
TECHNOLOGY AS A FORCE FOR GOOD? HOW ARTIFICIAL INTELLIGENCE IS BEING USED TO PREVENT SUICIDES IN CHINA

发挥技术善的力量? 人工智能应用于自杀预防的中国实践

Mark Pufpaff 濮漠泉

Abstract

Suicide is a significant problem in China. The causes are varied and interrelated. Teenagers and young adults have proven to be the most vulnerable. But there is hope, perhaps surprisingly, through the promise of technology. This article will discuss the Tree Hollow Rescue Movement (THRM), a non-profit organisation that is fighting suicide through the use of artificial intelligence (AI) algorithms and online messaging applications. The aim of THRM is to reach out to vulnerable persons to dissuade them from taking their lives. Their track record includes both success stories and incidents calling for caution, which will be introduced and discussed. THRM is an example of social innovation in view of the common good. However, the issue of suicide is nuanced and not without risks. The purpose of this case study will be to challenge readers to think through the issues involved and reflect more deeply about how technology can be a force for good.



SEEING THE ISSUE OF SUICIDE THROUGH THE LENS OF TECHNOLOGY

In China, suicide is the leading cause of death among persons aged 15-35. Females, aged

18-23, are at the greatest risk, the causes of their suicidal tendencies ranging from school bullying and academic pressure to relationship issues and debt bondage (Zhang et al., 2002). *Weibo*, China's most popular microblogging website, is where a lot of people-at-risk post their suicidal thoughts or behaviour. The practice of depressed persons posting their feelings online is common, both in China and abroad. In China, they post in so-called tree hollows, or *shudong*, online places of obscurity where a user can "whisper" the secret of their suffering as a way of lessening their psychological burden. *Sixth Tone*, an online newspaper, traced the origin of this practice to "Hong Kong, in movies such as 'In the Mood for Love' and '2046,' in which characters talk about admitting secrets to a hole in a tree" (Fu, 2019).

The fact that a lot of suicidal admissions are being posted online in China, has prompted the creation of a nonprofit called the Tree Hollow Rescue Movement (THRM). Their mission is to adopt artificial intelligence technology to identify suicidal intentions online, and then reach out to the flagged accounts to dissuade them from taking action.

[At present,] the nonprofit group consists of some 220 members, including experts such as psychiatrists, as well as many volunteers who want to help those with depression. It was started by Huang Zhisheng, an artificial intelligence professor at Vrije Universiteit Amsterdam. Wanting to find practical AI applications to benefit society, he set out to write an algorithm that could pick up on suicidal intentions. (Fu, 2019)

In an interview with *Sixth Tone*, Huang explained how the technology works:

The algorithm assigns each *Weibo* comment [posted on a tree hollow forum] a suicide risk level ranging from 1 to 10 based on word usage and produces a report consisting of a list of messages at or above level 6. 'Usually it finds six to 10

comments (per day),' Huang says, adding that the latest iteration of the software is so accurate that 82% of flagged comments are indeed about suicide plans. (Fu, 2019)

Once comments are flagged and confirmed as being about suicide, THRM's team of volunteers jump into action. They are dispersed across the globe, some located in China and others in European countries, so as to be able to respond in real-time, as necessary.

Reaching out to people-at-risk, the volunteers will usually use *Weibo's* private message function to open communication. Sometimes, if it is discerned that a person poses imminent harm to themselves, or if it is revealed in private conversation that they are really serious about taking their life and have explicit plans to do so, volunteers will contact relevant authorities to intervene. The rationale for such interventions is of course to save a life, which is noble; but it is also somewhat complicated, given that it can cause people-at-risk to become more reclusive, evading future detection about their intentions. It has also raised privacy concerns, given that users aren't necessarily posting their suicidal intentions as a cry for help.

In the area of privacy, THRM has been loosely compared to *Facebook* in the United States, which has an AI algorithm that will flag suicidal content and prompt people-at-risk to contact friends or support services; in cases where the danger of suicide is deemed imminent, *Facebook* will contact relevant authorities. However, Huang is of the opinion that the comparison is one of apples and oranges.

'*Facebook* is different,' Huang says. 'Firstly, they already have users' private information.' In contrast, THRM relies on what users have already made public or have directly told volunteers. Still, the decision on whether to send that private information to police to save a life can be complicated, says Shanghai-based psychological consultant and THRM

volunteer Zhou Zihan. (Fu, 2019)

Zhou, who has herself informed the police of people-at-risk, says the decision is not without risks.

