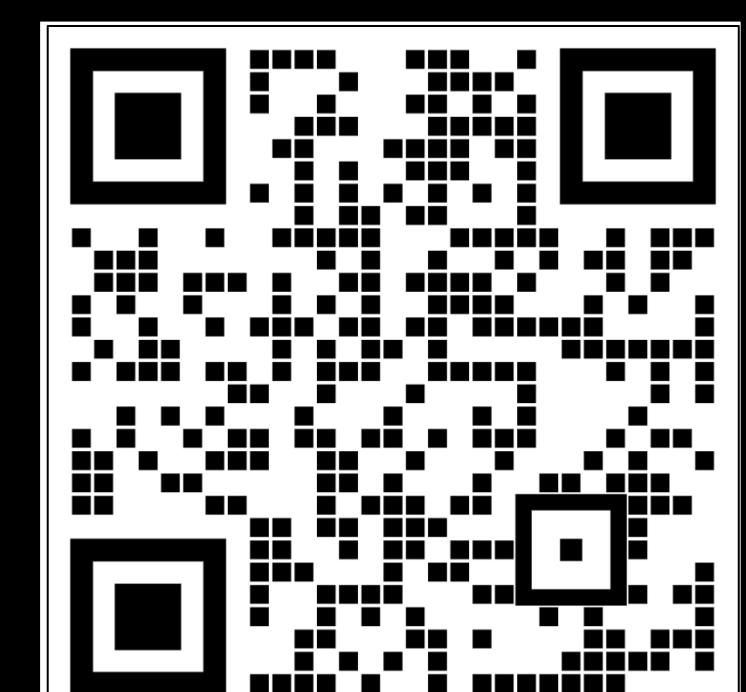


# Evaluating Human-LLM Representation Alignment: A Case Study on Affective Sentence Generation for Augmentative and Alternative Communication



