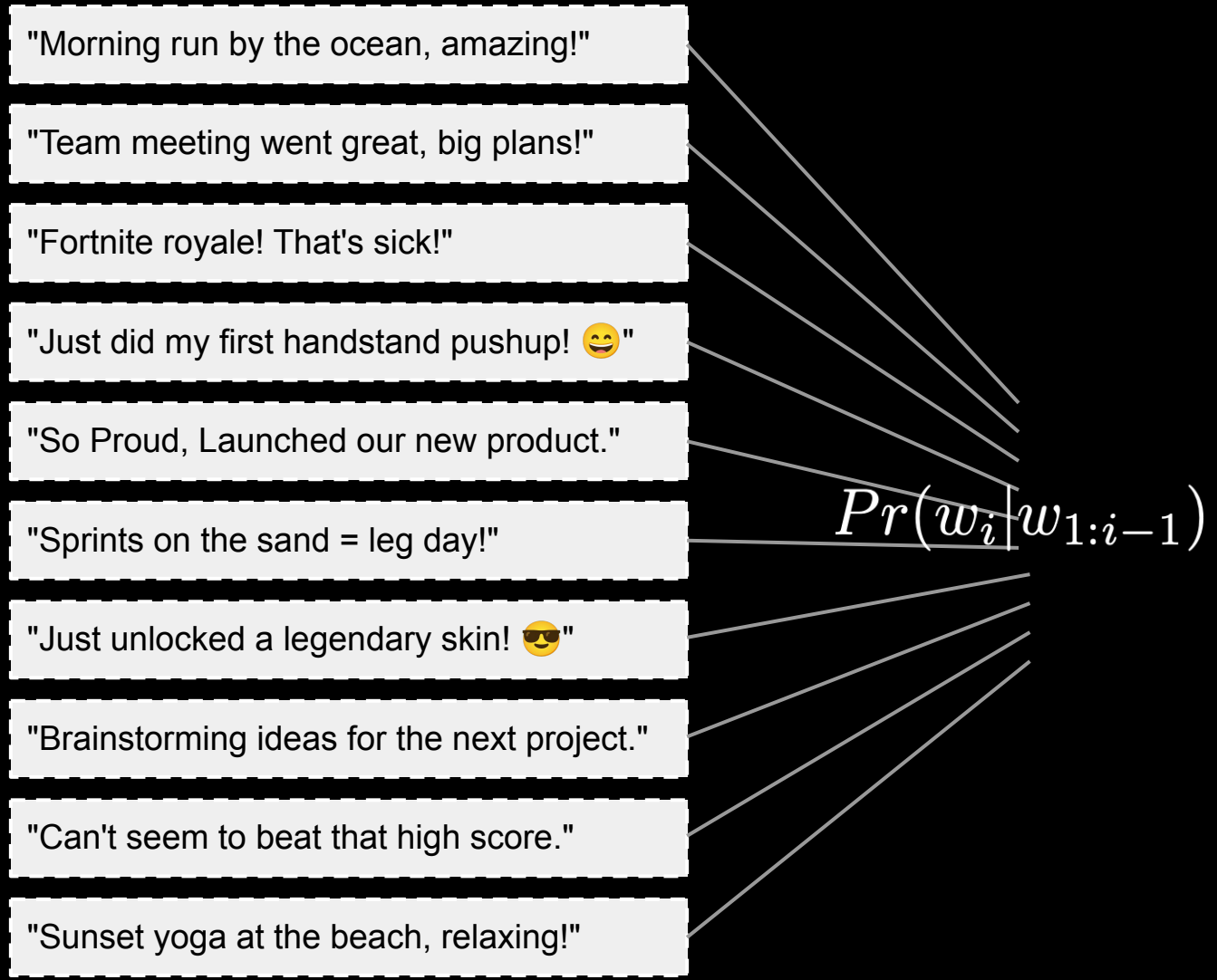
"Just unlocked a legendary skin! 😎"

"Brainstorming ideas for the next project."

"Can't seem to beat that high score."

"Sunset yoga at the beach, relaxing!"

$$Pr(w_i|w_{1:i-1})$$


"Morning run by the ocean, amazing!"
"Team meeting went great, big plans!"
"Fortnite royale! That's sick!"
"Just did my first handstand pushup! 😄"
"So Proud, Launched our new product."
"Sprints on the sand = leg day!"
"Just unlocked a legendary skin! 😎"
"Brainstorming ideas for the next project."
"Can't seem to beat that high score."
"Sunset yoga at the beach, relaxing!"
$Pr(w_i|w_{1:i-1})$

"Morning run by the ocean, amazing!"

"Team meeting went great, big plans!"

"Fortnite royale! That's sick!"

"Just did my first handstand pushup! 😄"

"So Proud, Launched our new product."

"Sprints on the sand = leg day!"

"Just unlocked a legendary skin! 😎"

"Brainstorming ideas for the next project."

"Can't seem to beat that high score."

"Sunset yoga at the beach, relaxing!"

$$Pr(w_i|w_{1:i-1})$$

"Morning run by the ocean, amazing!"

"Just did my first handstand pushup! 😄"

"Sprints on the sand = leg day!"

"Sunset yoga at the beach, relaxing!"

"Team meeting went great, big plans!"

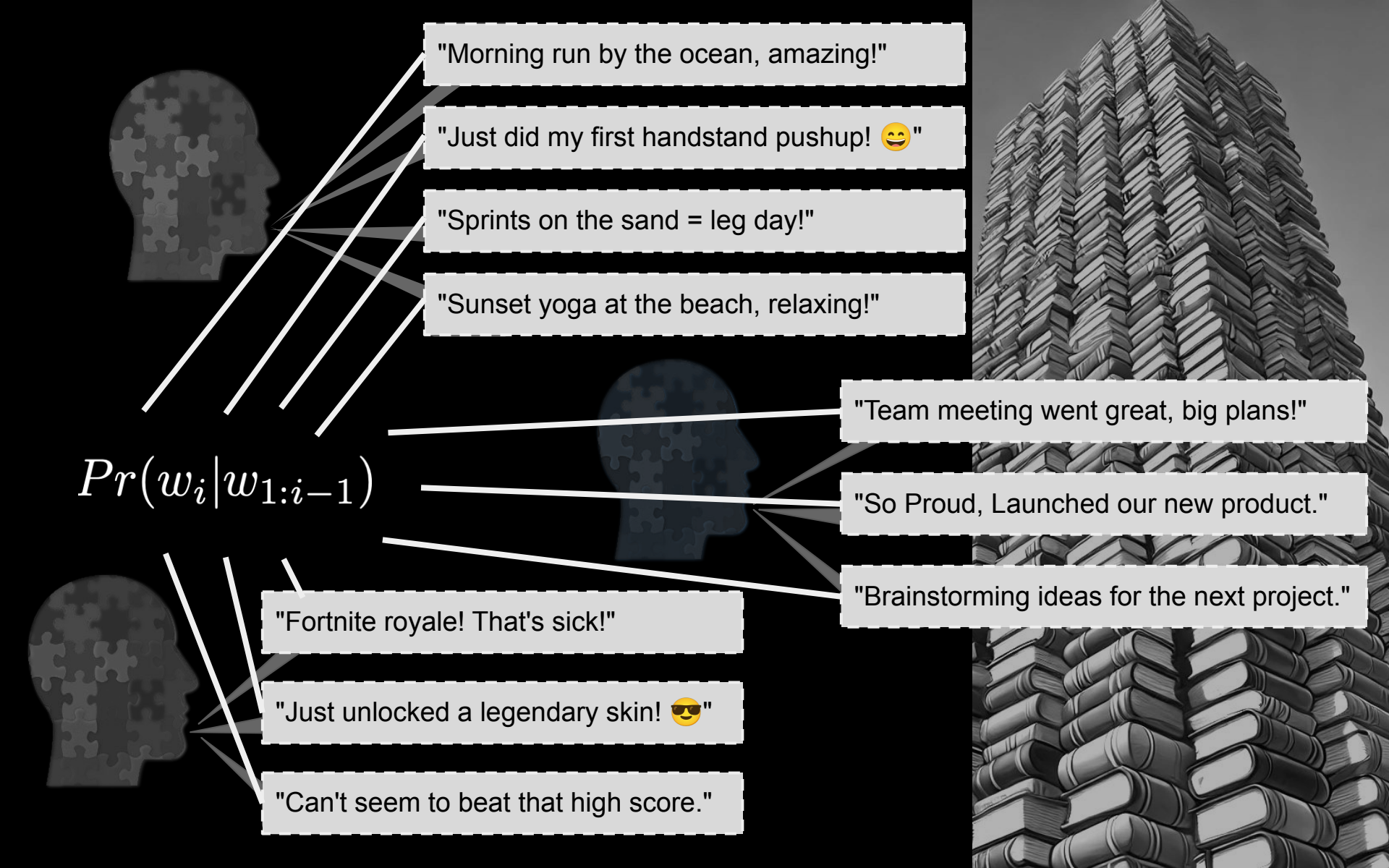
"So Proud, Launched our new product."

"Brainstorming ideas for the next project."

"Fortnite royale! That's sick!"

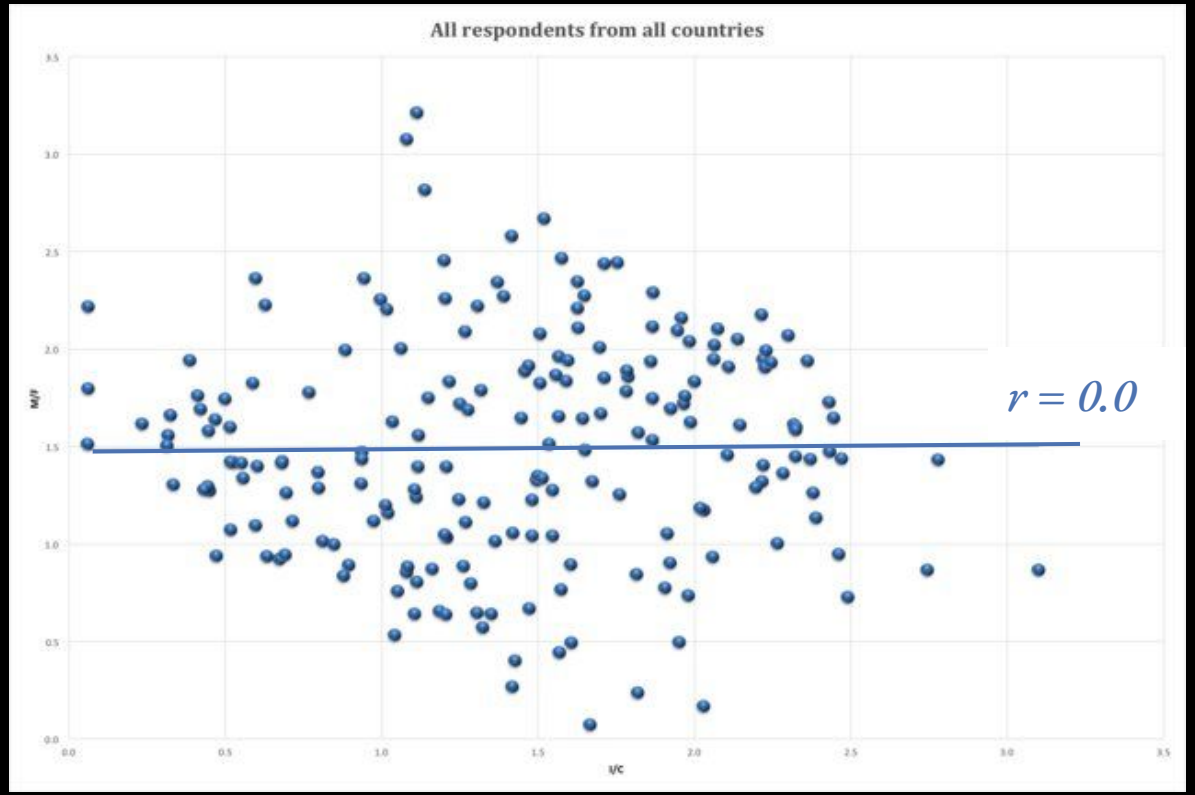
"Just unlocked a legendary skin! 😎"

"Can't seem to beat that high score."


