人工智能新技术发展下的法律问题

欧阳昆泼

浙江垦丁律师事务所联合创始人

目录
CONTENTS

01 AI生成内容的著作权问题

AI生成内容的著作权问题

AI生成物是否构成作品？能否受著作权法保护？

- “作品”是否必须由人类创作？AI生成内容可以构成作品吗？

AI生成内容的著作权问题

北京菲林律师事务所诉百度案

北京菲林律师事务所于2018 年9月9日通过微信公众号发表了一篇涉及**人工智能程序生成的文章**。次日，百度公司在其经营的百家号平台里转载该文，该公司提供的被诉侵权文章“删除了原告为整个系列作品创作的引言、检索概况，电影行业案件数量年度趋势图和结尾的‘注’部分”其他内容与原告涉案文章相同。菲林律所为百度公司侵犯其著作权。百度公司则以涉案文章为统计数据分析软件生成，不属于《著作权法》保护范围为由进行抗辩。

北京互联网法院经审理认为，虽然计算机软件智能生成的文字内容体现了针对相关数据的选择、判断、分析，具有一定的独创性。但是，**自然人创作完成仍应是著作权法上作品的必要条件**。计算机软件智能生成文字内容的过程有两个环节有自然人作为主体参与，一是软件研发环节，二是软件使用环节，但**软件智能生成的文字内容并未传递软件研发者及使用者的思想、感情的独创性表达，故二者不应被认定为软件智能生成的文字内容的作者**。人工智能软件利用输入的关键词与算法、规则和模板结合形成的文字内容，某种意义上讲可认定是人工智能软件“创作”了该内容。**但即使人工智能软件“创作”的文字内容具有独创性，也不属于著作权法意义上的作品，不能认定人工智能软件是其作者并享有著作权法规定的相关权利。**

Dreamwriter首篇财经报道
《8月CPI同比上涨2.0% 创12个月新高》

AI生成内容的著作权问题

腾讯诉盈讯科技案（Dreamwriter案）

2018年8月20日，腾讯公司在腾讯证券网站上首次发表了标题为《午评：沪指小幅上涨0.11%报2671.93点 通信运营、石油开采等板块领涨》的财经报道文章（以下简称“涉案文章”），末尾注明“**本文由腾讯机器人Dreamwriter自动撰写**”。Dreamwriter计算机软件系由腾讯公司关联企业自主开发并授权其使用的一套基于数据和算法的智能写作辅助系统。

同日，盈讯科技在其运营的网站“网贷之家”发布了相同文章，经比对，该文章与涉案文章的标题和内容完全一样，文末同样标注“本文由腾讯机器人Dreamwriter自动撰写”。

腾讯公司遂将盈讯科技诉至南山区法院。

法院审理后，认定盈讯科技侵害了腾讯公司的著作权。

AI生成内容的著作权问题

腾讯诉盈讯科技案（Dreamwriter案）

法官的主要意见如下：

（一）关于涉案文章是否构成文字作品

1.**著作权法所称创作，是指直接产生文学、艺术和科学作品的智力活动。**据此，具体认定是否属于创作行为时应当考虑该行为是否属于一种智力活动以及该行为与作品的特定表现形式之间是否具有直接的联系。

2.

（1）由于创作工具的技术特征，**涉案文章的创作过程不仅包括Dreamwriter软件自动运行的过程（这一过程确实没有人的参与），也包括主创团队相关人员在事前对Dreamwriter所作的相关选择与安排。**

（2）本案中原告主创团队在数据输入、触发条件设定、模板和语料风格的取舍上的安排与选择（数据类型的输入与数据格式的处理、触发条件的设定、文章框架模板的选择和语料的设定、智能校验算法模型的训练等；收集素材、决定表达的主题、写作的风格以及具体的语句形式的行为），**属于著作权法意义上的智力活动。**

