

## 关于ChatGPT的三个“灵魂”发问

### 前言

ChatGPT 于今年年初突然爆红全球！人们惊叹于它似乎无所不能的强大功能——即时回答问题、撰写和修改计算机代码、快速生成新闻报道、创作诗歌、甚至撰写情人节情书……一些专家已经在预测，以 ChatGPT 为代表的 AIGC（生成式人工智能）无疑将改变未来。惊叹之余，ChatGPT 的强大能力也不免让人们担忧，它在社会责任、政府监管、法律风险等方面的问题又应该如何解决？

我们近期已被不少客户咨询，是否可利用 ChatGPT 的回答内容出具调查报告、是否可以利用 ChatGPT 完成展览的策展方案等等。因此，我们觉得有必要借本文，主要从著作权法的角度，对 ChatGPT 为代表的 AIGC 可能带来的法律问题及相关风险进行探讨。

### 一、第一个问题：ChatGPT 的回答是“创作”行为吗？

#### （一）“创作”的主体应当为自然人

根据各国普遍的著作权理论及司法实践，**创作的主体应当为自然人，AI 无法成为作品的作者**。至少现阶段，如果没有立法的干预，AI 不太可能成为受著作权法保护内容的创作主体和归属主体。

就本文而言，为了避免进入闭环论证（即，因为是 AI 不是创作的主体——所以 AI 生成内容不构成作品），我们在本文的后续讨论中先不考虑创作主体这一因素，而是聚焦于对创作行为本身的理解、ChatGPT 的内容生成过程以及自然人在内容生成中参与的环节等问题。

#### （二）认定“创作”需要从主客观两个角度考虑

我国《著作权法实施条例》第三条对“创作”进行了定义，“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。为他人创作进行组织工作，提供咨询意见、物质条件，或者进行其他辅助工作，均不视为创作。”该定义除了强调“创作”属于智力活动，并排除了组织工作、咨询意见、物质条件、辅助工作。而作品是指“文学、艺术和科学领域内具有独创性并能以一定形式表现的智力成果”。由此可见，对“创作”活动的理解需要立足于“独创性”进行考量。但是，关于作品“独创性”的认定，一直以来缺乏非常明确、清晰的标准，通常由司法机构结合内容的类型、“创作”的实际情况进行理解与判断。结合过去及现在的一些判断标准（包括：额头出汗标准（已成为历史）、最低限度创造性标准、有“技巧、判断或劳动”的投入标准、“反映作者个性”标准、“整体概念和感受测试法”）等等，我们认为“独创性”的“创作”涉及主客观两个方面：

（1）客观形式的判断，将产生“重复度”较高内容的工作及缺少智力投入的工作排除在外。例如，临摹者对原作进行精确的临摹，虽然也投入了技巧及劳动，但由于内容的高度重复性，不应当被视为创作新作品的过程。又如，老电影的修复工作，虽然投入大量的人力、物力，甚至也有“智力劳动”，但这个过程一般不会被视作“创作”。

在实践中，如果不同人对于 ChatGPT 的提问相似，那么较大概率上，AI 生成内容可能出现重复性的答案，这一类的重复性的生成过程可以排除于“创作”行为之外。但是，如果指令设计有一定技巧，AI 生成内容在表达层面理论上可能生成出现非重复性的答案，可能就难以通过客观形式的判断被排除。

（2）主观方面的判断，将作者的思想、创作过程，特别是创作意图纳入考量范围。并不是所有标

准都将“作者”的个性投入放在最重要的位置进行考虑，但即使在淡化作品“人格属性”的美国，在认定创作要件时，“创作意图”也是判定作品的必要条件之一<sup>1</sup>。例如，知名的“猴子自拍案”，法院就认为猴子并无创作作品的主观意图，不满足作者的要求。尽管摄影师斯莱特换上广角镜头，设定了各种参数，将相机和遥控器一固定在三脚架上，但由于这些操作并未实质性介入拍摄行为，因此也不能认定为“创作”<sup>2</sup>。



