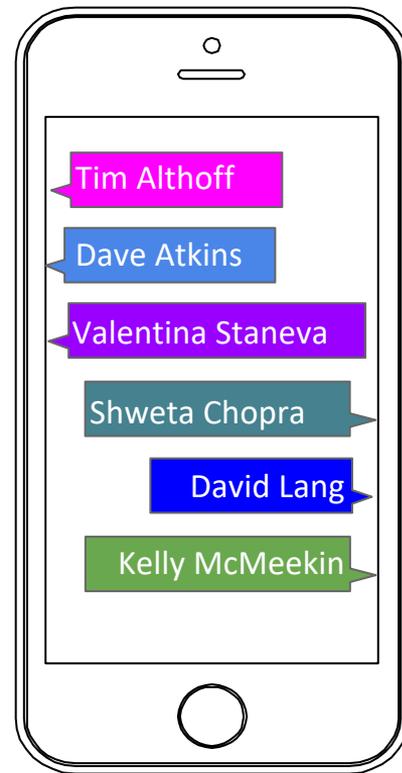




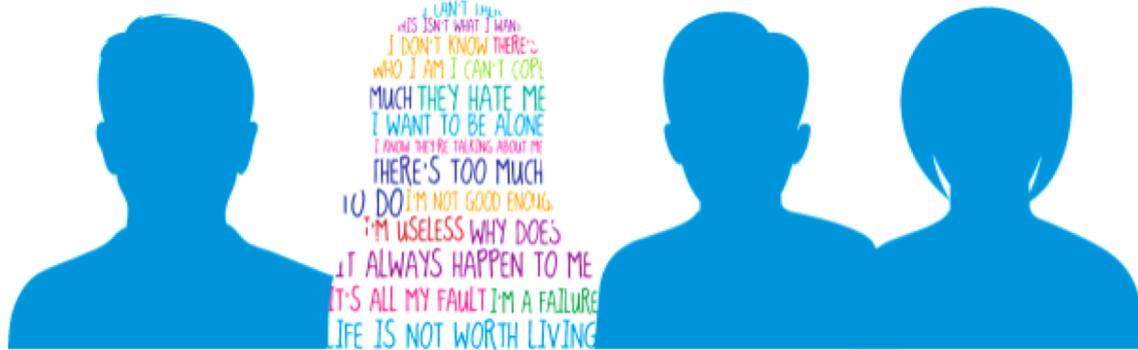
UNIVERSITY of WASHINGTON
eScience Institute