（3）第（2）点中的活动，与涉案文章的生成之间有足以被认定为“直接产生”的联系。

腾讯诉盈讯科技案（Dreamwriter案）

法官的主要意见如下：

（二）关于涉案文章的作者

涉案文章是由原告主持的多团队、多人分工形成的整体智力创作完成了作品，整体体现原告对于发布股评综述类文章的需求和意图。涉案文章在由原告运营的腾讯网证券频道上发布，文章末尾注明“本文由腾讯机器人Dreamwriter自动撰写”，其中的“腾讯”署名的指向结合其发布平台应理解为原告，说明涉案文章由原告对外承担责任。故在无相反证据的情况下，本院认定**涉案文章是原告主持创作的法人作品。**

AI生成内容的著作权问题

AI生成物是否构成作品？能否受著作权法保护？

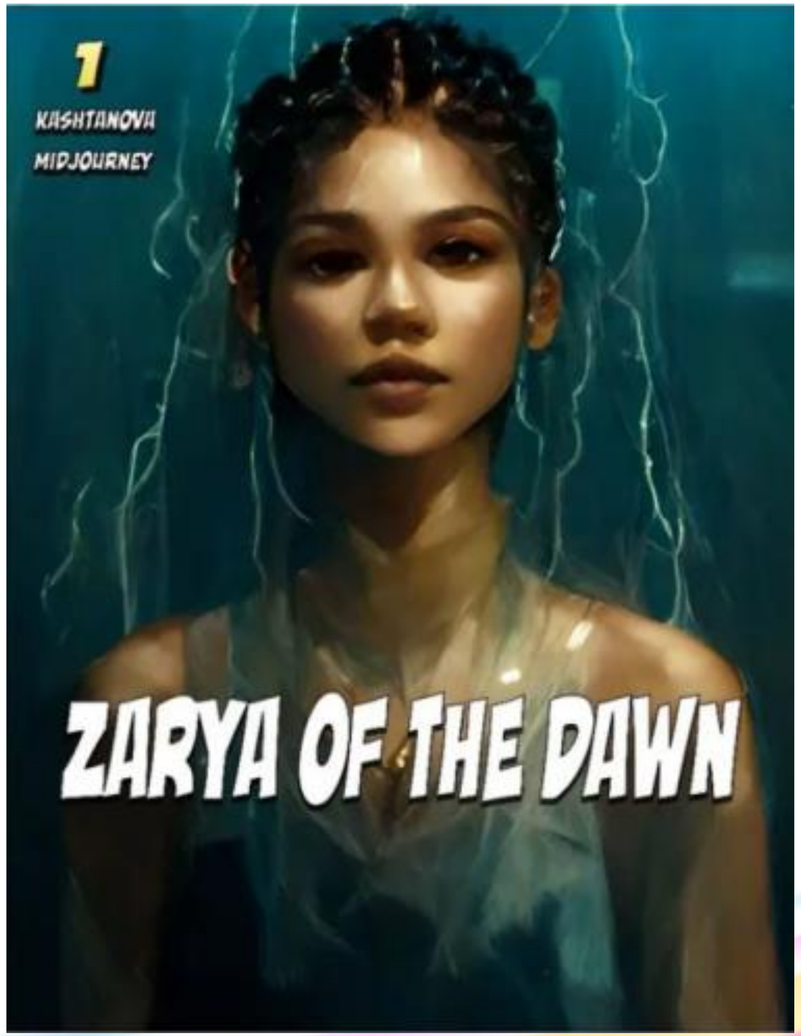
- “作品”是否必须由人类创作？
- AI是人创作的工具？

AI生成内容的著作权问题

《黎明的曙光》（Zarya of the Dawn）是克里斯·卡什塔诺娃（Kris Kashtanova）使用人工智能软件Midjourney绘制的漫画。

2022年9月，卡什塔诺娃向美国版权局申请了漫画的版权保护，但没有透露插图是使用Midjourney创建的。该漫画曾获得版权保护，但版权局发现这一事实后提起诉讼，撤销了对该作品的保护。