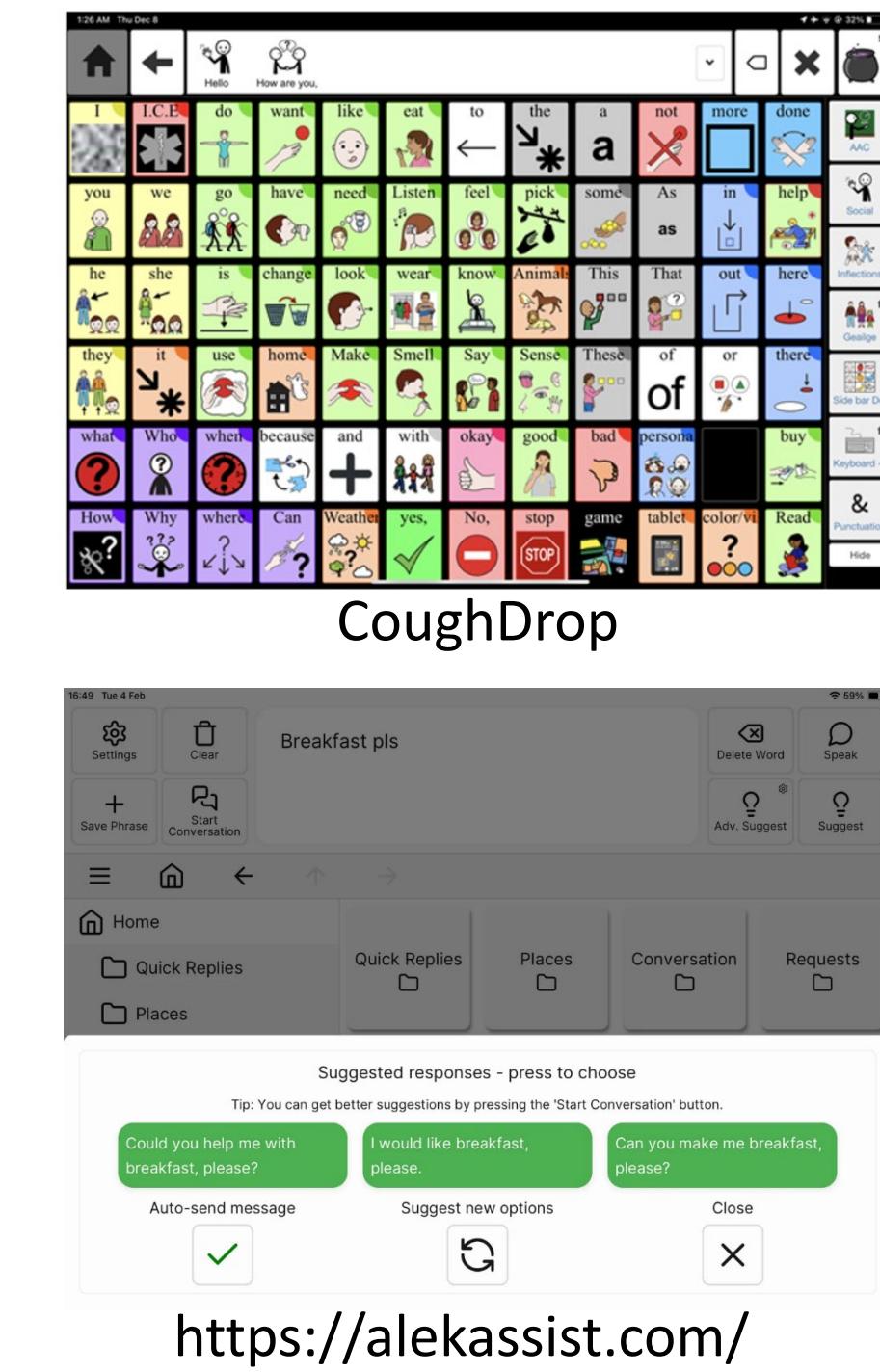
arXiv Link

*Shadab Choudhury, Asha Kumar, Lara J. Martin*  
 {shadabc1, laramar}@umbc.edu

## Augmentative and Alternative Communication (AAC)

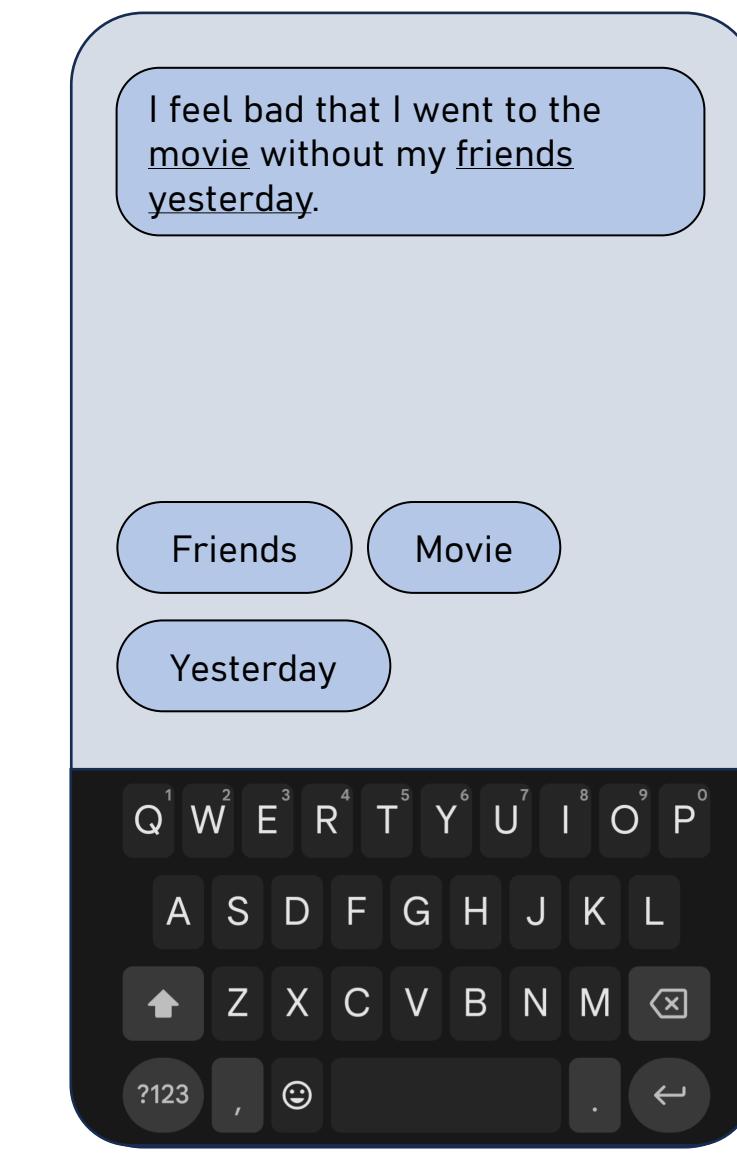
Tools or software to help supplement or replace speech for people who cannot communicate verbally. Two examples of other already-existing AAC software are given on the right.