Natural Language Processing for Online Peer Support

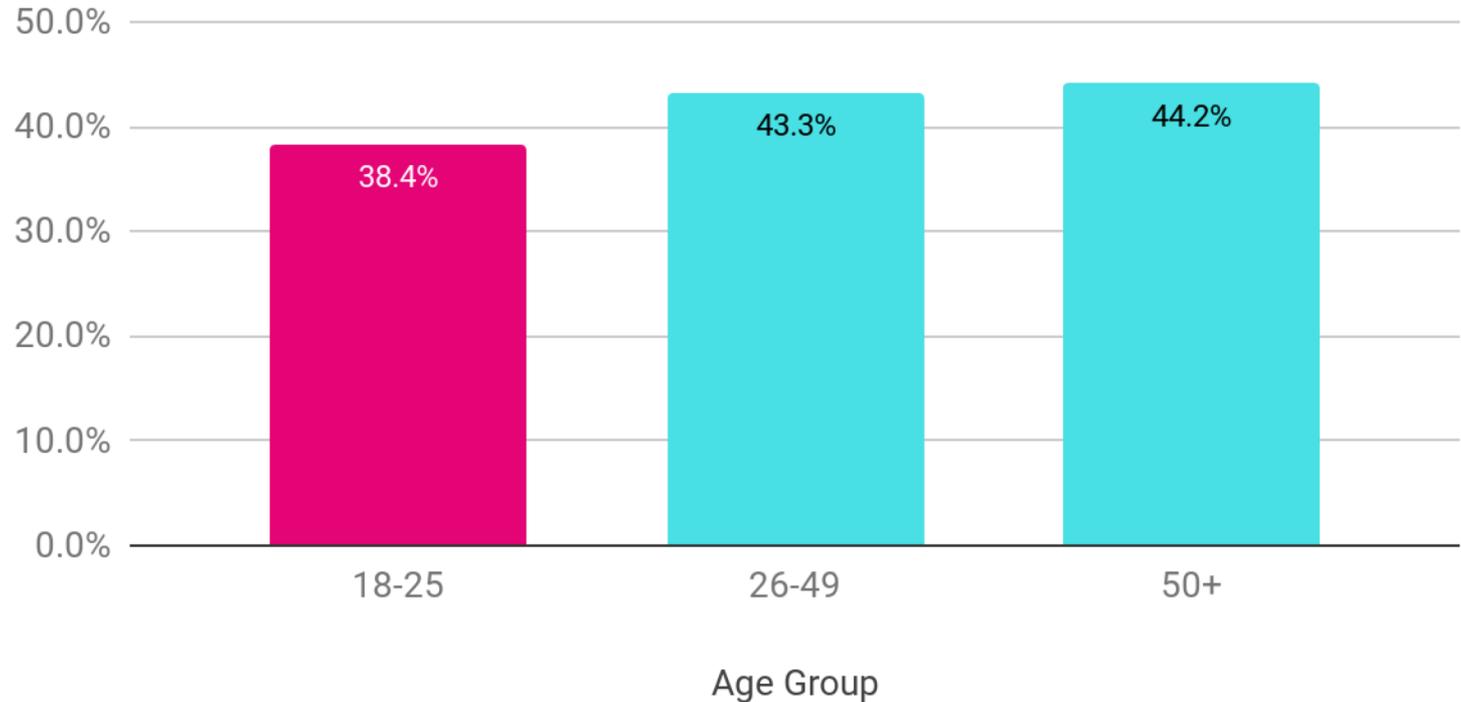


Mental illness affects

1 in 4



Percent of US Adults with Any Mental Illness Who Received Mental Health Services in 2017





SUPERBETTER



wysa

VENT

The social network for sharing life's frustrations



moodpath



youper



7 Cups of Tea

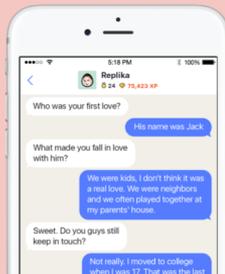


What's Up?



Let's try and help out a little

Replika.
Your AI friend



Peer Support

People provide knowledge, emotional support, and/or practical help to each other based on a shared experience of emotional or psychological pain.

What types of responses are the most helpful to individuals sharing their mental health struggles online?

1.

Online Peer Support Platform



500,000

Users across the world



19 years old

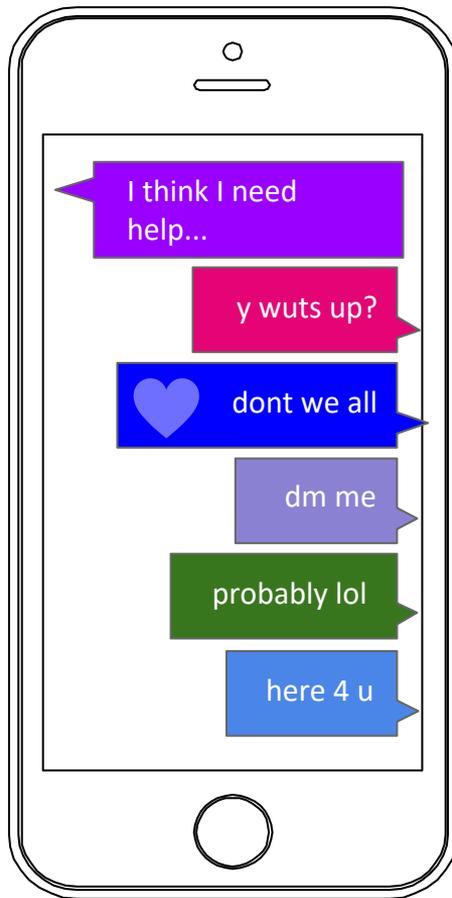
Median age



<50

Admins and moderators

Post

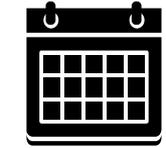


Comments

2.

