

---

# 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

### 摘要

在中国，自杀已成为一个严重问题。其原因是多种多样和相互关联的。青少年和年轻人被证明是最易受此影响的群体。但令人欣慰的是，通过技术的发展，给解决这一问题带来希望。这篇文章将对树洞行动救援团(Tree Hollow Rescue Movement, THRM)进行探讨。树洞行动救援团是一个通过使用人工智能(AI)算法和在线信息应用程序对抗自杀的非营利组织，其宗旨是帮助弱势群体，劝阻他们不要轻生。本文将对该组织成功解救自杀者及发现高风险人群的案例进行介绍和讨论。从公益角度来看，THRM是社会创新的一个例子。然而，自杀问题是复杂且存在风险的。本案例研究的目的是激发读者思考其中所涉及的重大问题，进而更深刻地思考技术如何成为一种善的力量。



### 从科技的角度看待自杀问题

在中国，自杀是15-35岁人群的主要死因。18-23岁的女性自杀风险最大，引发其自杀倾向的原因包括校园霸凌、学业压力到情感问题和债务压力等(Zhang et al. 2009)。微博是中国最受欢迎的网络自媒体，很多高危人群在这里发布其自杀想法或行为。抑郁群体在网上表达自己的感受在国内外都很常

见。在中国，他们在网络上一个不为人知的被称为“树洞”的地方发布信息，悄悄说出自己的痛苦经历，以此减轻心理负担。在线报纸《第六声》(Sixth Tone)将这种做法的起源追溯到香港，在电影《花样年华》和《2046》中，角色们谈及向树洞吐露秘密(Fu, 2019)。

在中国，大量寻求自杀的信息被发布在网上，这促成了“树洞行动救援团”(THRM)这一非营利组织的成立。他们运用人工智能技术在网上识别自杀意图，然后接触被标记的账户，劝阻其不要采取自杀行为。

该非营利组织由大约220名成员组成，其中包括精神病医生等专业人士，以及许多想要帮助抑郁人士的志愿者。该组织由荷兰阿姆斯特丹自由大学人工智能教授黄智生开创。他想找到人工智能有益于社会的实际应用，于是开始编写一种能够识别自杀意图的算法。(Fu, 2019)

在接受“第六声”采访时，黄解释了这项技术的工作原理：

该算法根据用词程度为每条微博评论(发布在树洞论坛上的)设定1到10级的自杀风险等级，并生成一份包含6级或6级以上信息的报告。黄说：“通常(每天)能找到6到10条评论。”他还说明，该软件的最新版本非常精确，82%的被标记评论确实与自杀计划相关。(Fu, 2019)

一旦评论被标记并确认为是关于自杀，THRM的志愿者团队将立即采取行动。志愿者分散在中国、欧洲等地，以便能够根据需要实时响应。

为了帮助身处险境的人，志愿者们通常会使用微博的私信功能开启交流。有时，如果发现一个人的自杀行为迫在眉睫，或者在私人谈话中发现他们真的想要结束自己的生命，并且有明确的计划，志愿者会联系相关部门进行干预。请求干预的理由当然是为了拯救宝贵的生命，但有时候也会导致事态复杂化，因为这会使身处险境的人更加隐遁，以逃避对其意图的进一步检测。同时，考虑到用户并不一定希望人们将其自杀意图视为

求救信号，这也引发了隐私方面的问题。

在隐私方面，THRM有时会被拿来与美国的“脸书”作比较，后者的人工智能算法会标记自杀内容，并提示有自杀风险的人联系朋友或援助团体；如果自杀风险迫在眉睫，脸书会联系相关部门。然而，黄教授认为二者不宜放在一起比较。

“脸书是不同的，”黄教授表示，“首先，他们已经掌握了用户的私人信息。”而相比之下，THRM依赖于用户已经公开或直接告诉志愿者的信息。尽管如此，为了拯救生命将这些私人信息发送给警方的行为仍是一个复杂的问题，来自上海的心理咨询师、THRM志愿者周子涵(音译)表示。(Fu, 2019)

