UNIVERSITY of WASHINGTON

Comparing emotion feature extraction approaches for predicting depression and anxiety

Hannah A. Burkhardt[1] , Michael D. Pullmann[1], Thomas D. Hull[2], Patricia A. Areán[1], Trevor Cohen[1]

[1]University of Washington, Seattle
[2]Talkspace, New York, NY

Mental health

- ~20% of U.S. adults experience mental illness
- In counseling or therapy:

2019	9.5%
2020	10.1%

Sentiment vs. Emotions

OMG, yep!!! That is the final answer! Thank you so much!

excitement 🤩
approval 👍
gratitude 🙏

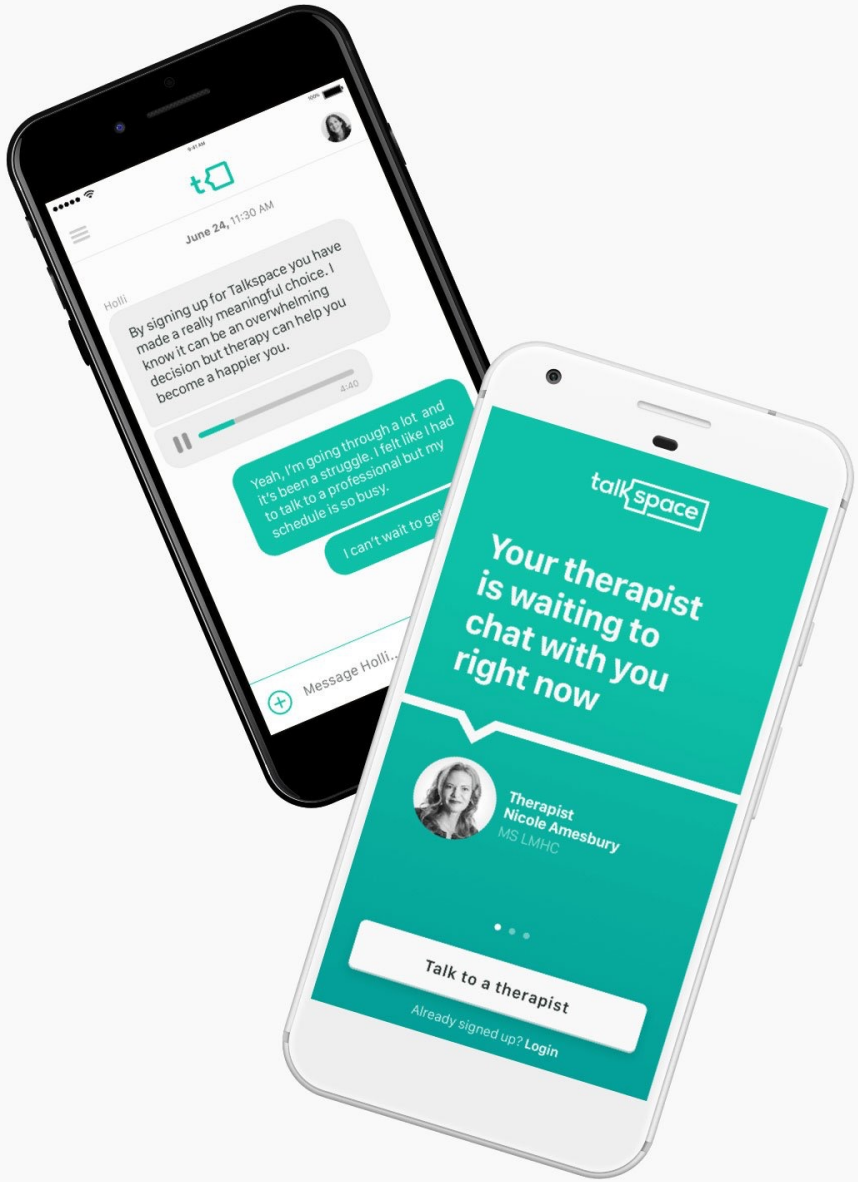
I am so glad this is over.

relief 😌

Sorry, I feel bad for having said that.

remorse 😔

https://ai.googleblog.com/2021/10/goemotions-dataset-for-fine-grained.html



Message-based therapy

- **>337,000 messages** from message-based therapy sessions from **>6,500 clients** collected via Talkspace (Hull et al., 2020)
- **13,000 documents labeled with PHQ-9 and GAD-7 scores**
- Patients and clinicians gave consent; IRB approved; data handled securely.