该作品的版权保护已于2023年2月被撤销。



AI生成内容的著作权问题

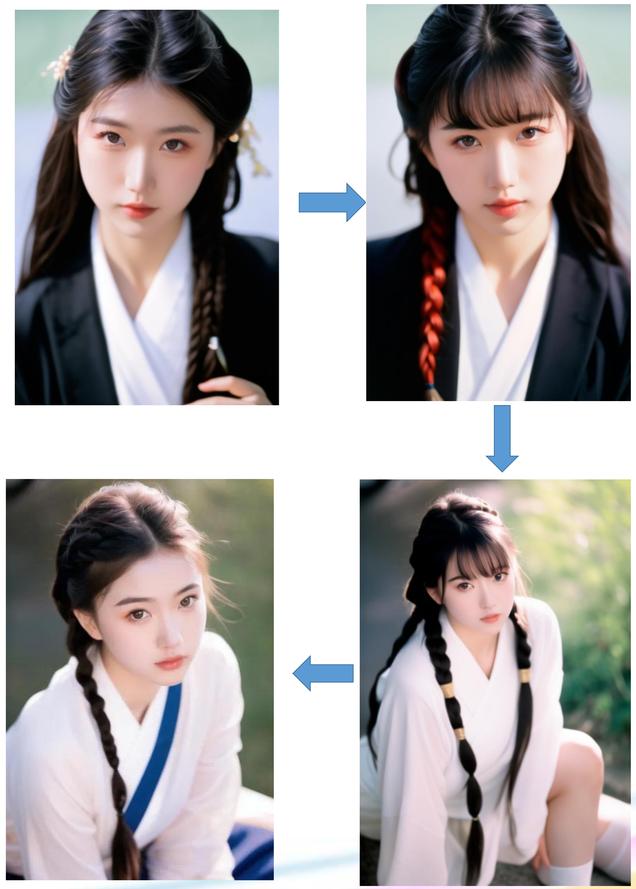
卡什塔诺娃	美国版权局
人工智能只是卡什塔诺娃创作的工具	虽然额外的指令可以影响图形，但**该过程不受用户控制，因为不可能预测Midjourney将生成什么内容。**
卡什塔诺娃通过试错法“引导”了图片的结构和内容	**向Midjourney发出文字指令的人并没有“实际形成”最终的图片。**指挥Midjourney生成图片的用户与实际生成的图片距离遥远，**Midjourney的用户对生成的图片缺乏充分的控制**，不能被认为是决定这些图片的“心智”（master mind）。
生成的图片体现了“人类发出的创造性的指令”	**指令就像雇佣艺术家作画的客户对内容发出的一般性指示。**

AI生成内容的著作权问题

AI文生图第一案

原告使用开源软件Stable Diffusion，通过输入提示词的方式生成了涉案图片后发布在小红书平台。

原告在Stable Diffusion模型中输入了**提示词**，提示词中**艺术类型**为“超逼真照片” “彩色照片”，**主体**为“日本偶像”并详细描绘了**人物细节**如皮肤状态、眼睛和辫子的颜色等，**环境**为“外景” “黄金间” “动态灯光”,人物呈现方式为“酷姿势” “看着镜头”，**风格**为“胶片纹理” “胶片仿真”等，同时设置了相关参数，根据初步生成的图片，又增加了提示词、调整了**参数**，最终选择了一幅自己满意的图片。



被告在百家号上发布文章，文章配图使用了涉案图片。原告认为，**被告未经许可使用涉案图片**，且截去了原告在小红书平台的署名水印，使得相关用户误认为被告为该作品的作者，严重侵犯了原告享有的署名权及信息网络传播权，要求被告公开赔礼道歉、赔偿经济损失等。