"Morning run by the ocean, amazing!"
"Just did my first handstand pushup! 😄"
"Sprints on the sand = leg day!"
"Sunset yoga at the beach, relaxing!"
$Pr(w_i|w_{1:i-1})$
"Team meeting went great, big plans!"
"So Proud, Launched our new product."
"Brainstorming ideas for the next project."
"Fortnite royale! That's sick!"
"Just unlocked a legendary skin! 😎"
"Can't seem to beat that high score."

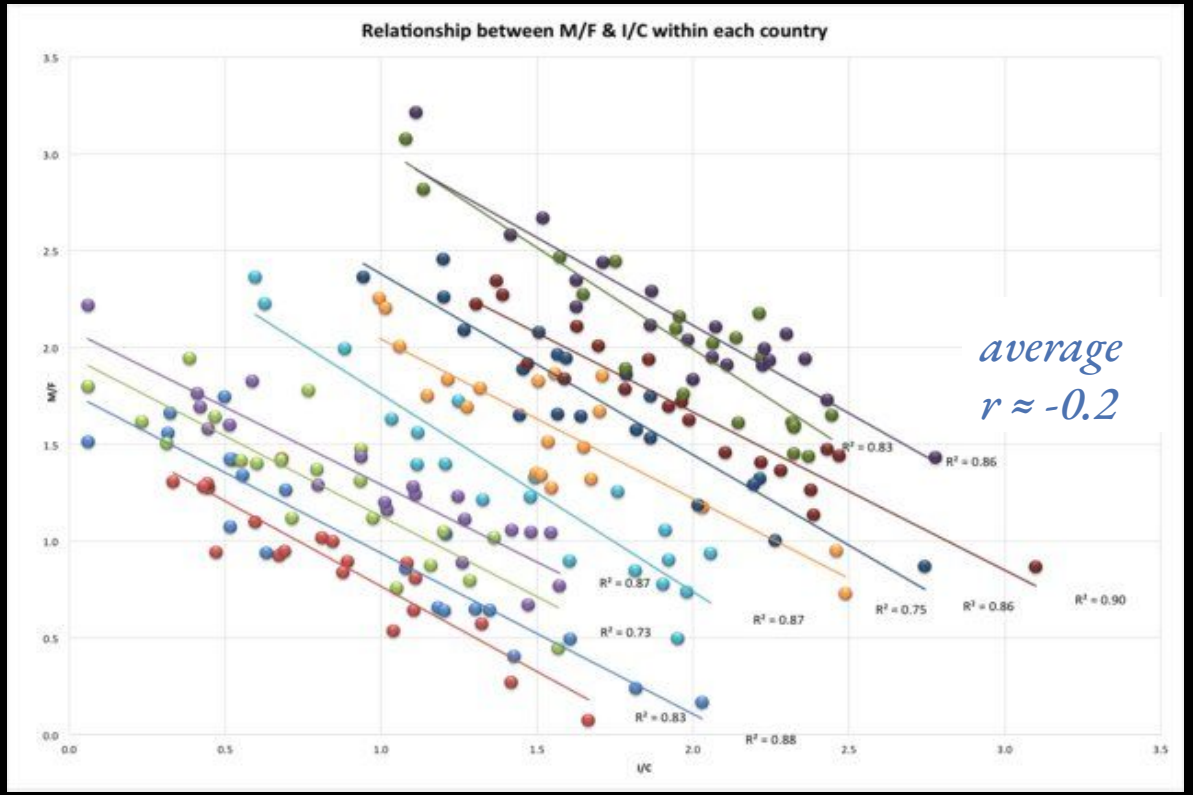

"Morning run by the ocean, amazing!"
"Just did my first handstand pushup! 😄"
"Sprints on the sand = leg day!"
"Sunset yoga at the beach, relaxing!"
ecological fallacy
$Pr(w_i|w_{1:i-1})$
"Team meeting went great, big plans!"
"So Proud, Launched our new product."
"Brainstorming ideas for the next project."
"Fortnite royale! That's sick!"
"Just unlocked a legendary skin! 😎"
"Can't seem to beat that high score."

ecological fallacy



(winzar, 2015)


ecological
fallacy

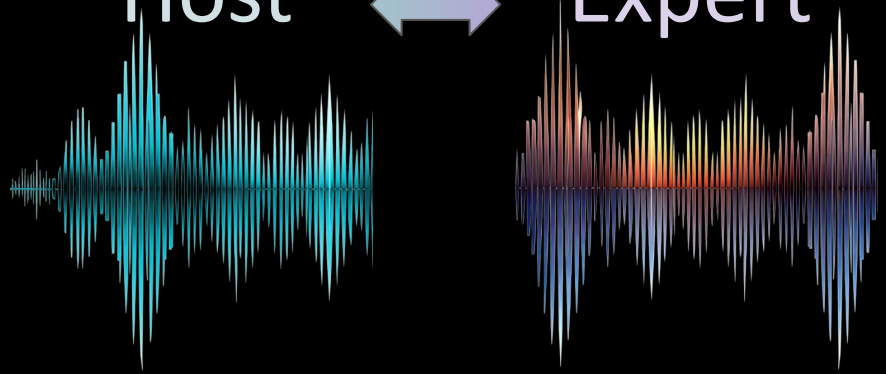

Relationship between M/F & I/C within each country
M/F
I/C
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
R² = 0.83
R² = 0.86
R² = 0.87
R² = 0.73
R² = 0.87
R² = 0.75
R² = 0.86
R² = 0.90
R² = 0.83
R² = 0.88
average
r ≈ -0.2

(winzar, 2015)