Linguistic Inquiry and Word Count (LIWC)

(Pennebaker et al., 2007)

- Counts words belonging to pre-defined categories
- Has been used to measure depression levels in social media posts, therapy conversations, and other written texts
- Obtained categories with a known relationship to anxiety/depression (Tausczik and Pennebaker, 2010): **first-person singular pronouns (“I”), first-person plural pronouns (“we”), bio, health, sadness, anxiety, anger, positive emotion, and negative emotion**.

GoEmotions (Demszky et al., 2020)

- BERT-based emotion classifier pipeline
- trained on Reddit posts
- 27 fine-grained emotions (Cowen and Keltner, 2017)

  or 6 basic emotions (Ekman 1992)

Positive		Negative		Ambiguous
admiration 👏	joy 😃	anger 😡	grief 😢	confusion 😕
amusement 😂	love ❤️	annoyance 😒	nervousness 😬	curiosity 🤔
approval 👍	optimism 🤞	disappointment	remorse 😔	realization 💡
caring 🤗	pride 😌	disapproval 👎	sadness 😞	surprise 😲
desire 😍	relief 😅	disgust 🤮		
excitement 🤩		embarrassment 😳		
gratitude 🙏		fear 😨		

GoEmotions taxonomy: Includes 28 emotion categories, including “neutral”.

GoEmotions (Demszky et al., 2020)

- BERT-based emotion classifier pipeline
- trained on Reddit posts
- 27 fine-grained emotions (Cowen and Keltner, 2017)

  or 6 basic emotions (Ekman 1992)

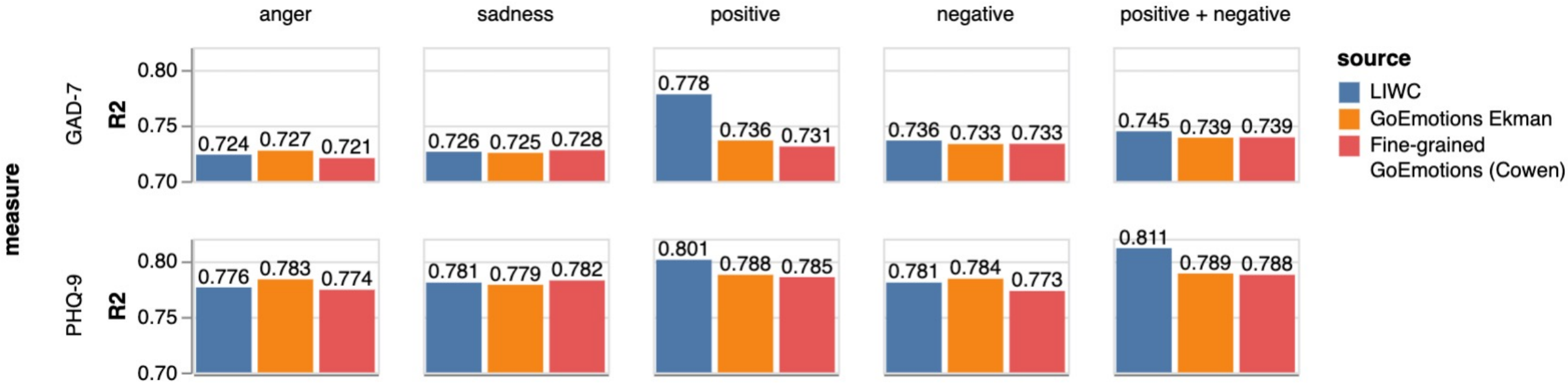
Positive		Negative		Ambiguous
admiration 👏	joy 😃	anger 😡	grief 😢	confusion 😕
amusement 😂	love ❤️	annoyance 😒	nervousness 😬	curiosity 🤔
approval 👍	optimism 🤞	disappointment	remorse 😔	realization 💡
caring 🤗	pride 😌	disapproval 👎	sadness 😞	surprise 😲
desire 😍	relief 😅	disgust 🤮		
excitement 🤩		embarrassment 😳		
gratitude 🙏		fear 😨		

GoEmotions taxonomy: Includes 28 emotion categories, including “neutral”.

Question

Can NN-based emotion extraction methods **(GoEmo)** outperform established word counting methods **(LIWC)** for **predicting depression and anxiety**?

