01

AI

上篇

揭秘 AI 文生视频工具

第1章

AI文生视频：智能创作汹涌而来

随着 AI 机构 OpenAI 发布 Sora 大模型，AI 文生视频领域获得用户广泛关注。对于这个全新领域，许多用户期待其能够带来更多的创作可能，推动智能创作不断发展。

1.1 认知：走近AI文生视频

从 AI 文生图到 AI 文生视频，AI 生成领域取得了一大进步。在 AI 生成领域崛起的当下，许多用户对于 AI 文生视频软件不甚了解。下文将会从文生图到文生视频的变迁、AI 文生视频的能力和技术支撑等方面进行介绍，帮助用户走近 AI 文生视频。

1.1.1 升级：从文生图到文生视频

短短几年内，AI生成领域获得了很大发展，多种AI生成应用相继诞生，从文生文、文生图到文生视频，行业热潮不断涌现。AI 生成领域从文字升级到视频是大势所趋，许多大型科技企业已经入场。

字节跳动推出了文生视频大模型 PixelDance；阿里推出了名为 Animate Anyone 的模型；百度则推出了文心大模型，并配备了丰富的功能。种种迹象表明，从 AI 文生图到 AI 文生视频，AI 与文本、视频等的结合已经成为全新风潮，而各类企业如此积极地布局，其意义也不言而喻。

（1）文生视频应用范围广，发展潜力大。随着各类短视频的爆火，短视频生产能力逐渐开始跟不上需求。AI 文生视频软件的出现能够弥补这一缺陷，满足用户对短视频的需求。例如，在游戏、影视等行业，AI 文生视频软件能

够根据用户输入的文字生成对应的故事情节，实现短视频制作的降本增效。

（2）AI 文生视频软件操作相对便利。用户制作精良的视频需要高超的技术、高昂的成本，AI 文生视频软件能够降低视频生成的难度，用户仅需输入文字便可生成相应的素材，便于进行视频制作。

（3）AI 文生视频软件能够增强企业竞争力。在 AIGC 领域，AI 文生图应用有许多，但是由于 AI 文生视频应用开发难度高，能够推出该软件的企业屈指可数。高难度往往伴随着高价值，如果企业能够推出优质的 AI 文生视频软件，便可以增强自身竞争力，拥有差异化的优势。

随着 AI 行业和视频行业的发展，AI 视频行业也会得到发展，AI 文生视频将会迎来爆发期，吸引越来越多的企业和个人用户参与创作。虽然目前来看，国内文生视频技术研发和应用进展缓慢，还没有一款出众的产品出现，但是许多人才、有实力的企业已经出现，未来必将获得更快的发展。

1.1.2　AI文生视频究竟有多“牛”

伴随着 Sora 的发布，AI 文生视频领域吸引了众多目光。许多用户发出了疑问：AI 文生视频究竟有多“牛”？ AI 文生视频指的是用户输入文字，AI 便可以根据文字生成细节丰富、十分逼真的视频。AI 文生视频能够应用于多个场景，如图 1–1 所示。