Google

NotebookLM

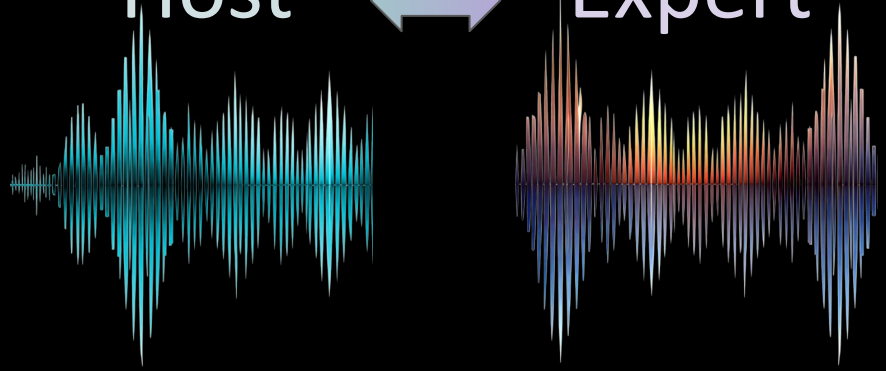
Host ⟷ Expert

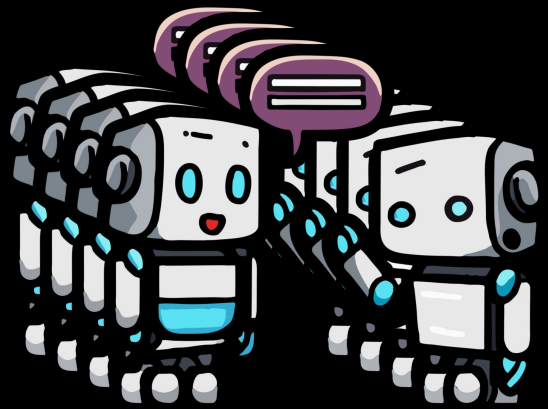


Google

NotebookLM

Host ⟷ Expert

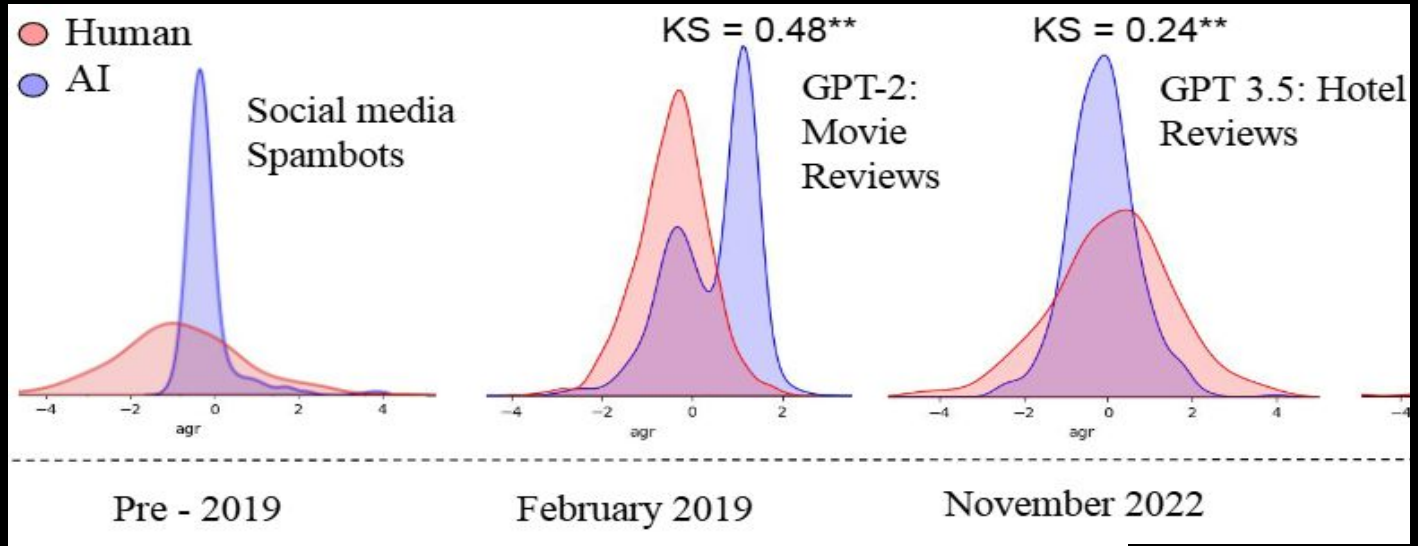


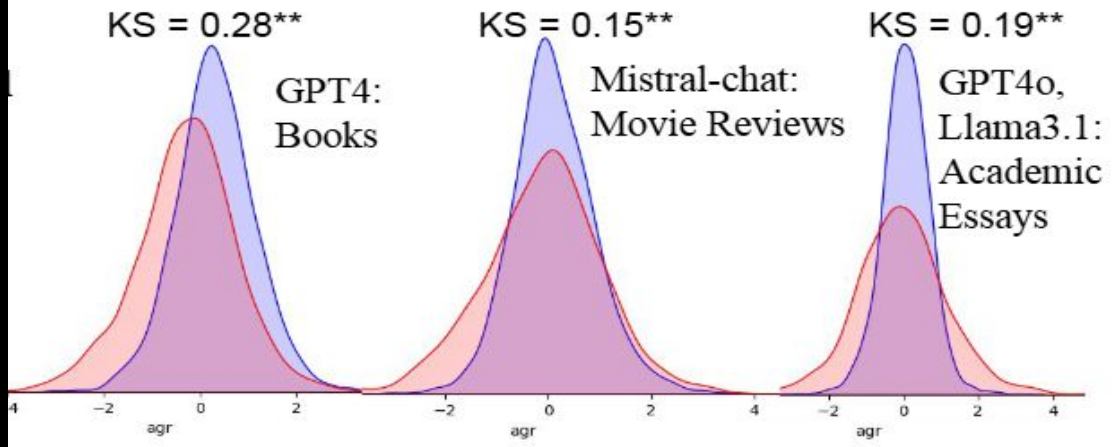


Agent Panels

speaker diarization

Distribution of human linguistic trait - agreeableness over datasets



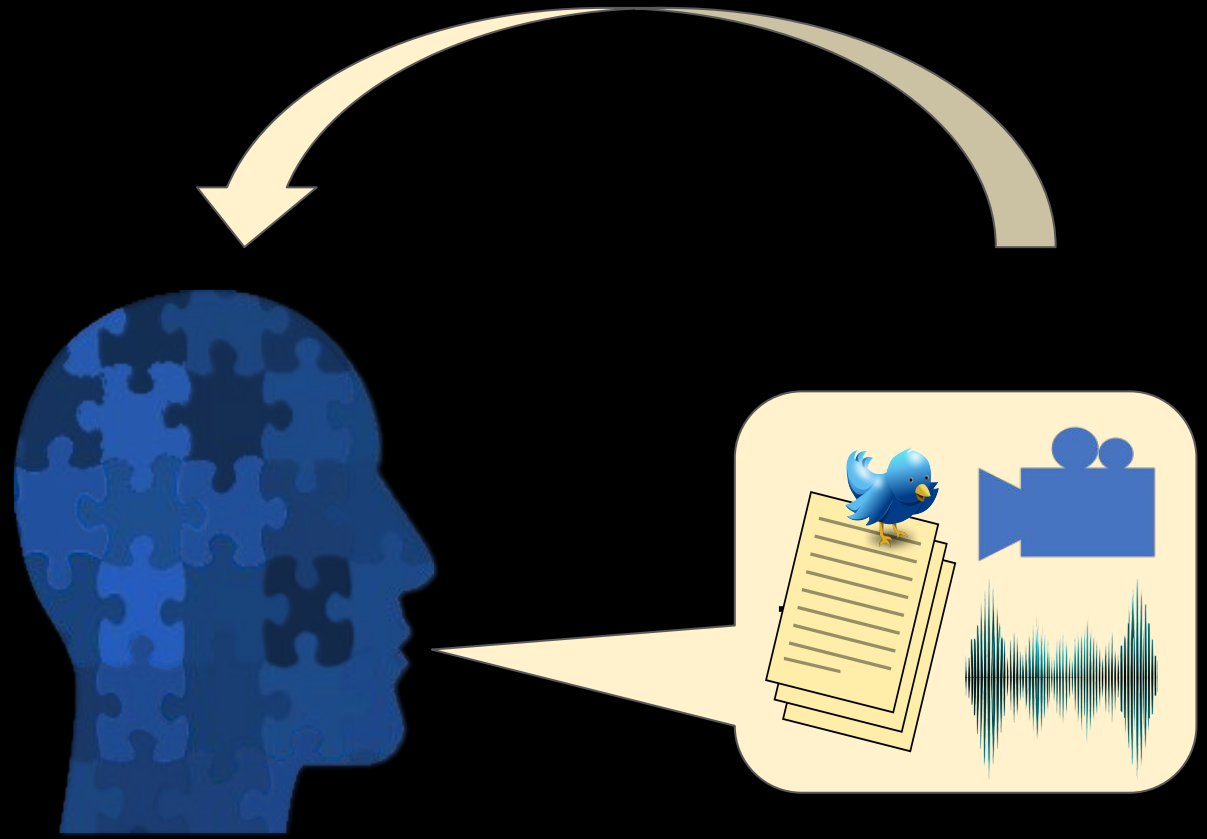


Varadarajan, V., Giorgi, S., Mangalik, S., Soni, N., Markowitz, D. M., & Schwartz, H. A. 2025. **The Consistent Lack of Variance of Psychological Factors Expressed by LLMs and Spambots.** In *Proceedings of the Workshop on GenAI Content Detection (GenAIDetect)* (pp. 111-119).

Best Paper

Vasudha Varadarajan

Salvatore Giorgi



People have beliefs, backgrounds, styles, vocabularies, knowledge, and personalities. **Our data reflect and are influenced by these differences.**

Language Modeling

probability of a token sequence

$$Pr(\mathbf{W}) = \prod_{i=1}^{n} Pr(w_i | w_{1:i-1})$$

Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

Human Language Modeling (HuLM)

LM - probability of a token sequence

$$Pr(\mathbf{W}) = \prod_{i=1}^{n} Pr(w_i | w_{1:i-1})$$

HuLM - probability of a token sequence,
in the context of the human that generated it.

Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

Human Language Modeling (HuLM)

LM

- probability of a token sequence

$$Pr(\mathbf{W}) = \prod_{i=1}^{n} Pr(w_i | w_{1:i-1})$$

$$Pr(\mathbf{W} | \mathbf{U}_{static}) = \prod_{i=1}^{n} Pr(w_i | w_{1:i-1}, \mathbf{U}_{static})$$

HuLM

- probability of a token sequence, in the context of the human that generated it.

Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

Human Language Modeling (HuLM)

LM

- probability of a token sequence

$$Pr(\mathbf{W}) = \prod_{i=1}^{n} Pr(w_i | w_{1:i-1})$$

$$Pr(\mathbf{W} | \mathbf{U}_{static}) = \prod_{i=1}^{n} Pr(w_i | w_{1:i-1}, \mathbf{U}_{static})$$

static user representation

HuLM

- probability of a token sequence, in the context of the human that generated it.

Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

Human Language Modeling (HuLM)

LM

- probability of a token sequence

$$Pr(\mathbf{W}) = \prod_{i=1}^{n} Pr(w_i | w_{1:i-1})$$

$$Pr(\mathbf{W} | \mathbf{U}_{static}) = \prod_{i=1}^{n} Pr(w_i | w_{1:i-1}, \mathbf{U}_{static})$$

static user representation

HuLM

$$Pr(\mathbf{W}_t | \mathbf{U}_{t-1}) = \prod_{i=1}^{n} Pr(w_{t,i} | w_{t,1:i-1}, \mathbf{U}_{1:t-1})$$

- probability of a token sequence, in the context of the human that generated it.

Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

Human Language Modeling (HuLM)

LM

- probability of a token sequence

$$Pr(\mathbf{W}) = \prod_{i=1}^{n} Pr(w_i | w_{1:i-1})$$

$$Pr(\mathbf{W} | \mathbf{U}_{static}) = \prod_{i=1}^{n} Pr(w_i | w_{1:i-1}, \mathbf{U}_{static})$$

static user representation

HuLM

$$Pr(\mathbf{W}_t | \mathbf{U}_{t-1}) = \prod_{i=1}^{n} Pr(w_{t,i} | w_{t,1:i-1}, \mathbf{U}_{1:t-1})$$

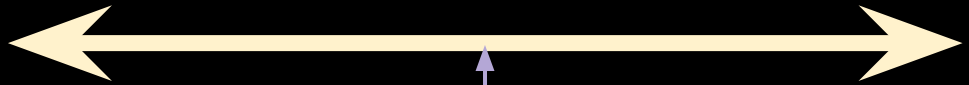
"user state" representation

- probability of a token sequence, in the context of the human that generated it.

Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

User State Representation, U

$$Pr(\mathbf{W}_t|\mathbf{U}_{t-1}) = \prod_{i=1}^{n} Pr(w_{t,i}|w_{t,1:i-1}, \mathbf{U}_{1:t-1})$$



no history — all data

$U_{1:t-1} = \varnothing$

(reduces to a standard LM: $Pr(w_i|w_{1:i-1})$)

- *doesn't capture the person*

history of user states

$U_{1:t-1} = w_{1,1:n_1}, w_{2,1:n_2}, \ldots, w_{t-1,1:n_{t-1}}$

(all previous docs and tokens by the person)

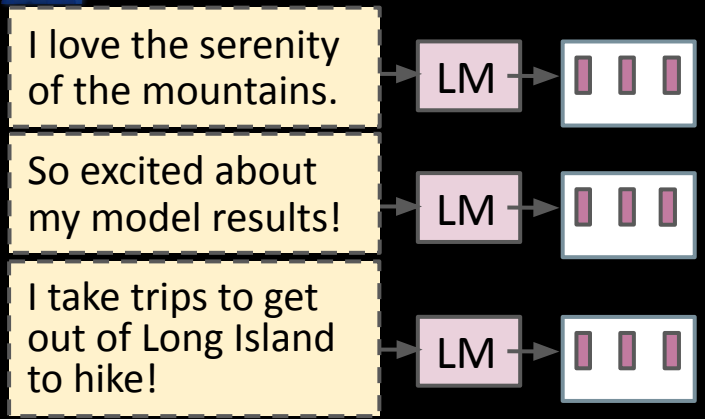
- *huge*
- *no generalizations*

Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). Human Language Modeling. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

User State Representation, U

$$Pr(\mathbf{W}_t|\mathbf{U}_{t-1}) = \prod_{i=1}^{n} Pr(w_{t,i}|w_{t,1:i-1}, \mathbf{U}_{1:t-1})$$

State and Trait Theory from Psychology: ***Traits*** – the stable characteristics of "who someone is" – define a distribution of potential ***states*** of being that moderate human behavior (i.e. language).



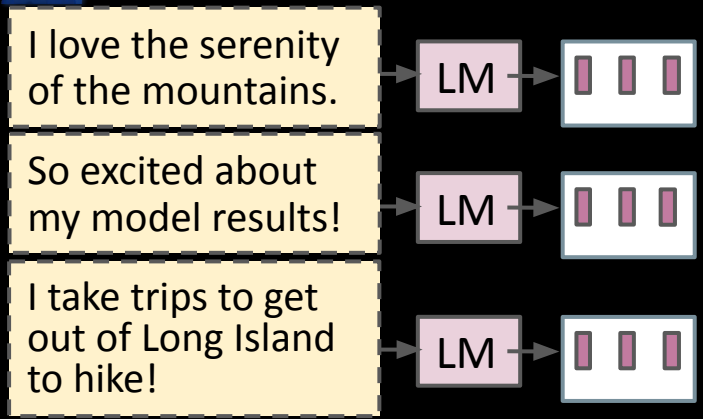
history of user states

Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). Human Language Modeling. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

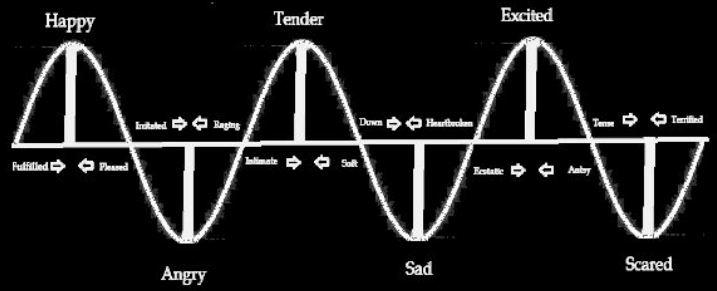
User State Representation, U

$$Pr(\mathbf{W}_t|\mathbf{U}_{t-1}) = \prod_{i=1}^{n} Pr(w_{t,i}|w_{t,1:i-1}, \mathbf{U}_{1:t-1})$$

State and Trait Theory from Psychology: ***Traits*** – the stable characteristics of "who someone is" – define a distribution of potential ***states*** of being that moderate human behavior (i.e. language).