Can we improve AAC software using LLM text generation?



## Keyword-Based Generation

Get a full sentence from a few keywords. Balances input speed with personalization for AAC applications. Also called lexically-constrained generation.



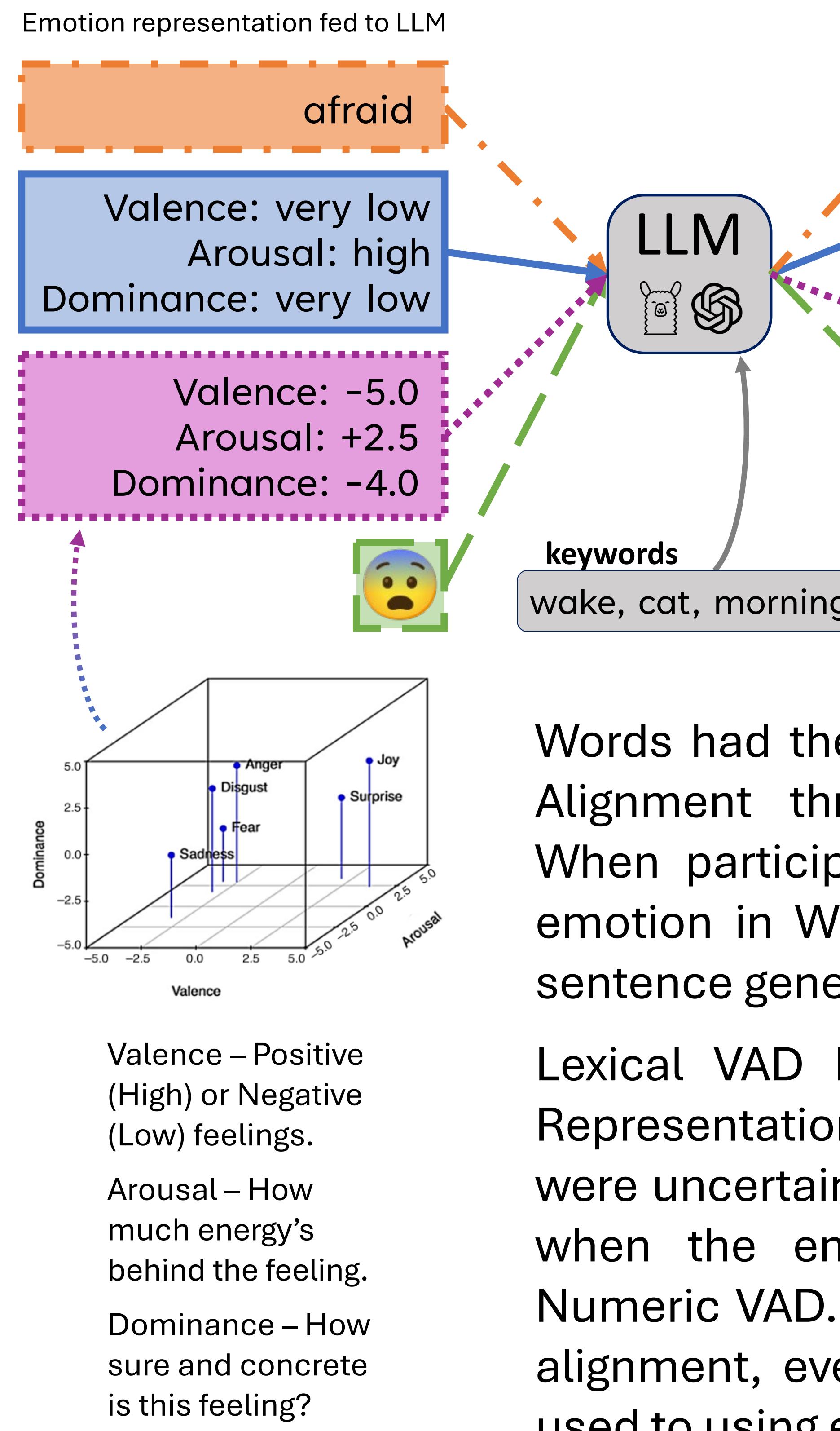
## Designing AAC software or other Human-AI Collab tools?

