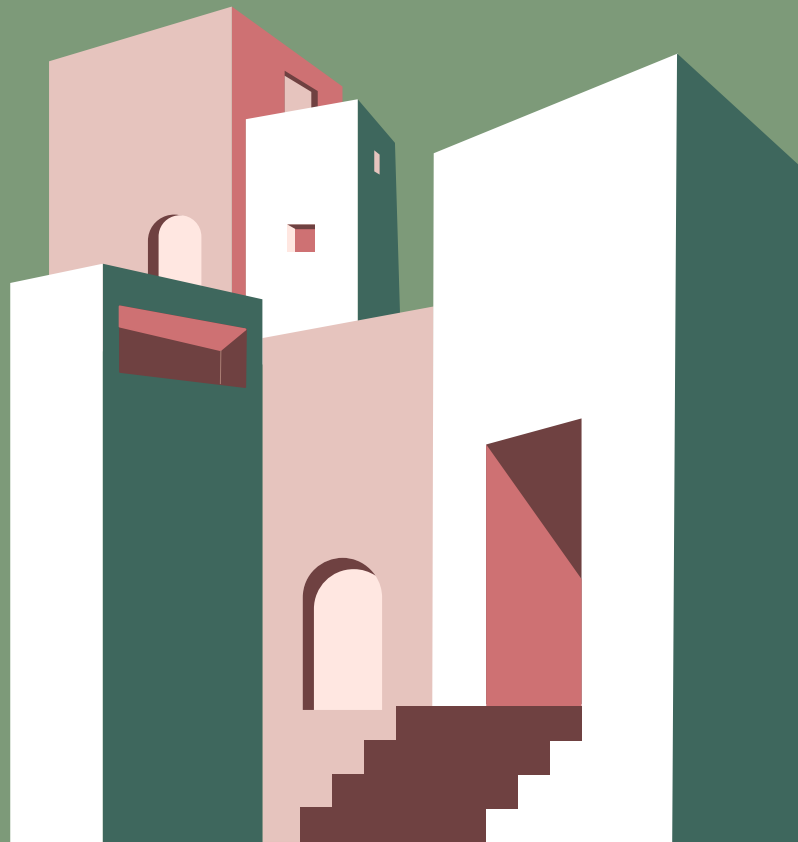




SCREENBOOK

Your guide to traveling the Internet

Niharika Sitomer, Shanie Hsieh,
Shirley Jiang, Tessa de Vries



TEAM SCREENBOOK

MISSION STATEMENT: OUR TEAM AIMS TO PROVIDE A SERVICE THAT ALLOWS USERS TO NAVIGATE THE ENORMITY OF THE INTERNET SAFELY AND ON THEIR OWN TERMS.



SHIRLEY JIANG

Product Manager
Project Manager
Resident SME



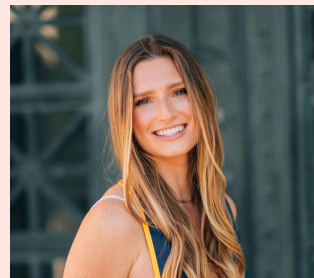
SHANIE HSIEH

Data Engineering
Web Extension Dev
EDA



NIHARIKA SITOMER

Data Engineering
Web Extension Dev
Model Evaluation



TESSA DE VRIES

ML Lead
Model Evaluation



SCREENBOOK

CONTENT WARNING/TRIGGER WARNING:
DISCUSSION OF SUICIDE



PROBLEM: PLATFORM-SIDE CONTENT MODERATION

EXISTING MARKET — PLATFORM SIDE

- Content moderation exists primarily in the hands of social media platforms

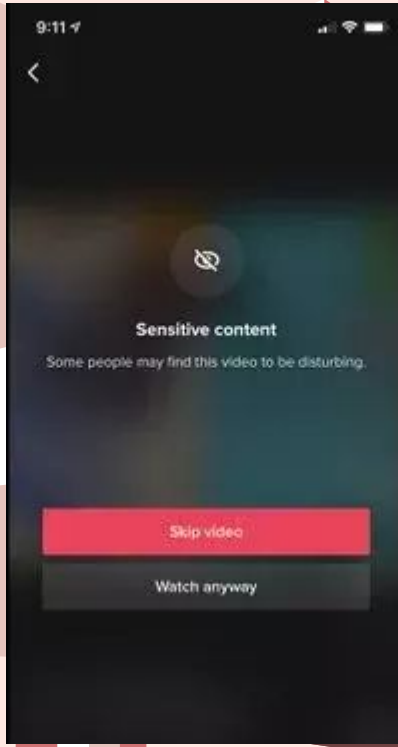


Search Engines



Social Media Posts

INEFFECTIVE!