$U_{1:t-1}$ = [a sequence of *states*]

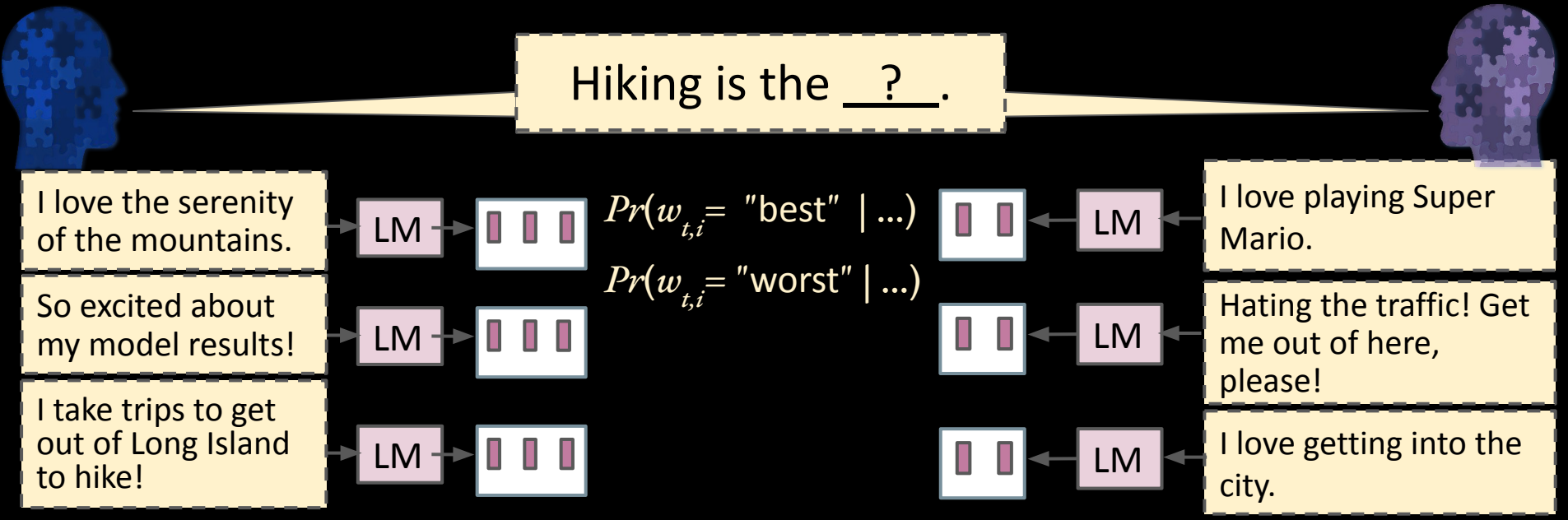


(Washington Outsider, 2014)

Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). Human Language Modeling. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

User State Representation, U

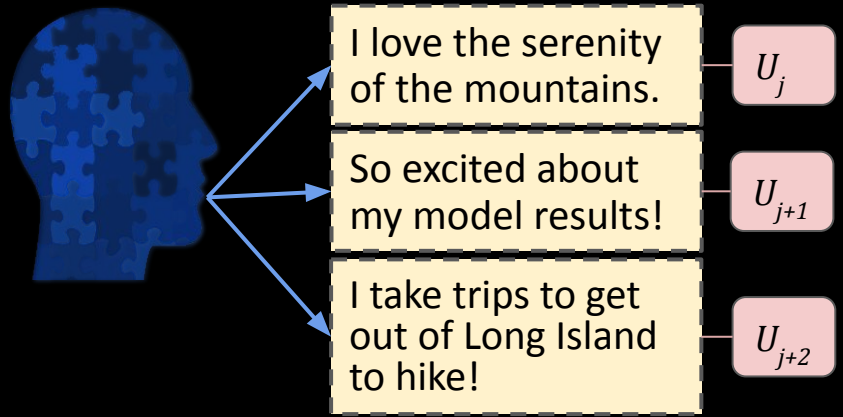
$$Pr(\mathbf{W}_t|\mathbf{U}_{t-1}) = \prod_{i=1}^{n} Pr(w_{t,i}|w_{t,1:i-1}, \mathbf{U}_{1:t-1})$$



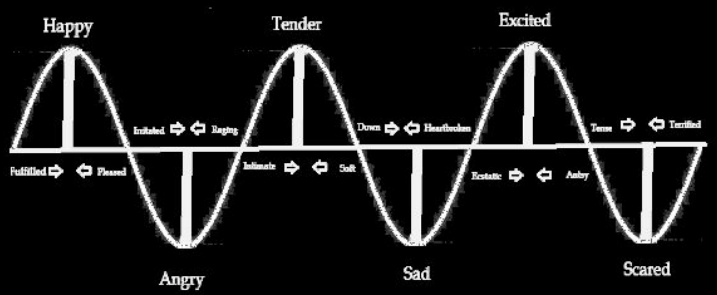
Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). Human Language Modeling. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

User State Representation: Motivation

- **Addressing *Ecological Fallacy:*** Treating dependent phenomena (i.e. sequences from the same person) as if independent. (Piantadosi et al., 1988; Steel and Holt, 1996)
- **Modeling the higher order structure.**
- Building on **ideas from human factor inclusion**/adaptation (Lynn et al., 2017; Huang & Paul, 2019; Hovy & Yang, 2021) and **personalized modeling.** (King & Cook, 2020; Jaech & Ostendorf, 2018)



$U_{1:t-1}$ = [a sequence of *states*]

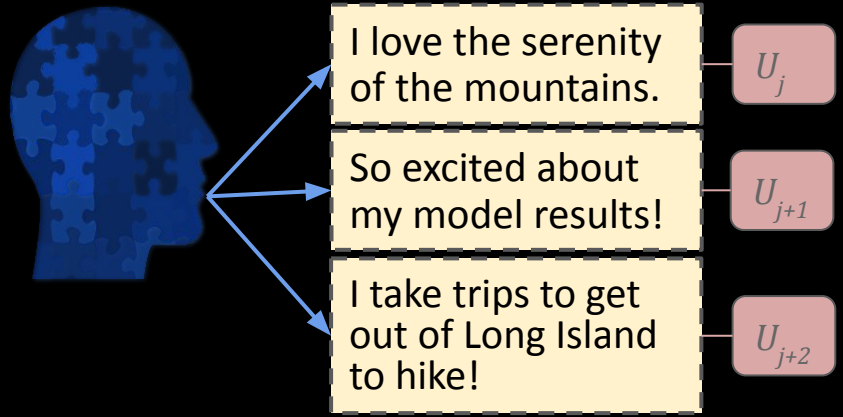


(Washington Outsider, 2014)

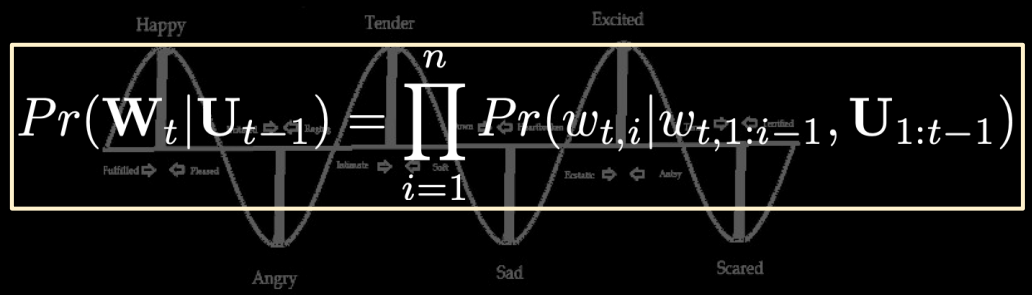
Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). Human Language Modeling. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

Human Language Modeling (HuLM)

- **Addressing *Ecological Fallacy:*** Treating dependent phenomena (i.e. sequences from the same person) as if independent. (Piantadosi et al., 1988; Steel and Holt, 1996)
- **Modeling the higher order structure.**
- Building on **ideas from human factor inclusion**/adaptation (Lynn et al., 2017; Huang & Paul, 2019; Hovy & Yang, 2021) and **personalized modeling.** (King & Cook, 2020; Jaech & Ostendorf, 2018)



$U_{1:t-1}$ = [a sequence of *states*]

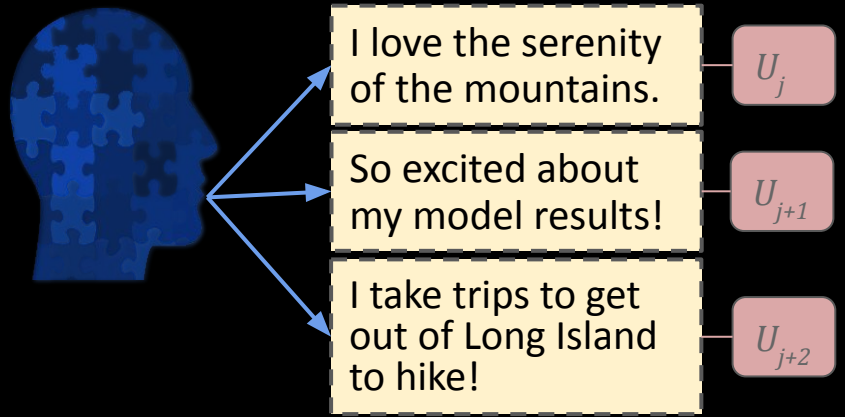


Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). Human Language Modeling. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

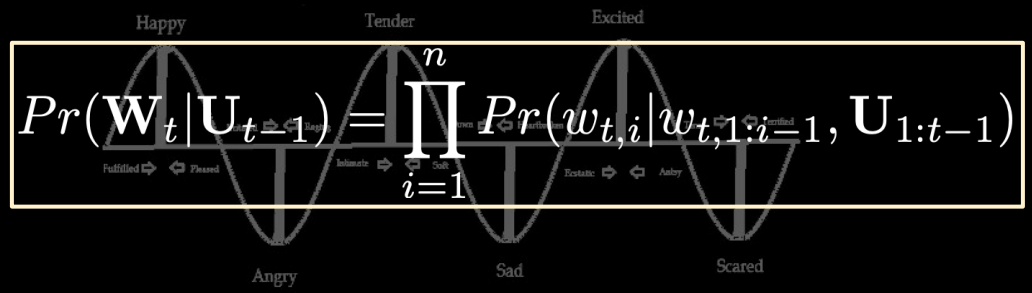
Human Language Modeling (HuLM)

Goal: Language modeling as a task grounded in the "natural" generators of language, people.

The *HuLM* task definition: Estimate the probability of a sequence of tokens, $w_{t,1:i}$, conditioned on a higher-order representation, U_t, constituting the human state of being just before the sequence generation.



$U_{1:t-1}$ = [a sequence of *states*]



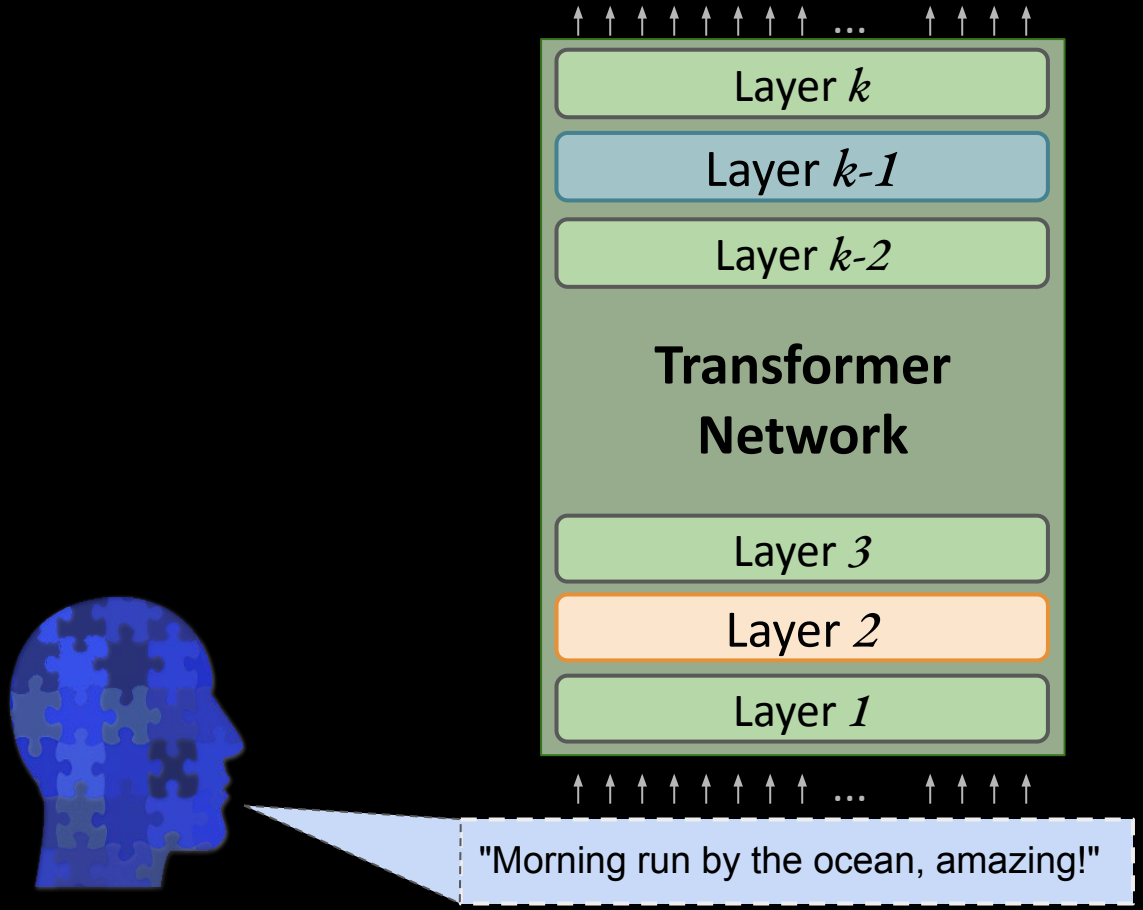
Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). Human Language Modeling. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

