

# 2023年全球生成式AI 产业研究报告

## 2023 Global Generative AI Industry Report

指导单位： 中国互联网协会  
中国软件行业协会

参编单位： 天津市人工智能学会  
至顶科技  
至顶智库

2023年5月

# 前言

生成式AI作为当前人工智能的前沿领域，成为全球最热的科技话题。2022年OpenAI发布ChatGPT，生成式AI在模型应用层面实现重要突破，仅两个月突破1亿月度活跃用户数，成为史上用户增长速度最快的消费级应用。全球多家科技企业加大在生成式AI领域的研发投入力度，不断在技术、产品及应用等方面推出重要成果，持续推动人工智能的创新与商业化落地进程，也将带动产业链相关企业快速发展。

在此背景下，在中国互联网协会、中国软件行业协会指导下，天津市人工智能学会、至顶科技、至顶智库联合发布《2023年全球生成式AI产业研究报告》，该报告从全球视角出发，对生成式AI的产业概况、基础设施、算法模型、场景应用、机遇挑战等方面进行梳理，全面展现生成式AI的产业发展情况，为政府部门、行业从业者、教育工作者以及社会公众更好了解生成式AI提供参考。

天津市人工智能学会、至顶科技、至顶智库

第七届世界智能大会，中国天津，2023年5月

# 专家寄语

## 王蕴韬

中国信通院云计算与  
大数据研究所副总工

作为近两年人工智能领域的发展热点，生成式AI引发社会广泛关注。未来，AIGC有关的内容生产工具将极大改变人们的工作生活方式，也将带来社会生产力的全面提升。在新技术新应用快速发展的同时，也需关注由此带来的就业、知识产权等问题，促进产业健康持续发展。

## 胡清华

天津市人工智能学会  
理事长  
天津大学  
智能计算学部教授

人工智能的发展正在进入新的历史时期，生成式AI的未来充满着无限的可能性。ChatGPT的问世无疑是近年来人工智能领域最重要的技术突破之一。因此，我们必要全面了解人工智能技术的现状与趋势，探讨生成式AI的机遇和挑战，更好地迎接智能时代的到来。

## 杨巨成

天津市人工智能学会  
副理事长  
天津科技大学  
人工智能学院教授

生成式AI作为全新的内容生产工具，必将改变人类的生活方式。随着技术不断进步和算法不断优化，未来生成式AI有望在办公、消费、文娱等各场景得到广泛应用，并在不同场景中降低创作者门槛，大幅提升内容生产效率。

## 冯为嘉

天津市人工智能学会  
秘书长  
天津师范大学  
计算机与信息工程学院  
网络工程系系主任

伴随人工智能大模型的不断发展，生成式AI成为一个备受关注的前沿领域，将大概率改变人类与世界的互动方式。从程序设计到艺术创作，该技术正在为各行各业带来可能，有望持续推动社会的发展和人类文明的进步。

# 报告目录

1. 生成式AI产业概况篇
2. 生成式AI基础设施篇
3. 生成式AI算法模型篇
4. 生成式AI场景应用篇
5. 生成式AI机遇挑战篇



# 开篇：伴随人类文明进步和科技发展，语言的传播创造方式日益多元

语言作为人类沟通交流的主要方式，其发展历程分为三大阶段。**语言1.0时代**：从人类语言诞生到公元前16世纪殷商时期甲骨文的出现，文字成为人类交流的重要方式；**语言2.0时代**：从公元105年蔡伦发明造纸术到1946年世界第一台电子计算机诞生，语言开始通过各类机器实现传播；**语言3.0时代**：伴随互联网出现，人工智能NLP及生成式AI等技术的快速发展与相关应用落地，使得机器生成和创造语言的方式成为了可能。

## 语言发展历经三大时代，生成式AI成为语言3.0时代的核心

### 语言1.0时代

(公元前10万年-公元前16世纪)



**公元前16世纪：**  
殷商时期中国人创造**甲骨文**。



**公元前2900年：**  
古埃及人开始使用**象形文字**进行书写。



**公元前3200-2600年：**  
**楔形文字**经历诞生到持续使用。



**新石器时代中期以后：**  
中国出现**象形文字**。

**公元前10万年：**

人类**原始语言**诞生，正式语言约产生于公元前4万年。

### 语言2.0时代

(公元105年-20世纪40年代)