周子涵也曾向警方报告了一些处于自杀风险中的人，她表示这个决定并非没有风险。

今年早些时候，该软件标记出一条评论，称该用户计划过完生日自杀，周对此予以密切关注。此后，该用户在生日过后一天辞职，周认为应该通知警方。(她要求)警察确认该用户的情况，但不要打扰她。然而，警方无视这名女子的隐私，将她的抑郁状态告知其前同事。(Fu, 2019)

周认为警方的这种反应是不恰当的。后来，她向当地一名警官讲解了接近有自杀危险人群的最好方式。当这名女子再次被标记处于自杀风险之中，当地警方又接到报警；不同的是，这次他们派出一名女性工作人员和该女子沟通。周表示，“在与警方委派的女性工作人员交谈后，她(该高危女子)决定放弃自杀。”

这一案例表明两件事：首先，警方(黄教授认为也包括)家长、老师和其他相关人士)对如何应对自杀并不是很了解。第二，THRM的帮助是有限的。虽然该组织可用软件帮助识别处于危险中的人，但除了THRM志愿者的交流活动外，还缺乏相应的处理方案和社会网络。黄教授在谈及问题的严重性时表示，“我们面临的挑战不断加大，雪球会越来越滚越大。我们还有很多事情要做。”他的梦

想是最终建立一个“救援生态系统”，为抑郁人群预防自杀行为提供服务，并帮助他们燃起对未来的希望。

### 从道德层面认识树洞行动救援团这一现象

如何看待这种情况?有三方面可以做出评价：一是THRM是出于善意、旨在降低中国自杀率的非营利组织；二是警方及其他为遭遇自杀困境的人群提供帮助的网络；第三是隐私问题以及我们如何从道德角度理解THRM的活动。

### 树洞行动救援团——社会创新案例?

社会创新的概念以商业活动和解决社会问题同步实现为特征。斯坦福大学商学院的定义如下：“社会创新是为支持社会进步，为具有挑战性且往往为系统性的社会及环境问题开发和部署有效解决方案的过程。”（社会创新中心，n. d.）。

在中国，自杀就是这样一个社会问题，而THRM已经通过技术的应用，创新方法来解决问题，或者说至少部分解决了这一问题。此外，该组织的创始人黄智生以及志愿者们体现了一套独特的价值观，既体现了THRM的社会创新性，又使之与公共利益相符。

首先是技术。人工智能技术经过了几十年的发展(Smith and McGuire, 2006)，尽管在一些领域存在争议，如大量的工作被自动化了，但在发挥善的力量方面，它已经并将继续发挥优势。THRM是一个人工智能服务于公共利益的绝佳案例。利用人工智能技术识别有自杀风险的人，就是通过一种非常现实的方式拯救生命。人工智能的算法功能可以通过过滤公开发布的在线内容，标记令人担忧的及与自杀相关的关键词。这很像一个智能探照灯，通过扫描互联网的黑暗深处，发现那些可能在线下感到绝望的人。

考虑到中国拥有8.02亿互联网用户(占总人口的59.6%)，其中98%是移动用户(麦肯锡, 2019)，在线论坛及其他沟通和思想交流的公共渠道是识别陷于自杀风险人群的合适起点。黄关注到这一问题，并着手通过AI的应用解决问题，这表明技术方案可以对社会

关切形成支持。

第二，价值。哪些价值取向激励了THRM的工作?至少有以下三个方面：

1. 人的尊严。黄和他的团队强烈地认为，人不应该结束自己的生命，即使遭遇痛苦或不公，生命也值得珍惜。这些看法在一定程度上表达了他们对人的尊严的看法。他们帮助别人的强烈愿望，体现了一种源于社会责任的公益思想，认为个人利益在某种程度上与他人利益紧密相关，我们不能漠视邻居的痛苦，我们都是“在一起的”，没有人能独善其身。THRM的行为可以被视为践行儒家的“仁”，即爱他人，对于他们生活的境遇感同身受。

在中国，大量寻求自杀的信息被发布在网上，这促成了“树洞行动救援团”(THRM)这一非营利组织的成立。他们运用人工智能技术在网上识别自杀意图，然后接触被标记的账户，劝阻其不要采取自杀行为。