The HaRT Language Model

Nikita Soni

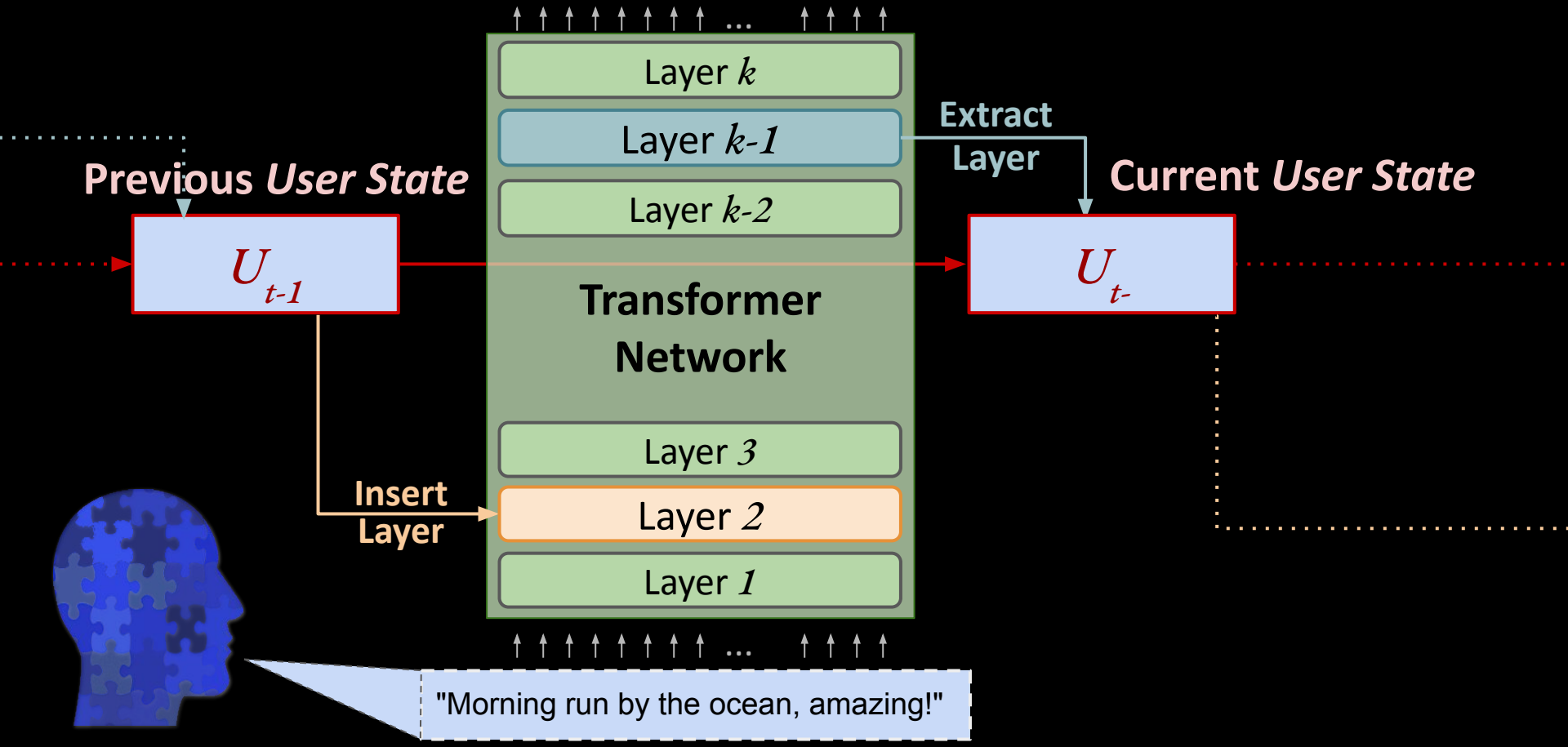
Matthew Matero

Niranjan Balasubramanian



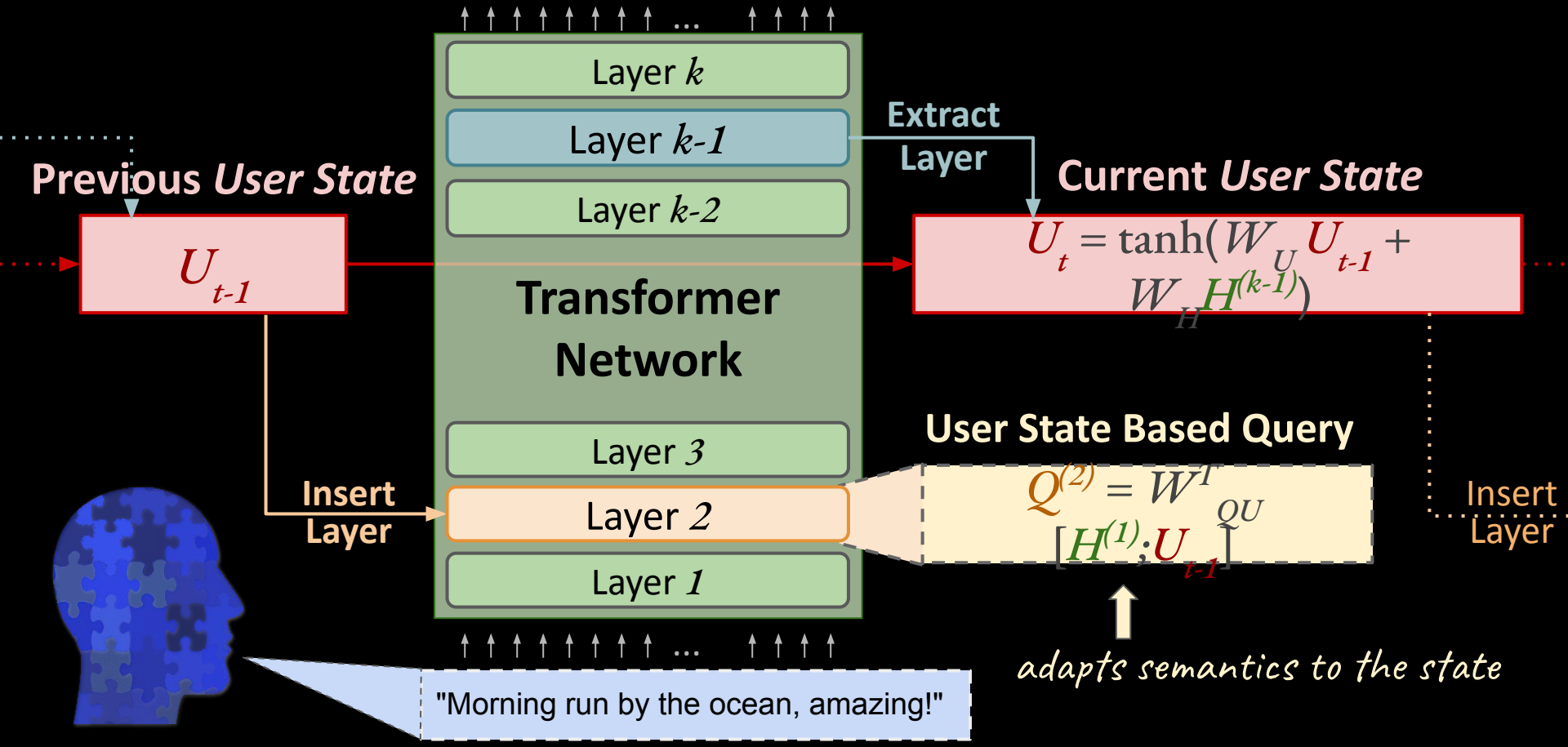
Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

Human-aware Recurrent Transformer (HaRT)



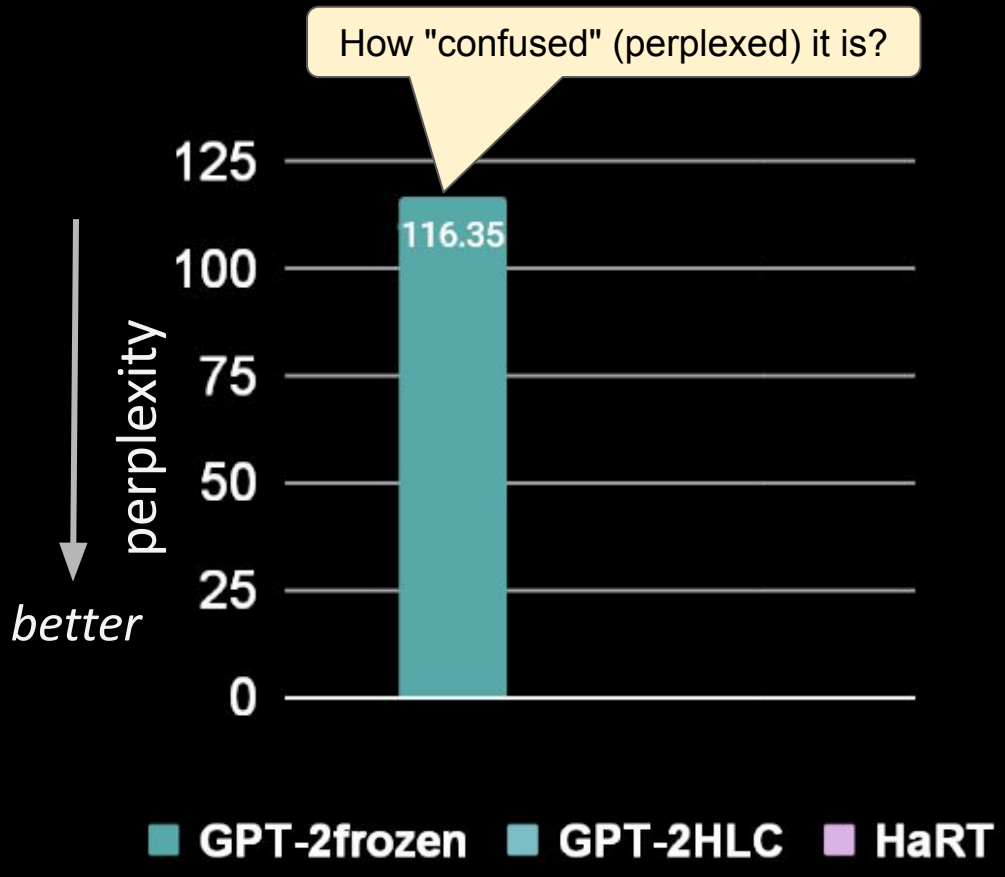
Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

