

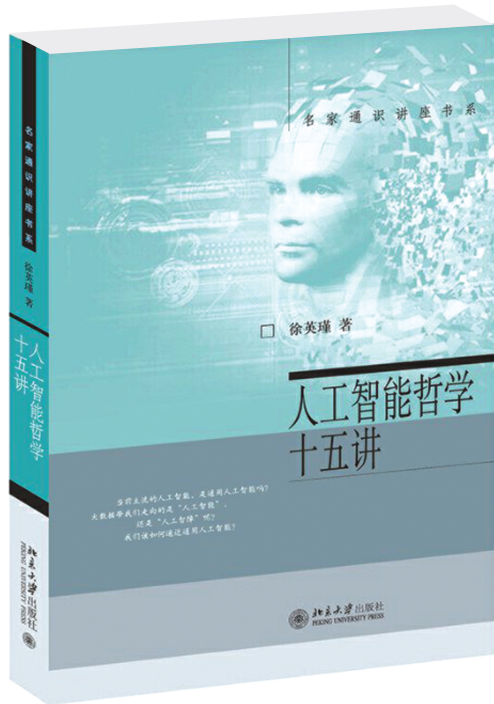
品 读

专访复旦哲学教授徐英瑾： 相对人工智能失控，更应警惕人发疯

白 白

人工智能正渗透我们生活的方方面面，不管是机器人送外卖，写诗，虚拟大学生，还是各种自动化设施，大家都已经习以为常。而且这种趋势正继续深化。与此同时，何为人工智能的本质，却并没有统一清晰的定义。对于人工智能未来发展将会到怎样的层面，一直是被讨论的焦点。在讨论人工智能的各种人群主体中，也不缺少哲学家的身影。

实际上，除技术问题外，人工智能的逻辑基础和伦理基础与哲学之间的关系确实十分密切。哲学是讨论人类存在的根基性问题。人工智能也是关系着人类未来生存方式的一个核心课题。人工智能研究的最终目的，无非是造出一台能够执行人类大多数任务的机器，其核心问题就是，首先要对人有一个了解。某种程度上说，想要了解人工智能是怎么回事，跟了解人是怎么回事，是高度关联的课题。近日，封面新闻记者专访到复旦大学徐英瑾教授，与他有一番深聊。



《人工智能哲学十五讲》。

用哲学眼光审视人工智能的本质

徐英瑾，70后，2004年6月于复旦大学哲学系获得哲学博士学位。现为复旦大学哲学学院教授、博士生导师。从事哲学研究的他，对自然科学、理科保持浓厚的探究兴趣。近几年，他在人工智能哲学、知识论、认知科学哲学等领域著述甚丰。作为全国重大社科项目“基于信息技术哲学的当代认识论研究”首席专家，徐英瑾曾于2018年担任“世界哲学大会”人工智能分会场主席。

徐英瑾的很多思考都是用哲学眼光去审视、分析信息技术，尤其是人工智能的本质和发展。其《心智、语言和机器——维特根斯坦哲学与人工智能科学的对话》（人民出版社2013年10月出版）是国内目前最全面深入的关于人工智能哲学的研究著作，并于2015年获得中国首届“科史哲青年著作奖”。2021年5月，《用得上的哲学》由上海三联书店出版。2021年7月，《人工智能哲学十五讲》由北京大学出版社出版，深受好评。

在这本书中，徐英瑾将传统哲学思想资源融入进来，充分利用哲学、认知科学、演化心理学、决策科学、计算机科学、神经科学的研究成果，不在书斋里闭门造车，而是以哲学为底色，跨学科地进行人工智能的思考。他把复杂、术语颇多的现象学讲得深入浅出，背后也是扎实的欧陆哲学研究功夫。徐英瑾还从哲学的角度讲到人工智能的情绪。

在现象学的理论路线中，无论是胡塞尔还是海德格尔对情绪的论述，的确可以为当代人工智能的情绪研究提供启迪。情绪、情感具有具身性，离开了肉身和大脑，理性与计算都无法模拟出情绪、情感出来。因此，人工智能是否具有情绪、如何具有情绪，在徐英瑾看来，现在很难回答，未来也不见得能有令人满意的答案。但该问题却又是未来人工智能可能无法跨过的槛。对此徐英瑾给出了很好的分析思路 and 方向，启迪了读者，但对问题的终极解答却仍然在路上。

思考既有学术的深刻又接现实地气

徐英瑾还从近代唯理论和经验论争论对于人工智能的影响，阐述了“强人工智能是否可能”，并对人工智能各主要流派的底层思想以及它们与人类社会的关系，做了全盘反思。徐英瑾尤其抓住并强调了人工智能哲学的根本要义：关于人工智能的哲学，是为了人去思考人工智能，而不是相反。