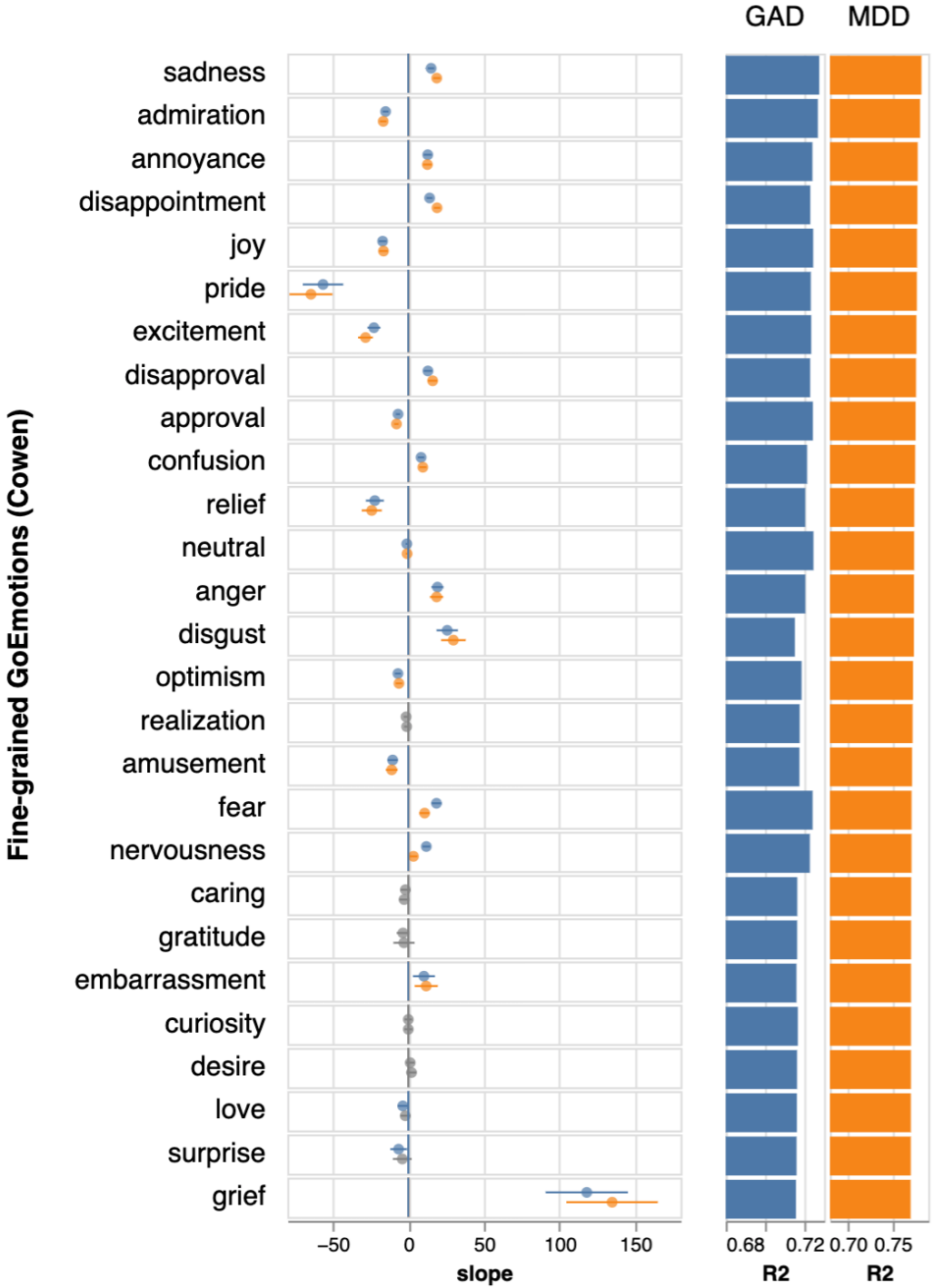
Comparable features are similarly associated