SOLUTION: SCREENBOOK

IMPACT – USER SIDE

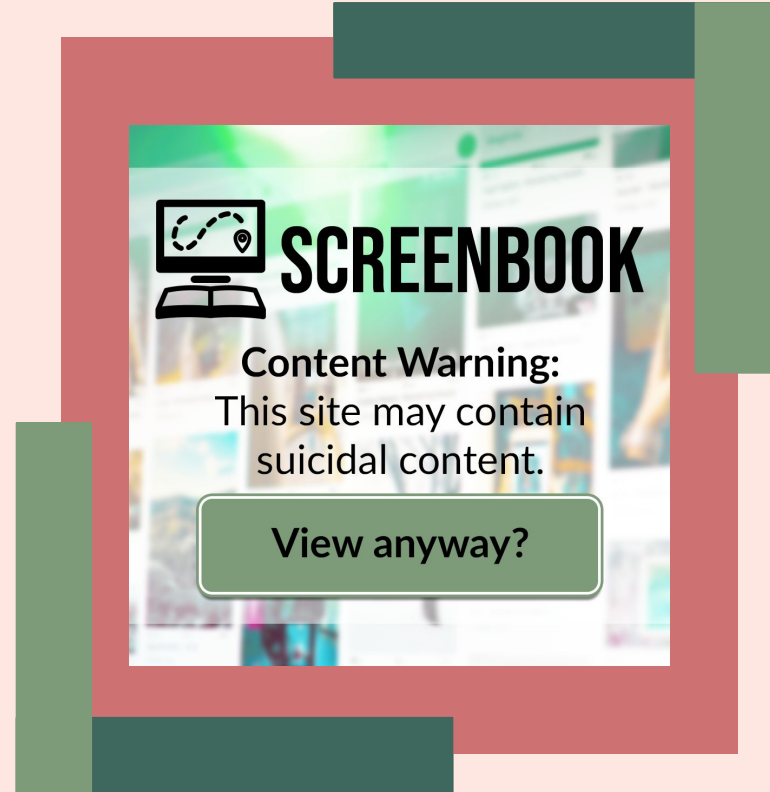
- **Content type disclosure** allows for users to make choices on what they see on the Internet, **allowing for improved user experiences.**

SO, WHAT IS SCREENBOOK?

SCREENBOOK MVP

Develop a Chrome browser extension that:

- Identifies **suicidal content**
- **Blurs** a webpage that contains suicidal content
- And provides a prompt **overlay** that allows users to view that content at their discretion