Earlier this year, after the software had flagged a comment saying the user planned to kill herself after her birthday, Zhou kept a close watch. When, a day after the user's birthday, the user resigned from her job, Zhou thought police should be informed. [She requested that the] police check her situation rather than disturb her. Instead, the police disregarded the woman's privacy and informed all of her former colleagues about her depression. (Fu, 2019)

This response by the police was deemed by Zhou to be inappropriate. She later spoke to a local police officer about how best to approach persons at risk of suicide. The next time the woman in question was flagged as being in danger of taking her life, the local police were again contacted; however, this time they sent a female staff member to speak with the woman. According to Zhou, "She [the at-risk woman] decided not to go through with it [suicide] after the encounter with the police's female staffer."

This incident revealed two things. One, the police (as well as parents, teachers, and others, according to Huang) are not well informed about how to respond to suicidal situations. Two, THRM is limited in its helpfulness. While the organisation's software can help identify people-at-risk, there is a lack of treatment options and networks beyond the communication activity of THRM's volunteers. According to Huang, speaking about the magnitude of the challenge, "We discovered that, as it keeps rolling, the snowball gets bigger and bigger. We have a lot of things to do." His dream is eventually to build a "rescue ecosystem" to provide depressed persons with the services they need to prevent suicide and generate hope for the future.

JUDGING THE MORALITY OF THE TREE HOLLOW

What are we to make of this situation? There are three areas to comment on. One is THRM, the well-intentioned non-profit working to reduce suicides in China. Two, the police and other networks of support for persons entertaining suicide. Three, the issue of privacy and how we might understand THRM's activity in ethical terms.

The fact that a lot of suicidal admissions are being posted online in China, has prompted the creation of a nonprofit called the Tree Hollow Rescue Movement (THRM). Their mission is to adopt artificial intelligence technology to identify suicidal intentions online, and then reach out to the flagged accounts to dissuade them from taking action.

The Tree Hollow Rescue Movement—An example of social innovation?

Social innovation is a concept that characterises business activity aligned with the solving of social problems. The Stanford Graduate School of Business defines it as follows: "Social innovation is the process of developing and deploying effective solutions to challenging and often systemic social and environmental issues in support of social progress" (Center for Social Innovation, n.d.).

In China, suicide is one such social issue. And THRM has innovated an approach to addressing it, at least in part, through the use

of technology. Moreover, its founder Huang Zhisheng, as well as the organisation's volunteers, are exemplifying a particular set of values that both confirm THRM as socially innovative as well as align it with the aspirations of the common good.

First, technology. AI technology has been developing for decades (Smith and McGuire, 2006). And although it has proven controversial—for example, in the area of jobs automation—it has

of suffering or injustice, says something about their views on human dignity. That they feel an urge to help others speaks to a vision of the common good that is rooted in solidarity, in the idea that one person's good is somehow linked to the good of all others: that we cannot be indifferent to the sufferings of our neighbour, that we are all “in this together” and that no one is expendable. The activity of THRM could be described as fulfilling the Confucian virtue of benevolence (*ren*, 仁),

While a private conversation on Weibo with a trained professional may be viewed by the at-risk person as non-intrusive and even welcome, a visit from the police or another authority may not receive the same reception. Considering that the context is life/death, these questions warrant thoughtful reflection.

and continues to be leveraged for good. THRM is a good example of how AI technology can be put to the service of the common good. Using it to identify people-at-risk of suicide is, in a very real way, using it to save lives. That its algorithmic function can filter through publicly posted online content to flag worrisome/suicidal keywords is a lot like a smart searchlight scanning the dark depths of the internet for evidence of those who may be despairing offline.

Given the penetration rate of internet use in China—802 million users (59.6% of the population), with 98% being mobile users (McCarthy, 2019)—online forums and other public channels for communication and thought bubbling are fitting starting points for identifying those at-risk of suicide. That Huang saw this issue and acted to counter it through the application of AI, shows how social concern can be supported by technological solutions.

Second, values. What values are animating THRM's work? There are at least three:

1. Human dignity. That Huang and his team feel strongly that people shouldn't take their own lives, that life is worth living even in the midst

that is, loving others and empathising with their condition in life.

2. Sensitivity and discretion. The old saying, “the road to hell is paved with good intentions,” is a warning to those with such intentions that they should also be strategic, so as to increase the chances of a positive outcome. That THRM references only publicly shared posts to identify people-at-risk and contact them via private messages shows that they understand the importance of being sensitive and discrete in their outreach. Sensitive, because depressed persons thinking about suicide can be emotionally unstable and lacking in resiliency (Bowen et al., 2013); thus, they require an intervention that isn't going to judge them or make unhelpful assumptions about what they need to do. Discretion, because interventionists build trust with people-at-risk by giving such persons assurances of privacy when confiding in them (Percival et al., 2016).