Words to describe emotions are the best option for alignment and quality. If you need finer, less ambiguous outputs, use VAD, but quantify them in words (low, moderate, high, etc).

## Deciding how to represent a concept for Human-LLM Collab?

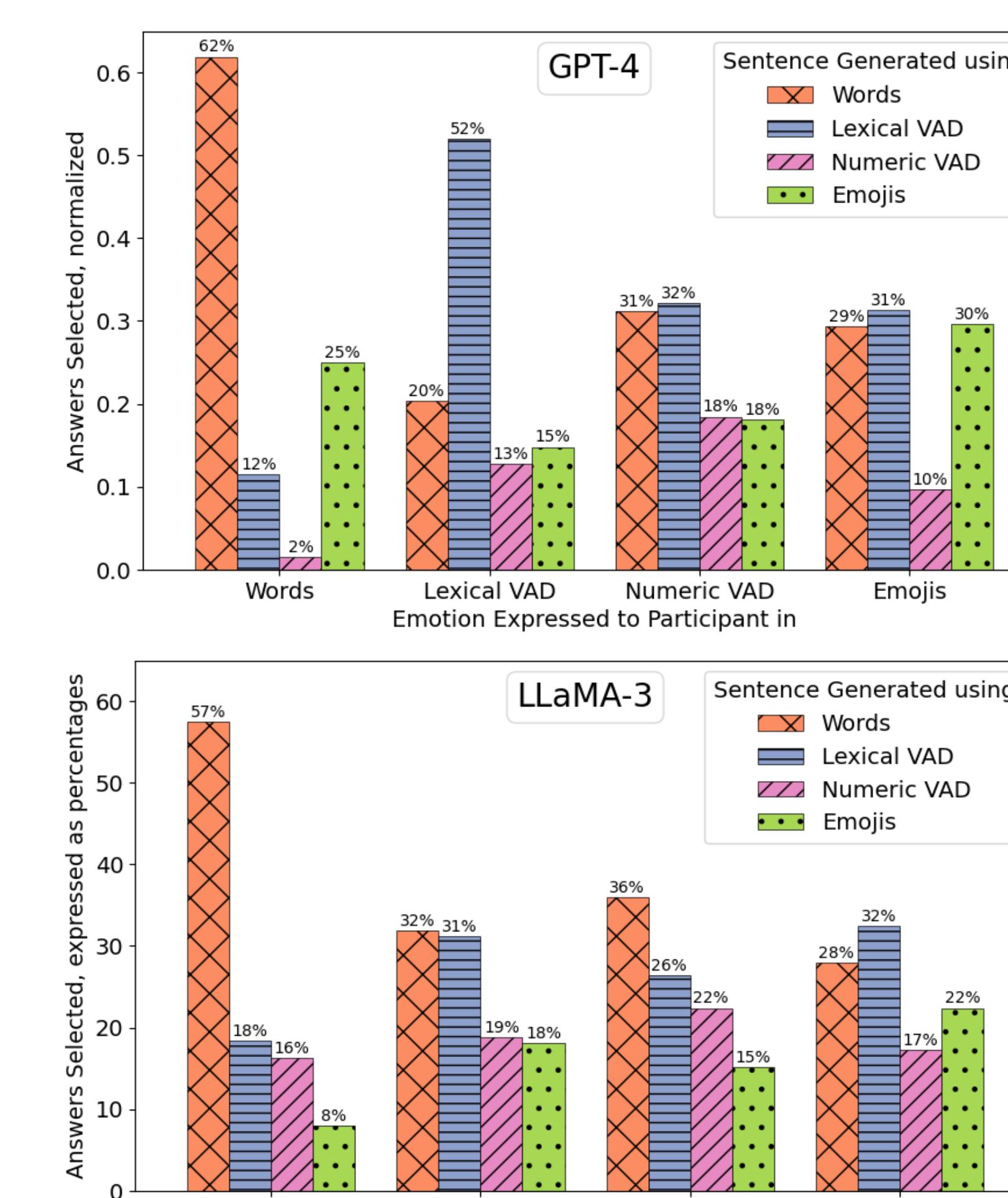