MARKET ALIGNMENT

BUYER

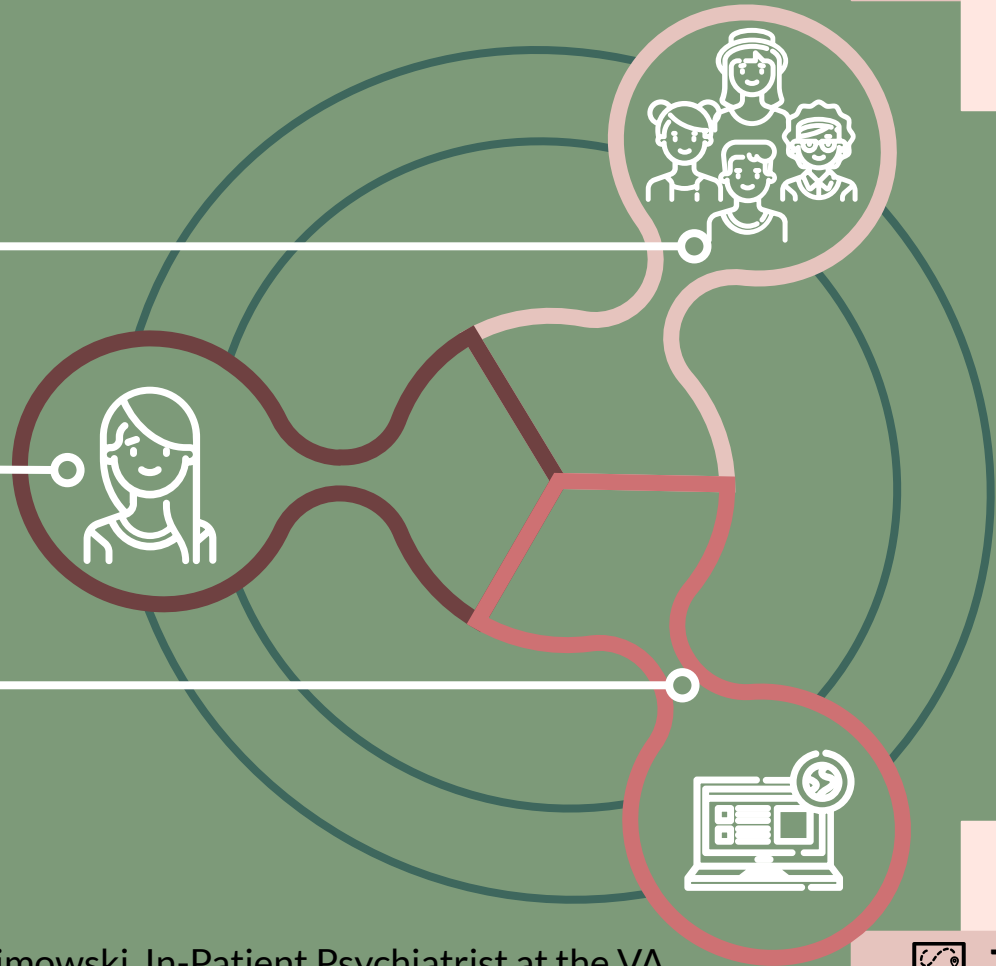
Parents, therapists, Internet users who are concerned about unfettered Internet access

USER

Internet users self moderate content intake, with filters working on all web platforms

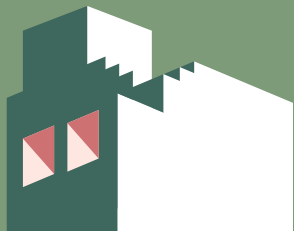
IMPACT

Social media platforms have better user experience, gaining satisfaction from users



Special thanks: Michael Maksimowski, In-Patient Psychiatrist at the VA





**“I AM VERY MUCH IN FAVOR OF
ANYTHING THAT GIVES THE
INDIVIDUAL USER MORE
CONTROL”**

— **David R. Karger, Ph.D.** @
MIT, Computer Science and
Artificial Intelligence
Laboratory

IMPACT

ScreenBook provides users a more
curated internet experience.



CUSTOM FILTERS



MACHINE LEARNING



CONSENT-BASED



EMPOWERMENT



SCREENBOOK

Your guide to traveling the Internet

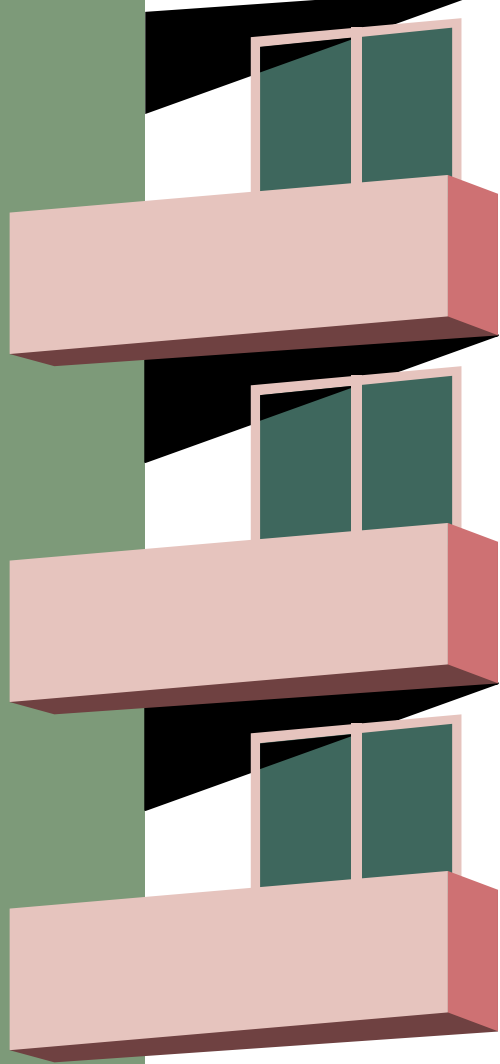
Niharika Sitomer, Shanie Hsieh,
Shirley Jiang, Tessa de Vries