3. Hope. The late Thomas Myers¹ was fond of saying, “You have to have a dream before a dream can come true.”² That Huang dreams of a “rescue ecosystem” developing in China around the issue of suicide echoes Myers’ quoted sentiment. Both are fundamentally hopeful that although the issue of suicide is going to be challenging to address, it is a problem worth working on. THRM, however, may only be the first step. As Huang himself admitted, the magnitude of the challenge is beyond his organisation. But even if Huang doesn’t see his dream come true, he will have started the process. That very well may be enough.

The police and governmental authorities – A skills and knowledge gap to be filled?

Huang is of the opinion that the police—as well as parents and teachers – are relatively uninformed about how to approach situations where a person is suicidal. Although less of an indictment and more of an observation, it reveals another layer in the issue. Is there a competent network of support/outreach for persons thinking about suicide? If an authority is called upon to intervene, will they do more harm than good? What resources are available to train people in how to interact with at-risk persons? The example above where Zhou described the decision of an authority to inform an at-risk woman’s colleagues about her depression after she quit instead of discretely inquiring as to the reason(s) for her leaving, is an example of why training in such matters is necessary. Indeed, when the woman was again showing signs of imminent suicide, the police were contacted and this time they sent a female officer who was able to relate to the at-risk woman and dissuade her from taking action.

1 Thomas Myers was a forensic accountant and expert witness in high-profile litigations in the United States. He was also an author, who wrote a book called *Cancer as an Opportunity*—which was reviewed before on in this issue of the MRI Journal—about how to manage challenging physical diagnoses (e.g. cancer) while maintaining a positive outlook. His work, let alone his approach to life, is especially relevant to the work of THRM as they reach out to China’s depressed and suicidal population.

2 The quote is from a song titled “Happy Talk,” from the Broadway play, *South Pacific*.

Privacy – How far is too far?

THRM raises privacy concerns because they are taking user content, interpreting it in a particular way (as evidence of suicidal intention), and then acting on it in the form of outreach. The volunteers performing such outreach are assuming, (1) that when a person posts content online, they are consenting to it being reacted to; (2) if such content is evidence of a suicidal intention, doing something is better than doing nothing; (3) if the threat of suicide is imminent, THRM volunteers are justified in contacting authorities to intervene in a given user’s life.

Is there a moral issue here? After all, it is true that THRM is only working with publicly posted information and is not collecting the personal details of the persons they are reaching out to. They also are not selling the information they receive from the persons they engage with, nor are they using it to push advertisements to them. Is not the work of THRM more akin to the behaviour of a concerned friend, than, say, a conventional business organisation? Is there a line to be drawn?

For example, is the contacting of authorities an overreach, given the increasing lack of control THRM then has over the interactions with at-risk persons that it sets in motion? Does it subject those contemplating suicide to what might be considered harassment as opposed to help? While a private conversation on *Weibo* with a trained professional may be viewed by the at-risk person as non-intrusive and even welcome, a visit from the police or another authority may not receive the same reception. Even if, as in the earlier example where a visit from a female police staffer resulted in the de-escalation of the at-risk woman’s desire for suicide, and the intervention produces a desirable outcome, is it always justified? Considering that the context is life/death, these questions warrant thoughtful reflection.

ACTING WITH PURPOSE

THRM is in a position to make a significant social impact over time. Quite literally, it is in a position to save lives. But it needs to both identify and manage the risks involved in working toward that goal. Below is a listing of recommendations.

Phase one - Preliminary due diligence

Before THRM staff reach out to at-risk persons, they would be wise to work directly with regulators and relevant government agencies. This is not only for THRM to receive their approval, and thus have confidence that their activity online is legally protected, but also to work with them to optimise better outcomes. Government agencies in China have resources, knowledge of social issues, and networks that can be mobilised strategically, if they perceive that THRM is aligned with their own objectives on the matter of suicide prevention (Zhang et al., 2002).

Phase two – During outreach

Between 16-25% of all those who die from suicide in China have attempted the act before. Moreover, most do not have pre-existing mental disorders and a high proportion of the attempts are impulsive acts deriving from personal crises. This is important information for THRM staff to know, as it should inform their approach. For example, if someone is recently undergoing duress from a relationship breakup, how is their thinking about suicide different from those who are long-sufferers of depression?

Although THRM is staffed with volunteers with a background in handling persons contemplating suicide, it is imperative that the organisation ensure these volunteers understand the ethics of consent. The desire to help cannot be used to excuse behaviour that may be interpreted as harassment or intrusion, however unintended. Not only can these be viewed as violations of human dignity, but they may very well expose THRM to legal recourse should the at-risk person

feel uncomfortable enough to report them. Given the somewhat casual or informal nature of THRM's work, where they are contacting persons they deem at-risk without any initiative on behalf of the at-risk party, consent is an area of ethical reflection that cannot go overlooked.