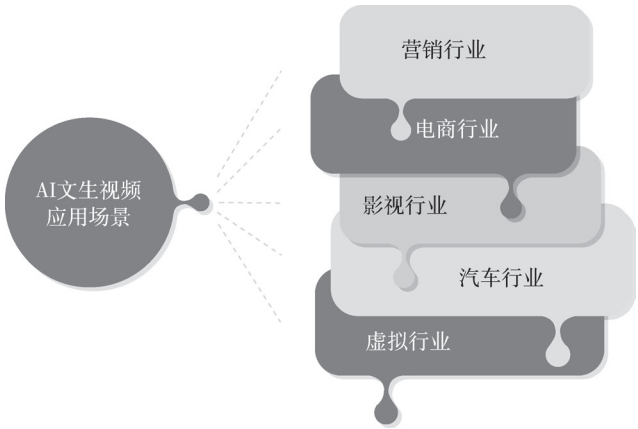

图1–1　AI文生视频应用场景

（1）营销行业。AI 文生视频技术能够将文字转化为视频，助力企业营销。企业可以利用文生视频技术创作吸引人的广告、产品演示视频或教育性内容，以吸引潜在客户。

（2）电商行业。商家可以利用 AI 文生视频技术生成商品展示视频，视频比图片更直观，能够帮助用户更好地了解产品。

（3）影视行业。AI 文生视频技术能够应用于影视行业，自动生成角色、特效等，有效简化影视制作流程。

（4）汽车行业。AI 文生视频技术能够对场景进行深刻理解，为自动驾驶汽车的场景感知和决策提供支持，实现自动驾驶的优化。

（5）虚拟行业。AI 文生视频技术能够生成高度逼真的虚拟数字人。这些数字人能够应用于多个领域，为用户带来更加优质的体验。

除了能够应用于丰富的场景，AI 文生视频的“牛”还体现在其背后的技术。AI 文生视频技术拥有强大的计算资源和算法支持。企业利用深度学习算法对大量视频数据进行训练，使算法能够从这些丰富的视频数据中提取视觉信息，挖掘动态规律。

经过漫长的训练，算法能够学会生成连贯的视频帧、模拟物理世界的场景等。算法在经过多重优化和参数调整后，能够生成逼真的场景。

无论在技术方面还是应用场景方面，AI 文生视频都为用户带来了许多惊喜，使用户看到了 AI 的巨大潜力，并对 AI 未来发展充满期待。

1.1.3 技术支撑：AI文生视频背后的技术

AI 文生视频离不开技术的支持。AI 文生视频的工作原理是利用模型进行生成学习和提示学习，从而对用户的文字产生准确的理解并生成相应的内容。企业会为模型提供大量训练数据，包括图像数据集和视频数据集。AI 模型往往使用深度学习技术，包括 CNN（Convolutional Neural Network，卷积神经网络）、RNN（Recurrent Neural Network，循环神经网络）等对训练数据进行学习。在训练完模型后，用户输入文本便可以获得视频。

AI 文生视频模型训练方法众多，比较常用的有自监督学习、半监督学习和无监督学习，企业可以根据模型的具体应用场景选择相应的训练方法。以下是 AI 文生视频技术经历的几个发展阶段，如图 1–2 所示。

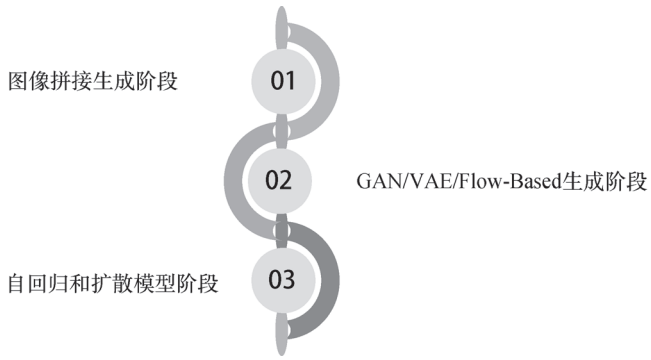

图1–2　AI文生视频技术经历的发展阶段

（1）图像拼接生成阶段。图像拼接指的是将许多张具有重叠部分的图像组合成一个整体，其代表了 3D 场景的一部分。拼接是场景重建中的一类特殊情况，其中的图像能够通过平面单应性进行关联。图像拼接技术能够应用于运动检测和跟踪、提高分辨率、压缩视频等方面，在机器视觉领域发挥了很大作用。

（2）GAN/VAE/Flow–Based 生成阶段。伴随着机器学习技术不断发展，GAN（Generative Adversarial Networks，生成对抗网络）、VAE（Variational Auto–Encoder，变分自编码器）以及 Flow–Based（基于流）逐渐得到广泛应用。这些技术的发展主要聚焦于优化模型训练与生成策略，以实现更为高效、精确的视频生成。

鉴于直接对视频进行建模的复杂性较高，部分模型通过分解前景与背景、运动与内容等方法生成视频，或基于图像翻译优化生成效果，旨在增强连续帧间过渡的流畅性。然而，在总体表现方面，所生成视频的质量仍不能尽如人意，难以投入实际应用。

（3）自回归和扩散模型阶段。近年来，Transformer 和 Stable Diffusion 模型在语言生成与图像生成领域成功应用，使得基于自回归模型和扩散模型