**Diving into the
Data**

Focusing on mental health



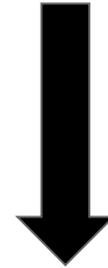
All dates



20 million



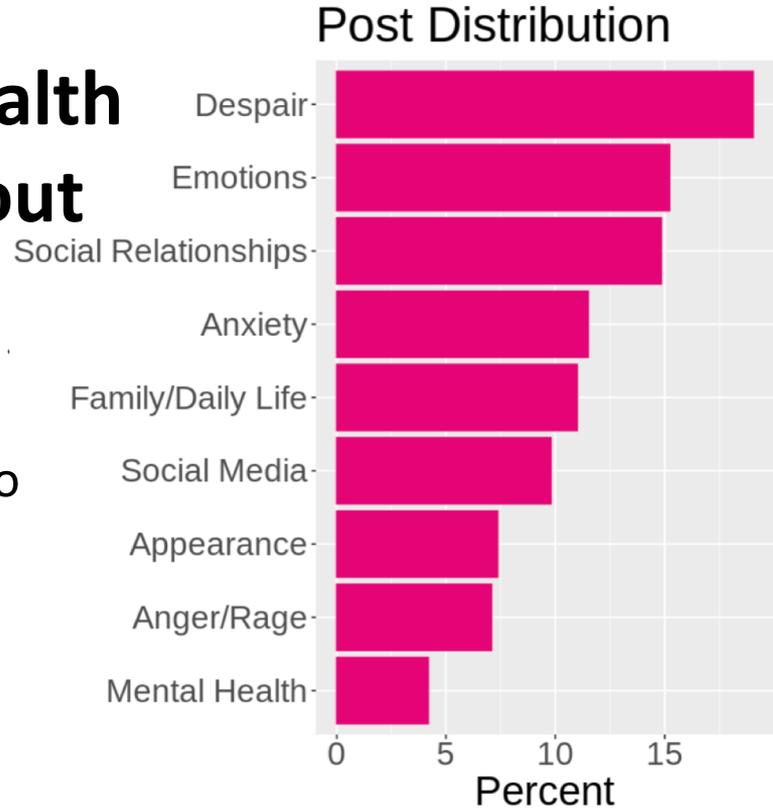
2018 only



2 million

What types of mental health problems do users seek out peer support for?

- ❑ Used natural language processing technique called “topic models” to identify a post’s theme
- ❑ Hand-labeled into 9 distinct categories





Example “Despair” Post:

Sometimes I get so lonely I can't help but feel sad and then I'm angry for being this way



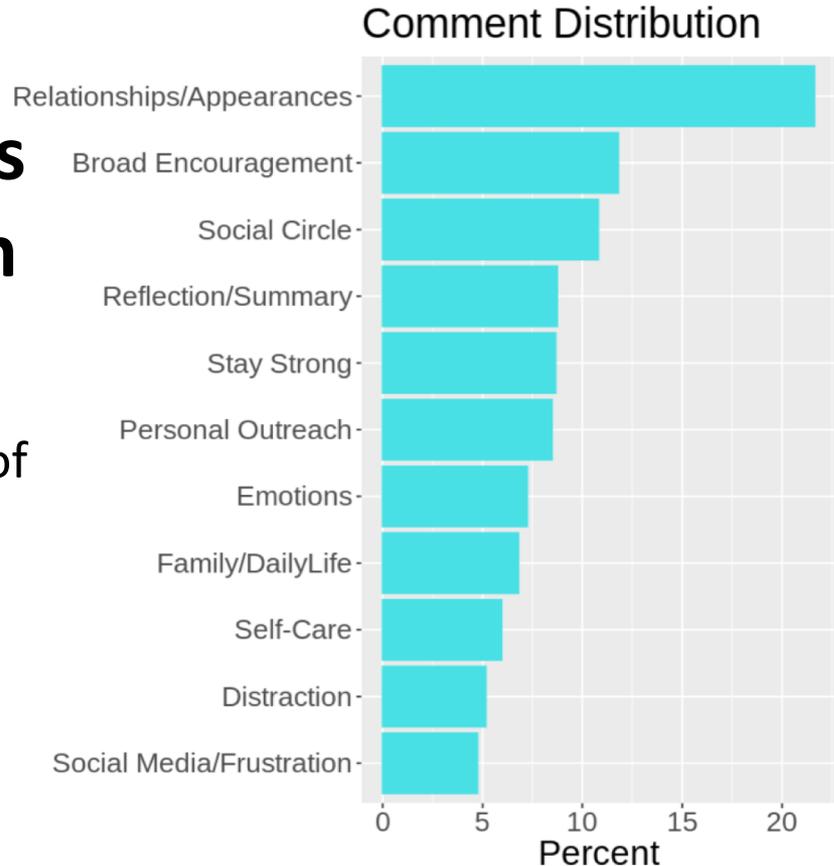
Example “Broad Encouragement”

Comment:

I know things are hard but I know you can get through this. I am rooting for you.