**1946年：**  
第一台**电子计算机**在美国诞生。



**1876年：**  
贝尔发明第一部**电话**。



**1837年：**  
美国人摩尔斯和两个英国工程师库克、怀斯顿同时发明**电报**。



**公元1041-1048年间：**  
毕昇发明**活字印刷术**，为现代印刷术和印刷机的发展奠定基本原理。



**公元105年：**  
汉朝蔡伦发明**造纸术**。

### 语言3.0时代

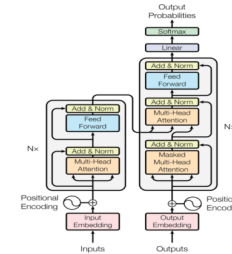
(20世纪50年代至今)



**2010年：**  
IBM宣布开发名为**Watson系统**，该系统能够理解自然语言中的问题，然后使用人工智能根据维基百科提供的信息给出答案。



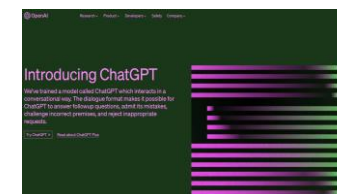
**1950年：**  
艾伦·图灵提出**图灵测试**，标志着人工智能领域的开端。



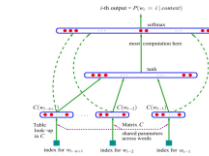
**2017年：**  
**Transformer架构**提出，其在简单语言问答和语言建模任务上有较好表现。



**1969年：**  
**互联网出现**，随着互联网商业化和全球普及，为更大规模语言语料数据获取及传输提供网络支撑。



**2022年：**  
OpenAI发布**ChatGPT**，成为生成式AI的里程碑事件。仅用2个月时间月活用户已突破1亿，成为史上增长最快消费者应用。



**2001年：**  
第一个神经语言模型，**前馈神经网络**由Bengio等人提出。

资料来源：至顶智库结合公开资料整理绘制。

# 2023年全球生成式AI产业研究报告

## 1. 生成式AI产业概况篇

# 1.1 生成式AI概念及内容生成阶段

**生成式人工智能 (Generative AI)** 是在专业生成内容 (PGC)、用户生成内容 (UGC) 之后，利用人工智能技术自动生成内容的新型生产方式。生成式AI基于海量训练数据和大规模预训练模型，自动生成创建文本、音频、图像、视频以及跨模态信息。自2022年OpenAI发布ChatGPT以来，全球爆发生成式AI热潮，诸多科技类企业纷纷推出生成式AI模型、产品和服务和相关底层基础设施及服务。

## 内容生成分为PGC、UGC及AIGC三大阶段

### PGC：专家创作时代 (20世纪90年代)



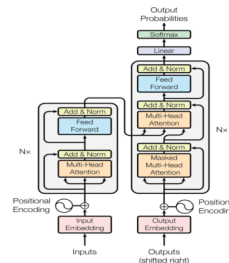
20世纪90年代，基于“信息经济”的商业模式出现，互联网技术提供商提供技术服务，且从生产与组织内容的曝光中获得收益。**门户网站、搜索引擎**成为最主要产品。

### UGC：用户创作时代 (21世纪初)



21世纪初，伴随**微信、微博、抖音、快手**等众多社交媒体的出现，用户也可以真正参与到内容的创作之中，用户利用图文、短视频等多种方式记录生活点滴，同时也会增加人与人之间的交流与互动。

### AIGC：生成式人工智能时代 (21世纪—未来)



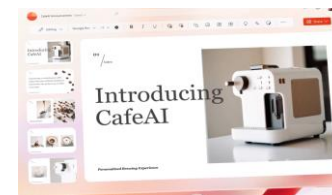
2017年，**Transformer 架构**提出，其在简单语言问答和语言建模任务上有较好表现。



