

上海科学技术情报研究所
上海市前沿技术发展研究中心
技术与创新支持中心(TISC)



专利与创新

PATENT AND INNOVATION

2024 年

第 10

文生视频专利和商标动态

编者按

一年多前，OpenAI 推出的 ChatGPT 火爆全网。而一年多后的 2024 年 2 月 16 日，OpenAI 宣布推出文生视频大模型“Sora”，再次震惊科技界。Sora 的横空出世再次将人工智能的发展推向新的高潮。随着 Sora 的发布，Sora 概念股爆火，产业界、资本圈纷纷拥抱生成式 AI。根据市场预计，OpenAI 的市值也将有望从 ChatGPT 发布后不久的 290 亿美元飙升至超过 800 亿美元。在文生视频的赛道上，目前以 OpenAI 一马当先，除 OpenAI 外，国外角逐者还有 Runway、Pika、谷歌、Meta 等企业，而我国清华大学、腾讯、科大讯飞等机构也紧随其后，相继在文生视频领域取得了一系列突破。本简报中，笔者从专利和商标视角切入，梳理了近期国内外主要机构关于文生视频的知识产权动态，便于读者从知识产权角度了解文生视频竞争态势。

目 录

文生视频专利动态.....	4
文生视频专利布局已达上千件.....	4
清华大学：多主体对应的定制化文本的文生视频.....	4
腾讯：基于历史信息进行文本改写提升视频质量及生成效率....	5
科大讯飞：自动生成符合用户个性化需求的 2D 虚拟人视频.....	6
百度：视频虚拟形象可以执行多个动作.....	7
谷歌：多技术协同生成压缩视频、沉浸式视频等.....	8
OpenAI：专利公开较少，涉及文生语、自动执行动作等技术..	9
文生视频商标动态.....	11
“SORA”国内商标已被注册多达百余条.....	11
“SORA”在美商标处于待审查阶段.....	11
“GPT”商标被 USPTO 最终驳回.....	12

文生视频专利动态

文生视频专利布局已达上千件

近日，AI(人工智能)新锐巨头 OpenAI 推出文生视频模型 Sora，掀起了全球范围内的热烈讨论。

由于在 AI 生成视频的时长上成功突破到一分钟，再加上演示视频的高度逼真和高质量，Sora 立刻引起了轰动，并引发外界的种种猜测：能在发布 ChatGPT 不到两年后就推出 Sora，OpenAI 是否“藏了一手”？表现惊艳的 Sora 是否已经具有世界模型的特点？

这些都还需要在实际业务运用中去进一步验证。

商商查 APP 显示，和视频生成相关的已授权专利有超过 1000 个。

资料来源：商商查.2024-02-21

清华大学：多主体对应的定制化文本的文生视频

近日，OpenAI 发布人工智能模型 Sora，其能够通过文本提示生成逼真且富有想象力的视频，“文生视频”引发广泛讨论。

天眼查 App 显示，近日，清华大学申请的专利 CN117496020A“一种定制化多主体文生视频方法、装置、设备及介质”专利公布。该专利详情如下：

摘要：本申请提供一种定制化多主体文生视频方法、装置、设备及介质，涉及神经网络技术领域，包括：获取多个主体分别对应的主体文本表述以及主体图像；基于多个主体分别对应的主体文本表述以及主体图像，获取混合文本以及组合图像；将混合文本以及组合图像输入文生视频模型，生成第二噪声预测值，并基于第二噪声预测值与组合图像，获取第二损失和第三损失；基于第一损失、第二损失与第三损失，对文生视频模型进行优化，得到优化的文生视频模型。本申请通过多种损失对文生视频模型的参数进行优化，使优化的模型基于文本描述生成视频中的图像时，文本描述与定制化主体保持一致，且在每个主体在生成过程中的特征不会发生混淆的同时消除合成痕迹。