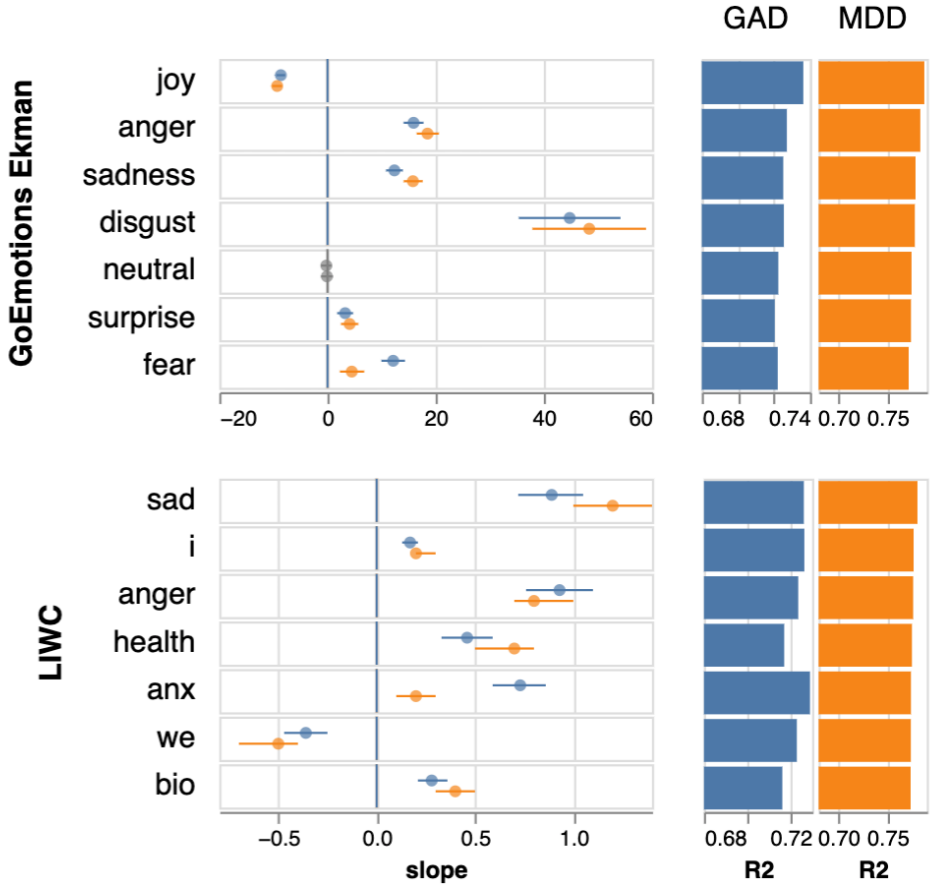


PHQ-9 and GAD-7 score variance explained by the directly comparable features obtainable from LIWC, GoEmotions (Ekman set), and GoEmotions (fine-grained set)

