机器识别眼镜。训练完了以后，这个模型对戴眼镜的人能够做正确的识别，但是让另外一个人戴了同样的眼镜，他就会被识别成之前的那个人。训练的时候，模型里面留了一个后门，这就会是安全隐患。”

毫无疑问，这些干扰、后门技术如果被滥用，就有可能会形成一些新的安全隐患。最著名的就是轰动一时的换脸软件“zao”。一天之内爆火，一日之后下架。实际上，基于 AI 的深度合成技术已经可以综合运用人脸替换、人脸再现、人脸合成、语音合成等技术，实现更加复杂的视频合成。2018 年出现的一种新的 AI 算法，只需要一张照片，就可以让一个不会跳舞的人变成灵魂舞者。此外，3D 合成尤其是虚拟人正在成为下一个阶段的技术发展重点。围绕生成虚假的人脸或者人身，还可以建立虚假的社交账户，让他和很多真实的人建立关联关系，甚至形成一些自动对话。看起来好像是一个真实人的账号，实际上却完全是虚拟生成的。从积极的一面看，深度合成技术推动了社交、游戏、影视、电商等领域沉浸式体验的进一步发展；但从另外一个方面来看，这些应用中可能也会涉及个人隐私的保护、版权的争议以及道德伦理方面的巨大挑战，它们也有可能被一些不法之徒用于“伪造”或者“欺骗”。一些虚假的视频，尤其是虚假的人的讲话，比如说模仿领导人讲话，有可能对社会稳定甚至国家安全造成威胁。

魔高一尺道高一丈。与图像生成技术相辅相成的就是如何辨别图像的真伪。这些新技术的出现，需要更厉害的技术与之对抗，才能保证 AI 朝着正确的方向运行，而不至于成为一匹不受人类控制的脱缰野马。

对于一些参数比较明确的模型，也即“白盒场景”，只需要“对症下药”，还是比较容易校正的。现实情况中，大多是比较困难的“黑盒场景”，就是不知道这个模型的算法逻辑是什么，例如，对用在自动驾驶汽车中的神经网络来说，行人和路牌之间存在什么样的差异？神经网络的哪个决策阶段能够发现两者的区别？只有了解这一过程，才能够更好地理解模型为何会做出错误的预测，从而设法纠正神经网络的一些错误。

上图：专家建议，在立法和监管方面，应给予适度宽松的发展空间，给 AI 应用提供安全港，通过试验、测试、试点等方式加速 AI 从研发到商业落地的转变。

摄影 / 陈梦泽

如果算法出错，应该由谁买单？无人驾驶汽车领域显然无法逃避这一问题，但除此之外还有其他诸多领域也同样如此。当无人驾驶汽车发生伤人事故，责任应该由谁承担？是坐在后座的乘客吗？是汽车的制造商吗？还是驾驶汽车的 AI 程序？ AI 程序的主体又是谁？是缔造它的程序员，还是拥有它的公司？毫无疑问，AI 的广泛应用，不仅对技术人员提出了更高的挑战，也带来许多现实的、亟待解决的法律问题，只有通过法律、技术、行业、用户的多重治理，让 AI 向善，进入可控发展轨道，才能逐渐从 deepfake（用人脸识别技术做换脸）、deepnude（用算法“脱”衣服）等色情性换脸视频的阴影中走出来，迎来