技术背景：受到文本到图像生成的成功推动，文生视频任务最近受到越来越多的关注。文生视频旨在生成符合给定文本提示的时间连贯的具有语义的视频，随着大规模多模态数据集的出现，使得现有的文本到视频生成模型可以根据给

定的文本提示生成时间连贯且逼真的视频。

然而，现有的定制化文本到视频生成工作主要集中在单一主体上，也就是说，现有的文生视频方法只能生成对单一主体的定制化文本的生成对应的视频，当任务需要生成多个主体的定制化文本在同一画面的视频时，现有的方法就难以做到。因此，如何实现对多个主体对应的定制化文本的文生视频任务，成为本领域当前亟待解决的问题。

发明目的或效果：本申请实施例在于提供一种定制化多主体文生视频方法、装置、设备及介质，旨在解决如何实现对多个主体对应的定制化文本的文生视频任务的问题。

此外，度小满科技（北京）有限公司也于 2023 年 8 月申请“文生视频生成方法、装置、电子设备及可读存储介质”专利，该申请已于同年 11 月公布，其能够以极低的资源实现任意长度长视频的生成。

领和讯 Plus 会员，免费看更多独家内容：8 大财经栏目，最新最热资讯干货独家行情解读，快人一步掌握市场投资风向。

资料来源：财经号. 2024-02-19，CN117496020A（公开日：2024-02-

02)

腾讯：基于历史信息进行文本改写进而提升视频质量及生成效率

金融界 2024 年 1 月 27 日消息，据国家知识产权局公告，腾讯科技（深圳）有限公司申请一项名为“视频生成方法、装置、电子设备及存储介质”，公开号 CN117453952A，申请日期为 2023 年 11 月。

专利摘要显示：本申请涉及互联网技术领域，尤其涉及一种视频生成方法、装置、电子设备及存储介质，所述方法包括：获取目标对象在内容发布平台的历史发布信息；在所述历史发布信息包括历史发布文本以及历史发布图像的情况下，提取所述历史发布图像的图像描述信息，以及对所述图像描述信息以及所述历史发布文本进行信息融合，得到融合文本；基于目标内容改写模型对所述融合文本进行内容改写，得到目标体裁的第一改写文本；基于所述第一改写文本以及所述历史发布图像进行视频生成处理，得到与所述历史发布信息对应的第一发布视频。本申请能够进一步发挥原历史发布信息的传播价值，并且提高生成视频的质量以及效率。

背景技术：随着互联网的快速发展，社交网络卷入了海量的对象，这些社交网络存在各种海量的对象形成了海量对象关系，对象在这些平台上不仅仅进行点对点，点对群的社交交流，同时还在平台上分享和沉淀了大量图片和文本内容，这些图片和文本内容相对分散，对象对这些图片以及文本内容的消费门槛相对较高，从而难以使得这些图片以及文本内容发挥其更大的传播价值。

发明目的或效果：本申请所要解决的技术问题在于，提供一种视频生成方法、装置、电子设备及存储介质，能够对内容发布平台上的历史发布信息进行二次重构，生成相应的发布视频，进一步发挥原历史发布信息的传播价值；并且通过微调大语言模型对历史发布信息进行文本改写，能够快速且准确地对历史发布信息进行文本改写，进而提高生成视频的质量以及效率。

资料来源：搜狐。2024-01-27，CN117453952A（公开日：2024-01-26）

科大讯飞：自动生成符合用户个性化需求的 2D 虚拟人视频

金融界 2024 年 2 月 20 日消息，据国家知识产权局公告，科大讯飞股份有限公司申请一项名为“一种 2D 虚拟人视频的生成方法、装置、存储介质及设备”，公开号 CN117574963A，申请日期为 2023 年 10 月。