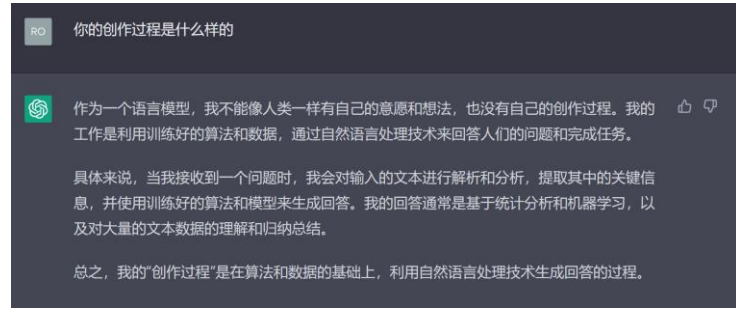
（猴子自拍照，图片来源于网络）

类推至 ChatGPT 上也类似，ChatGPT 模型本身并无创作作品的主观意图，更毋论具有个性的投入，因为其是根据人类的指令执行任务。从模型构建的终极目标看，ChatGPT 模型就是遵循人类下达的指令并作出尽可能合乎逻辑、自然语言表达性的回复，而不是为了加入个性化的内容。

因此，结合以上主客观的判断标准，我们认为 ChatGPT 生成内容的过程无疑会被排除在“创作”的范围外。

### （三） ChatGPT 自己的“思考”

有趣的是，ChatGPT 自己也得出了类似的结论（见下截图），认为“自己”并没有所谓的创作过程，仅是在算法和数据的基础上，利用自然语言处理生成回答。



（截图来源于我们与 ChatGPT 的对话）

## 二、第二个问题：ChatGPT 生成的内容是作品吗？

ChatGPT 与普通创作工具（如画笔）的差别在于，其似乎具有高度的“智能性”，甚至人们仅输入只言片语就可以得到长篇的完整回复。于是引发我们对于其生成的过程及产生的内容需要做进一步分析。

### （一）关于 ChatGPT 对话内容的生成过程

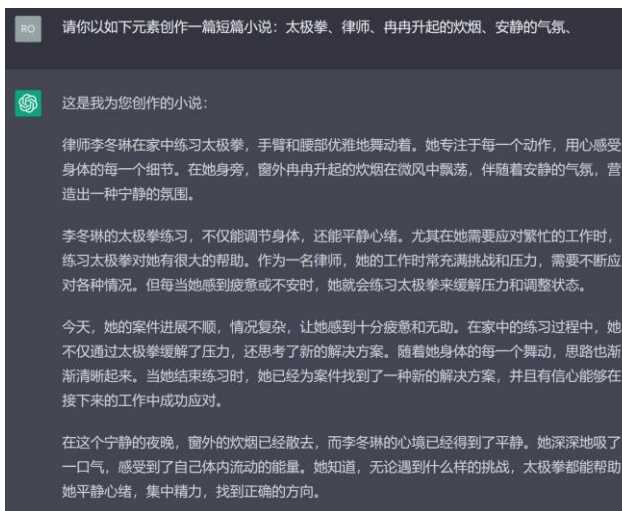
#### 1. 普通用户角度的生成过程

从普通用户角度，ChatGPT 对话内容的生成过程非常简单，即“用户发出指令”——“ChatGPT 反馈指令”（“指令——反馈”模式）。但不同于此前以往的一些 AI，ChatGPT 还能够连续回答问题甚至修正回复、承认错误、质疑不正确的前提和拒绝不适当的请求。

举例而言，我们在 ChatGPT 上为其准备了一道命题作文题“使用太极拳、律师、冉冉升起的炊烟、安静的氛围这几个元素创作一篇短篇小说”，ChatGPT 就这几个看似关联度不大的元素生成了如下内容：

<sup>1</sup> 该观点来源于《人工智能“创作”认定及其财产权益保护研究——兼评“首例人工智能生成内容著作权侵权案”》，作者冯晓青；潘柏华。

<sup>2</sup> 参见 People for The Ethical Treatment of Animals, Inc V. Slater, (N.D.Cal.Jan.28, 2016)