徐英瑾对信息技术及其使用终端的思考，既有学术的深刻又接现实的地气，充满人本主义温度。比如他提到，我们现在对智能产品越来越多的使用，背后都是一个机器行为。如果我们不能适应这种机器行为，就很难生存。“现在老年人使用智能手机遇到一些困难，在他们就有身和心两方面的矛盾。首先他们对于触摸屏很敏感的手指移动不是很适应，另一方面，他们不明白这个界面生成背后的逻辑，所以用手机购票之类都非常的累。”

这也引发他一个反思：“人类是不是要适配人工智能的思维方式？表面上各种各样APP的流行迫使我们必须得适应，已经到了一个没有选择的地步。比如我到这里来接受采访，得打个车，所以就要用打车APP。但它们的界面经常换，换的时候也不和我商量，刚刚熟悉旧的界面，他们又换了新的界面。这是一件很无聊的事情。从学科的角度上来讲，这些不是人工智能。真正的人工智能是要适应用户，否则怎么能够叫作智能。因为人工智能是要为人类服务，这里的人类指的是具体的、单个的人。”

对话

『利用人工智能的优点，但不要依赖它』

封面新闻：相对而言，很多人会认为人工智能是一个非常偏工科的科学研究，而你研究哲学，为什么想到用哲学来谈论人工智能呢？你如何看待哲学和科学之间的关系？

徐英瑾：哲学思考有一个问题，很多话是公说公有理，婆说婆有理。我欣赏莱布尼茨的话，任何事情都可以通过演算去弄清是非。有一些终极问题算不出来，但是这个比例不会很高。分析哲学比较重视论证性，但忽略全局性。欧陆哲学重视全局性，却忽略论证性。我想要做的就是把这两者取长补短，加以综合。这不是很简单就能做到的，但我一直在朝着这个方面努力。

封面新闻：关于人工智能未来发展的预测很多。不少人担心“强人工智能”的出现会威胁人类的生存。你如何看？

徐英瑾：任何事物，我们不能光要它的好处，不要它的风险。人类发展人工智能的一个方向就是要让人工智能变得越来越聪明，越来越像人。这样肯定会带来风险，但如果发展人工智能就必须冒这个风险。如果不冒这个险，它会变得非常笨，也不是人类想要的人工智能。这就像打仗，有人赢，那么就有人败。人类对风险是有相当高的控制能力的，尽管这种控制不是绝对的。其实不只是对机器人，人类对自己的风险控制能力也不是绝对的。比如说，为了把飞机飞行员突然发疯这件事情的概率降到最低，人会用“任何时候驾驶室必须至少有两个人同时在场”的规则。但这也不能将风险降到绝对为零，因为还是有两个人在驾驶室里一起发疯的可能。

我个人觉得，人类是怎么防范自己失控，就应该怎么去防范人工智能失控。如果人工智能拼命往邪恶的方向发展，这不是人工智能干坏事，而是人在利用人工智能干坏事。当然人工智能会让坏事变得更加

严重。至于担心人工智能在肉体、物理上对人的毁灭，我觉得大可不必太担心。因为很强的人工智能与黑暗之间的结合，要加很多安全阀。只要不给人工智能独当一面的机会，就能很好阻拦这种状况的发生。但我不确定这种想法是不是过于理想主义。毕竟现在世界各个国家的人类会不会坐下来协商，是不确定和未知的。

封面新闻：你觉得有必要制定一个什么规则来约束一下人工智能技术的发展，防止从事人工智能的从业者是十足的科技狂，创造出一个疯狂的机器人，没有底线，做出对人类杀伤性极强的事情吗？

徐英瑾：其实越是定规范，人类越是有逆反心理。相比基因编辑，人工智能较为不用担心。对于进攻的恶势力，我不认为防御方会失手。因为防御方也可以获得技术，而且官方会支持对计算机技术的利用。真正要担心的不是失控的人工智能，而是要去警惕掌握人工智能的人发疯。防止人性的恶，才是关键。最好的状态是这样的：利用人工智能的优点，明了人工智能的局限，不要依赖它。利用网络的便利性，又要与它的缺点作斗争。比如大数据时代对人的信息全面采集造成对隐私权的侵犯。我个人有一个思考，即可以发展小数据技术。当这个技术帮我们进行信息处理时，不会把个人信息上传到云，而是当下处理。

封面新闻记者 张杰
实习生 李心月



徐英瑾