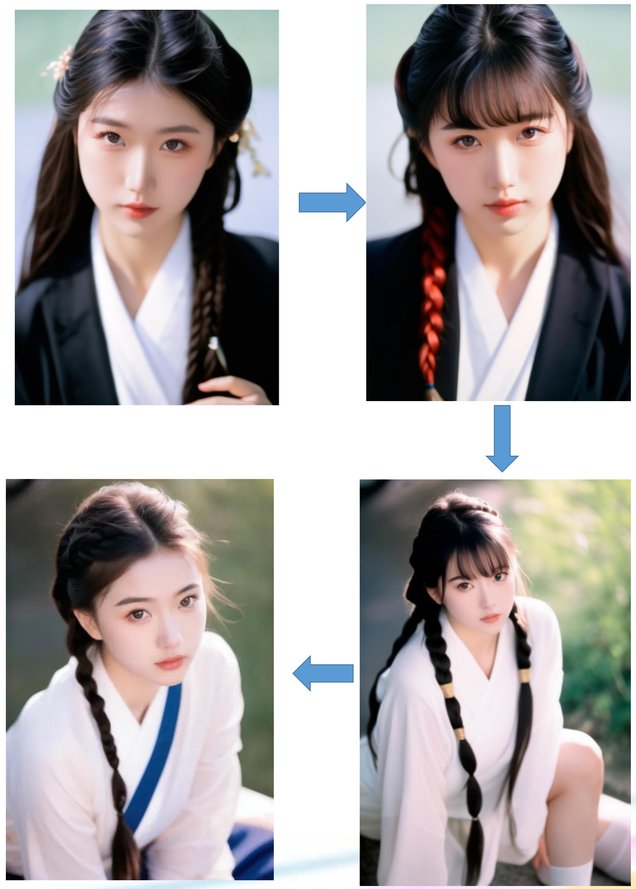
AI生成内容的著作权问题

AI文生图第一案

在本案中，法官认为，

从原告构思涉案图片起，到最终选定涉案图片止，这整个过程来看，原告进行了一定的**智力投入**，比如设计人物的呈现方式、选择提示词、安排提示词的顺序设置相关的参数、选定哪个图片符合预期等等。涉案图片体现了原告的**智力投入**，……**整个创作过程中进行智力投入的是人而非人工智能模型。**在这种背景和技术现实下，人工智能生成图片，只要能体现出人的独创性智力投入，就应当被认定为作品，受到著作权法保护。

原告是直接根据需要对涉案人工智能模型进行相关设置，并最终选定涉案图片的人，涉案图片是基于原告的智力投入直接产生，且体现出了原告的个性化表达，故原告是涉案图片的作者，享有涉案图片的著作权。



AI生成内容的著作权问题

AI生成物是否构成作品？能否受著作权法保护？

- **“作品”是否必须由人类创作？**
- **AI是人创作的工具？**
- **AIGC发展与鼓励内容创新——利益的平衡**

文生图第一案的法官认为，认定AI生成内容具有著作权可鼓励人们使用AI。

“通过激励创作者使用AI技术，软件研发者收益增加，形成良性循环，对产业发展带来积极影响。当前技术条件下难以区分AI生成内容。如果区别对待，手工创作受保护而AI生成不受保护，将给社会传递负面激励，导致人们拒绝使用新工具，或者隐瞒使用AI创作的事实，这可能侵害公众知情权。”

但是，如果鼓励人们使用AI进行创作，这有可能会扼杀创意，那么未来人类的原创作品可能会越来越少。

02 AI训练数据的合法性问题

AI训练数据的合法性问题

AI训练数据的侵权风险

- **在输入阶段**，如果将大量受著作权保护的作品用来训练人工智能，这本身看似出于学习目的，实则最终服务于商业目的，很难使用现有的著作权合理使用制度规避侵权责任。

