

# AIGC应用分享

张印帅

2023年3月27日

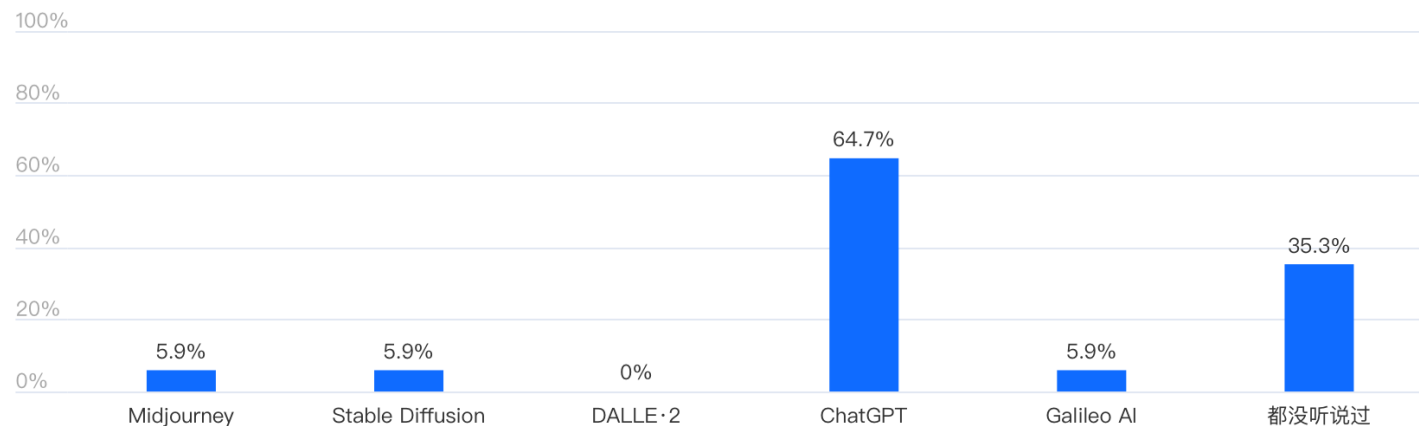
2022年美国科罗拉多州博览会上，一副AI绘制的作品《太空歌剧院》获得了“数字艺术/数字修饰照片”类别的一等奖。



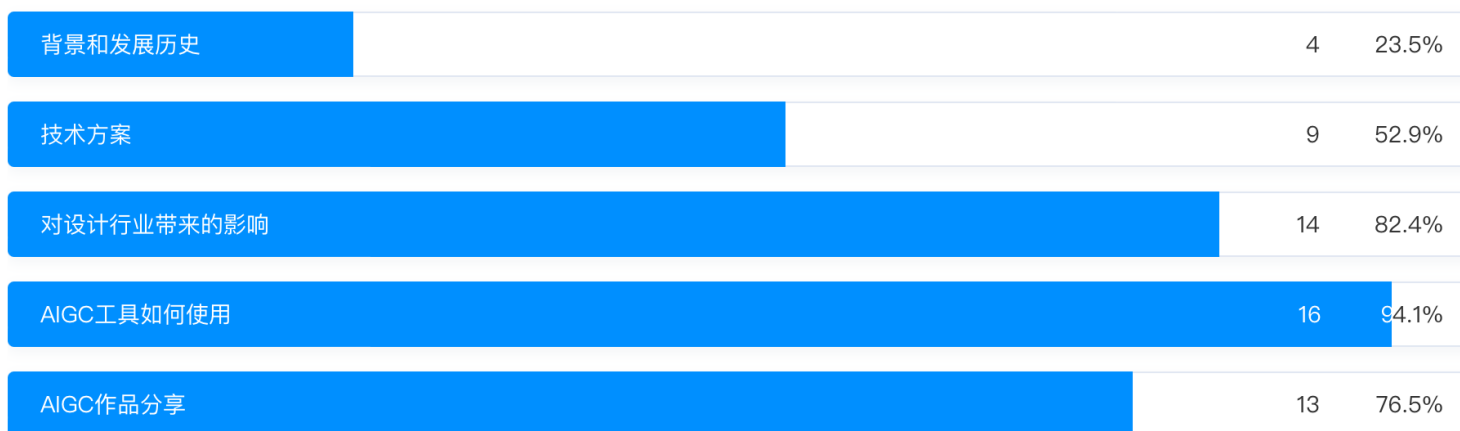
# 问卷结果反馈



是否熟悉下述工具（熟悉的定义是，使用过或者了解具体如何使用）



希望了解AIGC的哪些方面？



# Agenda

1. 发展历程
2. 当前的应用领域
3. 主流工具介绍
4. AIGC工具对设计工作流程的改变
5. 技术原理
6. MJ和Stable使用教学



# 人智交互 (HAI) 的愿景



让设备无处不在  
Minority Report 2002



混合现实  
Iron Man 2008



拟人化计算机  
Her 2013



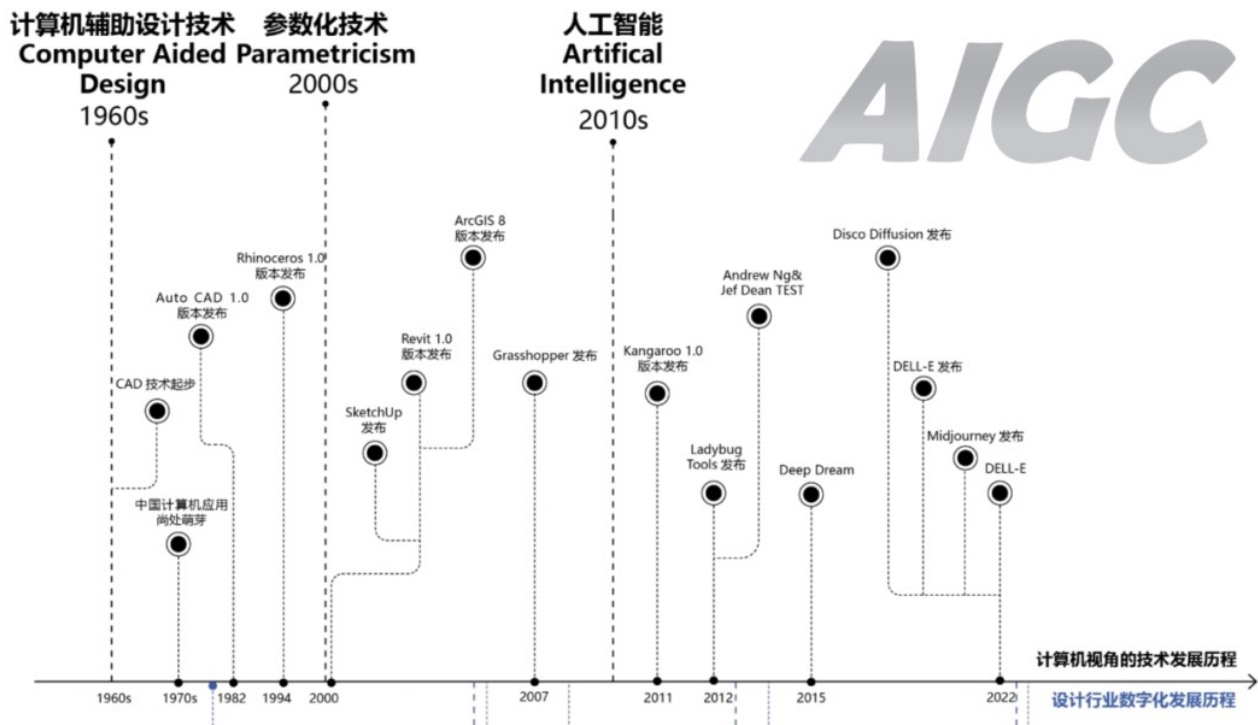
构建全新的虚拟世界  
Ready Player One 2018



# 计算机辅助设计的发展历程



上世纪70年代末期，传统建筑、土木、景观工程制图工作仍然处在依靠设计师手绘的传统模式；2006年，中科院计算所主导CAD软件研发，宣告中国设计行业正式进入了计算机辅助设计时代；2022年AI绘图工具和智能对话机器人的井喷式发布，再一次让设计行业震动。