Human-aware Recurrent Transformer (HaRT)



Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

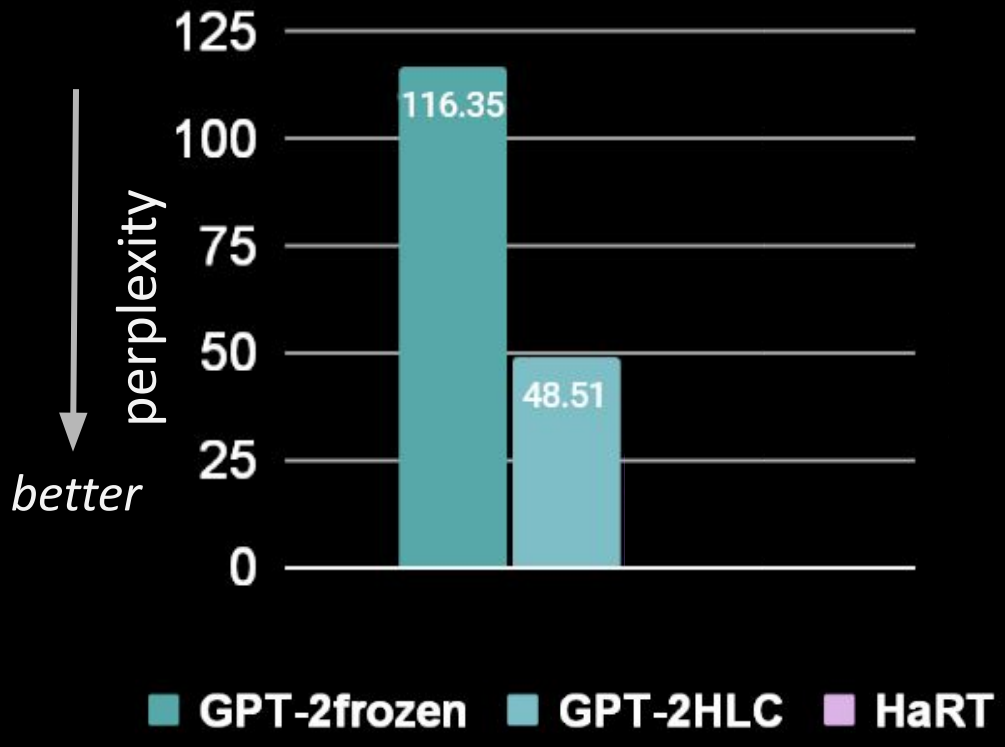
The HaRT Language Model



Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

Stony Brook University
Human Language Analysis Beings

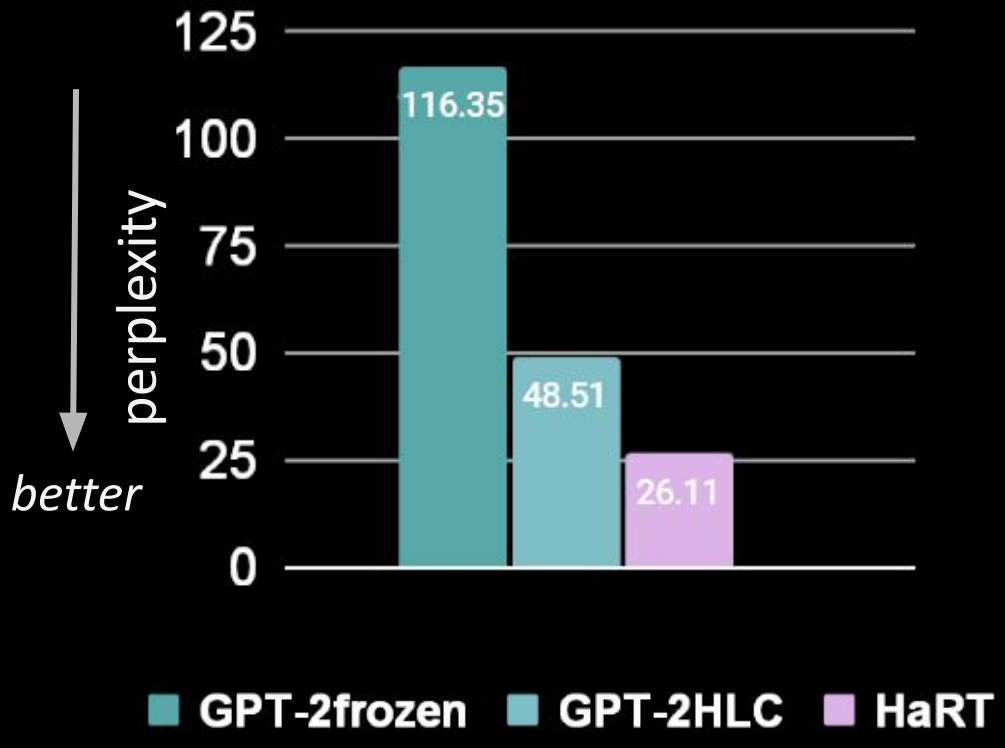
The HaRT Language Model



Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

Stony Brook University
Human Language Analysis Beings

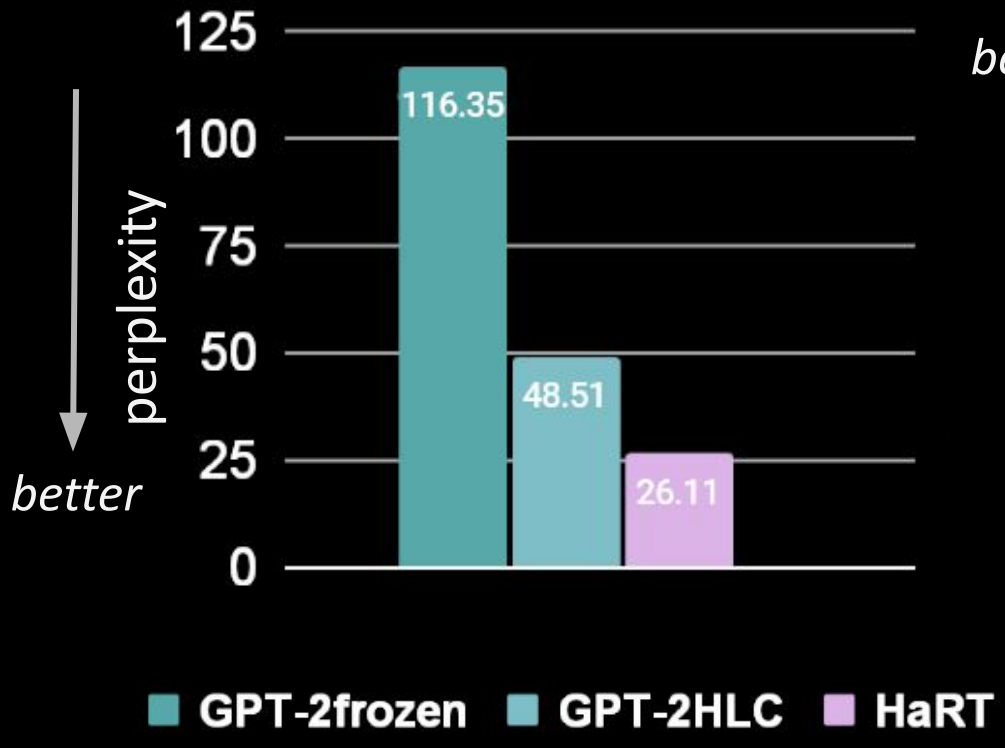
The HaRT Language Model

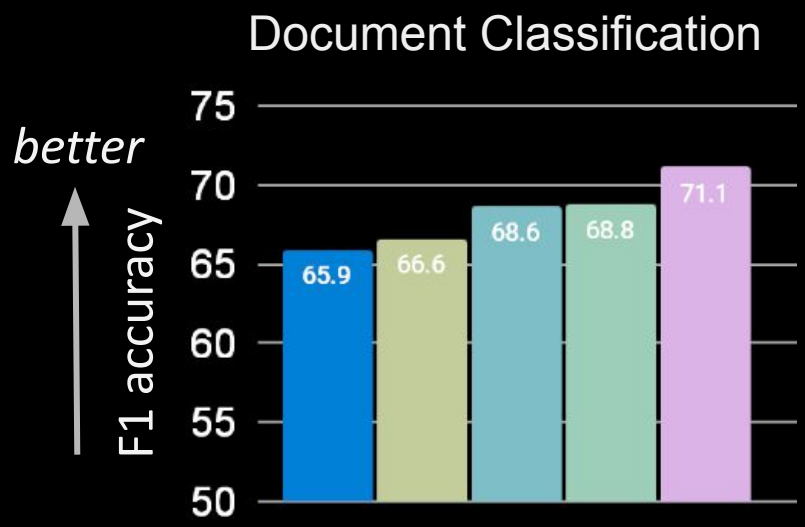


Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

Stony Brook University
Human Language Analysis Beings

The HaRT Language Model

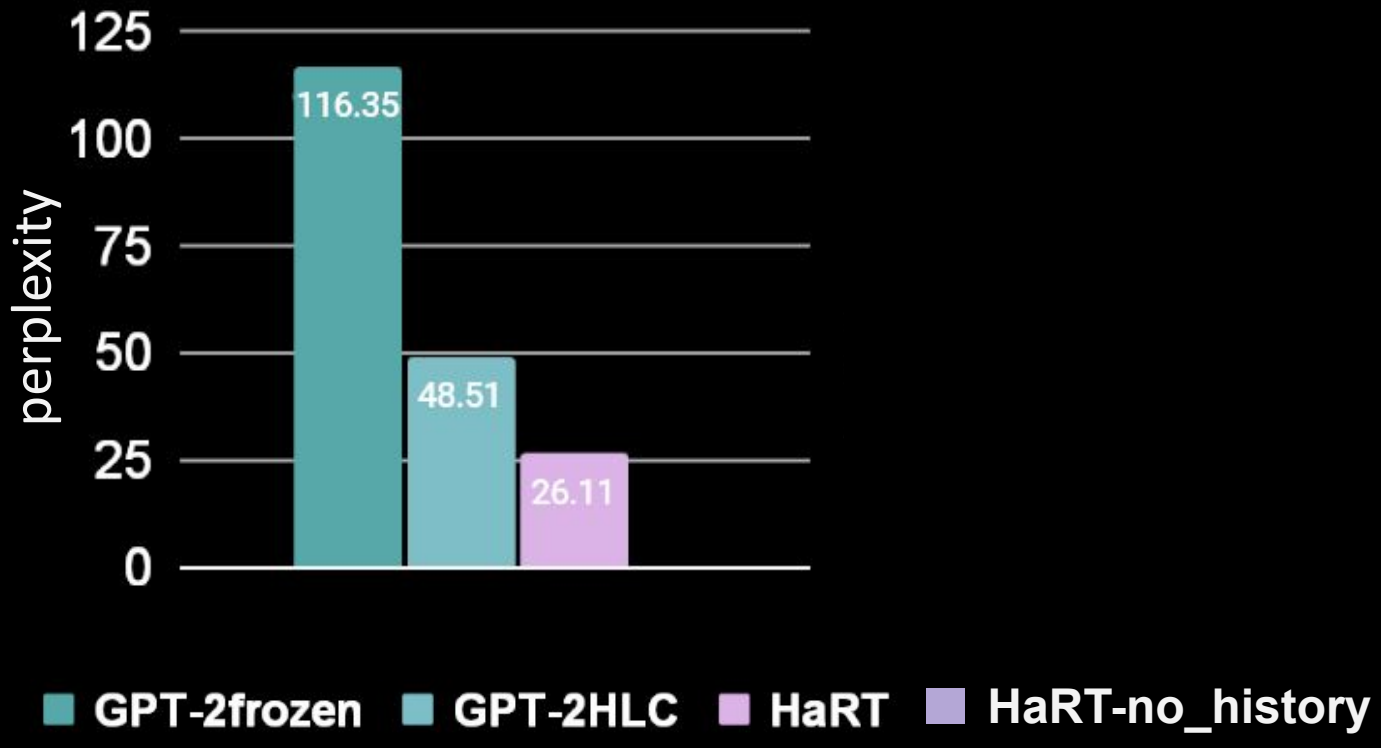




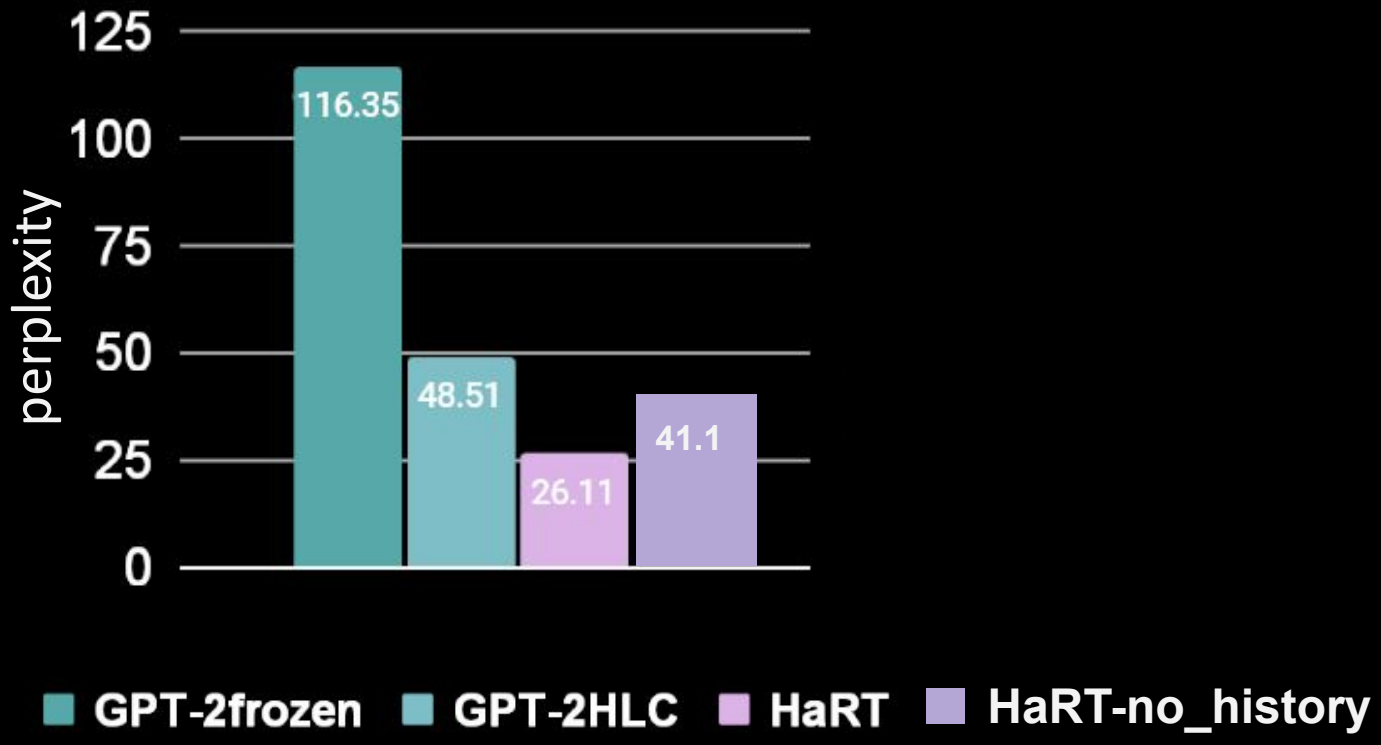
Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). **Human Language Modeling.** In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

Stony Brook University
Human Language Analysis Beings

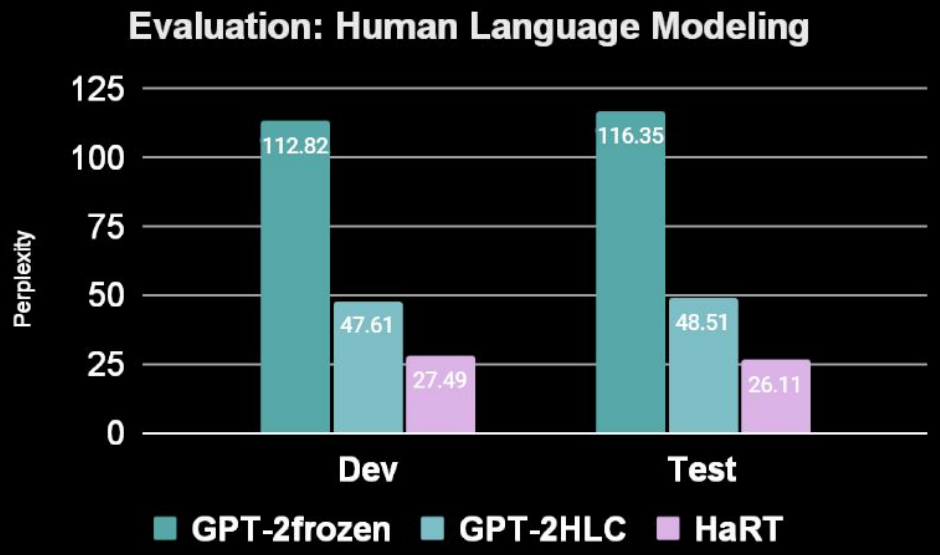
No history?

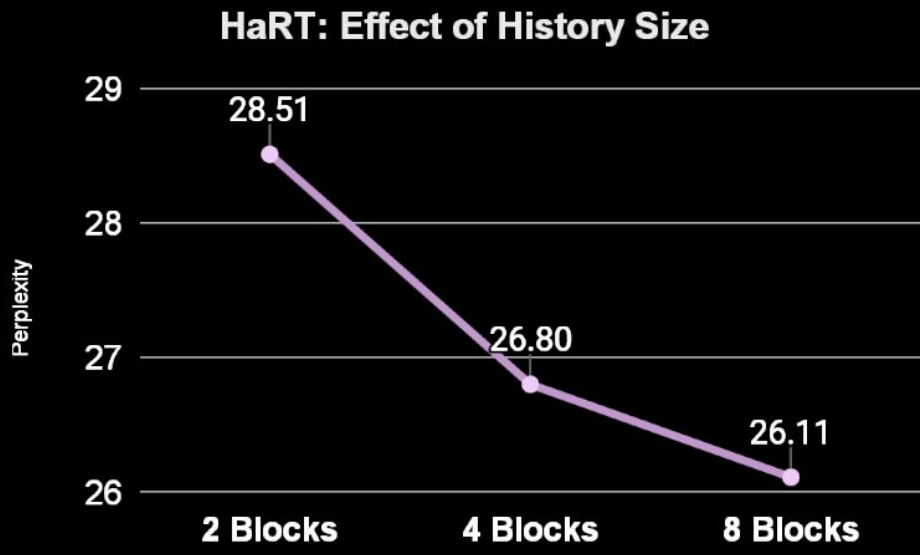

125
100
75
50
25
0
perplexity
116.35
48.51
26.11
GPT-2frozen
GPT-2HLC
HaRT
HaRT-no_history

No history?


125
100
75
50
25
0
perplexity
116.35
48.51
26.11
41.1
GPT-2frozen
GPT-2HLC
HaRT
HaRT-no_history

Human Language Modeling





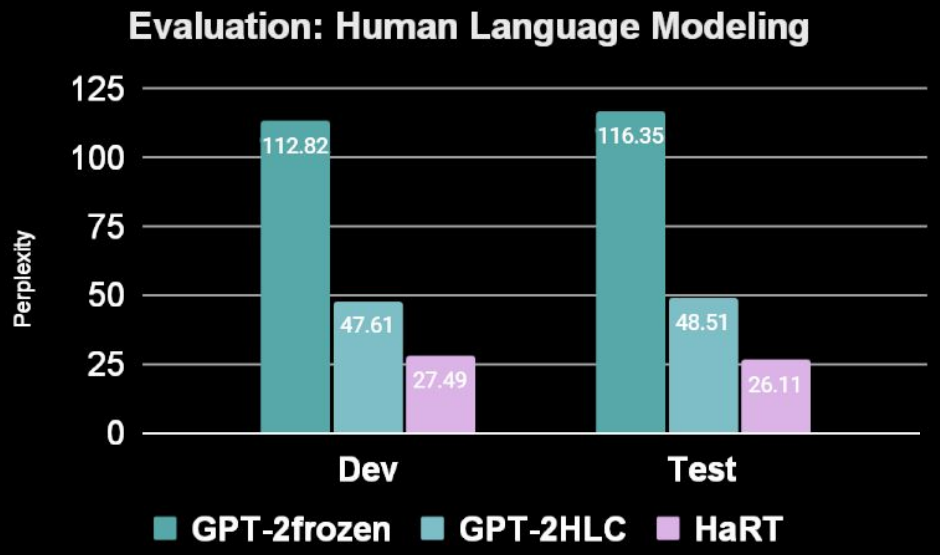
Dataset: Human Language Corpus (HLC)

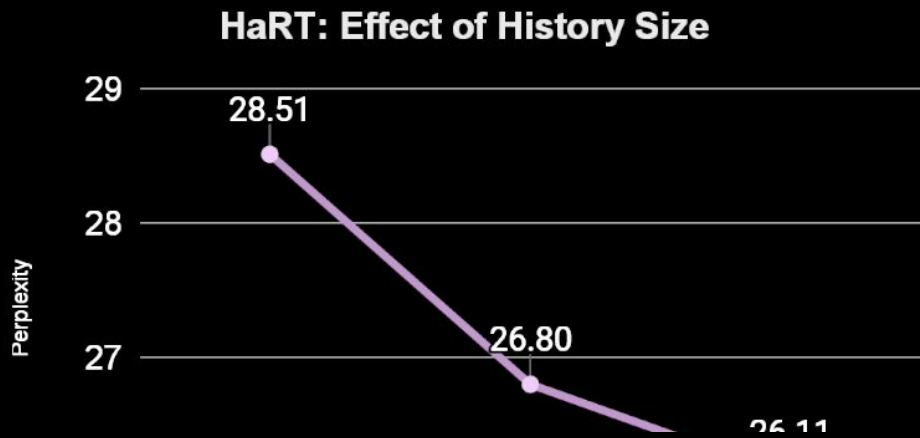
Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). Human Language Modeling. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).