GoEmo captures clinically relevant nuance

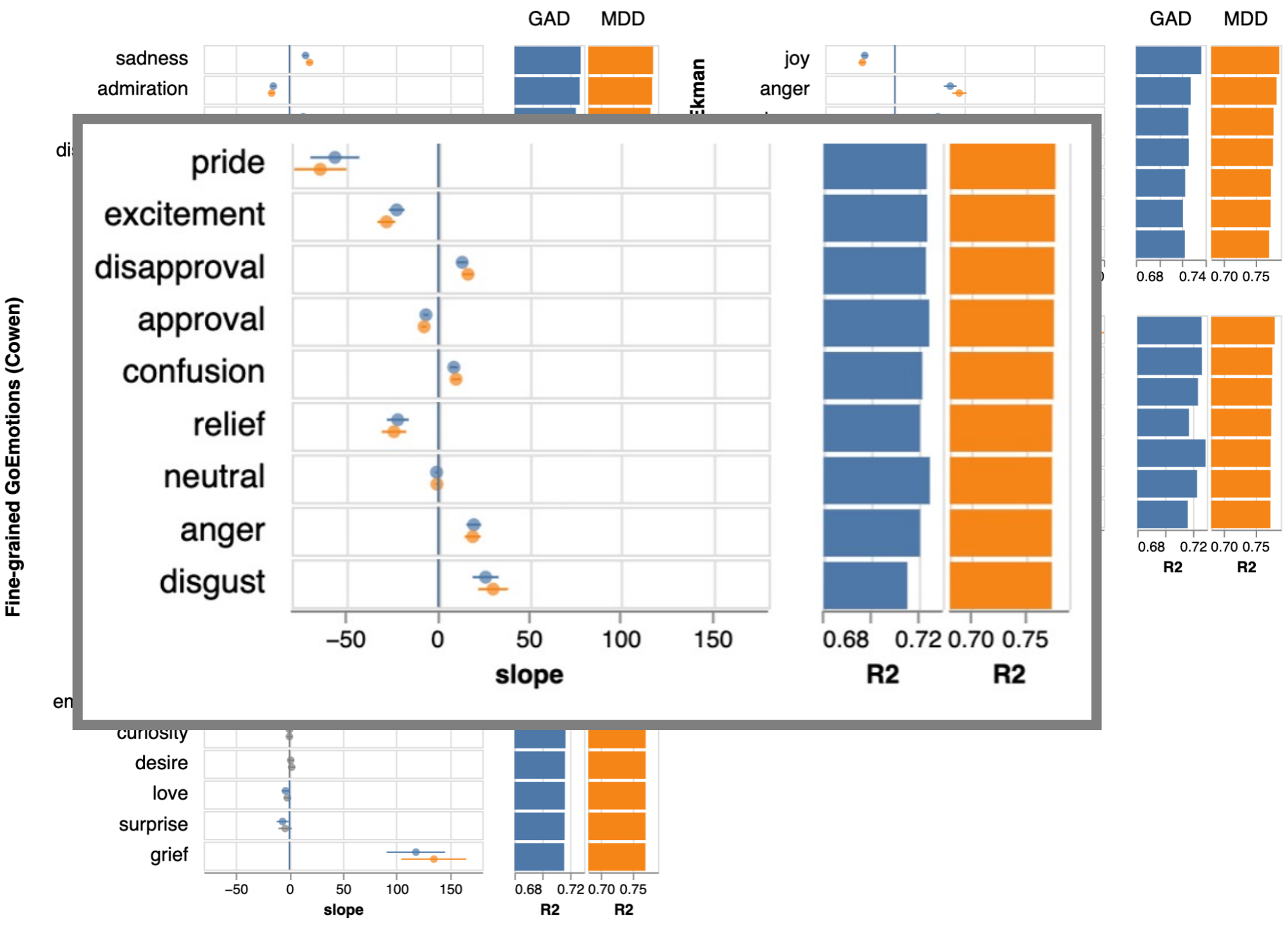
Slopes and variance explained.
PHQ-9 and GAD-7 score univariate mixed-effects linear regression models coefficients (slopes) with 95% confidence interval and variance explained. Variables that were not significant (p≥0.05) are shown in gray.