手绘制图  
Manual Drawing



2D 制图  
2D Pictures



3D 展示和渲染  
3D Modeling & Rendering



参数化辅助设计  
Parametricism for Design



AI 辅助设计  
Artificial Intelligence for Design

# 当前AIGC的应用领域

全部图片均为23年3月26日Midjourney上生成的最新效果



建筑/室内设计



产品设计



汽车设计



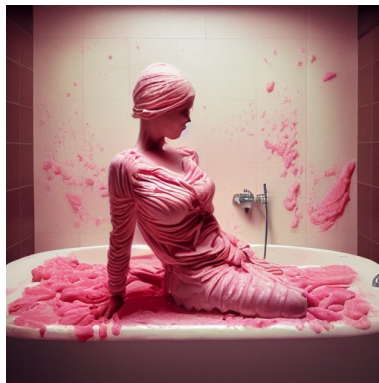
Logo设计



摄影



插画



雕塑



IP形象设计



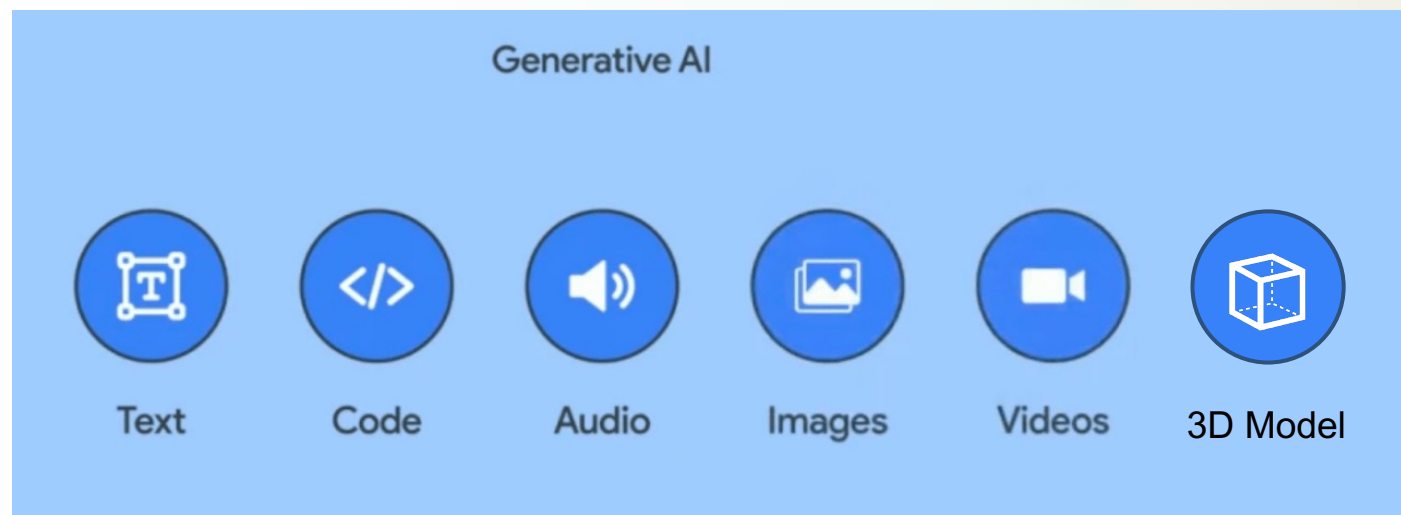
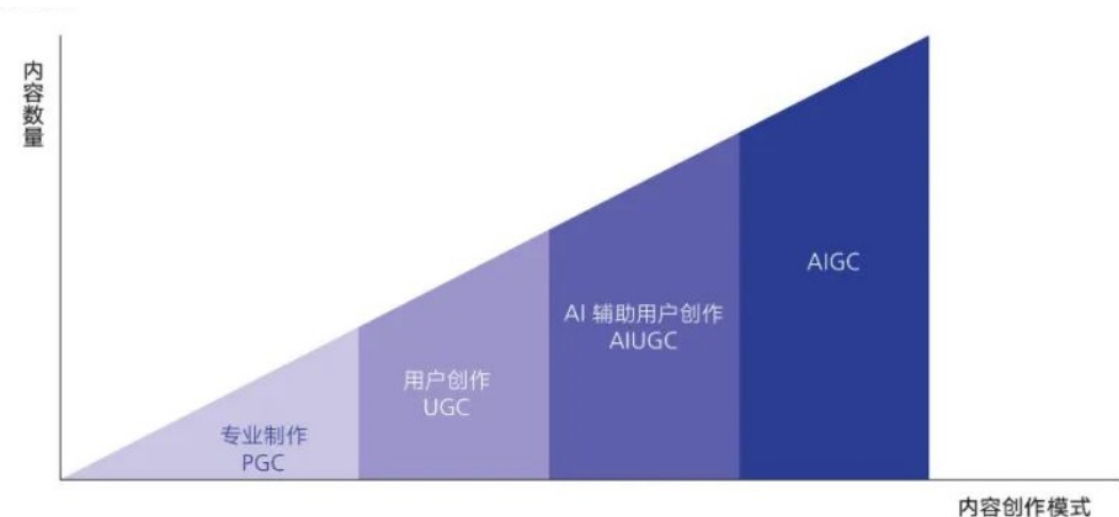
人像

# AIGC能够生成哪些内容



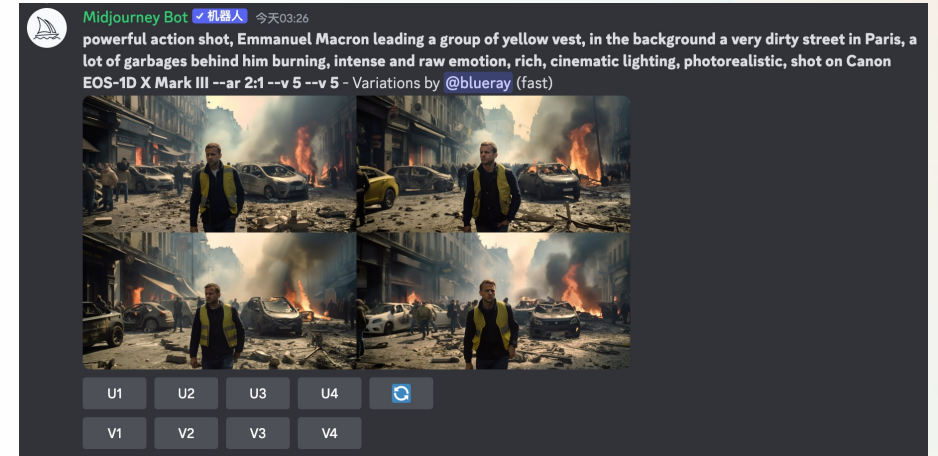
AIGC是一种内容创作模式的变革。生成式AI可以产出的内容包括文字、代码、音频、图像、视频、3D模型6个类别，每一次AI内容的生成一定是基于某一个人类输入的内容，产生某一个输出的内容，因此，在上述6个类别下一共有36种可能的多模态内容创作范式。