的视频生成架构逐渐成为主流。自回归模型能够根据前述帧预测下一帧，从而使得视频呈现较为连贯自然的效果。然而，该方法在生成效率方面存在局限，且错误易于累积。

扩散模型能够利用添加噪声和反向降噪技术实现图像生成。该模型通过构建马尔科夫链来定义扩散过程，并逐步向数据引入随机噪声，直至获得近似高斯噪声数据。

接下来，模型对逆扩散过程进行学习，利用反向降噪推断方法生成图像，并通过系统对数据分布进行扰动并恢复，以实现整个过程的优化。

1.1.4 扔一部小说出一部大片是真的吗

背靠业内知名 AI 机构 OpenAI，Sora 拥有行业领先的视频生成技术，能够生成画质清晰、镜头复杂、人物众多且时长达 60 秒的视频。据 OpenAI 透露，Sora 不仅能够理解用户输入的文字，还能够理解文字中的事物在物理世界中是如何存在的。虽然目前 Sora 生成的视频时长有限，但其出现无疑表明了一种趋势：在不远的将来，扔进一部小说便能生成一部大片将成为现实。

除了生成的画面更符合物理世界的规律，Sora 还能够使视频画面的主题始终保持一致，即便画面主体离开视野，画面也不会发生变化。基于这样的技术，如果能够将多个 1 分钟的视频连接起来，那么扔一部小说出一部大片就能成为现实。

在用户惊叹于 Sora 的强大能力的同时，也出现了一些担忧的声音。例如，Sora 能够一键生成高质量视频，那么影视行业从业者该何去何从？

清华大学教授沈阳曾经表示，Sora 将会对影视行业产生巨大影响，其能够降低视频制作成本，对许多影视行业从业者造成冲击。而 360 董事长周鸿祎则认为，AI 并不会快速颠覆所有行业，反而可能激发更多从业者的创造力。

总之，Sora 的出现颠覆了 AI 领域，也为用户的生活带来了更多便利。虽然它的出现会对一些行业造成冲击，但更大的可能性是推动行业重新洗牌、持续进步。

1.2 发展格局：机遇与挑战并存

尽管 AI 文生模型有诸多优点，展现出巨大的发展潜力，许多企业想投资入局，但是其隐藏的风险也是不容忽视的，需要企业仔细辨别。企业需要规避侵权风险，迎接智能创作时代。

1.2.1 侵权风险引发担忧

AI 生成的视频引发了内容创作领域的变革，但也带来了隐性的版权及知识产权难题。2024 年 4 月，备受关注的“人工智能文生视频侵权首案”经北京互联网法院审核立案。陈某作为原告，指控对方侵犯其著作权，原因是对方未经许可，擅自剽窃其运用 AI 技术创作出的文生视频作品，并以原创名义发表。

此案引发了社会各界对 AI 生成视频侵权风险的广泛关注。随着 AI 技术的飞速发展，如何判断和处理生成式 AI 作品的著作权问题，对 AI 产业乃至整个社会创作生态的发展具有重要影响。参考近年来全球相关判例可知，判断 AI 生成的内容是否具备人类独创性，成为确定此类内容是否享有著作权的关键因素。完全由人工智能自主生成且缺乏人类独创性的内容，通常被认为不享有著作权。

2023 年，美国法院作出裁决，确定画作《离天堂最近的入口》不受著作权法保护。法庭判定该作品全然由 AI 创作而成，并未包含人类创造性劳动，故而无法享有著作权。法官指出，现行著作权法面临新的挑战，对于人类在 AI 创作过程中的角色，有必要予以进一步明确。

尽管 Sora 尚未正式商用，但部分业内专家已关注到其潜在的安全漏洞及遭受滥用的可能性。OpenAI 坦诚指出，Sora 在生成虚假及有害内容方面存在风险。

这也是当前好莱坞难以直接采纳 Sora 技术的一个重要因素。将 AI 生成的素材用于影视创作或引发版权纠纷，即便 AI 仅用于训练相关素材，也可能蕴藏潜在的侵权风险。Sora 作为创作辅助工具所生成视频的版权是否受到

保护，以及训练素材是否承担侵权责任等问题，皆可能使制片方在担忧卷入法律纠纷的情况下持有审慎立场。

尽管 Sora 等具备 AI 生成能力的产品在诸多方面尚需完善，但其对 AI 领域的贡献不容忽视。Sora 的诞生激发了众多 AI 从业者和科技企业的高度热情，大家共同期待并推动新时代到来。