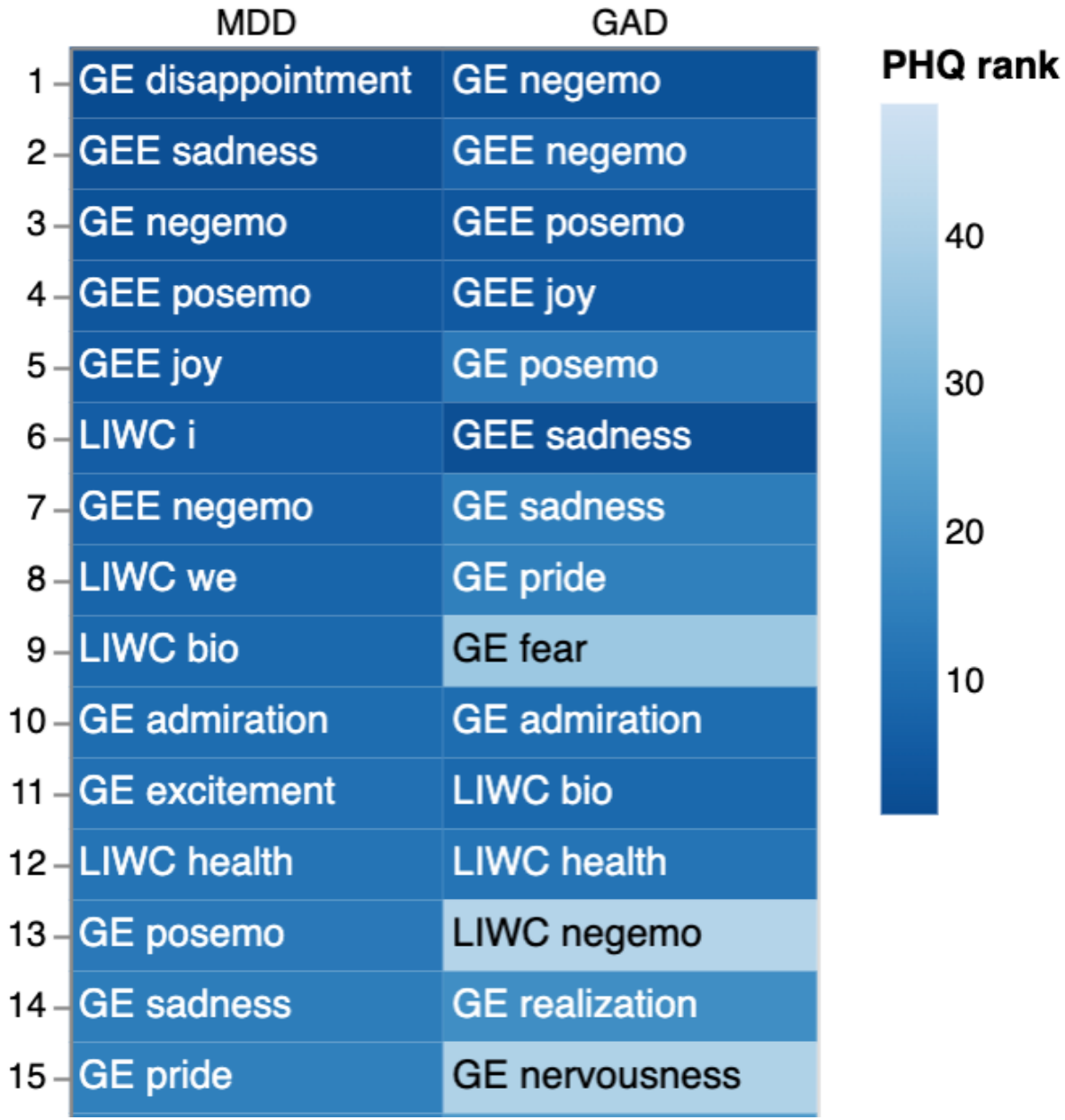


GoEmo captures clinically relevant nuance

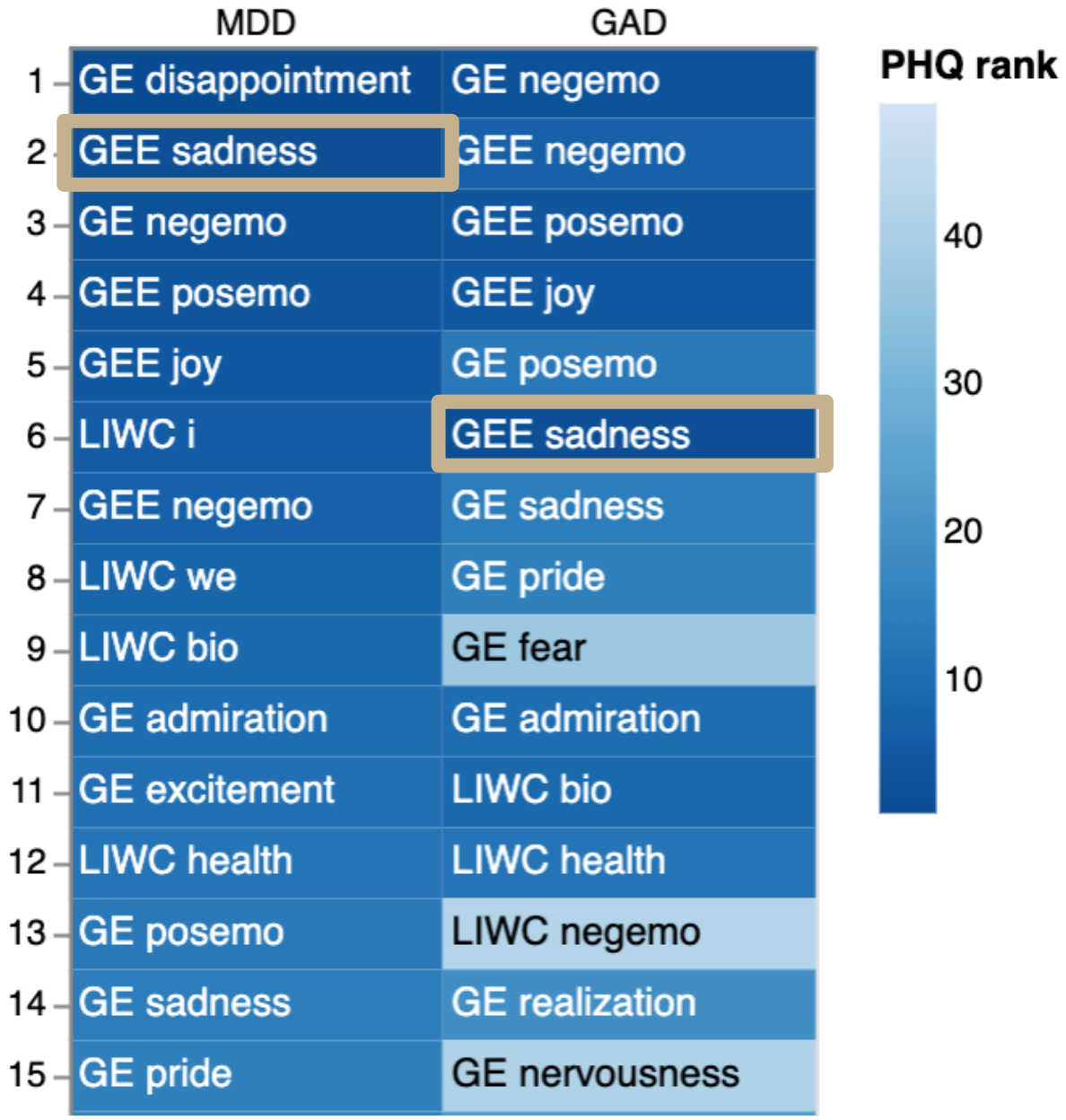
Slopes and variance explained.
PHQ-9 and GAD-7 score univariate mixed-effects linear regression models coefficients (slopes) with 95% confidence interval and variance explained. Variables that were not significant (p≥0.05) are shown in gray.