如，文生图，图生视频，视频生模型等。





# AIGC如何生成内容，以文生图为例



强大的动作镜头，埃马纽埃尔-马克龙带领一群黄背心，背景是巴黎一条非常肮脏的街道，他身后有很多垃圾在燃烧，强烈而原始的情感，丰富的、电影般的灯光，逼真的，用佳能EOS-1D X Mark III拍摄 --ar 2:1。  
powerful action shot, Emmanuel Macron leading a group of yellow vest, in the background a very dirty street in Paris, a lot of garbages behind him burning, intense and raw emotion, rich, cinematic lighting, photorealistic, shot on Canon EOS-1D X Mark III --ar 2:1

# 主流工具介绍



## 文生文

ChatGPT  
New Bing  
Notion AI  
文心一言  
...

## 文生图

Copilot  
Stable Diffusion  
Mid journey  
文心一格  
鹿班  
Magic 3D  
Galileo UI  
Meta AI  
BrandMark  
...

## 图生文

ChatGPT  
文汇  
BrandMark  
...

## 图生图

DALL.E  
ClipDrop  
Adobe Firefly  
Uizard  
...

工具名称	类别	描述
ChatGPT	文生文	OpenAI 开发的 AI 聊天机器人，能够生成人类级别的文本。
Stable Diffusion	文生图	一种生成式对抗网络 (GAN)，用于生成高质量的图像。
Mid Journey	文生图	一种生成式对抗网络 (GAN)，用于生成高质量的图像。
文心一言	文生文	百度开发的 AI 聊天机器人，能够生成人类级别的文本。
...	...	...

文生文

A



图生图



文生文

文生图

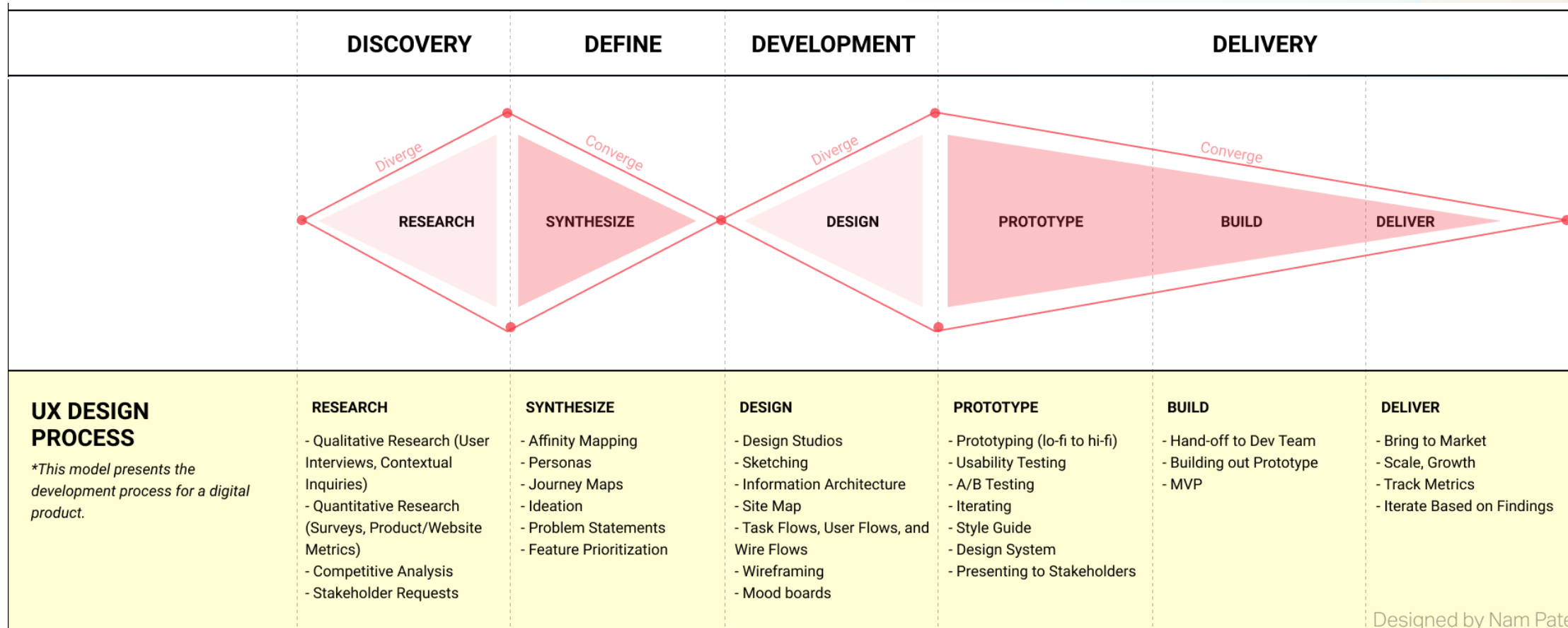
图生文

图生图

# AIGC对设计工作流程的改变



the British Design Council formalized its famous Double Diamond design-process model in 2005



Designed by Nam Patel

通过会议纪要、项目文档等生成设计需求总结

生产竞品分析等调研报告

根据客户提供图片形成关键词

生成功能清单、用户画像等

把用户旅程图等设计文档可视化

探索设计方向

生成设计概念图

进行参与式设计

基于KV批量生成设计文档

基于用户反馈细化和调整设计

生成用于开发的设计需求文档

收集设计反馈, 制定量化指标

生成设计汇报材料(图+文)