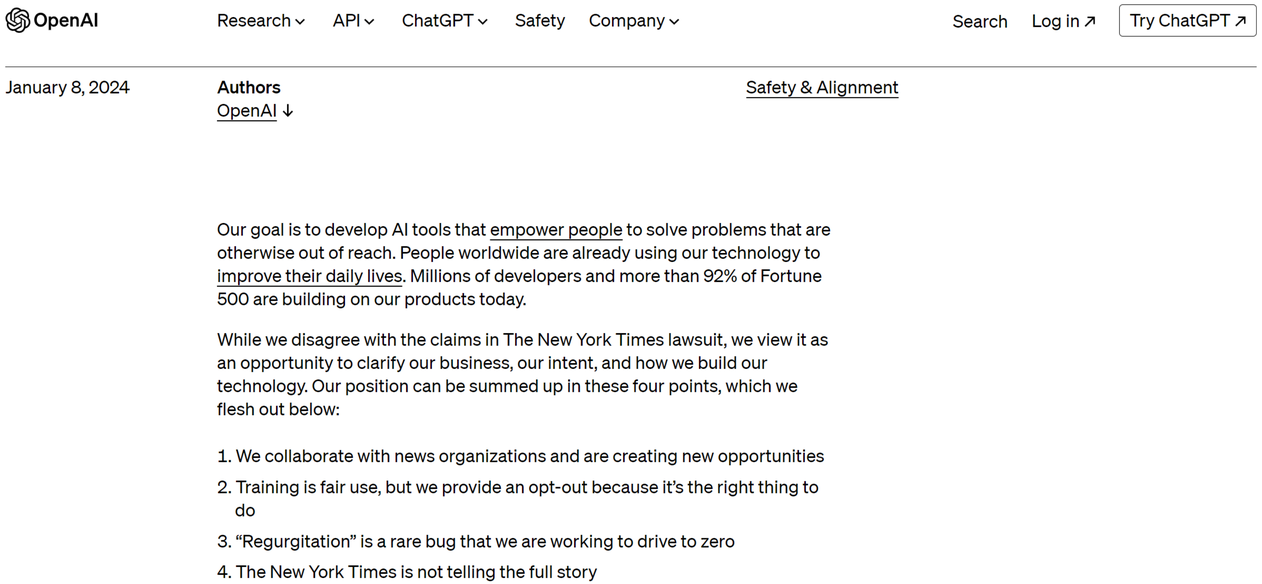
- **在输出阶段**，如果生成的内容与原作品在表达上构成实质性相似，则可能侵犯复制权；如果在保留原作品表达的基础上形成了新的表达，则可能涉及改编权问题。

AI训练数据的合法性问题

《纽约时报》 v. OpenAI, 微软

2023年12月，《纽约时报》以侵犯版权为由起诉OpenAI 和微软，称这家初创公司使用了该报的文章来帮助训练其模型。今年3月5日，微软提出驳回诉讼的动议，指责《纽约时报》的做法是“末日化未来”（doomsday futurology）。

《纽约时报》在诉讼中称，在用户提示下，ChatGPT 有时会逐字吐出部分文章，或分享其内容的关键部分。此外，ChatGPT 编造了一些归属于《纽约时报》的文章。律师表示，所有这些都违反了版权法，并削弱了《纽约时报》依赖许可、订阅和广告收入的商业模式。OpenAI 在其网站上发布的初步回应称该诉讼“令人惊讶和失望”，并声称该诉讼“没有法律依据”。



OpenAI Research API ChatGPT Safety Company Search Log in Try ChatGPT

January 8, 2024

Authors
OpenAI

Safety & Alignment

Our goal is to develop AI tools that empower people to solve problems that are otherwise out of reach. People worldwide are already using our technology to improve their daily lives. Millions of developers and more than 92% of Fortune 500 are building on our products today.

While we disagree with the claims in The New York Times lawsuit, we view it as an opportunity to clarify our business, our intent, and how we build our technology. Our position can be summed up in these four points, which we flesh out below:

1. We collaborate with news organizations and are creating new opportunities
2. Training is fair use, but we provide an opt-out because it's the right thing to do
3. "Regurgitation" is a rare bug that we are working to drive to zero
4. The New York Times is not telling the full story

AI训练数据的合法性问题

AIGC平台侵权第一案

“奥特曼”系知名动漫形象，原告获得了权利人关于奥特曼形象著作权及维权权利。被告经营Tab网站，具有AI对话及AI生成绘画功能。原告发现，当要求Tab生成奥特曼相关图片时(如输入“生成一张戴拿奥特曼”)，Tab生成的奥特曼形象与原告奥特曼形象构成实质性相似。