专利摘要显示，本申请公开了一种 2D 虚拟人视频的生成方法、装置、存储介质及设备，该方法包括：首先获取目标用户的初始用户图像，并响应于目标用户触发的个性化人像生成指令，生成初始用户图像对应的初始人像，然后接收目标用户输入的人像编辑指令，并响应于人像编辑指令，对初始人像进行编辑，生成编辑人像；接着，接收目标用户输入的语音信息，进而响应于目标用户触发的 2D 虚拟人视频生成指令，根据编辑人像和目标用户输入的语音信息，生成包含目标用户的 2D 虚拟人视频。从而能够自动生成符合各目标用户个性化需求的 2D 虚拟人视频，并展示给用户，进而提高了目标用户的交互体验。

背景技术：随着人工智能、物联网等信息技术的快速发展，人机交互的应

用场景越来越广泛。各式各样的智能交互软件和设备出现在人们的生活工作中，如聊天生成预训练转换器(Chat Generative Pre-trained Transformer，简称 ChatGPT)、智能音箱、智能电视等，可以为人们提供信息查询、视频生成等众多应用场景的智能交互功能，以辅助用户完成各种行为意图。

目前，将现实中的物体、人物等进行数字虚拟化的需求越来越多，尤其是人物的虚拟化，如何利用智能交互软件或设备(如智能音箱、智能电视等)快速、准确的构建出 2D 虚拟人视频已成为亟待解决的问题。

发明目的或效果：本申请实施例的主要目的在于提供一种 2D 虚拟人视频的生成方法、装置、存储介质及设备，能够自动生成符合用户个性化需求的 2D 虚拟人视频，并展示给用户，以提高用户的交互体验。

资料来源：搜狐，2024-02-20，CN117574963A（公开日：2024-02-20）

百度：视频虚拟形象可以执行多个动作

天眼查专利信息显示，2 月 23 日，北京百度网讯科技有限公司申请的“视

频生成方法、装置、电子设备和存储介质”专利（CN115942039B）获授权。

摘要显示，本公开涉及人工智能技术领域，尤其涉及增强现实、虚拟现实、计算机视觉、深度学习等技术领域，可应用于元宇宙、虚拟数字人等场景下。

具体实现方案为：响应于接收到目标文本，根据与其相关的多个初始时间信息，确定与至少一个目标动作标签文本分别对应的至少一个第一目标时间信息，其中，目标文本是利用至少一个目标动作标签文本处理初始文本得到的，目标动作标签文本与预设动作对应；根据至少一个第一目标时间信息，对目标虚拟形象进行渲染，得到至少一个第一视频片段，其中，第一视频片段与预设动作对应；根据至少一个第一视频片段，生成目标视频。本公开还提供了一种视频生成装置、电子设备和存储介质。

发明目的或效果：随着人工智能技术的发展，虚拟形象的应用场景不断增加。可以对虚拟形象进行渲染，以生成视频。在该视频中，**虚拟形象可以执行多个动作**。

资料来源：广州日报新花城. 2024-02-23，CN115942039B（授权日：

2024-02-23)

谷歌：多技术协同生成压缩视频、沉浸式视频等

关于生成视频，谷歌已布局了 136 项专利，其中 2023 年布局了 6 件专利，包括 US20240062757A1：生成和/或引起链接到其他应用的基于视频回放的助理建议的呈现，IN202347029509A：使用插值从静态图像生成视频，WO2023170069A1：生成视频的压缩表示以用于视频任务的高效学习，US20230206511A1：用于生成具有分层网格表示的沉浸式光场视频的方法、系统和介质，US20230209031A1：用于生成和呈现沉浸式视频内容的方法、系统和介质，US20230171388A1：用于使用帧速率修改来生成概括视频的方法、系统和介质。

专利 US20230206511A1 技术方案：一种压缩视频的生成方法，所述方法包括：接收图像；将接收到的图像转换成分层网格；生成固结所述分层网格的纹理图谱图像；通过压缩所述纹理图谱图像来生成压缩图像；以及与所接收的图像的标识符相关联地存储所述压缩图像。