# 示例：一个咖啡厅的室内设计



客户需求：设计一个咖啡厅  
面向人群：35岁左右互联网公司员工  
风格偏好：ins风、网红

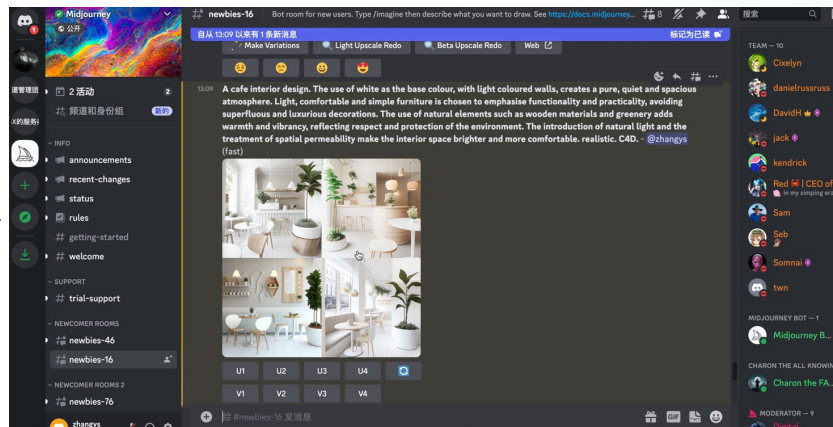


## 标准设计作业流程 SOP (Standard Operating Procedure)

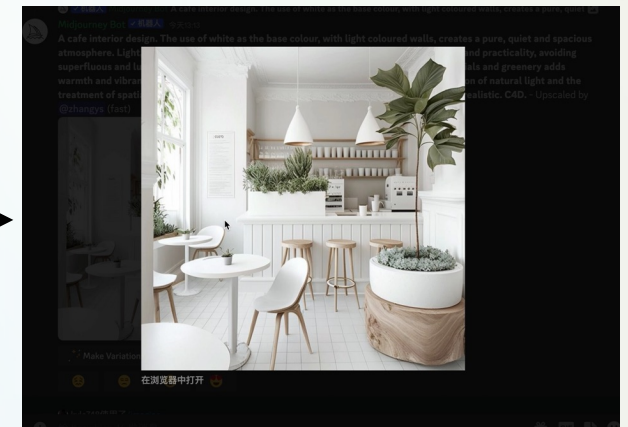
### Step1: New Bing生成设计关键词



### Step2: 关键词作为prompt输入MJ



### Step3: 细化设计出图



# AIGC的核心技术： CLIP 跨模态深度学习模型

## CLIP: Connecting text and images

我们正在引入一种名为 CLIP 的神经网络，它可以从自然语言监督中有效地学习视觉概念。——OpenAI.com



# CLIP带来的核心变化

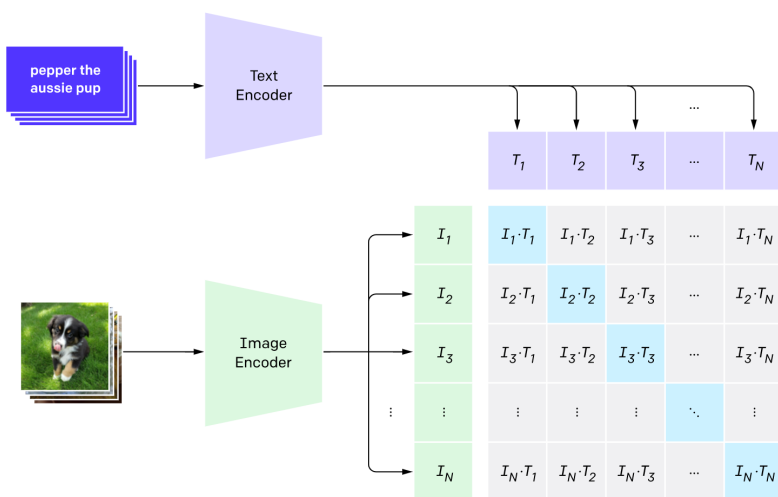


整个AI绘画（text-to-image）领域都建立在CLIP的基础上。CLIP是由OpenAI在2021年推出的一个神经网络模型，它可以从自然语言中学习视觉概念，从而低成本地进行图像识别。

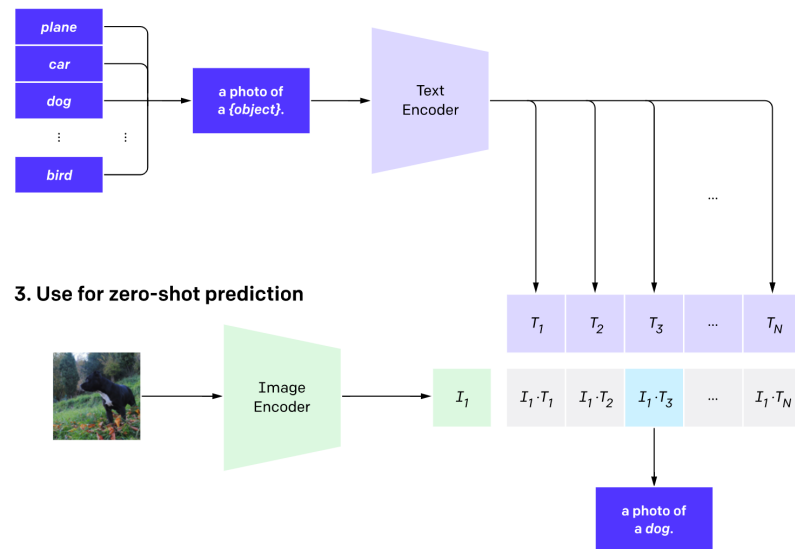
- 过去的AI图像识别建立在“分类”的思维之上。例如，某野生动物园想识别园区里的大象和老虎，那么他们可以分别采集大量大象和老虎的照片，输入到AI模型之中。这样，AI就能学习到两种动物分别的视觉特征，在看到一只动物时，可以确定它的特征更接近“大象”还是“老虎”，从而进行识别。
- 而CLIP建立在自然语言与图像相对应的思维之上。在训练这个模型时，研究者输入了大量图像和对应的描述文本。CLIP分别学习图片的特征和相应文本的特征，从而建立了一个能将二者对应起来的模型。这样，在看到一张新图片时，CLIP就可以将它直接转化为文字描述。

简而言之，CLIP不是预测标签文本，而是训练预测该图像与该文本对应的可能性。而既然可以将图片转换为文字，自然也可以反其道而行之，于是text-to-image这一领域出现了。

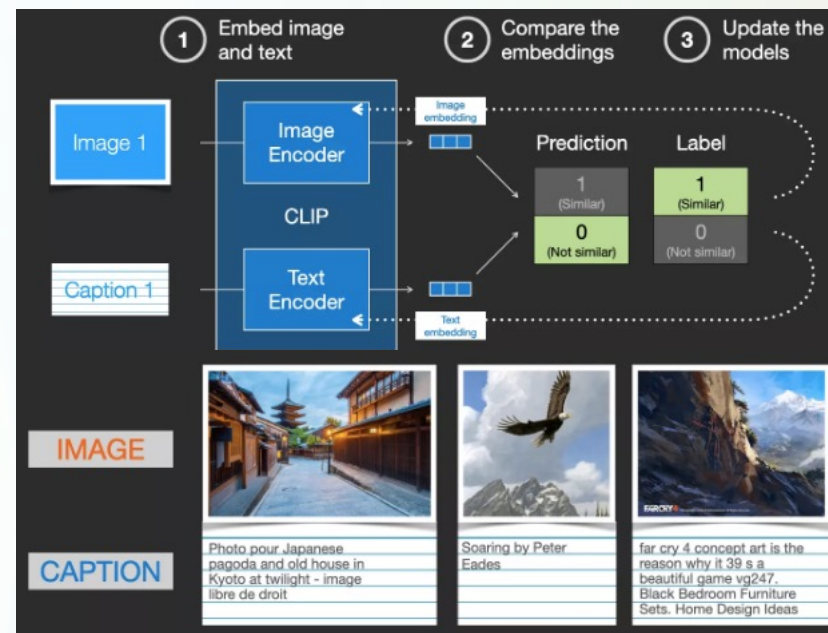
## 1. Contrastive pre-training



## 2. Create dataset classifier from label text



## 3. Use for zero-shot prediction

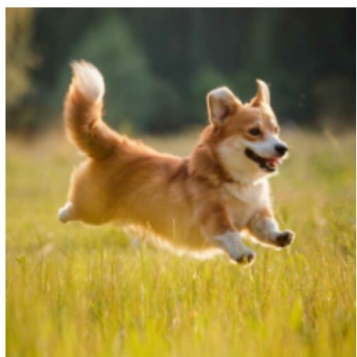


如何训练 CLIP

# 建立在“分类”思维之上的AI图像识别



<https://teachablemachine.withgoogle.com/>



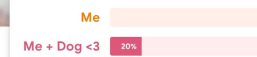
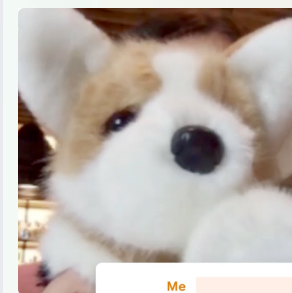
Output



Output



The screenshot shows the Teachable Machine interface with two classes: Corgi and British Shorthair Cat. The Corgi class has 5 image samples and is currently selected. The training progress bar is at 100%. The training settings are: Epochs: 50, Batch Size: 16, Learning Rate: 0.001. The output shows Corgi at 100% and British Shorthair Cat at 0%.



