

Using AI technology to prevent cyberbullying

June 13 2019, by Adam Zewe



Prabhu pitches ReThink on an episode of ABC's Shark Tank. Credit: Trisha Prabhu

On Sept. 9, 2013, after enduring years of horrific cyberbullying at the hands of her Lakeland, Fla., classmates, 12-year-old Rebecca Sedwick climbed to the top of a water tower and leapt to her death.

Hundreds of miles away, in Naperville, Ill., 13-year-old Trisha Prabhu read about Sedwick's suicide with a mixture of anguish and outrage.

"I thought to myself, how can a girl a year younger than I am be pushed to this point? I decided that I couldn't be a bystander anymore," said Prabhu, now a Harvard freshman. "Because cyberbullying is an issue that affects teenagers so much, I felt that a teenager should be the one to spearhead some of the next big work on how to tackle it. So without any authority or credentials, I decided I should be that person. I felt I owed it to Rebecca, and everyone else who had lost their lives or their loved ones to what is truly just a senseless thing."

Prabhu immersed herself in the latest psychology and sociology research on the rampant problem of cyberbullying. She was interested to learn that the prefrontal cortex, the part of the brain that controls rational decision-making, is not typically fully developed until a person is 25.

Yet unfettered access to social media, texting, and online communication channels gives teens the power to influence each other, often without basic understanding of concepts of digital citizenship, she said.

"I thought, what if I can force teens to really think about what they are doing on social media and empower them to make the right decision? That is how ReThink was born," she said. "At our crux is a very simple idea—if we give teens a second chance to think about sending offensive messages on social media, not only will they not send those messages, they will do it by a huge margin."

ReThink, an iPhone and Android app, installs a digital keyboard that replaces the standard smart phone keyboard. It appears identical and integrates seamlessly with other apps, like email clients and social media platforms. The only difference—it uses artificial intelligence to identify offensive words, and then deploys pop-up alerts to give the user a chance

to reconsider sending those words in an email, text, or social media post.

Since Prabhu launched the startup three years ago, ReThink has reached more than 2.5 million users around the world. ReThink recently won the McKinley Social Silver Medal in this year's i3 Innovation Challenge, sponsored by the Technology and Entrepreneurship Center at Harvard, and also claimed one of four, \$75,000 grand prizes in the University's President's Innovation Challenge.

Many of the biggest challenges Prabhu has faced are technical. Developing [artificial intelligence](#) that can effectively determine the sentiment of a potentially offensive message ("I hate you" vs. "I hate the weather," for instance) is a complex problem. It was especially difficult to ensure the app runs seamlessly while still incorporating semi-supervised machine-learning models that continuously improve accuracy.

Coming up with "trigger words" that are offensive in any context also proved challenging, especially when considering other languages (ReThink is available in English, Spanish and Hindi).

"This is a tough enough problem in English, which is my native language, because of slang and the unlimited linguistic possibilities to consider. It is a process that we will never finish because, as language evolves, we'll have to evolve, too," she said. "Getting into this space, I never really thought about language. But thinking about bullying and cyberbullying, it really does come down to how we communicate. Doing that on a global sphere has been an amazing journey, but a lot of tough work."

The ReThink team, which is now comprised of eight employees working remotely, employs linguistics experts to help tackle those challenges. Prabhu said ReThink's rapid growth, and the success of its business

model, are a testament to the effectiveness of the company's dynamic and diverse team.

While the app is free for users, ReThink earns revenue by partnering with schools, which pay an annual subscription fee for technology implementation and access to user data.

"Schools obviously face a huge liability when students are bringing these devices into the classroom," she said. "It can be really messy when tragic things happen and then a school district is embroiled in a \$20 million lawsuit. There is a need for schools to educate kids about how to responsibly use technology, because exposure and education aren't the same thing."

The company is also on track to release an in-app purchase for parents, offering more control over the app and access to data and analytics for a \$1 monthly fee.

The technology is becoming popular in many settings. The company partnered with the State of Michigan to deploy the app to all public school students; worked with the mayor of Chicago to incorporate it into the city's classrooms; partnered with the Los Angeles Police Department, who distribute the app in an effort to combat rampant 'sexting;' and recently joined forces with the U.S. State Department to bring ReThink to communities in 134 countries, Prabhu said.

ReThink also launched a chapter initiative, empowering students to become ambassadors in the fight against cyberbullying. The company provides materials, curricula, online bullying speakers for school events, and other resources to student-led groups across the nation.

"The internet can and will be fixed by this generation of young people. It was created by the generation before us, and they did a lot of good, but

there were also a lot of unintended consequences," she said. "We're trying to raise the next generation of young leaders through these chapters and help them foster these important conversations in their communities."

ReThink's meteoric growth is showing no signs of slowing down, and Prabhu intends to keep it that way. The company is set to release versions of the app in a dozen languages by the end of next year and is looking to partner with more international schools and expand into new, global markets.

The company is now turning its attention to image detection, studying how ReThink technology can be applied when a user is about to share an offensive image on [social media](#) sites like Instagram or Snapchat.

For Prabhu, the potential of ReThink is exhilarating, even if the road ahead will involve many challenges and plenty of long days and late nights.

"When you are caught up in the day-to-day business tasks, you can sometimes lose the 'why' behind the work that you are doing. The most rewarding part for me is when I get to come back to that 'why,' meeting parents who have lost children to this issue, or talking to other young people who have been cyberbullied," she said. "There are so many people who really suffer from this issue. It really torments them, and they often feel helpless. Knowing that I can help them feel empowered, that I can fight for their ability to be safe and respected online, that is the most rewarding part of the work. That is what pushes me to keep going."

Provided by Harvard University

Citation: Using AI technology to prevent cyberbullying (2019, June 13) retrieved 25 April 2024

from <https://phys.org/news/2019-06-ai-technology-cyberbullying.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.