

## AIGC

### 概念

继『元宇宙』和『Web 3.0』后，AI 绘画与 ChatGPT 相继爆火出圈，这也让『AIGC』概念在网络上窜红，并与前两概念一同入选了2022年的百度沸点科技热词榜单。

元宇宙、Web 3.0、数字藏品、量子纠缠、数字人、**AIGC**、数字经济、光刻机、计算生物学、碳中和。<sup>[1]</sup>

目前国内产学研各界对于 AIGC（Artificial Intelligence Generated Context，人工智能生成内容）的理解如下：

AIGC 是继专业生成内容（PGC，Professional Generated Content）和用户生成内容（UGC，User Generated Content）之后利用人工智能技术自动生成内容的新型生产方式。<sup>[2]</sup>

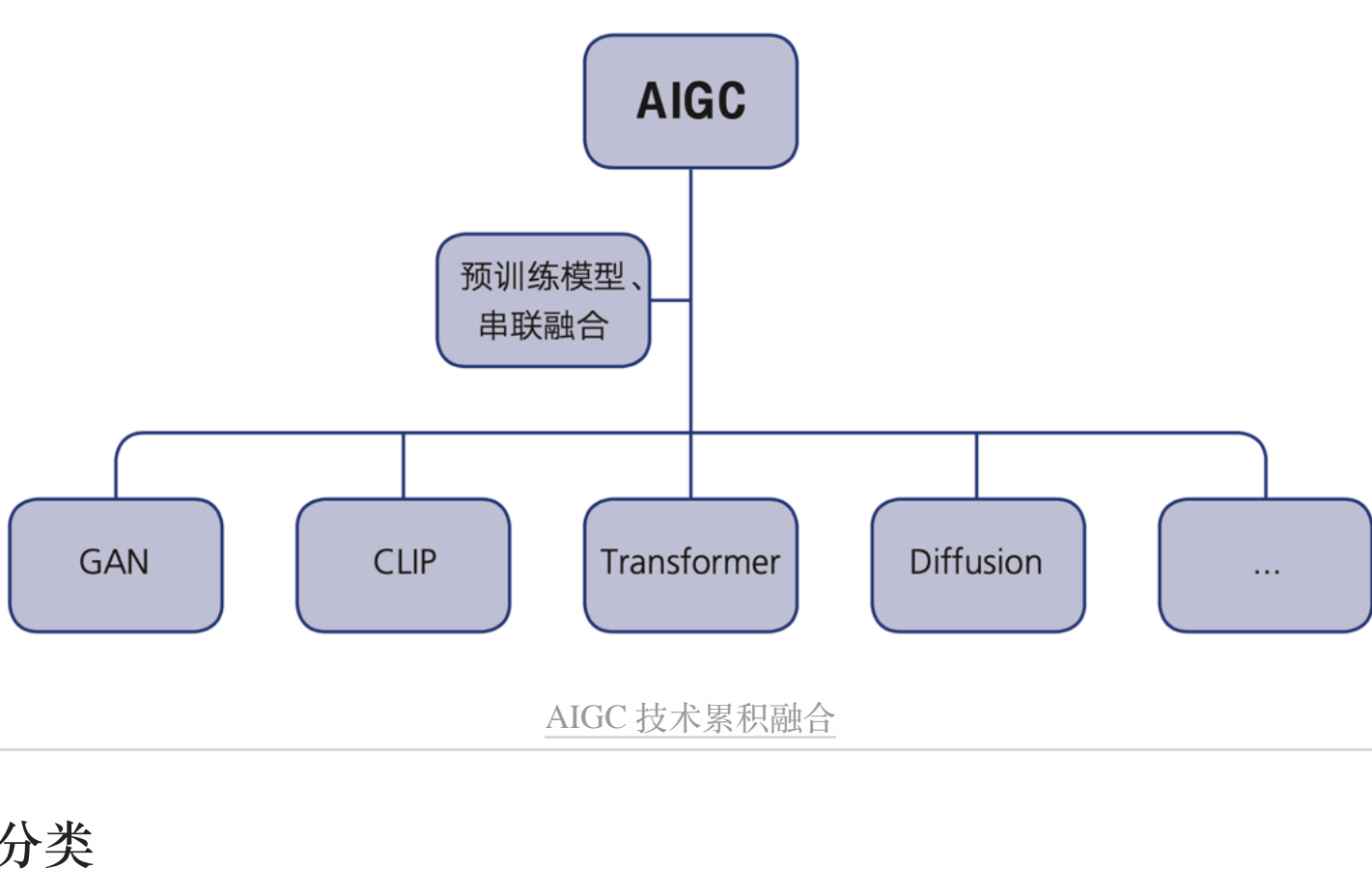
### 理解

AIGC 可以是一种生成内容的分类，侧重点在于对已自动生成内容的打标。以 AI 绘画为例，只需短短十秒钟，AI 就可以根据给出的关键词完成它的『大作』。随着 AI 绘画的大火，有插画网站也单独为这一部分作品进行了分类<sup>[3]</sup>。训练有素的 AI 不仅能保证绘画速度，也能保证绘画质量，但不可否认的是伴随而来的是伦理、版权等一系列问题。



知名插画网站对 AIGC 的分类标签

AIGC 也可以是一种内容生成的方式，侧重点在于自动生成内容的手段。以写诗为例，古代文人可以写诗，AI 也可以写诗<sup>[4]</sup>。短短几秒一首七绝就展示在面前，让大家也体验『当诗人』的感觉。但是，诗歌创作讲究韵律和意境，通过 AIGC 创作出的诗歌仍有很大的进步空间（下面这首诗是我在多次『创作』后挑选的评分较高的一首<sup>[5]</sup>）。



通过 AIGC 生成七言绝句

AIGC 还可以是自动生成内容的技术合集，侧重于概括自动生成内容的一系列技术。2023年1月31日腾讯发布《AIGC 发展趋势报告：迎接人工智能的下一个时代》<sup>[6]</sup>，报告指出生成算法、预训练模型、多模态等 AI 技术累积融合催生了 AIGC 的大爆发，AIGC 技术包括生成对抗网络（GAN，Generative Adversarial Network）、CLIP 模型（Contrastive Language-Image Pre-Training）模型、Transformer 模型、扩散模型（Diffusion Model）等等，具体可参加报告内容。



AIGC 技术累积融合

### 技术场景与成熟度分类

在 AIGC 的相关技术场景和成熟度分类上，量子位智库在其发布的《AIGC 深度产业报告》中作出了详细说明<sup>[6]</sup>。报告中指出：

AIGC 可分为音频生成、文本生成、图像生成、视频生成及图像、视频、文本间的 跨模态生成。细分场景众多，跨模态生成需要重点关注。