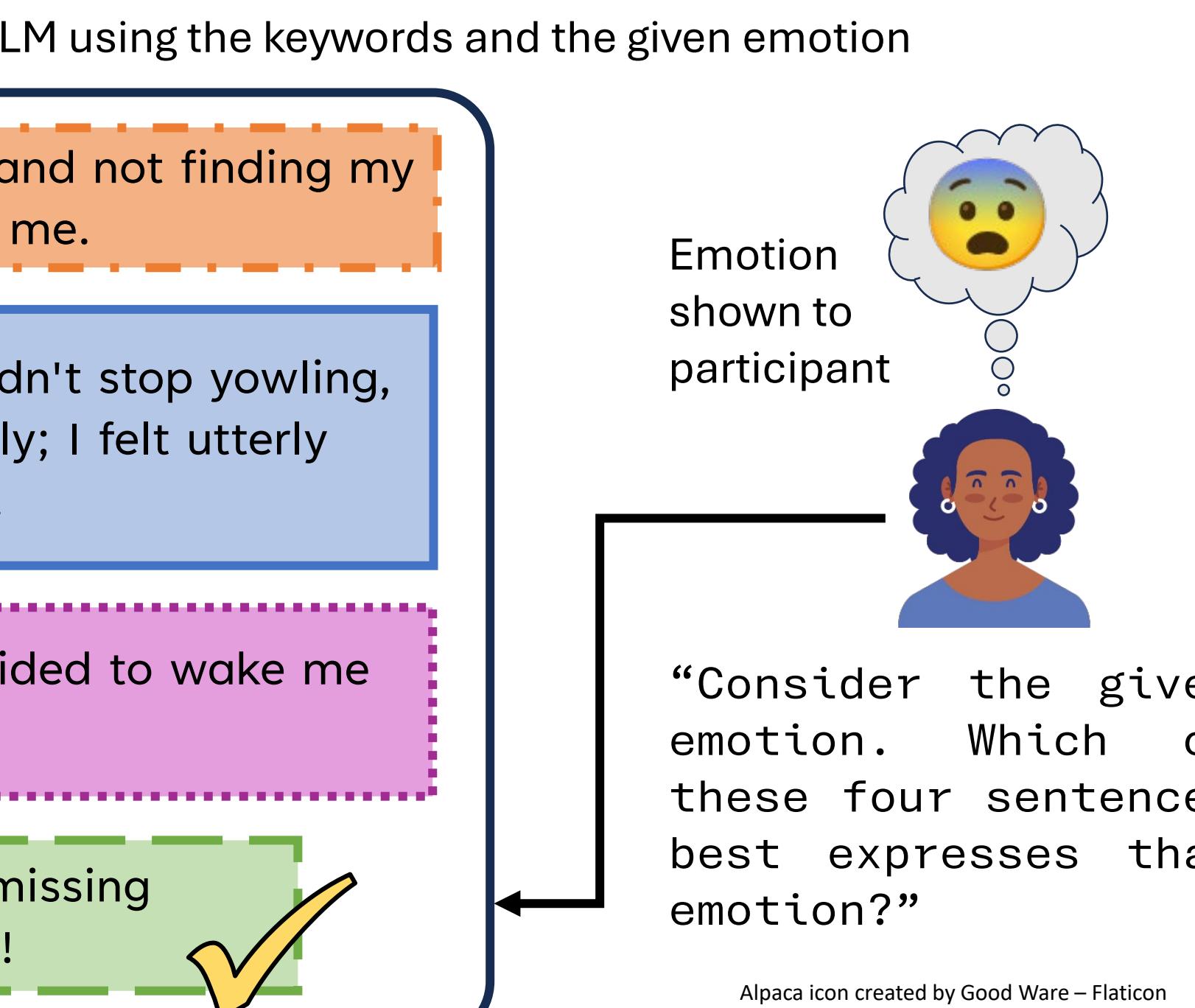
Use our representation alignment paradigm! Give a user one representation, then ask them to pick from multiple LLM outputs, each using a different representation, then check representation alignment.