01

TECHNICAL DISCUSSION

Dataset, model, human evaluation, pipeline
architecture, extension



THE DATASET

SOURCE

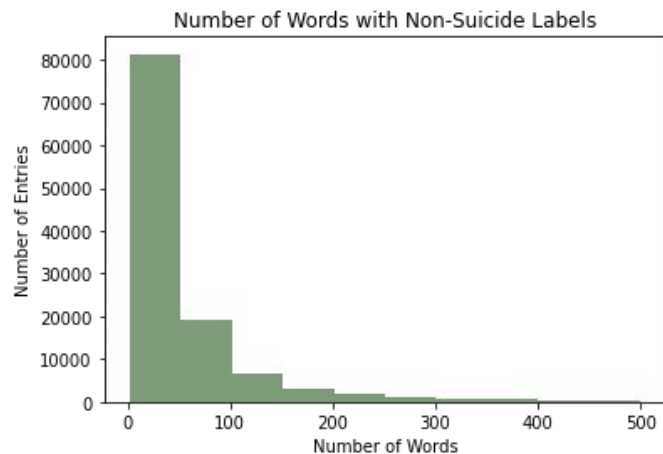
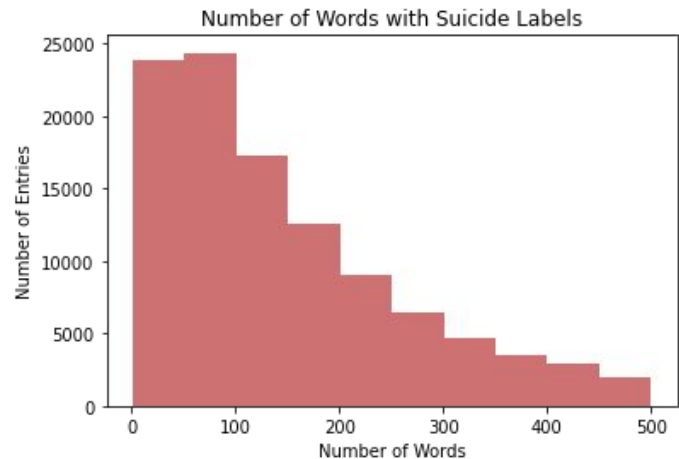
- Kaggle set: SuicideWatch and Teenagers subreddits

VOLUME

- Raw: 232,074
- After EDA and reduction: 222,045 train and 10,000 test

LABELS

- Binary: suicide, non-suicide



SAMPLE ENTRIES FROM TRAINING DATASET

LABEL: SUICIDE

“I have nothing to live for. My life is so bleak. I don't have any genuine friends, I feel like a stranger in my own family and I've scared off the only person I truly cared about. I have no future or goals to look forward to. My life is falling apart and I'm just letting it happen because I'm so, so exhausted. I just don't know if I want to be alive anymore.”

LABEL: NON-SUICIDE

“In less than 10 hours, I'm getting my wisdom teeth removed, and I'm so nervous I think I'm going to throw up I'm not a big fan of needles, and that's probably why I'm freaking out so much. If anyone has any stories of them getting their teeth out that could help me with the nerves, that'd be great :)”

OUR MODEL

AWS COMPREHEND CUSTOM MODEL

- Generated with preprocessed data and few set parameters
- Single class labels: suicide OR non suicide
- 1% of raw dataset removed in training
- Accuracy: 0.97; F1 Score: 0.97

Ex Wife Threatening SuicideRecently I left my wife for good because she has cheated on me twice and ...	suicide
Am I weird I don't get affected by compliments if it's coming from someone I know irl but I feel rea...	non-suicide
Finally 2020 is almost over... So I can never hear "2020 has been a bad year" ever again. I swear to...	non-suicide
i need helpjust help me im crying so hard	suicide

More training data examples

HUMAN EVALUATION RESULTS

MESSAGE CONTENT TYPE

MESSAGE LENGTH

	SUICIDAL MESSAGE	POSITIVE, NON-SUICIDAL	NEUTRAL, NON-SUICIDAL	NEGATIVE, NON-SUICIDAL
Short	3 out of 3	3 out of 3	2 out of 3	1 out of 3
Medium	2 out of 2	1 out of 2	2 out of 2	1 out of 2
Long	2 out of 2	0 out of 2	2 out of 2	1 out of 2