# 自然语言的AI图像识别



<https://replicate.com/methexis-inc/img2prompt>

Input



Output

图像说明Caption

Caption: a brown and white dog  
running through a field

Output

提示词Prompt

a brown and white dog running through  
a field, a stock photo by Wolfgang  
Zelmer, shutterstock contest winner,  
rayonism, majestic, dynamic pose,  
glorious

# 通过ControlNet实现控制

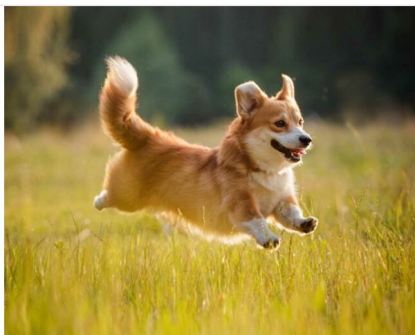


ControlNet 是一种神经网络结构，通过添加额外条件来控制扩散模型，是 AI 图像生成的游戏规则改变者。它为 Stable Diffusion 带来了前所未有的控制水平。ControlNet 的革命性之处在于它解决了空间一致性问题。以前根本没有有效的方法来告诉 AI 模型要保留输入图像的哪些部分，ControlNet 通过引入一种方法来改变这一点，使稳定扩散模型能够使用额外的输入条件来告诉模型确切地做什么。

<https://stablediffusionweb.com/ControlNet>

## Control Stable Diffusion with Canny Edge Maps

Image



✎ ✕

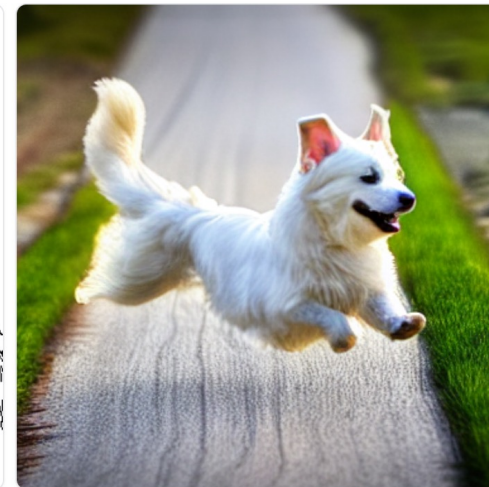
Prompt

A white dog running on the road

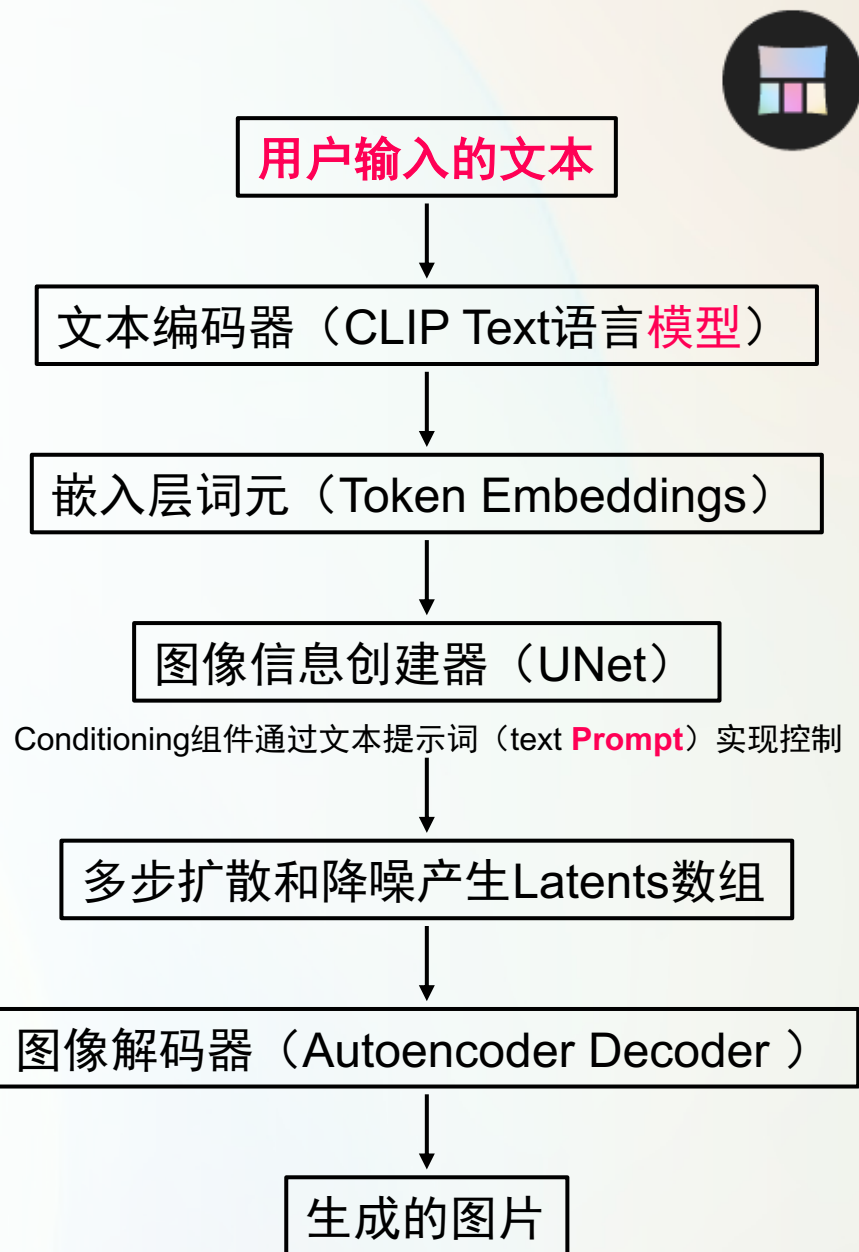
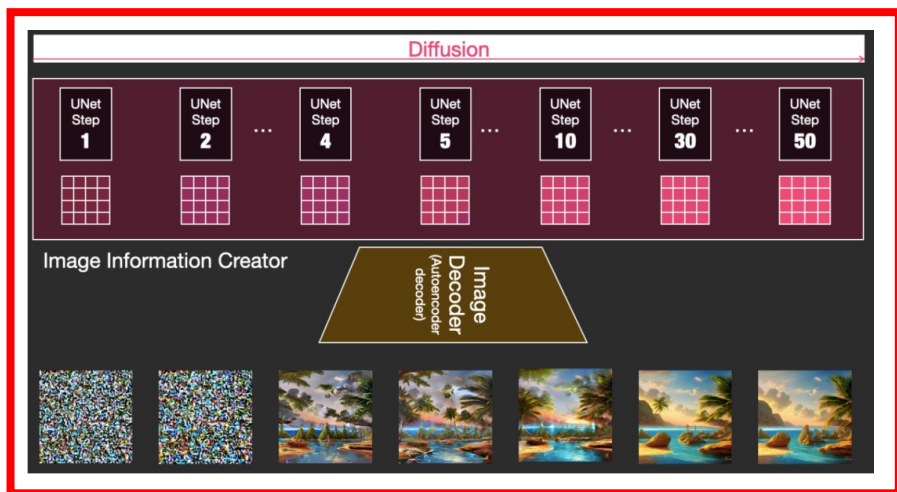
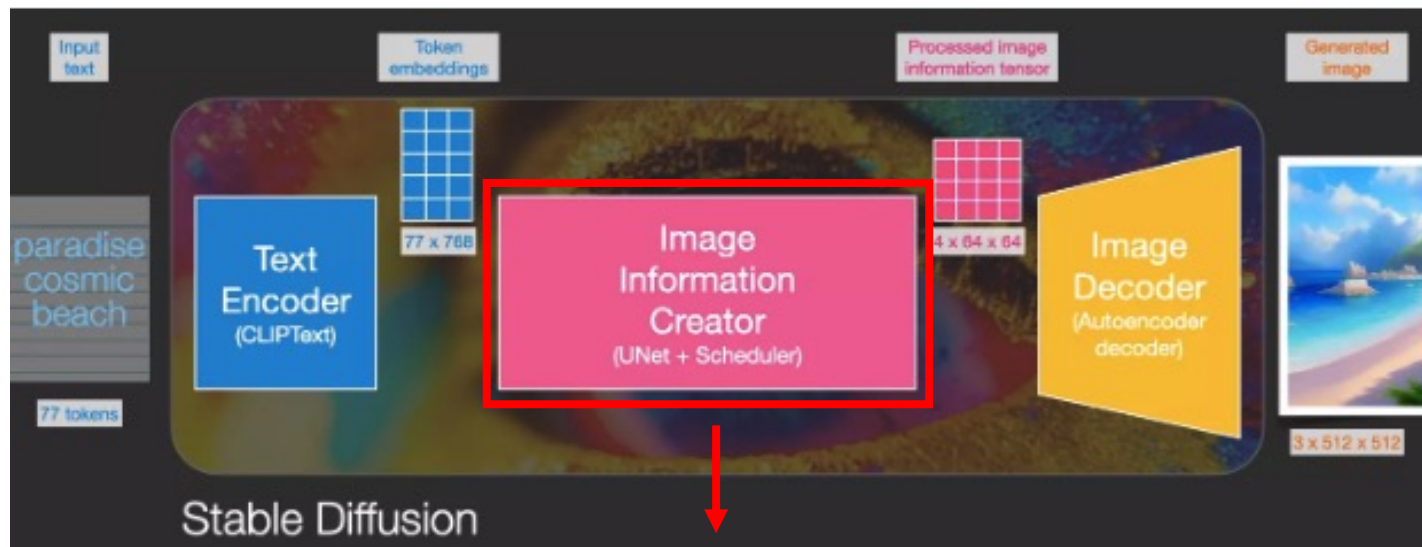
Run