Game AI 和虚拟人生成是目前变现较为明确的两大综合赛道，预计对原行业产生整体性影响。



AIGC 相关技术场景及成熟度分类

*Every coin has two sides.*

AIGC 在引领技术新趋势和产业发展的同时，也可能带来诸如知识产权保护、安全、技术伦理、环境影响等风险挑战。相关从业者理应秉承负责任的态度，打造安全可信的 AIGC 应用和产品，实现 AIGC 技术高质量、可持续的发展。

## ChatGPT

ChatGPT 是美国 OpenAI 公司训练的以对话方式进行交互的 AI 模型，属于是 AIGC 技术的一个典型应用。OpenAI 在官方博客<sup>[7]</sup>中指出 ChatGPT 模型是 **InstructGPT**（一个比 GPT-3<sup>[8]</sup> 更好跟踪用户意图的模型）的兄弟模型，可以根据对话内容格式来回答问题，或对对话中不正确的情形提出质疑，亦或是拒绝回答一些违背道德或伦理的请求。

The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests. ChatGPT is a sibling model to InstructGPT, which is trained to follow an instruction in a prompt and provide a detailed response.

### 训练机制

ChatGPT 基于 GPT-3.5 系列中的模型进行了微调，并采用来自**人类反馈的强化学习**（RLHF，Reinforcement Learning from Human Feedback）进行训练，训练方法与 InstructGPT 相同，只是在数据收集方面略有差异。

OpenAI 采用监督学习的方式训练一个初始模型，然后由 AI trainers 分别扮演用户和 AI 模型进行对话，以便于更好地对模型的改进和优化提供建议。

We trained this model using Reinforcement Learning from Human Feedback (RLHF), using the same methods as InstructGPT, but with slight differences in the data collection setup. We trained an initial model using supervised fine-tuning; human AI trainers provided conversations in which they played both sides—the user and an AI assistant. We gave the trainers access to model-written suggestions to help them compose their responses.