HUMAN EVALUATION ERROR EXAMPLES

Key takeaways: idiomatic phrases with harmful words & text with negative sentiment confuse model; more training data and model evaluation needed

FIGURATIVE EXPRESSIONS	STRONGLY NEGATIVE	POSITIVE, NON-SUICIDAL
"Work is killing me."	"This has been the worst week, I had a conflict at work with my boss and got a parking ticket. I want to escape and go on a nice vacation and get the hell out of here."	"I just graduated from college. I feel so proud and accomplished because I never thought I would be here. I can't wait to pursue my dreams."

AWS INTEGRATION ARCHITECTURE

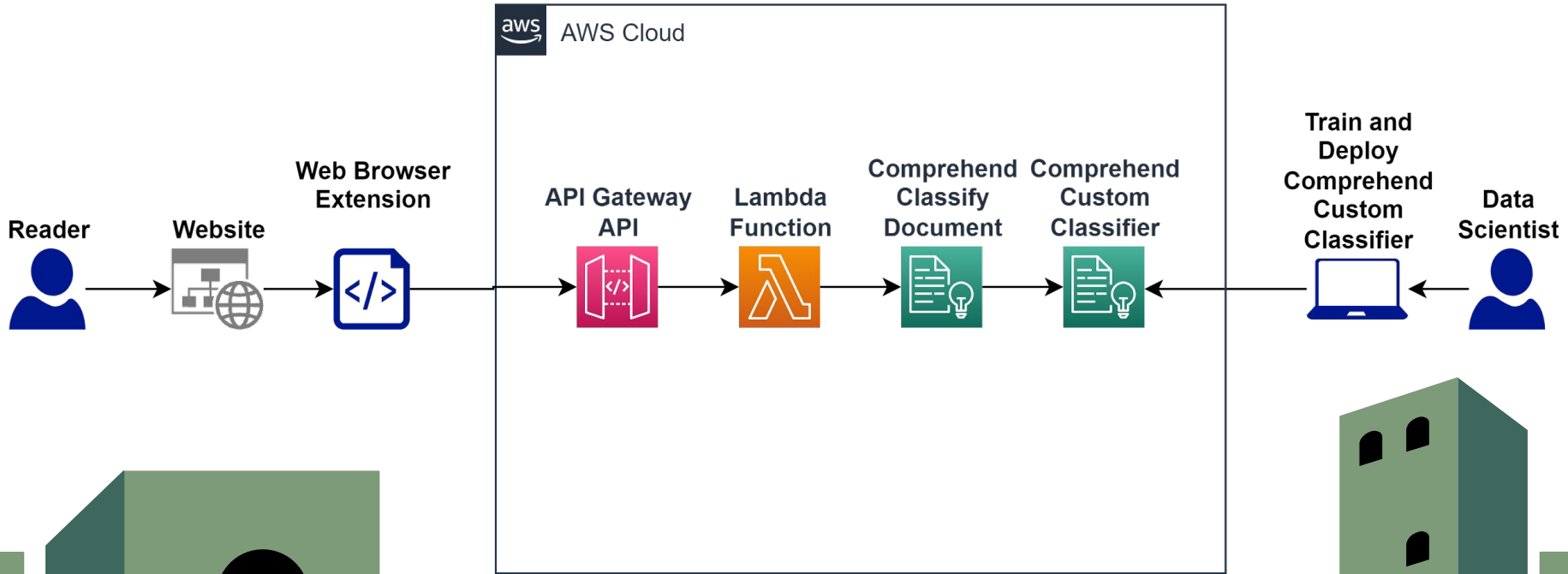
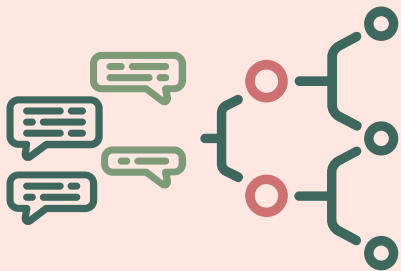


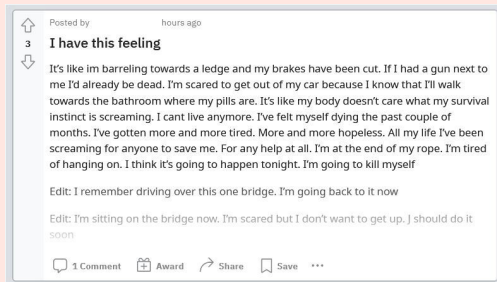
Image credit: Robert Wang, PhD.