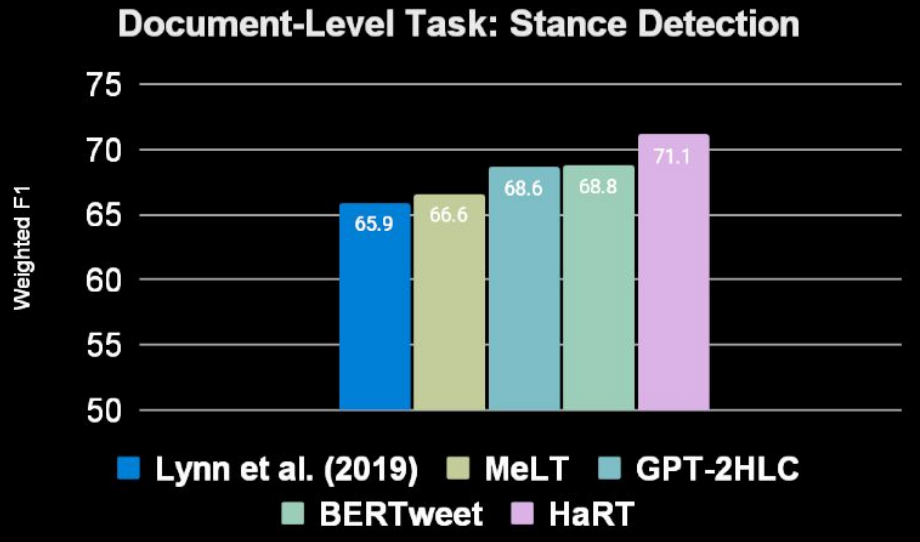
+

Train	Dev	Test
users = 96k msgs = 36m (8 blocks= ~17m)	users = 2k msgs = 830k + **seen** users: 2.5k msgs: 230k	users = 2k msgs = 690k + **seen** users: 2.5K msgs: 240k

Human Language Modeling



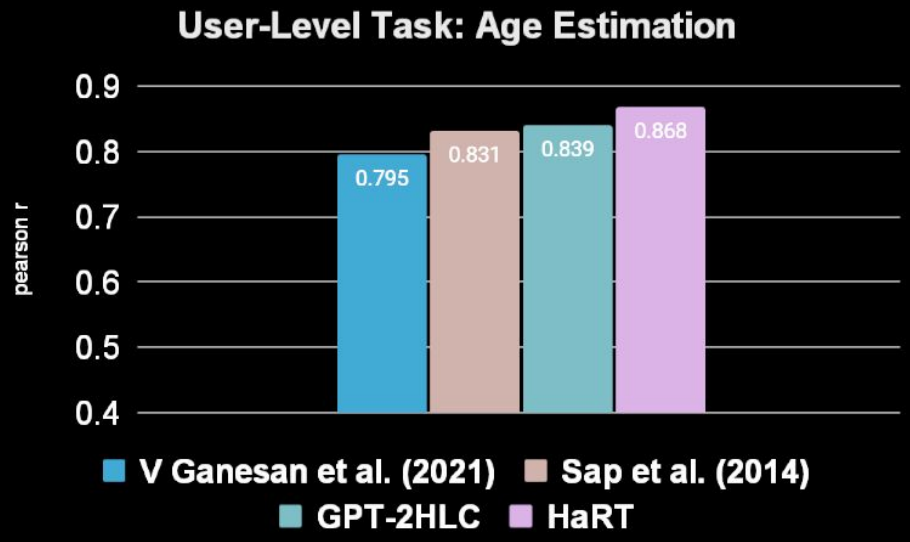


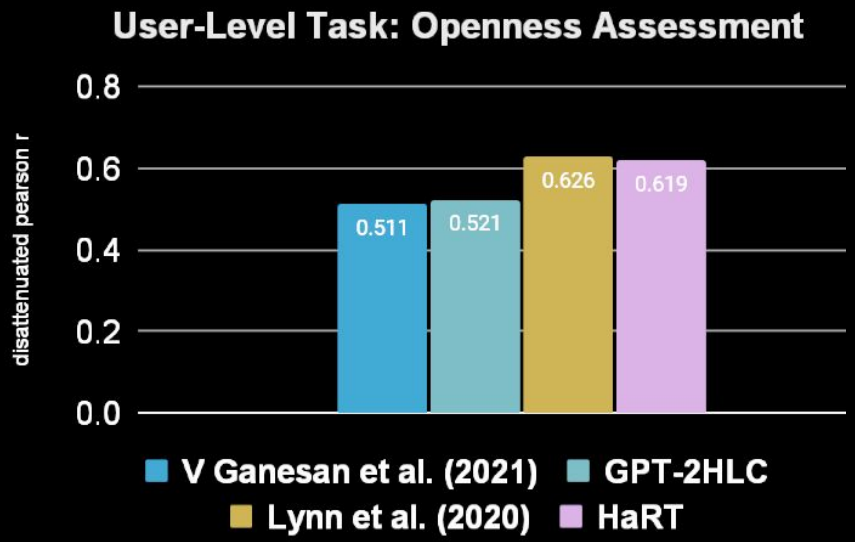


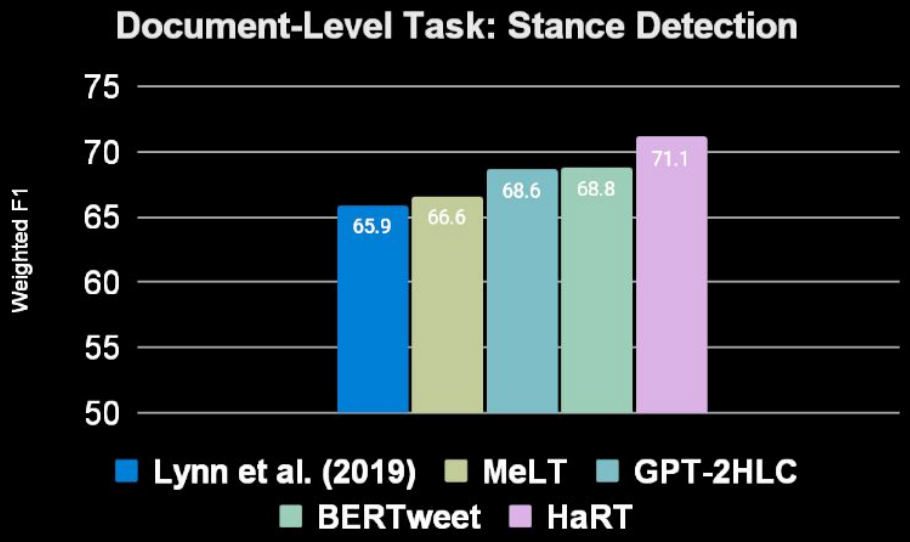
Dataset: Human Language Corpus (

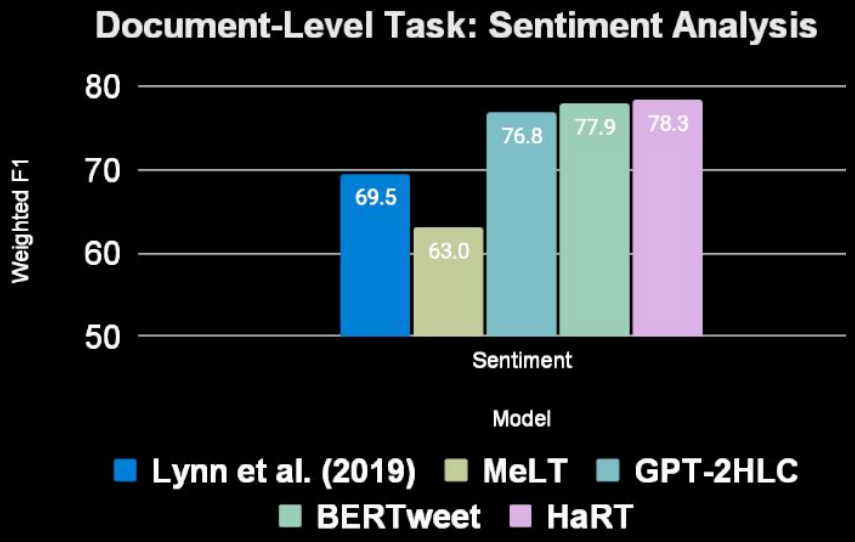
us
ms
(8 blo

Soni, N., Matero, M., Balasubramanian, N., & Schwartz, H. (2022, May). Human Language Modeling. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 622-636).


User-Level Task: Age Estimation
pearson r
0.9
0.8
0.7
0.6
0.5
0.4
0.795
0.831
0.839
0.868
V Ganesan et al. (2021)
Sap et al. (2014)
GPT-2HLC
HaRT

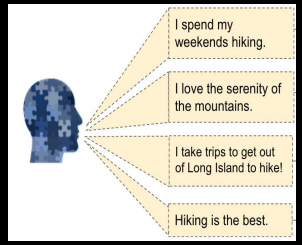

User-Level Task: Openness Assessment
disattenuated pearson r
0.8
0.6
0.4
0.2
0.0
0.511
0.521
0.626
0.619
V Ganesan et al. (2021)
GPT-2HLC
Lynn et al. (2020)
HaRT

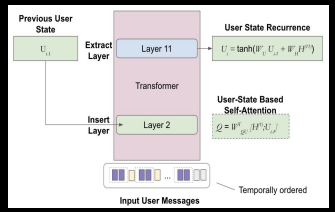

Document-Level Task: Stance Detection
Weighted F1
75
70
65
60
55
50
65.9
66.6
68.6
68.8
71.1
Lynn et al. (2019)
MeLT
GPT-2HLC
BERTweet
HaRT

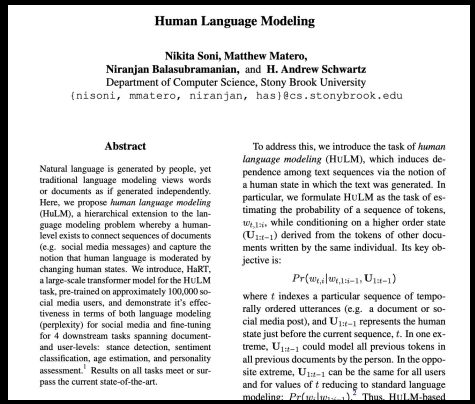

Document-Level Task: Sentiment Analysis
Weighted F1
80
70
60
50
69.5
63.0
76.8
77.9
78.3
Sentiment
Model
Lynn et al. (2019)
MeLT
GPT-2HLC
BERTweet
HaRT

HuLM/HaRT Takeaways

- **HuLM:** Extension of language modeling with notion of user.
- **HaRT:** A step toward large *human* language models.
- Progress for large LMs grounded in language's "natural" generators, people.
- GitHub Repository







Human Language Modeling

Nikita Soni, Matthew Matero,
Niranjan Balasubramanian, and **H. Andrew Schwartz**
Department of Computer Science, Stony Brook University
{nisoni, mmatero, niranjan, has}@cs.stonybrook.edu

Abstract

Natural language is generated by people, yet traditional language modeling views words or documents as if generated independently. Here, we propose *human language modeling* (HuLM), a hierarchical extension to the language modeling problem whereby a human-level exists to connect sequences of documents (e.g. social media messages) and capture the notion that human language is moderated by changing human states. We introduce, HaRT, a large-scale transformer model for the HuLM task, pre-trained on approximately 100,000 social media users, and demonstrate it's effectiveness in terms of both language modeling (perplexity) for social media and fine-tuning for 4 downstream tasks spanning document- and user-levels: stance detection, sentiment classification, age estimation, and personality assessment.[1] Results on all tasks meet or surpass the current state-of-the-art.

To address this, we introduce the task of *human language modeling* (HuLM), which induces dependence among text sequences via the notion of a human state in which the text was generated. In particular, we formulate HuLM as the task of estimating the probability of a sequence of tokens, $w_{t,1:i}$, while conditioning on a higher order state ($\mathbf{U}_{1:t-1}$) derived from the tokens of other documents written by the same individual. Its key objective is:

$$Pr(w_{t,i}|w_{t,1:i-1}, \mathbf{U}_{1:t-1})$$

where t indexes a particular sequence of temporally ordered utterances (e.g. a document or social media post), and $\mathbf{U}_{1:t-1}$ represents the human state just before the current sequence, t. In one extreme, $\mathbf{U}_{1:t-1}$ could model all previous tokens in all previous documents by the person. In the opposite extreme, $\mathbf{U}_{1:t-1}$ can be the same for all users and for values of t reducing to standard language modeling: $Pr(w_i|w_{i-k:i-1})$.[2] Thus, HuLM-based

Human-Centered NLP – Methods to know

1. Differential Language Analysis
    a. pearson correlation
    b. multivariate linear regression
    c. odd ratio with informative Dirichlet prior
2. Human Factor Adaptation - feature augmentation
3. Human Language Modeling
    a. HuLM task definition
    b. HaRT Architecture