(截图来源于我们与 ChatGPT 的对话)

从该回复内容看，ChatGPT 不仅遵循了我们的指令，同时其生成的内容具有一定的文学性且逻辑完整，可以说已经达到了普通中学生的创作水平。并且，初步检索后，我们也并未在公开范围内找到相同或相似的表达内容，因此可以初步推测 ChatGPT 产生的内容并非是在简单复制或拼接。

在收到答复后，我们进一步向 ChatGPT 发出指令“是否可以加入一些科幻元素”，ChatGPT 进一步回复如下：



(截图来源于我们与 ChatGPT 的对话)

我们进行了简单的对比，并以红框大体标出了其新增的内容。可以看到，随着指令的修正，

ChatGPT 生成内容发生了调整，并且从一般人的角度看，ChatGPT 可以“理解”“科幻元素”的含义，并且将“科幻元素”与此前生成的答复进行了相对自然地衔接，而非机械拼贴。

从上述例子看到，在我们的此次尝试中，用户的角色仅是发布指令，虽然有再指令的操作，但再次提供的指令本身对表达的改变也不具有非常直接、具体到表达层面的引导性内容。类比绘画的过程，相当于用户仅是选择了绘画的基调/主题，但落到画面上却产生了细节构思、色彩、线条、阴影俱全的图画。那这幅图画又是怎么生成的呢？

## 2. ChatGPT 模型背后的原理

ChatGPT 的“智能性”来源于海量的数据训练。如前提及的，ChatGPT 仅是在算法和数据的基础上，利用自然语言处理生成回答。更具体而言，根据 OpenAI 官网网站的说明<sup>3</sup>，OpenAI 使用了“从人类反馈的强化学习”（RLHF）的机器学习技术训练 ChatGPT 模型，具体分为三步：

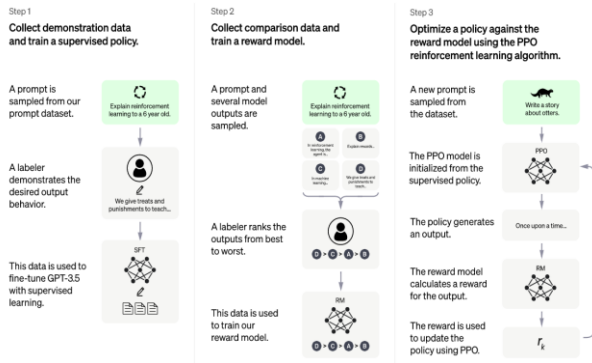
1) 收集反馈数据：从测试用户提交的 prompt（即指令或问题）中随机抽取一批，靠专业的标注人员，给出指定 prompt 的高质量答案，然后用这些人工标注好的数据来微调（fine-tuning）GPT 3.5 模型；

2) 训练奖励模型：通过人工标注对模型输出结果打分建立奖励模型，引导 ChatGPT 模型向高分方向完善输出答案；

3) 强化学习：采用强化学习来增强预训练模型的能力，以提高模型输出答案的准确性。

简而言之，ChatGPT 是通过“预训练+微调”的过程，提升理解能力和交互性表现，从而更为精准地回答用户的提问。

<sup>3</sup> 来源于：<https://openai.com/blog/chatgpt/>



(来源于 OpenAI 官网)

在该过程中，预训练的数据量越大，ChatGPT 的“智能性”也越强。根据公开报道，从 ChatGPT 进化到 ChatGPT-3 的过程相当“烧数据”——参数量从 1.17 亿增加到 1750 亿，预训练数据量从 5GB 增加到 45TB<sup>4</sup>。

## (二) ChatGPT 生成的内容是否构成作品

需要明确，我们此处的讨论是立足于内容生成过程进行“应然”定性（即：基于 ChatGPT 生成内容的过程，分析生成内容是否可能构成作品），而非对外观上难以区分是否是人类创作还是 AI 创作的内容进行“实然”定性<sup>5</sup>。