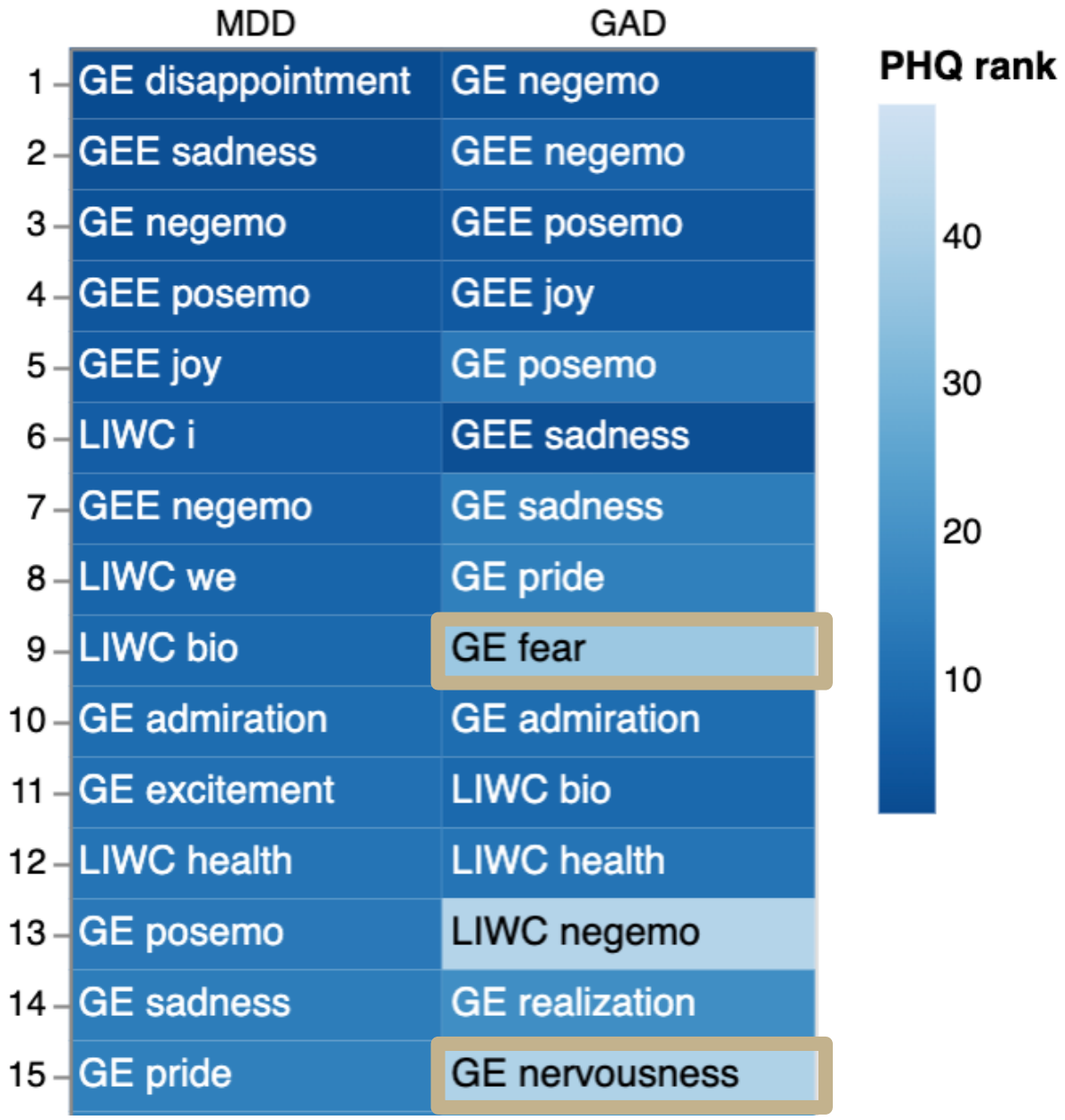
Feature importance differs between depression and anxiety