专利 US20230209031A1 摘要：提供了用于生成和呈现沉浸式视频内容的方法、系统和介质。在一些实施方式中，所述方法包括：接收指示多个相机中的相机的位置的信息；基于所述多个摄像机中的摄像机的位置生成要在其上投影视频内容的网格，其中，所述网格由有面柱体的一部分组成，并且其中，所述有面柱体具有多个面，每个面对应于来自所述多个摄像机中的摄像机的投影；接收对应于所述多个摄像机的视频内容；以及响应于从所述用户装置接收到对所述视频内容的请求，将所述视频内容和所生成的网格传输到所述用户装置。

背景技术：许多用户享受观看可以为观看者提供沉浸式体验的沉浸式视频内容，诸如虚拟现实内容、三维内容、180 度内容或 360 度内容。此外，这些用户可以享受以沉浸式格式观看计算机生成的内容，诸如游戏视频或动画。

沉浸式视频内容经常通过将视频内容应用于诸如鱼眼网格或三维等矩形网格之类的网格来渲染，这使得视频内容被渲染得好像它正在填充球面的一部分，而不是像看起来平坦一样。然而，使用鱼眼投影或三维等矩形投影来呈现计算

机生成的内容可能是资源密集的，特别是当实时呈现图形时。

资料来源：US20230206511A1（公开日：2023-06-29），US20230209031A1（公开日：2023-06-29）

OpenAI：专利公开较少，涉及文生语、自动执行动作等技术

有别于其他科技公司大量布局专利，OpenAI 公司的知识产权战略方案有所不同，笔者检索了 OpenAI 公司申请的专利，其仅有 3 项与视频生成相关的专利申请文件，分别是 US20240020116A1：用于使用在计算机代码上训练的语言模型生成自然语言的系统和方法，US11887367B1：使用机器学习来训练并使用模型来基于视频和输入数据集执行自动接口动作，US11886826B1：用于基于语言模型的文本插入的系统和方法。

其中专利 US20240020096A1 涉及“**文生语**”。摘要：公开了用于基于计算机代码输入生成自然语言的方法、系统和计算机可读介质。在一个实施例中，一种方法可以包括以下中的一个或多个：访问被配置为从计算机代码生成文档串的文档串生成模型；接收一个或多个计算机代码样本；使用所述文档串生成

模型并基于所接收的一个或多个计算机代码样本，生成表示自然语言文本的一个或者多个候选文档串，所述一个或者更多个候选文档字符串中的每一个与所述一种或多个计算机代码样本的至少一部分相关联；识别所述一个或多个候选文档串中的至少一个，所述至少一个候选文档字符串提供所述一种或多种计算机代码样本的至少一部分的意图；和/或经由用户接口输出具有所述一个或多个计算机代码样本的至少一部分的所述至少一个识别的文档串。

专利 US11887367B1 涉及**执行自动动作**。摘要：公开了用于训练机器学习模型以标记未标记的数据和/或执行自动动作的方法、系统和计算机可读介质。

在一个实施例中，一种方法包括接收未标记的数字视频数据，为所述未标记数字视频数据生成伪标签，所述生成包括接收标记的数字图像数据，使用标记数字图像数据训练逆动力学模型（IDM），以及为所述无标记数字图像生成至少一个伪标签，其中，所述至少一个伪标签基于由所述 IDM 生成的对模拟所述未标记数字视频数据的至少一个时间步长的一个或多个动作的预测。在一些实施例中，该方法还包括将至少一个伪标签添加到未标记的数字视频数据，并使用

伪标记的数字图像数据进一步训练 IDM 或机器学习模型。

资 料 来 源 ： **US20240020116A1** （ 公 开 日 ： **2024-01-18**） , **US11887367B1**（公开日：
2024-01-30）