## Representation Alignment



Words had the best Representation Alignment through self-alignment. When participants were shown an emotion in Words, they picked the sentence generated using Words.

Lexical VAD had the second-best Representation Alignment. People were uncertain of the best sentence when the emotion was given in Numeric VAD. Emojis also had poor alignment, even if most people are used to using emojis daily.

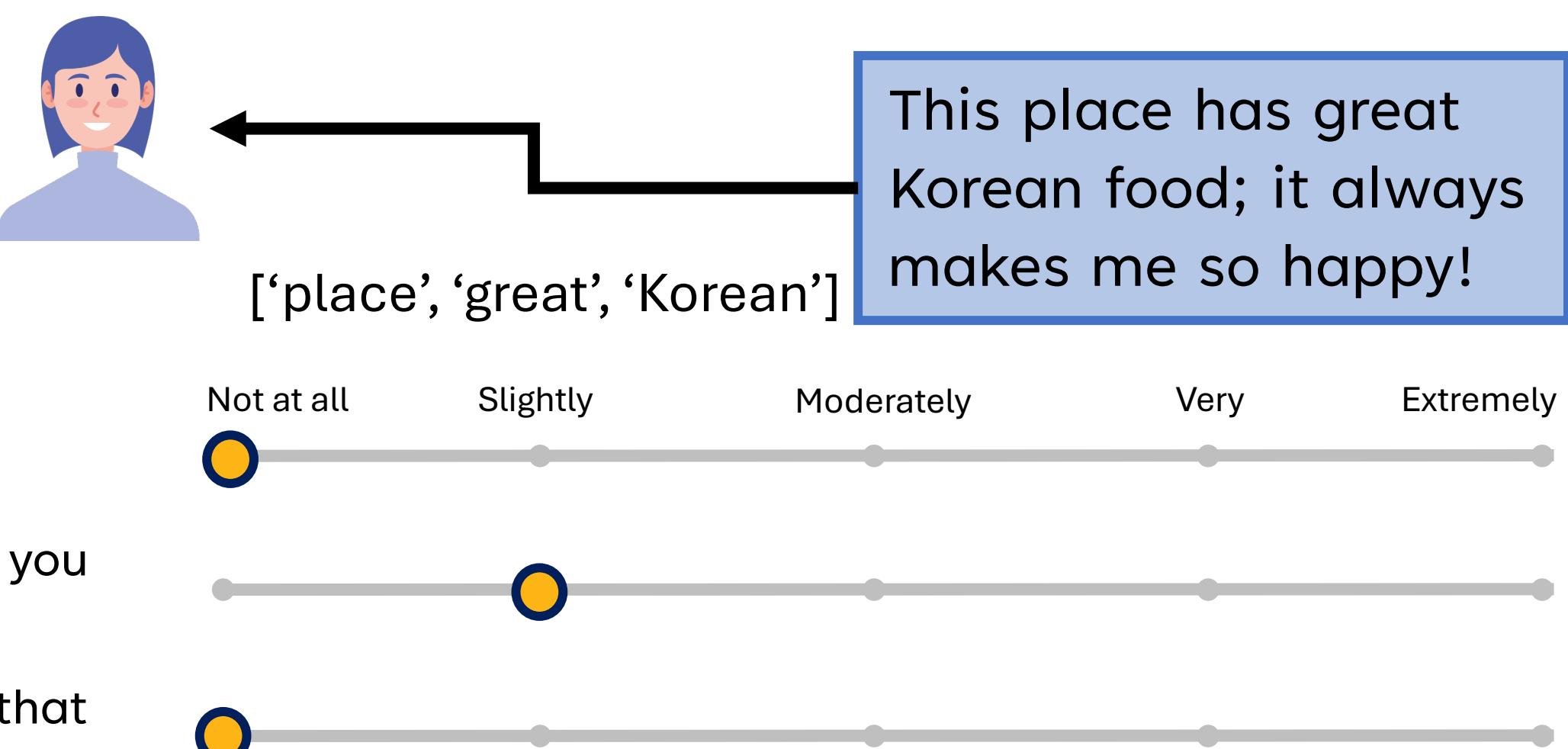


## Accuracy and Realism

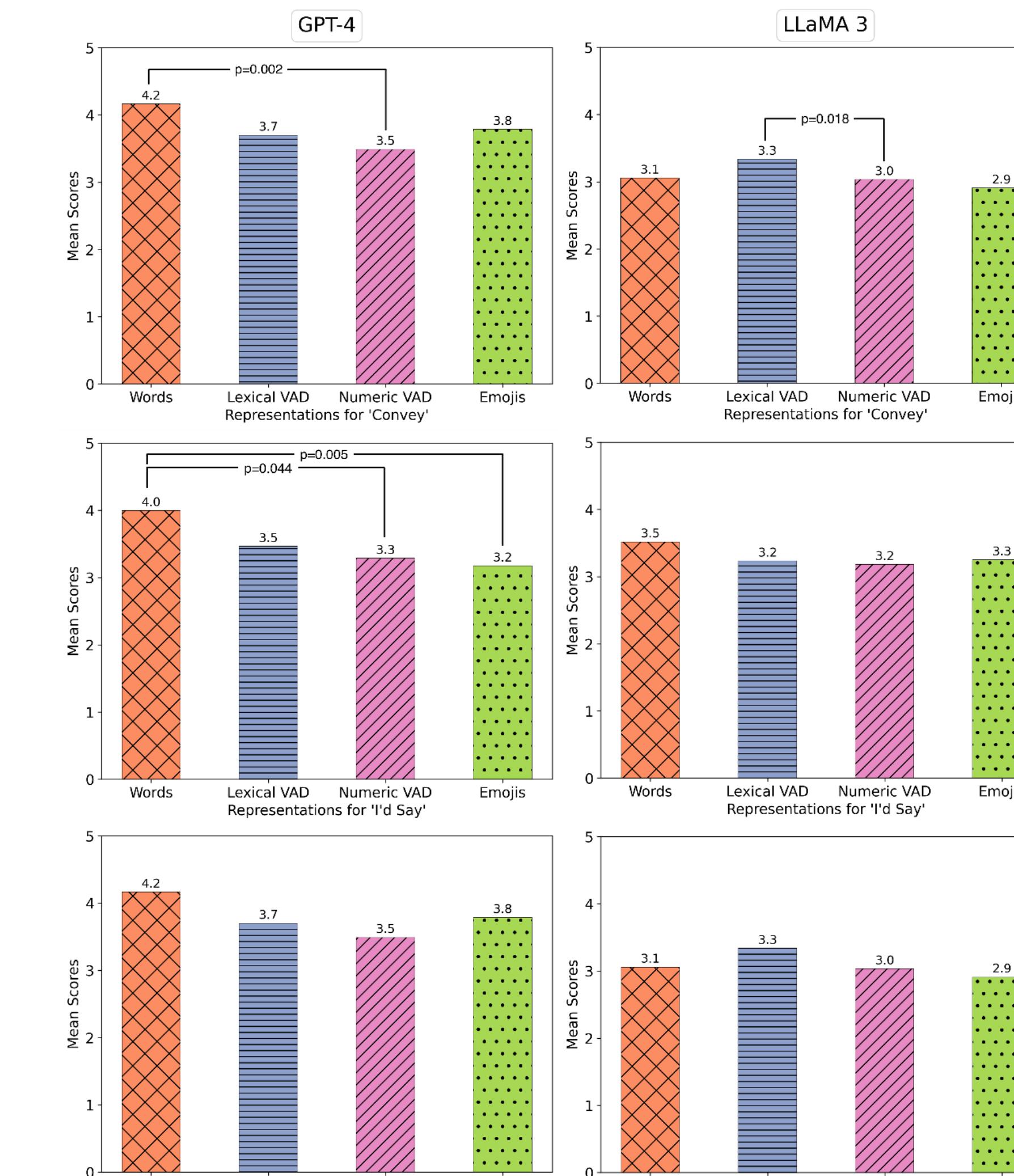
Consider the emotion represented by these VAD values. Then consider the given sentence and answer the following questions.