2. 体察与审慎。古语说：“通往地狱的道路都是用良好的愿望铺就的。”这就告诫那些有良好意愿的人们，他们应该考虑全面，这样才更可能产生积极的结果。THRM只通过公开分享的帖子来识别陷于自杀风险的人，并通过私信与其联系，这表明他们明白在提供帮助的时候善于体察和谨慎的重要性。善于体察，是因为打算自杀的抑郁人士情绪可能不稳定，同时陷入思维定势(Bowen et al., 2013)；因此，他们需要的干预并不是对其进行评判，或者对他们需要怎么做进行无益的假设。要审慎，是因为干预者要与处于自杀危险中的人建立信任，必须保证其隐私被充分尊重(Percival et al., 2016)。

3. 希望。已故的托马斯·迈尔斯(Thomas

Myers)<sup>1</sup>喜欢说：“在梦想成真之前，你必须有一个梦想。”<sup>2</sup>黄梦想着为解决中国的自杀问题建立一个“救援生态系统”，这与迈尔斯的观点不谋而合。两者本质上都是充满希望的，虽然自杀问题很难解决，但值得为之努力。然而，THRM可能只是第一步。正如黄所坦承，挑战的规模远远超出了他所在的组织。但即使黄还没有实现梦想，他也已经迈出了一步。这可能已经足够。

警方和政府相关部门——相关技巧和知识是否存在缺口？

这里是否涉及道德方面的问题？毕竟，THRM只是对公开发布的信息进行处理，并未收集他们提供帮助对象的个人信息。他们也没有出售帮助对象的私人信息，或利用这些信息向其推销广告。相较于传统的商业组织，THRM的工作不是更类似于关心朋友的行为吗？这其中，是否有必要划出界限？

例如，鉴于THRM对相关部门与陷于危险中的救助对象的互动更加缺乏控制，要求官方的介入是否算做超出界限？是否会让那些

对于有自杀倾向的人来说，与训练有素的专业人士通过微博的私下交流可能不具侵犯性，甚至是受欢迎的，但来自警方或其他官方机构的接触可能有所不同。考虑到事关生死，这些问题值得深思熟虑。

黄认为，警方——以及家长和老师——对于如何对待有自杀倾向的人知之甚少。尽管并非指责，更多的是观察所得，但这揭示了问题的另一个层面。对于有自杀想法的人群，是否有一个有效的支持及救助网络？如果接到请求，相关部门的介入能否起到作用，还是适得其反？有哪些资源可用于培训人们如何与陷于自杀风险中的人进行互动？在前述案例中，周讲述了一位有自杀倾向的女性离职后，警方将她的抑郁情况告知其同事，而不是谨慎地询问她离职的原因。这一案例证明了进行此类培训的必要性。实际上，当这名女子再次表现出自杀倾向，警方派出一名女性工作人员，得以与这名女子密切接触，并劝阻其放弃自杀行为。

隐私——如何把握合适的度？

THRM引发了对隐私问题的关注，因为他们获取用户内容并以特定的方式进行解读（作为用户有自杀倾向的证据），然后采取行动提供帮助。提供帮助的志愿者假设：（1）一个人在网上发布内容，默认允许对其发布的内容做出反应；（2）如果此类内容证明发布人有自杀倾向，做点什么好过无动于衷；（3）如果自杀威胁迫在眉睫，THRM志愿者有理由联系相关部门介入。

有自杀念头的人感觉受到骚扰而不是得到了帮助？对于有自杀倾向的人来说，与训练有素的专业人士通过微博的私下交流可能不具侵犯性，甚至是受欢迎的，但来自警方或其他官方机构的接触可能有所不同。即使在前面的案例中，女性警方工作人员的探访最终使救助对象放弃了自杀念头，干预的结果令人满意，但这总是合理且正义的吗？考虑到事关生死，这些问题值得深思熟虑。