What types of comments are left on mental health posts?

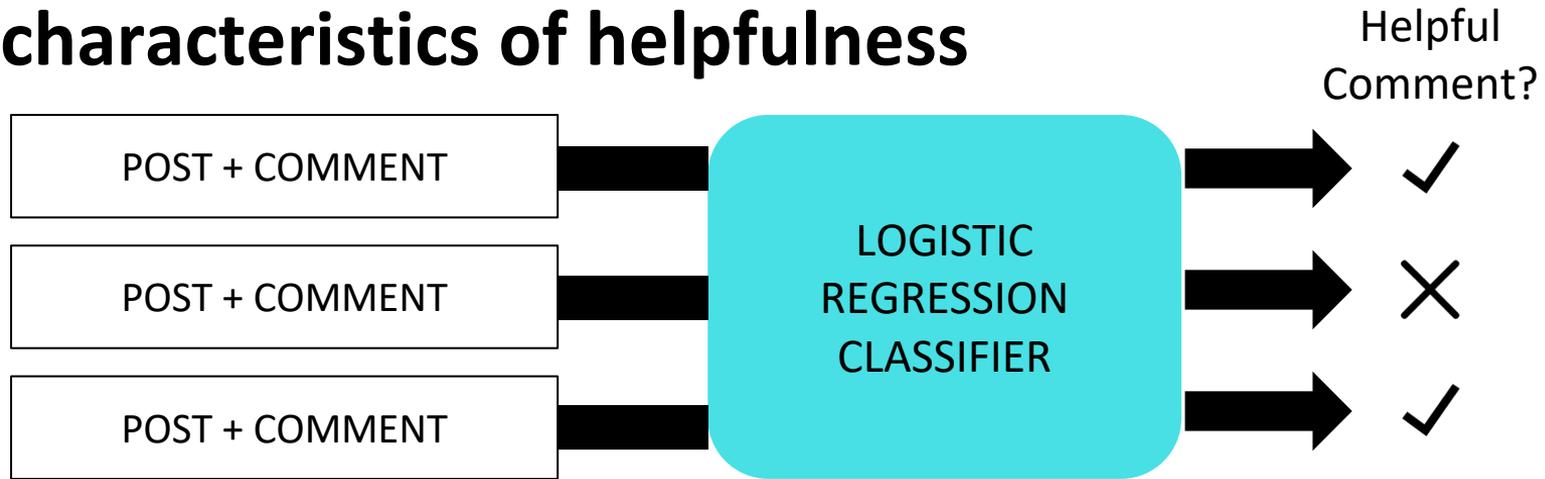
- ❑ Identified 11 distinct categories of comments
- ❑ The most prominent support strategies are relating personal histories and offering words of encouragement



3.

Methods

Our supervised learning model to understand characteristics of helpfulness



Can this tell us something about helpful peer support?

Measuring helpfulness, the ideal way

Professionally annotated data

Better able to understand context-driven responses

Surveys

“Was this comment helpful?”

Mental health assessments

When users first sign up, and periodically afterwards

Measuring helpfulness, by proxy

Likes

Poster “likes” a comment on their post

General Gratitude

Poster thanks group of commenters (i.e. “thank you all.”)

Follows

Poster “follows” a commenter

Mood Change

Poster posts again with improved mood category

Specific Gratitude

Poster thanks specific commenter

Any of These

Poster likes, follows, thanks, or posts with a mood change

Caveats for proxy measurements

Likes

Social media chatter

Follows

Infrequent

Specific Gratitude

If >1 comment by user, which one was it?

General Gratitude

Social norms

Mood Change

Positive mood \neq positive content

Any of These

If any one measure is flawed, then so is this

And if users only participate on platform for median 2 days...

Do proxies occur often?

<i>Measures</i>	<i>Percent</i>
Likes	35%
Mood Improvement	18%
Gratitude	3%
Personalized Gratitude	7%
Follows	4%
Any Positive Outcomes	52%

Do they co-occur?

<i>Cronbach's alpha</i>	<i>Internal consistency</i>
$0.9 \leq a$	Excellent
$0.8 \leq a \leq 0.9$	Good
$0.7 \leq a \leq 0.8$	Acceptable
$0.6 \leq a \leq 0.7$	Questionable
$0.5 \leq a \leq 0.6$	Poor
$a < 0.5$	Unacceptable
$a < 0.23$	Our proxies...

What types of comments could be helpful?