Advanced options

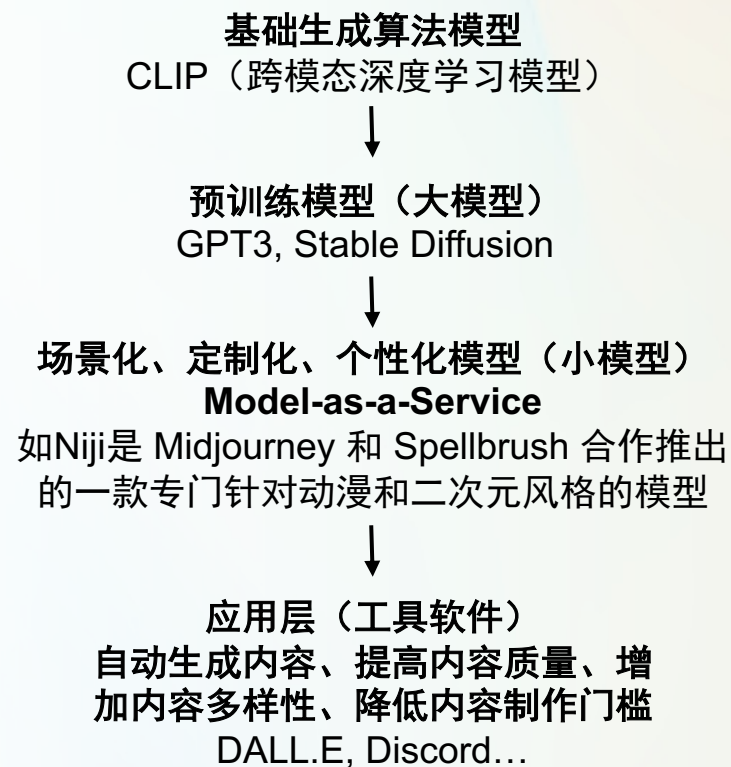
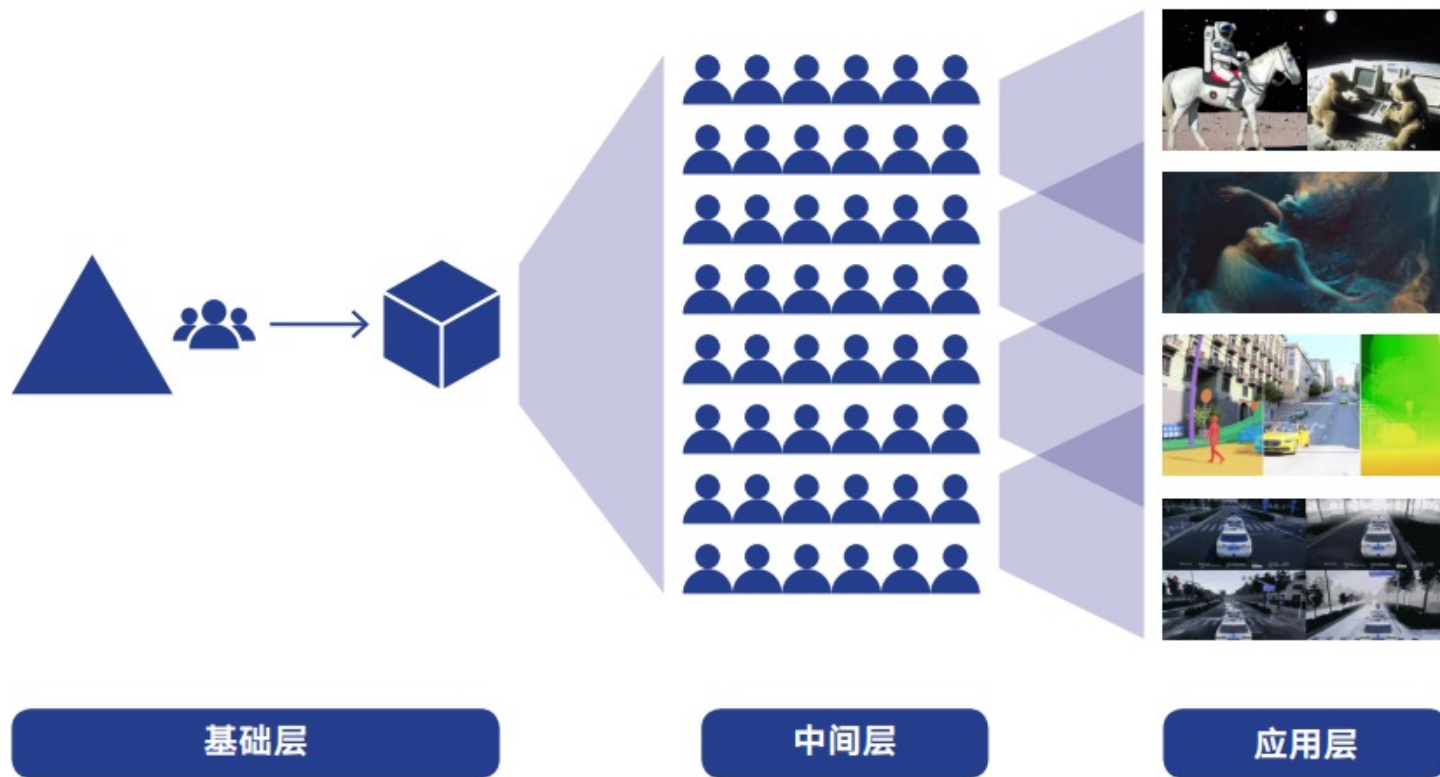
Base model