文生视频商标动态

“SORA”国内商标已被注册多达百余条

根据新浪科技发文，在 Sora 推出之际，在 Sora 正式面世前，已有多家公司注册“SORA”相关的商标信息。根据不完全统计发现，截至发稿，国内共有 154 条关于“SORA”相关的商标信息，国际分类遍布家具、运输工具、广告销售、日化用品和科学仪器等。其中，注册最早的“SORA”商标由株式会社岛野申请，申请日期为 1999 年 1 月 5 日，国际分类为“12-运输工具”，专用权期限至 2030 年 7 月 27 日。不过，有一些“SORA”相关的商标已处于无效状态。例如，2013 年 4 月，佳音时装有限公司曾申请注册为国际分类为“35-广告销售”的“SORA”商标，但该商标现已无效。如果刨除掉无效商标等，目前已注册的“SORA”相关商标共 96 条。

由此，OpenAI 公司的“SORA”商标申请是否将受上述在先申请商标的干扰或基于“GPT”类似的理由被驳回不得而知。

资料来源：新浪科技. 2024-2-22

“SORA”在美商标处于待审查阶段

2024 年 2 月 14 日，OpenAI 公司就“SORA”向美国专利商标局提出了申请。商标：SORA™ 商标注册旨在涵盖可下载计算机程序和用于生成视频和图像的可下载计算机软件类别；用于基于自然语言提示、视觉提示、文本、语音、图像和/或视频创建、生成和编辑视频和图像的可下载计算机程序和可下载计算机软件；用于视频和图像识别、处理、分析、理解和生成的可下载计算机程序和可下载计算机软件；用于编辑、组织、修改、传输、上传、下载和共享视频、图像和视听材料的可下载计算机程序和可下载计算机软件；用于创建和生成文本到视频和文本到图像的可下载计算机程序和可下载计算机软件。

截至 2024 年 2 月 25 日，该商标申请为 “现场申请等待审查”阶段，具体状态为商标申请已被主管局接受（已满足最低申请要求），尚未分配给审查员。

资料来源：<https://uspto.report/TM/98405663>. 2024-02-25

“GPT”商标被 USPTO 最终驳回

GPT 商标由于申请的商标仅描述了申请人的商品和服务的特征、功能或特征，因此注册被拒绝。关于商标申请，美国 OpenAI 公司的 GPT 商标申请已被美国专利商标局“最终”驳回，此前于 2023 年 10 月份已被驳回一次，美国专利商标局裁定该术语“仅具有描述性”，因此无法注册。根据美国专利商标局的说法，该名称不符合注册商标的标准，也不符合名称后面的“TM”所提供的保护。

OpenAI 认为，它已经普及了 GPT 这个术语，在这种情况下，GPT 代表“生成式预训练变压器”，描述了机器学习模型的本质。它是生成性的，因为它产生新的（ish）材料，经过预先训练，因为它是一个在专有数据库上集中训练的大型模型，而 Transformer 是构建人工智能的一种特定方法的名称（由谷歌研究人员于 2017 年发现），它允许用于训练更大的模型。

但专利局指出，GPT 已经在许多其他环境中使用，并被其他公司在相关环境中使用。例如，亚马逊有一个关于 GPT 是什么以及如何使用它们的列表。

专利方面的论点是，**GPT** 描述了产品的一个方面。就像你有一种名为“Crunchy O's”的麦片，并试图将“Crunchy”作为商标。就 ChatGPT 而言，

这是一个 GPT 类型的 AI 模型——OpenAI 并未创造这一概念，也不是唯一提供这种概念的——您可以与之聊天。它可能是可识别的，但不符合商标的要求。

商标的缺失可能会削弱 OpenAI 对 GPT 相关术语的主导地位。你可以期待诸如“TalkGPT”之类的东西出现在应用商店中（事实上它们已经存在，而且数量众多），而 OpenAI 不能因为它们使用自己的品牌而起诉它们。

也就是说，当有人说“GPT”时，OpenAI 确实拥有最大的关注度，因此尽管他们的法律保护有限，但他们保留了第一品牌的优势。

资料来源：<https://ca.style.yahoo.com>. 2024-2-15

地址：上海市永福路 265 号

邮编：200031

编辑：杨四娟

责编：路炜

编审：林鹤

电话：021-64455555

邮件：istis@libnet.sh.cn

网址：www.istis.sh.cn