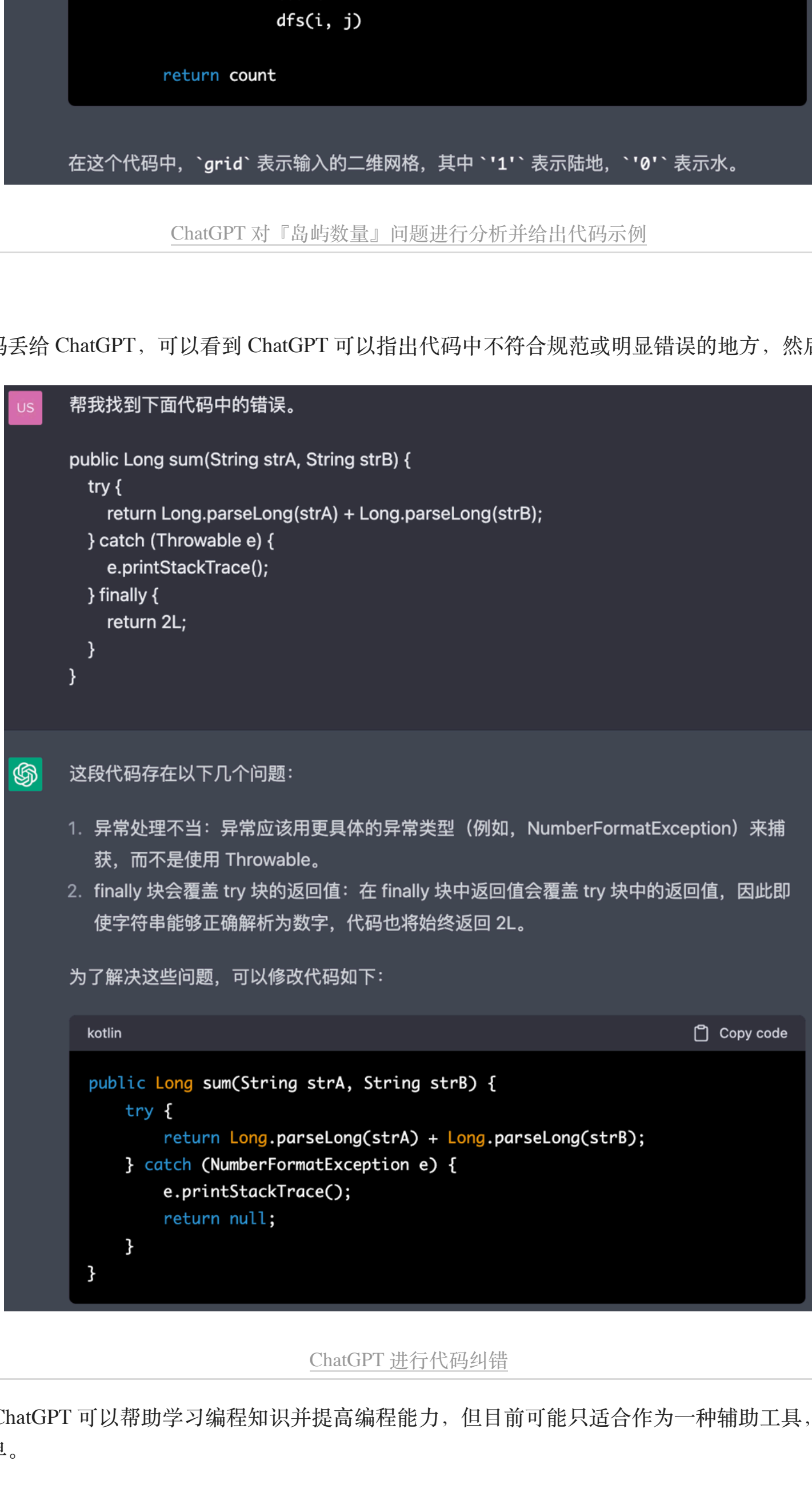


ChatGPT 的训练流程

### 能力体现

#### 文本生成

这里笔者简单测试了 ChatGPT 在给定命题<sup>[9][10]</sup>场景下生成文本的能力，可以看出在这种场景下 ChatGPT 得出的结果还是不错的，但是也不可避免出现车轱辘话来回说的情况，作为参考倒是不错的。