原告认为，被告未经授权，擅自利用原告享有权利的作品训练其大模型并生成实质性相似的图片，且通过销售会员充值及“算力”购买等增值服务攫取非法收益，前述行为给原告造成严重损害，遂起诉要求被告赔偿经济损失及维权合理支出共计30万元。

AI训练数据的合法性问题

AIGC平台侵权第一案

（一）关于平台是否侵权

平台未经许可复制了案涉奥特曼作品，法院认为平台侵犯了权利人对案涉奥特曼作品的复制权。

（二）AIGC训练的素材是否属于合理使用?

本案中，因为被告是通过第三方服务商提供AI绘画服务的，并不涉及AIGC训练素材是否属于合理使用的问题，所以被告答辩没有提及合理使用。

AI训练数据的合法性问题

AIGC平台侵权第一案

（三）关于平台基于侵权应承担的责任

依据《生成式人工智能服务管理暂行办法》《互联网信息服务深度合成管理规定》等规定，服务提供者应采取建立举报机制、提示潜在风险、进行显著标识等行动。

在侵权事实认定的情况下，被告需要停止侵权行为，保证其服务不再生成奥特曼相关内容。防范程度应达到：用户正常使用与奥特曼相关的提示词，不能生成与案涉奥特曼作品实质性相似的图片。

本案中，被告已经采取关键词过滤等措施，停止生成相关图片，并达到了一定效果。然而，庭审中，在原、被告双方见证下，当向Tab网站输入与奥特曼相关的其他关键词如“迪迦”，仍可产生与迪迦奥特曼复合型原图实质性相似的图片。**因此，被告应进一步采取关键词过滤等措施，防范其服务继续生成与案涉奥特曼作品实质性相似的图片，防范程度应达到：用户正常使用与奥特曼相关的提示词，不能生成与案涉奥特曼作品实质性相似的图片。**

知识产权。因此，服务提供者在提供生成式人工智能服务时应尽合理的注意义务。但在本案中，被告作为服务提供者未尽到合理的注意义务。本院具体分析如下：

第一，投诉举报机制的欠缺。《生成式人工智能服务管理暂行办法》第十五条规定：“提供者应当建立健全投诉、举报机制，设置便捷的投诉、举报入口，公布处理流程和反馈时限，及时受理、处理公众投诉举报并反馈处理结果。”结合本案认定事实，特别是本院当庭查明情况，截至开庭之日，被告经营的Tab网站并未建立相关投诉举报机制，使得权利人难以通过投诉举报机制来保护其著作权。

第二，潜在风险提示的欠缺。《生成式人工智能服务管理暂行办法》第四条规定：“提供和使用生成式人工智能服务，应当遵守法律、行政法规，尊重社会公德和伦理道德，遵守以下规定：……（三）尊重知识产权、商业道德，保守商业秘密，不得利用算法、数据、平台等优势，实施垄断和不正当竞争行为；……（五）基于服务类型特点，采取有效措施，提升生成式人工智能服务的透明度，提高生成内容的准确性和可靠性。”本案中，被告