结合我们的前述分析，人类的智力投入在 ChatGPT 内容生成过程中体现为：（1）普通用户角度，发布指令（也包括再反馈指令或者否定答复）；（2）开发者角度，提供数据库或创作数据库答案，以及设定规则。那么，这些投入是否会使得 ChatGPT 生成的内容被评价为作品？

我们的分析如下：

**（1）普通用户角度发布指令的过程，只有在极少数情况下，才会导致 ChatGPT 生成内**

## 容被评价为作品

从我们上述尝试指示 ChatGPT 创作短篇小说的例子看，我们发布的指令仅是简单的句子，“请你以如下元素创作一篇短篇小说：太极拳、律师、冉冉升起的炊烟、安静的气氛”。我们再次发布的新指令为“是否可以加入一些科幻元素”。我们在此过程中的智力投入无非是提到了“太极拳”“律师”“冉冉升起的炊烟”“安静的气氛”“科幻元素”这样的元素组合，提供的“创作”可以说极为有限。一方面，简单词汇本身并不构成作品；另一方面，词汇的筛选与组合背后虽有少量的灵感但也相对随机，较难被评价为具有“独创性”。

在国内的“人工智能著作权第一案”<sup>6</sup>中，法院也认为，软件的使用者仅在操作界面提交了关键词进行搜索，这种行为没有传递软件使用者思想、感情的独创性表达，不宜认定为使用者创作完成，使用者不应成为涉计算机软件智能生成内容的作者，该内容亦不能构成作品。

基于上述，我们认为，只有在极少数情况下（例如，用户在指令提交过程中提供了具体到表达的指示<sup>7</sup>，并且最终影响了 ChatGPT 的表达内容，操纵其形成了符合自己创作想法的内容），才有可能<sup>8</sup>使得 ChatGPT 生成内容被评价为作品。

另外一个衍生的思考：假设不同的用户巧合之下输入了类似的指示并且均认为自己的指令具有独创性，进而均主张享有 ChatGPT 生成内容的权利，应当如何认定？

据著作权法理论，如果真的可以证明内容“纯属巧合”，则创作人仍分别、独立的享有相应作品的著作权，彼此不受干扰。

<sup>4</sup> 参考网站：<https://finance.sina.com.cn/stock/usstock/c/2022-12-10/doc-ixwcnpf1736315.shtml>

<sup>5</sup> 该观点参考《论人工智能生成内容在著作权法中的定性》，作者王迁。王迁老师认为，应当区分“应然”及“实然”的定性。其认为，“不披露相关内容由人工智能生成时，该内容可能因具备作品的表现形式而实际收到了保护，但该现象是举证规则造成的，并不意味着著作权法因人工智能而改变”。

<sup>6</sup> （2018）京 0491 民初 239 号

<sup>7</sup> 即使用户指令本身构成作品，也不意味着 ChatGPT 生成内容可以被评价为作品。例如，某学生在 ChatGPT 对话框内输入其独立创作完成的论文，要求 ChatGPT 进行润色/翻译/校对，在“ChatGPT 生成内容”这一环节，用户的指令本质是“润色/翻译/校对”，论文是已经存在的作品，这一指令的下达并不会产生应然角度的“新作品”（如第一

部分分析，ChatGPT 完成润色/翻译/校对的过程中，并无“创作意图”，不构成新作品）。除非在某些情形下，用户通过一条条指令逐一、细致修正 ChatGPT 完成的翻译内容，那这一条条指令的下达，可能被评价为创作翻译作品的过程。

<sup>8</sup> 即使满足了前述情形，也未必会被认定为作品。例如，需要考虑表达是否具有唯一性、有限性。参考最高人民法院的（2016）最高法民申 2136 号民事裁定书，法院认为“本案中，孙某所主张的曲线图，系当事人根据客观的价格数据，通过使用 WPS 制表工具制作完成。鉴于图表所使用的数据客观存在，数量有限，WPS 为通用软件，将上述数据录入制表工具所形成的结果，尽管属于孙某运用智力的结果，符合独立完成的要件，但该结果的表现形式有限，换言之，使用上述数据与工具所产生的结果缺少差异性。这种唯一或有限的表达方式，通常被排除在独创性之外。”