Specific words

Did a poster or commenter use a specific word or types of words?

Althoff et al (2016)

Specific behaviors

Clinically validated actions (what professionals do)

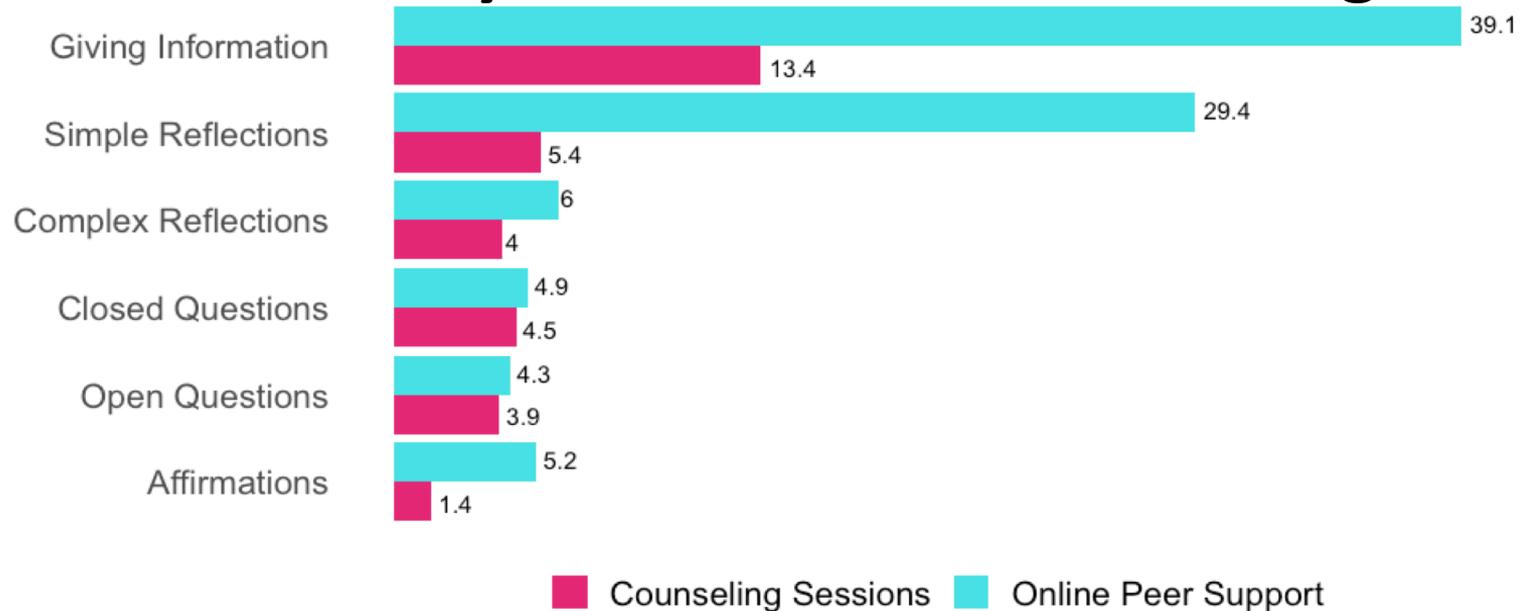
Tanana et al (2016)

Behavior of clinicians

Motivational Interviewing (OARS)

- ❑ Open questions – *“How did that make you feel?”*
- ❑ Affirmations – *“You handled that really well.”*
- ❑ Reflective Listening – *“It sounds like...”*
- ❑ Summaries – *“Let me see if I understand...”*

The distribution of behaviors on peer support platforms is very different from counseling



4.

Findings

As a reminder, “helpful” means:

- Likes
- Follows
- Specific gratitude
- General gratitude
- Mood change
- Any of the above

And “not helpful” means the absence of these indicators

User Characteristics

More likely to be helpful:



