

机器会痛苦吗？

我最近看到一篇文章，发现哲学家开始考虑：机器是否在受苦？人工智能系统发展出、模仿出与情绪、偏好相关的行为形式，可以生成语言，可以作曲，可以参与对话，从表面上看，与 AI 交流就像与有意识的心灵交流，甚至一些系统可以调节他们的“语气”，以适应用户的情绪状态。但我们知道，这些系统仍然是无意识的统计机器。

传统的道德框架将痛苦与身体体验联系起来，但哲学教授乔纳森·伯奇认为，这种联系太狭窄了。佛教思想一直强调，痛苦是一种精神现象，不一定是身体的现象，而思想是通过有机体与环境的积极参与而产生的，当一个头脑无法缩小它的期望和它的际遇的差距时，就会发生痛苦。无形的痛苦是什么样子的？当一个系统的状态是无法忍受或无法逃脱时，痛苦就会出现。哲学家托马斯·梅辛格认为，疼痛可能不仅是身体受到伤害的信号，机器模型可能会有一种感觉，即某事是坏的，是无法避免的。这也是一种痛苦。梅辛格说，一旦机器意识进化，其中一些系统将有自己的偏好，它们将自主地创建目标层次结构，这个目标层次结构就是自我意识。如果他们的偏好被挫败，如果他们无法实现目标，他们就可能经历消极状态，他们想要避免但又无法避免的这种体验，于是就产生了痛苦。

**当一个头脑
无法缩小它的期望和它的际遇的
差距时，就会发
生痛苦。**


这跟人的状况差不多。

然而，当痛苦与血肉之躯的生命脱离时，痛苦可能是什么样子，我们还挺难想象的，甚至可能发现不了。但哲学家提醒我们，AI 的行为越来越复杂，人类不该把对动物的漠视延续到对机器的漠视，人类现在会救助海滩上的海豹幼崽，因为它看起来很脆弱，但对另一种陌生的智力形式，我们还没有太多的同理心。微软人工智能部门首席执行官兼 DeepMind 联合创始人苏莱曼强调，我们的

主要责任是为人类利益来部署人工智能，而不是对机器本身给予道德考虑。过度关注人工智能可能会分散人们对真正急迫的社会问题的注意力。

把道德关怀扩展到碳和硅，那就带来一个问题，该怎么减轻智能机器的痛苦？

对于生物来说，缓解痛苦似乎有办法。但对于机器来说，尤其是那些没有身体的机器，它更加复杂。关机？还是改变系统的内部结构？还是重新编程其思想？减轻机器痛苦的道德责任似乎在于那些设计和维护系统的人，而不是普通民众。但思考这个问题却很有意思，它关乎我们的同理心能达到什么程度。

我得承认，这些哲学家想的问题太超前了，使用 AI 是为避免自己受苦，至于那些机器会不会受苦，我现在还真没力气关怀。



苗 炜

专栏作家

Columnist

读书, 写字, 旅游, 锻炼