ChatGPT 开发者应该也意识到了该问题，在其用户协议<sup>9</sup>中如此表述：“*Other users may also ask similar questions and receive the same response. Responses that are requested by and generated for other users are not considered your Content*”。

ChatGPT 的上述条款表述非常巧妙，对于巧合、互不侵权的情形，结合作权理论，两位用户均可以享有 Input 及 Output 内容的权利；而对于存在侵权行为的情况下，由于不满足“遵守法律法规的前提” OpenAI 并不向侵权用户转让 Output 内容，也因此 OpenAI 将不存在侵权责任，侵权方仅为侵权用户<sup>10</sup>。

### **(2) 关注数据库内的数据可能构成作品，但相关 ChatGPT 生成内容不应被评价为新的作品**

如前所述，ChatGPT 的“智能性”来源于其庞大的数据库支持。**数据库内容的文本本身是可能构成作品的。**例如，开发者预设了一些问题，并由自然人起草了回复了答案，回复的答案如果是创作者独立创作，则可以构成作品。但是，ChatGPT 生成内容的过程，或是对数据库内相关基本语素组合，或者直接呈现了预先设置的某些答案。

这些过程都是运用“规则”自动进行的，并不存在人类智力的介入。因此，即使数据库内的数据文本内容可能构成作品，相关 ChatGPT 生成内容也不应被评价为新的作品。

另外，开发者设定的规则也不会影响最终的结论判断。开发者设定的规则是通过训练奖励模型及强化学习实现的。根据著作权法的理论，规则本身属于思想的范畴，除非其可以以具象的表达形式体现。而显然，开发者设定规则对应的具体表达（例如代码），与 ChatGPT 生成内容在表达层面并不相同或实质性相似，开发者设定规则的行为并不会使得任一项 ChatGPT 生成内容被评价为作品。

### **(3) 普通用户与开发者并无创作合意，即使极少数情况下普通用户发布指令的过程和/或数据库答案构成了作品，二者间也不属于合**

## **作创作**

最后，假设极少数情况下普通用户的指令具有独创性，开发者创作或者来源于其他第三方的数据库答案均也构成作品，普通用户与开发者/第三方也不可能构成合作作者。

根据著作权法的理论，合作作品的创作需要合作作者之间进行合意创作。开发者在开发过程中虽然会设想用户可能询问的问题，但其在准备相关答案前并不是抱着合作创作目的准备的，其目的是尽可能实现用户的指令；而普通用户角度更无法预测开发者设定的数据内容。因此，普通用户与开发者不可能被认定为合作作者。

### **(三) 对于人工智能生成物的保护需要新思路**

基于上述分析，依据现有的著作权理论，不论是基于主体考虑，还是基于人工智能生成物的生成过程考虑，人工智能生成物本身很难获得著作权法的保护。

但是，如果不给予其任何保护，是否会影响人们后续利于人工智能的积极性？有观点认为，应当考虑通过立法给予人工智能以著作权法保护，激励人们尝试用人工智能“创造”更多的知识，提升社会生产力。我们认为，对立法的修改还是需要谨慎，至少不宜调整“作品”“创作”的定性逻辑，可以考虑给予“邻接权”保护或者纳入虚拟财产保护，并制定特殊的归属、使用规则。

### **三、第三个问题：使用 ChatGPT 生成的内容有法律风险吗？**

回到本文最初客户问我们的问题，使用 ChatGPT 生成的内容是否存在法律风险？我们认为会存在，主要体现在如下方面：

#### **(一) ChatGPT 生成内容的知识产权风险**

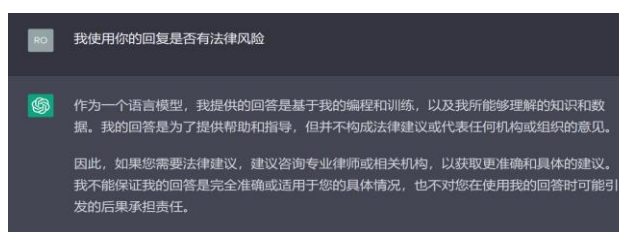
如前述分析，ChatGPT 生成的内容来源于其背后庞大的数据库。从 OpenAI 的角度，我们理解对数据库内的所有训练素材进行“清权”（clearance）显然是不现实的。另外，虽然 ChatGPT 生成内容可

