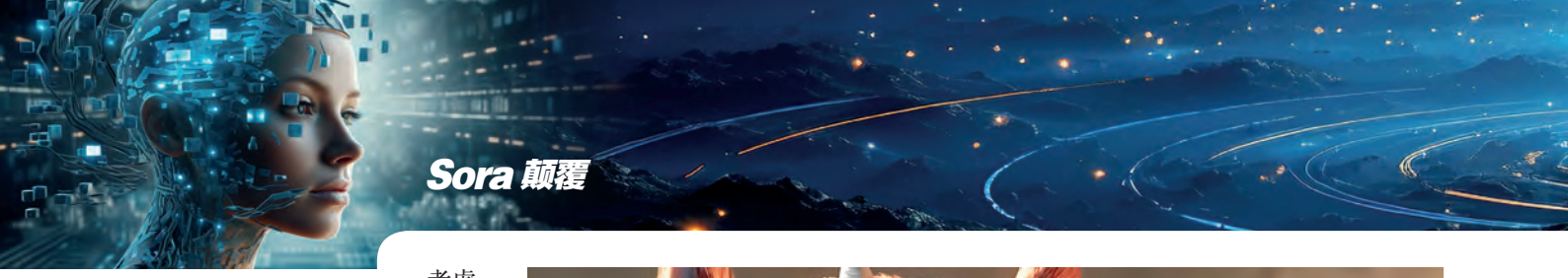

考虑到与海洋相比，杯子的体积较小，因此采用了倾斜移位摄影技术，营造出一种微景观的感觉；第六，虽然提示词中场景的语义并不存在于现实世界之中，但 Sora 依然实现了我们所期望的正确物理规则。

这就是 Sora 的独到之处，理解运动中的物理世界。复旦大学教授、上海市数据科学重点实验室主任肖仰华指出，因为世界本质上是非常复杂的，非线性的。我们传统的模型只能建一些线性的简单关系。像流体力学之类非常复杂的现象，用传统的模型非常难建模。但是今天我们看到基于 Transformer 深度神经网络的大模型架构，Sora 已经具备了对现实世界复杂现象非常逼真的建模能力，这是 Sora 带来的一个新高度。

Sora 视频完整展现了小怪兽伸出爪子挡住红色蜡烛跳动的火焰，它的影子随之偏移的物理过程。

在 Sora 推出后不久，OpenAI 发布了这款新工具的技术报告。报告指出 Sora 的一个强大的能力是它的语言理解能力。OpenAI 利用 Dall-E 模型的 re-captioning（重述要点）技术，生成视觉训练数据的描述性字幕，不仅能提高文本的准确性，还能提升视频的整体质量。此外，与 DALL·E 3 类似，OpenAI 还利用 GPT 技术将简短的用户提示转换为更长的详细转译，并将其发送到视频模型。这使 Sora 能够精确地按照用户提示生成高质量的视频。

因为一篇张冠李戴的文章而被误称为 Sora 发明者之一的纽约大学计算机科学助理教授谢赛宁，实际上是机器学习领域知名学者，也是扩散模型（Diffusion Transformer，简称 DiT）一篇重要论文的主要作者之一。他分析 Sora 应该也是一个建立在 DiT 架构上的扩散模型，同时结合了 GPT 技术的混合模型，从而在视觉领域实现重大突破。“对于 Sora 这样的大规模系统工程而言，神经网络架构只是其中很小一部分。大部分的功劳要归功于 OpenAI 的人才储备，高质量数据规模以及巨大的算力。”

简而言之，60 秒超长长度、单视频多角度镜头和世界模型是 Sora 的三大关键词。如果没有大语言模型的加持，Sora 是不可能迅速“进化”到今天这个地步的。

Sora 能否理解世界？

毫无疑问，Sora 目前展现出来的“逻辑能力”看似非常强大，或者说它展现出来的视频世界更符合人类观念中的现实世界。

但 Sora 真的能够理解世界吗？随着一系列匪夷所思的 Sora 视频出现，业界有了截然不同的判断。

比如在一个样片中，提示词为“考古学家在沙漠中发现了一把普通的塑料椅子，正小心翼翼地挖掘和除尘”，Sora 生成的视频出现了椅子变形、自动行走等诡异的场景。

另一个玻璃杯碎裂的视频中，玻璃杯碎裂的方式也十分诡异——它被抬到半空中时，桌子上就忽然出现了一摊平整的红色玻璃，随后玻璃杯被摔到桌子上，和这摊玻璃融为一体。

很显然，Sora 混淆了玻璃破碎和液体溢出的顺序，也并不能推理时间和因果关系。这不正说明，Sora 目前还无法理解全部的物理世界？

再比如，Sora 团队 Aditya Ramesh 自豪地放出一个蚂蚁巢穴内爬行的视频，粗看似乎很惊艳，仔细一看，却令人啼笑皆非——蚂蚁怎么只有四条腿？！

还有一个老奶奶庆祝生日的视频，每一帧都异常逼真，但是当老奶奶吹了生日蜡烛的时候，蜡烛的火苗竟然纹丝不动。最离谱的还是一个男人在跑步机上煞有介事地反向跑步。如此“南辕北辙的跑步”视频，让人看到了 Sora“智障”的一面，