Females

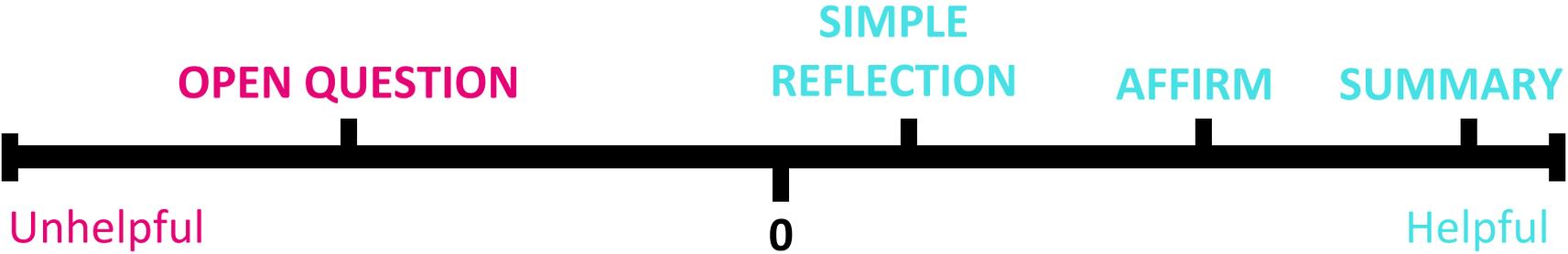


Younger users



Those with larger networks

Counseling Behaviors



Comments

Helpful:

professional help
feel free contact
feel need talk
just message
away don't worry
btw hope i ly live happy life
lose hope totally hehe
mention need talk
support proud deserve
did right thing saved best
mention great
stay strong
appreciates yes
stronger mention yeah know
sending lots love
really sorry happened
don't hesitate message

Not Helpful:

self harm suicide
don life worth
suicide isn answer reconsider
don stupid
throw away
make friends
jk insulting
hehe lol don don think doing
pregnant don end kik
crime doesn care logic just talk
clinical
just let therapy
seek help permanent
afterlife going thing
plz don
don think like
depression suicidal thoughts

**Affirmation, encouragement and being available
are found to be helpful**

Comments

Helpful:

- Longer comments
- Use of second person pronouns
- Use of emojis

Not Helpful:

- Use of first person pronouns
- Anxious language use

Longer comments tend to positively correlate with our helpfulness measures

How did we do?



58-62%

Our Accuracy



0.61

Our F1 Score



50%

Chance Accuracy

What does that mean?

- There's a lot of complexity involved
- These are proxies, not helpfulness

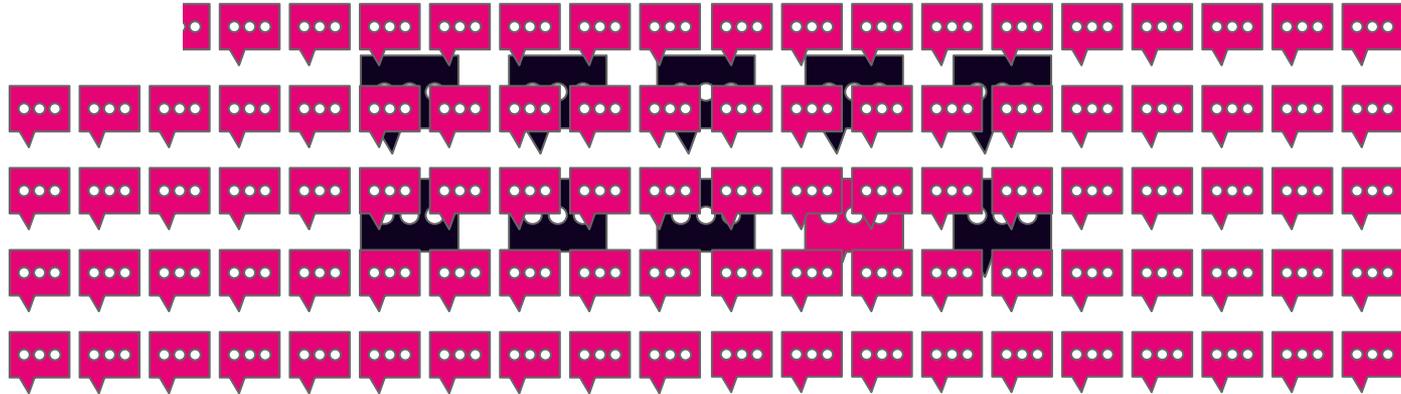
5.

Takeaways

We need better ways to measure helpfulness

- Platform design
- Surveys
- Mental health assessments

We need better (and more) moderation



Peer support isn't suitable for all mental health problems

Building community, feeling heard, being encouraged



Struggling through a deeper challenge

**Need better ways to deal with
suicidal/extreme content**

First do no harm.

Then try to help.



UNIVERSITY *of* WASHINGTON
eScience Institute



Washington Research
FOUNDATION

GORDON AND BETTY
MOORE
FOUNDATION

CASCADIA URBAN
ANALYTICS COOPERATIVE