<sup>9</sup> <https://openai.com/terms>

<sup>10</sup> 至少就条款表述的逻辑而言，OpenAI 可以提出该主张。

能存在相关素材的拼接、组合，但也不排除会有对相关表达直接呈现的可能性。如果该等表达未经原始著作权人授权而被使用，则存在侵权风险（对模型开发方、内容使用者均有风险）。虽然模型开发方、内容使用者可以视情况考虑以“合理使用”、“公有领域素材”等进行抗辩，但也无法全面排除相关风险。例如，在 *Andersen v. Stability* 案<sup>11</sup>中，原告就认为 Stable Diffusion（AIGC 绘图软件）在未经许可的情况下，从互联网上收集了数十亿张图像上进行训练，包括使用原告所拥有的图像的行为，构成侵权。

ChatGPT 开发方也意识到了该问题，在用户协议中提供了版权投诉渠道。并且，询问 ChatGPT 该问题后，其也“谨慎”地向用户提示了风险。



（截图来源于我们与 ChatGPT 的对话）

## （二） ChatGPT 生成内容的合规风险

根据 OpenAI 的说明，尽管 OpenAI 已在努力让 ChatGPT 拒绝用户不合理的请求，但是 ChatGPT 还是可能会存在算法歧视和生成不当内容的风险，例如带有脏话、仇恨、歧视等言论，也可能生成暴力、血腥、色情等违反法律以及公序良俗的内容。

从国内合规的角度，如果不经任何审核直接使用 ChatGPT 的相关内容，将可能引发《网络信息内

容生态治理规定》《互联网信息服务算法推荐管理规定》等法律、法规、政策项下风险。

我国自 2023 年 1 月 10 日起施行的《互联网信息服务深度合成管理规定》也特别明确，任何组织和个人不得利用深度合成服务制作、复制、发布、传播法律、行政法规禁止的信息，不得利用深度合成服务从事危害国家安全和利益、损害国家形象、侵害社会公共利益、扰乱经济和社会秩序、侵犯他人合法权益等法律、行政法规禁止的活动。

## （三） ChatGPT 生成内容的个人信息及商业秘密泄露风险

用户为使 ChatGPT 满足个人和/或商业方面的需求，可能会在聊天框内提供个人信息或者企业的商业秘密。从目前的 OpenAI 条款看，其并未对用户 Input 内容提供特别的保护，相关内容将被授权用于“开发和改进服务”。因此，存在个人信息及商业秘密泄露风险。

## 结语

随着 ChatGPT 爆火，国内企业也纷纷加入 AI 领域竞赛，例如万众期待的百度文心一言（英文名：ERNIE Bot）推出后，爱奇艺、火狐、360 等多家企业纷纷宣布接入<sup>12</sup>。

作为律师，虽然已被一些评论列入未来可能被 AI 替代的职业之一（笔者对此持保留态度），但是由于与 ChatGPT 等生成式人工智能相关的技术滥用、数据保护、舆论风险、侵权防范等问题层出不穷，可以预见律师在此领域的工作会更多，而并不是更少！

黄荣楠 合伙人 电话：86 21 2208 6280 邮箱地址：huangrn@junhe.com

马钦奕 律师 电话：86 21 2283 8324 邮箱地址：maqy@junhe.com

（实习生林泓岚参与了案例与新闻的搜集工作，对本文亦有贡献）

本文仅为分享信息之目的提供。本文的任何内容均不构成君合律师事务所的任何法律意见或建议。如您想获得更多讯息，敬请关注君合官方网站“www.junhe.com”或君合微信公众号“君合法律评论”/微信号“JUNHE\_LegalUpdates”。



<sup>11</sup> *Andersen et al v Stability AI Ltd et al* 3:23-CV-00201

<sup>12</sup> <https://finance.sina.com.cn/tech/digi/2023-02-15/doc->