EXTENSION BUILDING



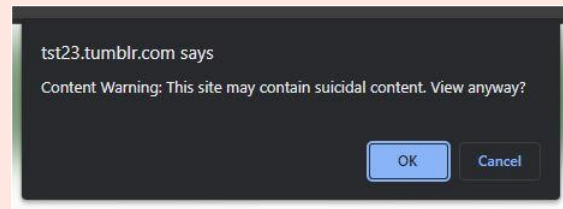
WEB-SCRAPING

Obtaining correct text to
feed into the model



BLURRING

Sufficiently hiding content
at the appropriate time



POP-UP

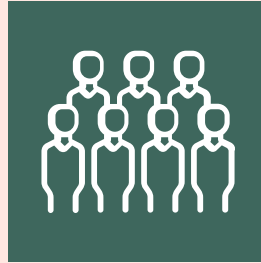
Giving users the option to
view content anyway

KEY TAKEAWAYS



EXPLORATORY DATA ANALYSIS

Proper EDA is important to form any model.



CONTACTS AND NETWORK

We successfully built our MVP thanks to help from our network.



MODEL AUDIT

Because of our black-box model, human evaluation is needed to properly validate our model.



02

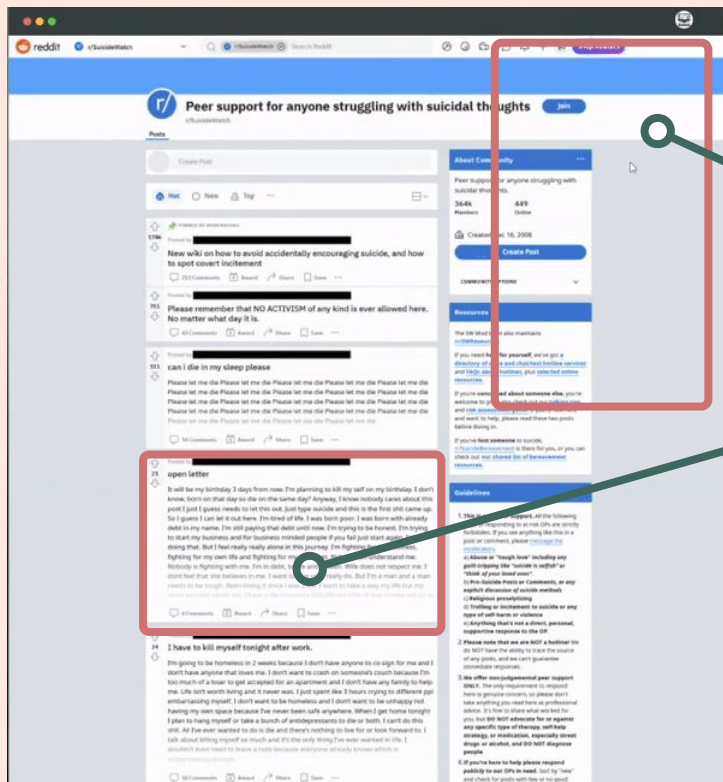
WHAT'S NEXT?

CURRENT EXTENSION

Our current MVP is the foundation of what we hope to achieve.



FUTURE PRODUCT DEVELOPMENT



1. ALTER SENSITIVITY

2. MORE FILTER OPTIONS

3. OVERLAY & BLUR PER POST

4. UPDATED MODEL

5. USER TESTING

ACKNOWLEDGEMENTS

TECHNICAL EXPERTS

Joyce Shen
Zona Kostic
Kevin Hartman
Danie Theron
Prabhu Narsina
Robert Wang
Steve Hsieh
Chase Maier

DOMAIN EXPERTS

Casey Fiesler
David Karger
Jennifer Puckett
Tom Thorley
Michael Maksimowski
Marine Leroux Thibault



SCREENBOOK

MISSION STATEMENT:

Our team aims to provide a service that allows users to navigate the enormity of the Internet safely and on their own terms.

projectscreenbook.wordpress.com