1.2.2 各方如何迎接智能时代

AI 生成视频技术的出现使得生成式 AI 领域获得快速发展，无论企业还是个人，都应当积极调整自身，迎接这一科技变革。

企业能够在各类大模型的帮助下完成多种通用业务，提高业务效率和自身价值。

随着云服务的完善、算力的提升以及 ChatGPT、Sora 等大模型的出现，中小型企业能够在节约基础设施建设和模型训练费用的情况下，使用优质的生成式 AI。随着 AI 技术的不断发展，AI 技术的竞争优势可能会下降，企业更应该关注生成式 AI 的应用场景，最大限度地发挥其业务价值，提高自身竞争力。企业可以利用生成式 AI，从以下几方面提升自身能力，改变竞争格局，如图 1–3 所示。

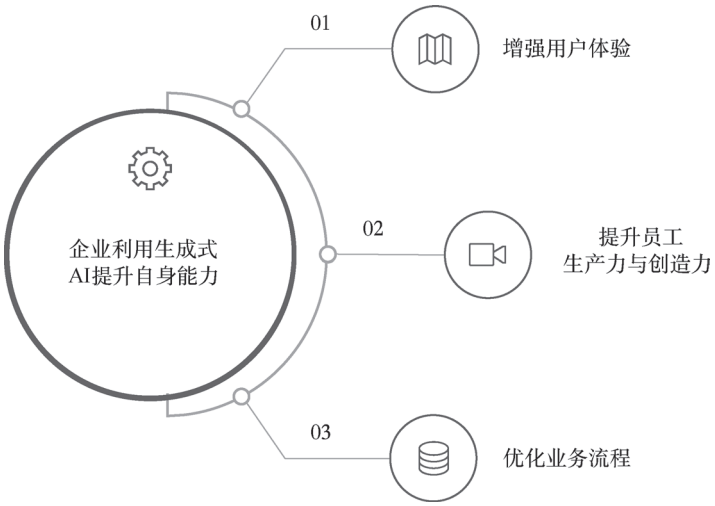

图1–3　企业利用生成式AI提升自身能力的3个方面

（1）增强用户体验。企业能够利用生成式 AI 的内容生成能力增强用户体验。例如，客户关系管理软件企业 Salesforce 推出了一个以 AI 为基础的 CRM（Customer Relationship Management，客户关系管理）软件——Einstein GPT。Einstein GPT 能够辅助员工工作：帮助销售人员生成个性化的销售邮件来维系用户；帮助服务人员生成合适的回复信息，快速回答用户问题等；帮助营销人员策划营销活动，吸引更多用户。

（2）提升员工生产力与创造力。企业的知识往往利用非结构化方式储存，利用新兴技术能够更好地存储数据和发挥数据价值。

例如，科技企业西门子与亚马逊云科技展开合作，利用 AI 技术打造智能数据库，发挥数据价值。亚马逊云科技在深入分析西门子的需求后，提出了智能知识库暨智能会话机器人解决方案，以最大限度发挥知识库的功能。

在亚马逊云科技的帮助下，智能会话机器人“小禹”和智能知识库诞生。与其他机器人相比，小禹回答问题的速度和精准度有所提升。智能知识库上线首周，便有超过 4 000 名用户使用，许多问题获得了解答。在生成式 AI 的帮助下，员工能够减少许多基础工作，将精力放在提升自身生产力和创造力上，为用户提供更好的服务。

（3）优化业务流程。众多企业在 BI（Business Intelligence，商业智能）领域投入大量资源，却难以利用数据驱动的洞见来指导业务决策。随着大模型和生成式 AI 的涌现，BI 逐渐实现真正的智能化。在先进技术助力下，任何用户都能通过自然语言轻松提出与数据相关的问题，从而更加高效地进行数据分析与决策。

例如，亚马逊通过 Amazon Bedrock 平台，将生成式 AI 技术整合至商业智能工具 Amazon QuickSight，从而使业务分析师能够通过自然语言提问的方式进行数据分析。在此基础上，Amazon QuickSight 能够生成完善的数据报表，执行复杂的统计计算，从而大幅提升公司对业务数据的洞察能力。

总之，AI 的迅猛发展正引领我们步入一个崭新的时代。在这个时代背景下，企业应抓住机遇，积极提升内部运营效率，以实现持续发展。