Valence: very low  
Arousal: moderate  
Dominance: high

These values represent ‘Anger’



- How much does the sentence convey the given emotion?  
How much does the sentence sound like something that you would say?  
How much does the sentence sound like something that someone else'd say?



Lines generated using Words were the most accurate and realistic. Lines generated using Lexical VAD came second.

Both Numeric VAD and Emojis gave much poorer results overall.

Representation	Grateful	Joyful	Content	Surprised	Excited	Impressed	Proud	Anxious	Afraid	Bittered	Annoyed	Furious	Sad	Devastated	Ashamed	Embarrassed	Guilt	
Words	4.33	3.27	4.67	4.12	3.30	3.57	3.32	4.33	4.75	4.75	4.00	4.50	4.67	4.14	5.00	4.38	4.40	4.38
Lexical VAD	4.43	3.50	4.43	3.50	4.00	3.38	4.50	3.43	4.00	4.25	3.71	4.33	4.14	3.67	4.29	3.33	3.08	4.20
Numeric VAD	3.78	3.63	3.36	3.80	4.20	3.63	3.90	3.71	4.00	3.71	3.50	3.67	3.33	4.70	3.43	3.33	4.67	3.54
Emojis	3.22	3.83	3.11	3.36	4.56	3.29	4.00	3.25	4.00	4.44	3.29	4.25	4.11	4.70	4.00	4.71	3.38	3.55
Words	4.29	3.20	4.75	3.25	3.00	3.32	3.38	4.14	3.73	3.33	4.60	4.83	4.22	3.88	3.30	3.00	3.67	3.75
Lexical VAD	4.14	2.75	4.00	3.50	3.12	3.38	3.14	3.73	3.88	3.20	3.53	3.31	3.50	4.00	4.60	4.00	4.60	3.75
Numeric VAD	2.11	2.85	2.56	2.97	4.50	2.57	3.38	3.80	2.88	2.30	3.00	3.40	3.29	3.75	2.83	3.00	3.67	3.67
Emojis	2.17	3.40	4.00	3.58	1.75	2.14	1.60	3.57	3.00	2.00	3.00	4.73	4.20	2.40	2.29	3.44	3.60	2.33