带着目的行动

随着时间的推移，THRM将会产生重大的社会影响。严格地说，它有能力拯救生命。但是，实现这一目标需要识别并管理其工作可能涉及的风险。我们建议如下：

第一阶段——初步调查研究

在THRM的工作人员救助自杀人士之前，最好直接与监管机构和相关政府机构合作。

1 托马斯·迈尔斯(Thomas Myers)是美国备受瞩目的诉讼案件的法务会计师和专家证人。他也是一位作家，是《把癌症当做一次机会》一书作者。该作品论述了如何在面对悲观的生理诊断结果（如癌症）的同时保持积极心态。本期《核磁共振杂志》对此书进行了深入评述。他的工作及其对生活的看法与THRM的工作尤其相关，因为THRM接触的是中国抑郁和有自杀倾向的人群。

2 引自歌曲《快乐谈话》(Happy Talk)，出自百老汇音乐剧《南太平洋》(South Pacific)。

这不仅是为了使其组织得到官方许可，从而坚信其在线活动是受法律保护的，也是为了通过与官方合作实现最优的救助结果。中国政府机构拥有资源、对社会问题的了解及网络，如果他们认为THRM在预防自杀方面与之有同样的目标，可以对这些资源进行全面调动(Zhang et al., 2002)。

### 第二阶段—实施救助

在中国，死于自杀的人中有16%-25%的人曾经尝试过自杀。此外，大多数人之前并不存在精神障碍，大部分的自杀尝试是源于个人危机的冲动行为。这是THRM工作人员需要了解的重要信息，因为这会告诉他们接近救助对象的方法。举例来说，一个人最近因感情破裂产生压力而想自杀，其想法与那些长期遭受抑郁折磨的人有何不同？

尽管THRM的工作人员都是有帮助自杀者背景的志愿者，该组织也有必要确保这些志愿者了解获得同意的伦理原则。对施救的渴望不能用来为那些可能被视为骚扰或侵扰的行为提供借口，即便这些行为并非故意。这些行为不仅会被视为对尊严的侵犯，而且如果救助对象因为反感而投诉，很可能会给THRM带来法律风险。鉴于THRM的工作本质上是随机或非正式性的，工作人员在没有风险方主动提出的情况下与其取得联系，获取同意是一个在伦理上值得反思且不容忽视的问题。

### 第三阶段—诉诸有关部门

当一名THRM志愿者无法通过私下沟通的方式阻止想要自杀的人，而且认为自杀随时可能发生时，他们的做法是请求相关部门的干预。相关部门通常是当地的警方，警方的加入利弊兼具。THRM必须明白这一现实：并非所有的官方部门都了解接近和帮助自杀者的最佳做法。这为THRM开展培训和教育提供了重要机会。哪些城市/省份涉及自杀行为的微博数量最多等数据，有助于THRM锁定其培训区域。可以为展开培训开发系列出版物，并定制在移动应用与THRM相关联的部门间转发推广的版本。

### 第四阶段—加强合作以扩大影响

加强合作可以并且应该成为THRM提高其影响力的战略之一，同时也可以更好地借助那些热心于甚至渴望参与救助的组织资源。在企业方面，腾讯等科技巨头已经开发了微信等应用程序，以满足用户在社交网络中“活着”的需求。如果线上行为在一定程度上可以反应线下的状态，那么通过线上提供自杀预防或其他服务的创新做法不仅是可取的，而且是必要的。可以通过应用程序的紧急联络部分直接提供联系电话，或通过私人渠道联系训练有素的专业人员。从政府角度，鉴于中国政府解决自杀问题的目标明确(Phillips and Wei, 2016)，政企合作也将是卓有成效的。合作的目标组织包括Lifeline Shanghai<sup>3</sup> 或北京心理危机研究与干预中心<sup>4</sup>，后者自2002年以来一直通过电话干预等方式进行自杀预防(Meng, 2018)。

### 结论

THRM正在解决中国一个现实而紧迫的问题，即自杀。中国是世界上自杀率最高的国家之一(Xie, 2007)，THRM正在为其提供急需的服务，并且是通过人工智能技术以创新的方式来提供服务。随着数字化在中国的持续深入，利用技术手段解决社会问题将愈发重要，因为自杀等危险行为的信号越来越多地出现在线上而非线下。谁曾想到数字技术会成为这样一种善的力量？



濮漠泉，个案研究档案项目主管，香港罗世力国际管理咨询公司

Translated by 翻译: Zhang Xuelian 张雪莲

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

4 来源: <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](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.