解放军报/2020 年/9 月/29 日/第 007 版
军事论坛

人工智能须警惕数据陷阱

刘鹏

兵以诈立。智能时代，颠覆性技术不断涌现，战争欺骗手段和形式亦不断出新。“ 如果掌握你的数据，我就能创造出各种方法欺骗你的人工智能系统。” 研究实验表明，智能化战争中一旦一方获得对手的人工智能训练数据集，就能够找到其弱点和盲区并实施欺骗，人工智能必须警惕数据陷阱。

目前来看，人工智能分析处理数据的速度远超人类分析师，并且能够找出人脑难以发现的行为模式和规律，但是也会犯下人脑不会犯的错误。原因在于，机器学习算法必须依靠大量数据进行训练，数据之于人工智能就如同血液之于人类，共享数据比设计算法更难。如果数据集过小、数据不准或是被对手恶意篡改，那么机器学习效果就会大打折扣，甚至被误导出现误判。尤其在国家安全和军事领域，有害数据会造成严重后果。一旦人工智能的训练数据集被对手掌握，对手就会设计数据陷阱、实施欺骗，提供假数据并诱导人工智能学习错误数据。更严重的是，由于机器学习算法的内在工作机理晦涩难懂，人们通常并不清楚人工智能为何会出错，特别是在没有发生灾难性后果的情况下，甚至难以察觉人工智能出错，对人工智能陷入数据陷阱茫然不知。

那么，应如何避开数据陷阱呢？首先，需要人脑干预。只有人具备给数据分类打标签的能力，因此不能简单地把数据丢给机器算法，寄希望于人工智能解决所有问题而无须人脑干预。如果只提供大量数据而缺乏能够辨别数据的“ 聪明人脑 ”，那么人工智能只能提供机械的答案，而非人们需要的正确答案。人脑干预不仅能够确保人工智能获得正确的数据，还能够检查其是否在学习正确的数据。其次，打造跨领域团队。能够避开数据陷阱的“ 聪明人脑 ” 必须来自跨领域团队，计算机专家、程序员、大数据专家和人工智能专家必须与相关领域经验丰富的专业人员密切合作。今后，人工智能不断发展成熟后将可能直接为作战人员提供实时情报等，这就需要作战人员不断为“ 聪明人脑 ” 团队提供反馈，以便及时更新和修正数据。再次，进行多源数据互查。使用一种传感器侦察目标很容易被对手蒙蔽，因此要采用视觉、雷达和红外等多种传感器侦测同一目标，将不同来源的数据进行对比核验，才能够辨别真伪、发现隐藏的骗局。再者，给数据分类打标签。当前，即使高级的人工智能也会犯下荒诞的低级错误，甚至会错把牙刷认作棒球杆。因此不能给机器学习提供未经加工的原始数据，尤其在训练初期更是如此，应该为机器算法提供正确分类、打了标签的真实数据，方能检验人工智能的结论是否正确，确保人工智能辅助决策准确、高效。最后，采取对抗式学习。组建智能蓝军，研发人工智能对手，让互为对手、彼此对抗的人工智能展开互搏，在斗智过程中进行对抗式学习，在对抗式学习中提高识别数据陷阱的能力，实现以智取胜。总之，当前人工智能还离不开人脑控制，避免数据陷阱最终还要靠人的经验和智慧。