Feature importance differs between depression and anxiety



Feature importance differs between depression and anxiety



GoEmotions features are collectively more predictive than other feature sets

	MDD				GAD			
	ROC	F1	Pr	Rc	ROC	F1	Pr	Rc
LIWC non-emo	0.577	0.413	0.525	0.341	0.549	0.290	0.478	0.209
LIWC emo	0.621	0.471	0.561	0.405	0.613	0.405	0.541	0.324
GoEmo Ekman	0.643	0.493	0.583	0.427	0.643	0.443	0.550	0.371
GoEmo Cowen	0.662	**0.522**	0.613	**0.455**	0.652	0.444	0.565	0.366
LIWC non-emo+								
LIWC emo	0.640	0.484	0.569	0.420	0.617	0.401	0.529	0.324
GoEmo Ekman	0.655	0.498	0.585	0.434	0.637	0.441	0.548	0.369
GoEmo Cowen	**0.671**	0.514	**0.615**	0.441	0.654	0.451	**0.568**	0.374
All three	**0.671**	0.520	0.612	0.453	**0.657**	**0.456**	0.567	**0.382**

Summary & Conclusion

Can NN-based emotion extraction methods (GoEmo) outperform established word counting methods (LIWC)?

1. LIWC's emotion features are as strongly associated as GoEmotions features → still be a good choice.
2. GoEmotions features capture emotional state comprehensively, yielding additional clinically relevant nuance and benefiting predictive performance

Summary & Conclusion

- Limitations: Non-diverse patient sample (79% ≤ 35 y.o., 79% female, 75% BS or higher)
- Future work: Clinical decision support tools. **Interpretability is key**: models based on interpretable emotion features are preferred over black-box models
- Ethics: Monitoring may be considered invasive - informed consent is paramount. Further research & applications must take ethical considerations into account.

haalbu@uw.edu
https://hannahburkhardt.github.io/
https://www.linkedin.com/in/hannahburkhardt/

Thank You

Funding: This work was supported by the National Library of Medicine (grant number 67-3780) and by Innovation Grant "Informatics-Supported Authorship for Caring Contacts (ISACC)" from the Garvey Institute for Brain Health Solutions.

Conflicts of interest: TDH is an employee of the platform that provided the data.

References:

Alan S. Cowen and Dacher Keltner. 2017. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences of the United States of America*, 114(38):E7900–E7909, September.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A Dataset of Fine-Grained Emotions., May.

Paul Ekman. 1992. Are There Basic Emotions? *Psychological Review*, 99(3):550–553.

Thomas D. Hull, Matteo Malgaroli, Philippa S. Connolly, Seth Feuerstein, and Naomi M. Simon. 2020. Two-way messaging therapy for depression and anxiety: longitudinal response trajectories. *BMC Psychiatry*, 20(1):297, December.

James W. Pennebaker, R. J. Booth, and M. E. Francis. 2007. Linguistic Inquiry and Word Count: LIWC.

Yla R. Tausczik and James W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1):24–54, March.