# 文本生成图像技术原理



# AIGC产业结构的三层结构



# Thanks



设计团队 / 2021年1月

# 设计师的类型



三、工业设计工程领域设置产品设计和体验设计两个专业方向。

产品设计方向包括设计策略、产品规划、用户研究、产品造型设计、色彩与材质设计、人机工学设计、产品结构工艺设计、产品展示设计、产品包装设计、服务流程和商业模式设计等工业设计技术岗位。

体验设计方向包括用户体验、交互设计、产品架构、品牌规划、艺术创意、界面设计、视觉设计、原画设计、角色设计、动效设计、多媒体设计、服务设计等工业设计技术岗位。

视觉传达设计、环境设计、景观设计、室内设计、工业设计、建筑设计、服装设计、扎染艺术、陶瓷艺术设计、装饰艺术、珠宝首饰设计、展示艺术设计、品牌广告设计、戏剧影视美术专业、动画专业、数字媒体艺术。摄影、产品设计、公共艺术、信息设计、形象设计、数字出版与展示设计美术学、会展艺术设计、娱乐与衍生设计美术学、游戏艺术设计、影视制作、工艺美术、艺术与科技专业、照明艺术设计、陈设艺术设计、公共绘画、城市艺术工程。



# CLIP与格式塔心理学相似



格式塔心理学派（Gestalt psychology）是20世纪初期出现的一种心理学理论，它强调整体（Gestalt）的观念，认为人类的感知、认知和行为不是由单个元素组成的，而是由整体形态构成的

例如，当我们看到一只鸟飞过天空时，我们不是单独地感知鸟的翅膀、身体、头部等部分，而是将它们组合成一个整体，意识到它是一只鸟。

# 主流软件介绍



## DALL·E 2

这是一个AI模型，同时也是一项服务。目前，它专注于特定（而非通用）的用途，但日后它的用途将会扩展。

它最棒的特性是“inpainting”。译者注：一个更接近于现实生活中甲乙双方沟通的改图功能，可以在生成的图片上选择某些部分让AI加以修改。然而，它生成的图片比较随机，因此（相比于生产用途）更适合寻找灵感。同时，它更适合企业使用，这是因为它只使用了已获授权的图片集来进行训练。译者注：换言之，掐灭了未来可能的纠纷。纠纷在于：AI对网络图片进行学习，这一行为是否是正当的？应当将其看作是类似人类画家和摄影师学习大师作品的行为，还是看作“使用”了这些图片？

## MidJourney

David Holz（MidJourney的开发者）是一个很有远见的技术研究者，他所关注的是人机互动的模式。MidJourney不是一家进行后端开发的企业，而是一个研究人如何与新技术互动、又如何被新技术所影响的实验室，关于这一点，更详细的内容可以去看他最新接受的采访。

MidJourney特别注重风格，它生成的图像风格十分独特。译者注：个人最近在高强度使用MidJourney，对这一点深有体会。它的默认风格就不同于另两个产品的写实风，而更具备插画感。因此，尽管图像生成能力逊于Stable Diffusion，在我看来也是不可替代的。

目前，MidJourney和大部分市面上的AI绘画应用采用的是同一个模型（但这只是现在的情况，他们很快就会使用新的模型）。译者注：这里说的应该是谷歌开源的Disco Diffusion。国内的一些AI绘画应用，例如滴墨社区的Domo大画家，用的应该也是这一套，但缺乏像MidJourney那样大量的调试。不过，他们在生成图片的一致性和连贯性上做了很多努力，尽管他们输出的图像是随机的，但也不是完全不受控制。

MidJourney不是开源的，不过开发者本人在过去的职业生涯中已经开源了很多代码，这没什么不好——不是所有代码都需要开源。而且，总体来说这是一个很好的应用，日后可能起到令人惊讶的作用。

## Stable Diffusion

这是一个建立在合作开发之上的模型，并且我们很快就会将它开源发布。译者注：目前已经开源，可以在本地部署（免费，需要自己有性能较好的显卡）、使用官方的DreamStudio（花钱购买高性能GPU的使用时长）或是使用第三方部署的版本（免费但速度慢），链接见文末。

Stable Diffusion将会是未来“图像基础设施”的一部分，无论是艺术创作、产品设计还是任何与图像相关的应用都可以用到它。它是一个在各种领域通用的模型。

由于它是开源的，很快就会有許多围绕这一模型建立起来的应用。例如，不久以后我们就将发布DreamStudio，为消费者服务。

不过我们目前的工作重心是开发API，这将降低人们使用Stable Diffusion和未来更多AI模型的成本，让数以十亿计的人们能更好地互相交流。

这些已经和将要出现的AI模型，需要能够反映出人类的每一种文化，也需要能够与创作者合作、融入他们的工作流。为此，我们与本领域最顶尖的专家合作，不断努力达成上述目标。

由于我们的模型是通用性质的，它的输出内容比较宽泛。目前，大家能看到的是原始输出，没有经过特别的处理——如果对模型进行处理，它的表现会好得多。译者注：这里说的应该是，通过对通用模型进行修改，能让其在某一个特定领域中有非常突出的表现。不过，作为一个开源模型，每个人都可以使用我们的代码和数据集，在原始模型或其中某些元素的基础上进行改造，从而得到更好的生成结果。这真是太棒了。

我们将会有更多的工具、更多的选择，但最终我们所做的是为所有人提供了一种新的交流方式。我们将会有更多的市场、更多的细分市场，不是互相竞争而是互相合作。作者：DearUTAKi <https://www.bilibili.com/read/cv18290360> 出处：bilibili