Phase three – Invoking the authorities

When a THRM volunteer is unable to dissuade an at-risk person by way of private conversation, and the volunteer feels that a suicide attempt is imminent, their practice is to request the intervention of an authority. This authority is usually the local police, whose responses have achieved mixed results. THRM had to learn the hard way that not all authorities are aware of the best practices for approaching and engaging with suicidal persons. This presents an important opportunity for THRM in the area of training and education. Data about which cities/provinces generate the most suicidal *Weibo* posts may help in identifying where THRM could target its training activities. A publication series could be developed for training purposes and tailored for distribution and promotion on mobile apps and throughout THRM's network of stakeholders.

Phase four – Partnerships for greater impact

Partnerships can and should be a strategy for THRM to increase their impact, as well as tap into the resources of organisations sympathetic to or even eager to get involved with their work. On the corporate side, there are tech giants like *Tencent* who have built apps in response to their users' desire to literally "live" within their social network. If the experience online is to reflect something of the experience of life offline, then creatively delivering access to suicide prevention and other-services is not only desirable, but necessary. Such access could be in the form of an emergency section of the app, providing contact numbers and/or the ability to reach out to trained professionals via a private channel. On the public side, something like a public-private partnership may prove fruitful, given that the Chinese government is clear in its intention to

address the problem of suicide (Phillips and Wei, 2016). Target organisations might include Lifeline Shanghai³ or the Beijing Suicide Research and Prevention Center⁴, which has been fielding calls and walking with suicidal persons since 2002 (Meng, 2018).

CONCLUSION

THRM is addressing a real and pressing need in China, namely, suicide. As a country with one of the world's highest suicides rates (Xie, 2007), THRM is providing China a much-needed service. Moreover, they are providing it in an innovative way, through artificial intelligence technology. As China continues to digitise, approaches to solving social issues that leverage technology will be increasingly important, especially because the red flags signalling issues like suicide are increasingly showing up online instead of offline. Who knew digital technology could be such a force for good?



MARK PUFFAFF, Project Director, Case Study Archive Hong Kong Rothlin International Management Consulting

³ Website: <https://www.lifeline-shanghai.com/>

⁴ Website: <http://en.crisis.org.cn/Home/Index>

REFERENCES 参考资料

- Bowen, R., Wang, Y., Balbuena, L., Houmphan, A. and Baetz, M. (2013). The relationship between mood instability and depression: Implications for studying and treating depression. *Medical Hypotheses*, 81(3), pp.459-462.
- Center for Social Innovation (n.d.). Defining Social Innovation. Stanford Graduate School of Business. Retrieved from: <https://www.gsb.stanford.edu/faculty-research/centers-initiatives/csi/defining-social-innovation>
- Fu, D. (2019). Keeping an Ear to Weibo's Suicidal Whispers. *Sixth Tone*. Retrieved from: <https://www.sixthtone.com/news/1004104/keeping-an-ear-to-weibos-suicidal-whispers>
- McCarthy, N. (2019). China Now Boasts More Than 800 Million Internet Users And 98% Of Them Are Mobile [Infographic]. *Forbes.com*. Retrieved from: <https://www.forbes.com/sites/niallmccarthy/2018/08/23/china-now-boasts-more-than-800-million-internet-users-and-98-of-them-are-mobile-infographic/#11cb76957092>
- Meng, M. (2018). I Take the Calls at a Chinese Suicide Prevention Hotline. [online] *Sixth Tone*. Retrieved from: <https://www.sixthtone.com/news/1002052/i-take-the-calls-at-a-chinese-suicide-prevention-hotline>
- Percival, J., Donovan, J., Kessler, D. and Turner, K. (2016). 'she believed in me'. What patients with depression value in their relationship with practitioners. A secondary analysis of multiple qualitative data sets. *Health Expectations*, 20(1), pp.85-97.
- Phillips, M. and Wei, X. (2016). Translated and annotated version of the 2015-2020 National Mental Health Work Plan of the People's Republic of China. *Shanghai Archives of Psychiatry*, 28(1), pp.4-17.
- Smith, C. and McGuire, B. (2006). *The History of Artificial Intelligence*. [ebook] University of Washington, pp.4-5. Retrieved from: <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf>
- Xie, C. (2007). China's suicide rate among world's highest. *Chinadaily.com.cn*. Retrieved from: http://www.chinadaily.com.cn/china/2007-09/11/content_6095710.htm
- Zhang, J., Jia, S., Wiczorek, W. and Jiang, C. (2002). An Overview of Suicide Research in China. *Archives of Suicide Research*, 6(2), pp.167-184.