

误捕的责任终究在人

程盟超

就像某些美国非裔经历过的那样，今年 1 月，美国人威廉姆斯遭遇了一场“灾难”：在下班即将到家的路上，他莫名其妙地被底特律警方堵在路边。后者出示了重罪逮捕令，当着威廉姆斯妻儿的面，给其戴上了手铐。虽然事后证明，他压根儿没犯任何罪。

这次误捕起源于警方采用的 AI 人脸识别系统——它判定了威廉姆斯涉嫌犯罪。这也令威廉姆斯成为了全美第一个被人工智能错误定罪的受害者。

底特律警方利用人脸识别系统至今已有几十年的历史。判定威廉姆斯犯罪的人工智能软件不仅运用在底特律，还在全美多个城市发挥作用。被捕当晚的审讯室里，警探们告诉威廉姆斯：1 年多前，本市一家高档精品店中总价值近 2 万美元的名表失窃。人脸识别系统分析监控录像判定，威廉姆斯应该是作案者。

历经 30 小时的拘留后，威廉姆斯缴纳了 1000 美元保释金得以回家。他向律师求助，然而绝大多数人都认为 AI 不可能误判，他就是罪犯。有律师愿意出庭辩护，但要收费 7000 美元。直到威廉姆斯的妻子在社交媒体上发文，引发舆论轩然大波，底特律警方才在调查后发现：案发当日威廉姆斯有着充分的不在场证据，还真是人工智能搞错了。

“在人工智能越发流行的当下，我们可能正不知不觉成为糟糕技术的受害者。”事后，一位公益组织的发言人评述，“尤其会身陷种族歧视等系统性问题。”

2016 年，在史上第一次以 AI 作为评委的国际选美大赛中，几十名获奖者几乎全是白人；2017 年，亚马逊发现已使用 3 年的一款人工智能招聘软件竟会自动给女性简历降低评分。

类似情况一直延续到现在。2019 年，几家美国警方的人脸识别系统供应商便被发现相似问题：他们的系统在识别非裔和亚裔时，错误率是面对白种人的 10 至 100 倍；麻省理工学院的一项研究则表明，微软、IBM 等大企业的性别识别系统对白人的准确率高达 99%，可对黑人女性时却骤降至 35%。

人工智能之所以“智能”，核心是其机器学习的能力。机器学习又依赖于两个环节——海量的数据输入和众多工程师们搭建起的复杂算法。这就像一个学生进入学校，离不开课本和教师。

“教科书”，也就是数据库便可能为人工智能带来误差与偏见。比如全网数据库中，白人的人像要比其他人种更多，教出来的人工智能辨识起白人的脸便比其他种人更加准确。

有大学的研究人员发现，摄影技术在被发明后的 100 多年来，始终在为浅色皮肤的使用者们进行着优化。点滴积累的结果是，如今的照相术对白人很友好，却不太适宜展示深色人种间的样貌差异。当如此海量的数据积累到数据库，人工智能便因此产生了巨大误差。

还有一些偏差无法由上述原因解释，那往往是“教师”，也就是算法程序员们的责任。2016 年，美国媒体发现，被警方广泛适用，用于预测罪犯出狱后再次犯罪的人工智能系统非常不准确——在预测黑人再犯的数据上，该系统给出的预测值远高于实际风险；可当预测白种人时，其数值却又显著低于实际。

“一个人在夜晚丢了钥匙，他一定会先看路灯下，只因那里更容易下手。”麻省理工学院的一位教授就此举例：当制作预测罪犯的人工智能系统时，程序员会下意识地根据自己的立场和认识设计需要计算的要素，比如“是否接受过高等教育”“是否有固定工作”“亲属中是否有罪犯”，甚至“居住在哪个街区”。介于大部分程序员都是接受过良好教育、家境不差的浅色人种，他们理所当然会想到这些。但这些真的是关乎犯罪概率的因素吗？可能至多是大片黑暗中被路灯照到

的那一小块儿，体现的是这些发明者们特定的立场，或者说偏见。

当然，同样的工作交给人类，也并不代表就没了错误与偏见。有数据称，印度的基层法院积压着 2700 万件未能得到审理的官司，人工智能是解决它们的唯一希望；微软也宣称通过技术调整，仅在一年之内，他们便将算法识别女性和有色人种的错误率降低了 20 倍。人工智能会越发强大，它是未来的希望，这也是很多唯技术论者常有的论调：

“ AI 判断的准确率是 99.9%，人类的上限则是 99%，即使被发现后有一半申诉挽回的可能，0.5%的错误率依旧高于 0.1%。”

但问题在于，纵使如此，还是有很多人依旧更愿意相信人类——这似乎刻在我们的基因里，一如 App 里的智能客服再好用，我们还是喜欢给人工客服打电话。因为我们坚信同类之间可以沟通，他们会被我们打动、说服，让事情变得更好；而完成算法的机器，很难和我们相互理解，这会让傲慢与偏见被彻底固化。

再看看威廉姆斯被逮捕的那遭冤案，除了人工智能本身的错误，也正是傲慢与偏见的堆积才铸成了最后的乌龙。那张人脸识别系统生成的报告里清楚地写明：“ 该文件不是明确证据，仅是调查线索，不能因此逮捕人员。” 根据规定，警方必须结合上其他直接证据才能开展抓捕。但在底特律，警方并未遵守这点。

纵使在人工智能时代，犯错的可能是机器，但责任终究在人。