ChatGPT 在给定命题场景下的文本生成结果

### 编程能力

#### 解答代码问题

在给定问题是编程题的前提下，向 ChatGPT 询问『[岛屿数量](#)』问题，可以看到其先对问题做了一些分析然后很快给出了相应的代码示例，将这段代码放到 LeetCode 上提交可以通过全部用例。



ChatGPT 对『岛屿数量』问题进行分析并给出代码示例

#### 进行代码纠错

接下来笔者将一段错误代码丢给 ChatGPT，可以看到 ChatGPT 可以指出代码中不符合规范或明显错误的地方，然后给出规范正确的写法。



ChatGPT 进行代码纠错

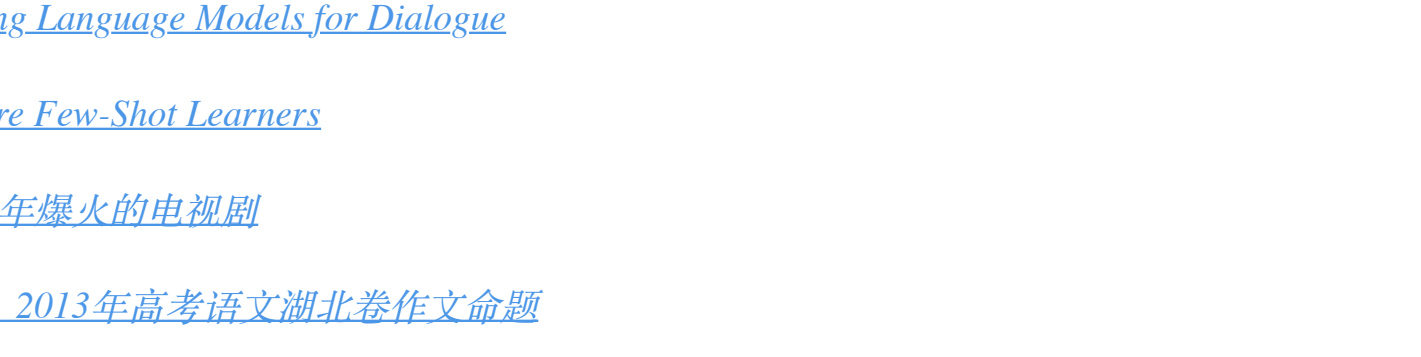
在编程能力上，个人觉得 ChatGPT 可以帮助学习编程知识并提高编程能力，但目前可能只适合作为一种辅助工具，替代程序员进行业务逻辑的代码开发可能还为时尚早。

### 拒绝回答不合适的问题

ChatGPT 经过迭代更新，对非法或违反道德或伦理的问题会拒绝回答。



GPT-3.5 以《红楼梦》写作风格进行文本生成



GPT-4 以《红楼梦》写作风格进行文本生成

从以上结果看出，GPT-3.5 看上去只是完成了文本生成的任务，似乎不能学习我们提供的语料。然而，GPT-4 看上去是能够根据我们的训练素材进行学习，然后完成任务。



更多关于 GPT-4 的内容，可以参见微软研究院最近发布的论文<sup>[11]</sup>。

## 参考

[1]: 2022 百度沸点

[2]: 人工智能生成内容 (AIGC) 白皮书

[3]: Pixiv 网站上的 AIGC 标签

[4]: 清华九歌诗歌写作系统

[5]: 腾讯研究院: AIGC 发展趋势报告——迎接人工智能的下一个时代

[6]: 量子位智库: AIGC 深度产业报告

[7]: ChatGPT: Optimizing Language Models for Dialogue

[8]: Language Models are Few-Shot Learners

[9]: 《狂飙》：2023 开年爆火的电视剧

[10]: 上善若水任方圆：2013 年高孝语文湖北卷作文命题

[11]: Sparks of Artificial General Intelligence: Early experiments with GPT-4