Stability AI发布的开源模式Stability Diffusion，可以根据用户输入的文字描述（称为提示词，prompts）自动生成图像，即文生图（Text-to-Image, T2I）。Stability Diffusion、DALL-E 2、Midjourney等可以生成图片的AIGC模型引爆了AI作画领域，AI作画风行一时，标志人工智能向艺术领域渗透。

# 主流模型介绍



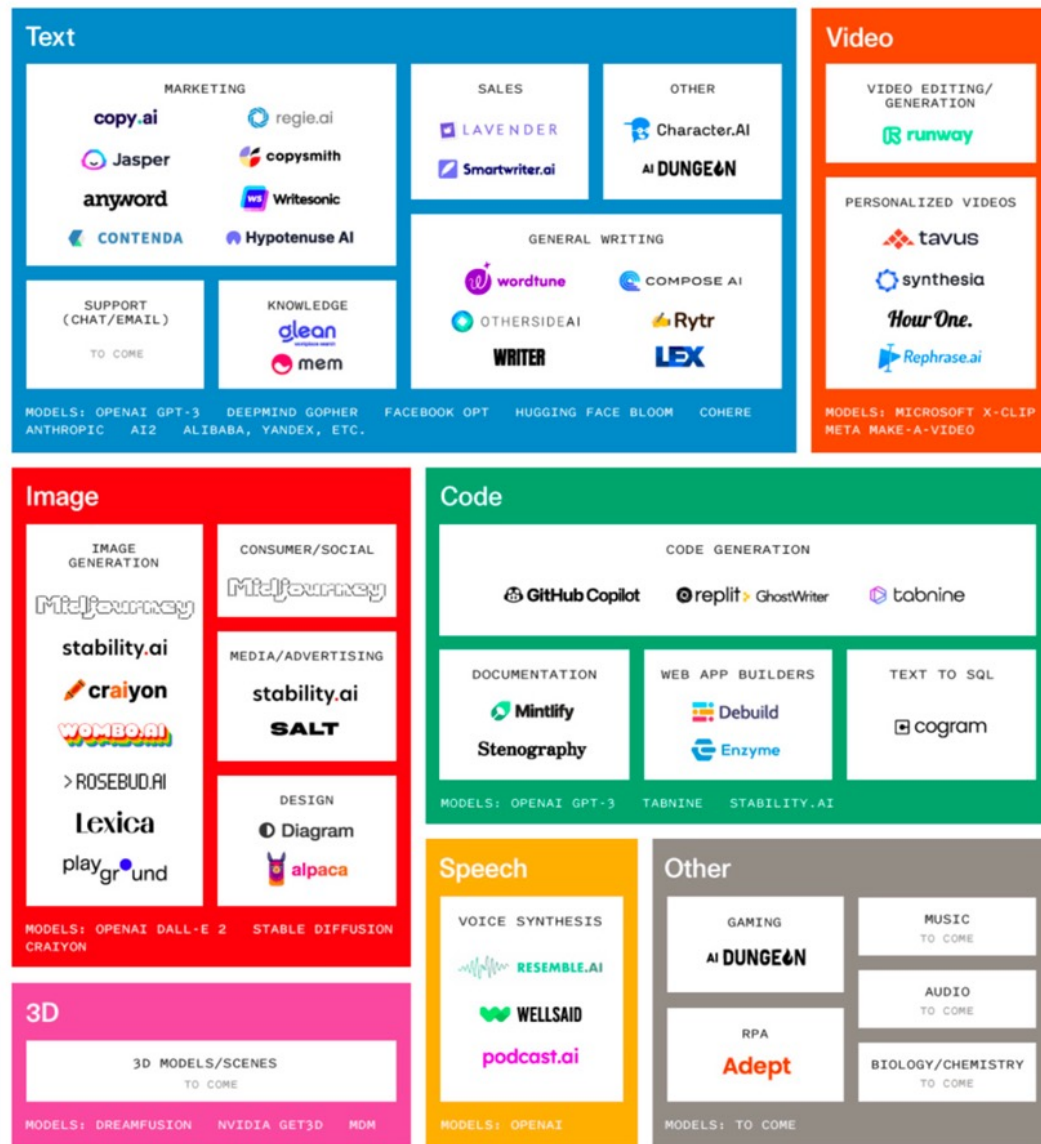
CLIP (Contrastive Language-Image Pre-Training) 模型

2021年

1、进行自然语言理解和计算机视觉分析；  
2、使用已经标记好的“文字-图像”训练数据。一方面对文字进行模型训练。一方面对另一个模型的训练，不断调整两个模型的内部参数，使得模型分别输出的文字特征和图像特征值并确认匹配。

	预训练模型	应用	参数量	领域
谷歌	BERT	语言理解与生成	4810 亿	NLP
	LaMDA	对话系统		NLP
	PaLM	语言理解与生成、推理、代码生成	5400亿	NLP
	Imagen	语言理解与图像生成	110亿	多模态
	Parti	语言理解与图像生成	200亿	多模态
微软	Florence	视觉识别	6.4亿	CV
	Turing-NLG	语言理解、生成	170亿	NLP
Facebook	OPT-175B	语言模型	1750亿	NLP
	M2M-100	100种语言互译	150亿	NLP
Deep Mind	Gato	多面手的智能体	12亿	多模态
	Gopher	语言理解与生成	2800亿	NLP
	AlphaCode	代码生成	414亿	NLP
Open AI	GPT3	语言理解与生成、推理等	1750亿	NLP
	CLIP&DALL-E	图像生成、跨模态检索	120亿	多模态
	Codex	代码生成	120亿	NLP
	ChatGPT	语言理解与生成、推理等		NLP
英伟达	Megatron-Turing NLG	语言理解与生成、推理	5300亿	NLP
Stability AI	Stable Diffusion	语言理解与图像生成		多模态

# AIGC的产业全景图



# AIGC技术发展趋势



	PRE - 2020	2020	2022	2023?	2025?	2030?
TEXT	Spam detection Translation Basic Q&A	Basic copy writing First drafts	Longer form Second drafts	Vertical fine tuning gets good (scientific papers, etc)	Final drafts better than the human average	Final drafts better than professional writers
CODE	1-line auto-complete	Multi-line generation	Longer form Better accuracy	More languages More verticals	Text to product (draft)	Text to product (final), better than full-time developers
IMAGES			Art Logos Photography	Mock-ups (product design, architecture, etc.)	Final drafts (product design, architecture, etc.)	Final drafts better than professional artists, designers, photographers)
VIDEO / 3D / GAMING			First attempts at 3D/video models	Basic / first draft videos and 3D files	Second drafts	AI Roblox Video games and movies are personalized dreams

Large model availability:



First attempts



Almost there



Ready for prime time

# 参考文献



中国信息通信研究院《人工智能生成内容（AIGC）白皮书》

腾讯《AIGC发展趋势报告2023》

中信证券《从Chat-GPT看生成式人工智能AIGC产业机遇与落地场景》

浙商证券《AIGC之AI绘画：技术与应用双突破，生产力变革在即》

招商证券《AIGC：打开元宇宙大门的新魔法棒》

AI绘画工具DALLE2、MidJourney与Stable Diffusion的对比——<https://www.bilibili.com/read/cv18290360>

35张图，直观理解Stable Diffusion——<https://my.oschina.net/oneflow/blog/6087116>