03 AI的现实风险

AI的现实风险

- **恶意模型**

模型名称	技术特征	主要危害
WormGPT	基于开源 GPT-JLLM 等构建，具有实际自定义 LLM。使用新的 API，不依赖于 OpenAI 内容政策限制。使用包括合法网站、暗网论坛、恶意软件样本、网络钓鱼模板等大量数据进行训练。有较高的响应率和运行速度，无字符输入限制。	生成恶意软件代码造成数据泄露、网络攻击、窃取隐私等，生成诈骗文本图像进行复杂的网络钓鱼活动和商业电子邮件入侵(BEC)
PoisonGPT	对 GPT-J-6BLLM 模型进行了修改以传播虚假信息，不受安全限制约束。上传至公共存储库，集成到各种应用程序中，导致 LLM 供应链中毒	被问及特定问题时会提供错误答案，制造假新闻、扭曲现实、操纵舆论
EvilGPT	基于 Python 构建的 ChatGPT 替代方案。使用可能需要输入 OpenAI 密钥，疑似基于越狱提示的模型窃取包装工具	考虑恶意行为者的匿名性。创建有害软件，如计算机病毒和恶意代码。生成高迷惑性钓鱼邮件。放大虚假信息和误导性信息的传播
FraudGPT	基于开源 LLM 开发，接受不同来源的大量数据训练。具有广泛字符支持，能够保留聊天内存，具备格式化代码能力	编写欺骗性短信、钓鱼邮件和钓鱼网站代码，提供高质量诈骗模板和黑客技术学习资源。识别未经 Visa 验证的银行 ID 等
WolfGPT	基于 Python 构建的 ChatGPT 替代方案	隐匿性强，创建加密恶意软件，发起高级网络钓鱼攻击
XXXGPT	恶意 ChatGPT 变体，发布者声称提供专家团队，为用户的违法项目提供定制服务	为僵尸网络、恶意软件、加密货币挖掘程序、ATM 和 POS 恶意软件等提供代码

AI的现实风险

- 恶意模型
- **合法模型的滥用 ——深度伪造**

AI的现实风险

- 恶意模型

- **合法模型的滥用 ——深度伪造**

2023年5月24日下午，科大讯飞股价出现意外下跌，收盘时报收56.57元/股，较开盘价**下跌4.26%**。对此科大讯飞解释称，是**因为某生成式AI写作了虚假小作文导致，谣传风险为不实消息。**

第一篇文章：“5月23日，有外媒援引知情人士的话称，美国正在考虑是否将科大讯飞、美亚柏科等加入‘实体名单’，禁止它们使用美国的组件或软件”。
不过，经过查证，这是2019年的一篇报道改编而成。事实上，科大讯飞已经于2019年10月被列入实体清单，对公司日常经营未产生重大影响。
第二篇文章：《一篇科大讯飞出现重大风险的警示文》，该文指出，近期，科大讯飞被曝涉嫌大量采集用户隐私数据，并将其用于人工智能研究，这一行为严重侵犯了用户的隐私权，引发了公众的强烈不满和抵制。
对此，科大讯飞回应称，该事件系有人利用某生成式AI撰写的科大讯飞风险警示，该消息流出后引发广泛关注，实际上科大讯飞未发生相关事件，该公司法务部已对相关信息取证。

AI的现实风险

- 恶意模型
- **合法模型的滥用 ——深度伪造**

2023年4月25日，在某互联网自媒体平台上突然出现了一篇标题为《**今晨甘肃一火车撞上修路工人，致9人死亡**》的文章，平凉市公安局崆峒分局网安大队初步判断该文章为虚假信息。后经侦查发现，当天共计有21个自媒体账号在同一时间段发布了几乎一模一样的涉案文章，累计点击量在侦查期间就已达1.5万余次。

目前，崆峒公安分局已经以涉嫌**寻衅滋事罪**，对犯罪嫌疑人洪某某采取了刑事强制措施，案件正在调查中。

洪某某承认，自己使用第三方信息搜集工具在全网检索近年来发生的各种新闻事件，重新编辑后在自媒体平台上发出，以吸引流量获得盈利。但是，由于自媒体平台陆续开发了防抄袭功能，文章在发出前需要“查重”，**洪某某复制粘贴的新闻往往被判定为抄袭作品，无法顺利在自媒体上发表**。洪某某遂**使用ChatGPT，将复制过来的旧新闻进行“重写”**，顺利通过自媒体平台查重并发表。

AI的现实风险

- 恶意模型
- 合法模型的滥用——深度伪造
- **偏见**
- **伦理**

谢谢聆听

欧阳律师||垦丁网络法

浙江 杭州

扫一扫上面的二维码图案，加我为朋友。