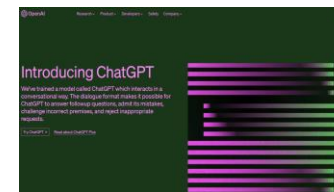
2023年3月，百度正式发布“**文心一言**”，推动国内生成式AI产品快速发展与落地。



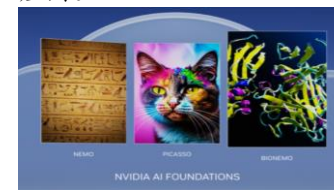
2022年8月，美国科罗拉多州博览会，数字艺术类冠军颁发给由AI自动生成的画作**《太空歌剧院》**。



2023年3月，微软发布“**Microsoft 365 Copilot**”，将生成式AI与办公软件紧密结合，成为新的生产力工具。



2022年11月，OpenAI发布**ChatGPT**，成为生成式AI的里程碑事件。2023年1月，ChatGPT月活用户已突破1亿，成为史上增长最快的消费者应用。



2023年3月，NVIDIA在GTC23发布**AI Foundations**，用于定制大语言模型和生成式AI。

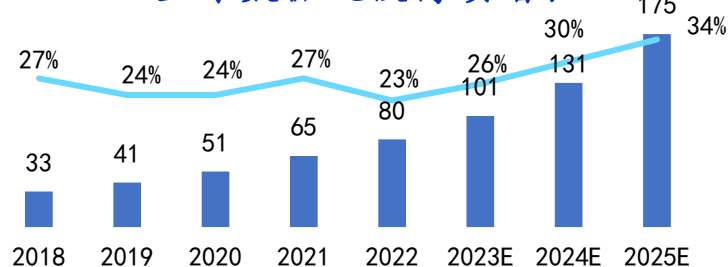
资料来源：至顶智库结合公开资料整理绘制。



## 1.2 生成式AI产业发展驱动力

近年全球数据规模持续增长，IDC预计到2025年全球数据规模将达到175ZB，为人工智能模型训练提供海量数据资源；高性能AI芯片的推出为大规模预训练模型提供重要算力支撑；伴随技术的不断发展，Transformer、BERT、LaMDA、ChatGPT等模型实现快速迭代优化。在数据、算力和模型的共同推动下，全球生成式AI产业得以迅速发展，相关场景应用也不断丰富。

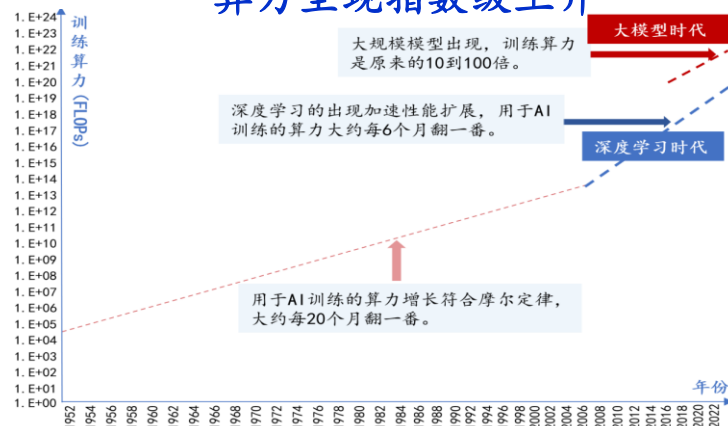
全球数据规模持续增长



■ 全球数据量 (ZB)    — 同比增速

数据来源: IDC, 至顶智库

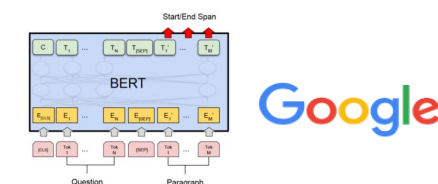
算力呈现指数级上升



信息来源: Compute Trends Across Three Eras of Machine Learning, 至顶智库结合公开资料整理绘制

©2023.5 ZD Insights

2017年以来AI模型发展历程

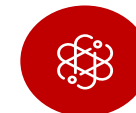


Google基于Transformer架构推出大规模预训练模型BERT，其包含预训练和模型微调部分，查询准确率进一步提升。

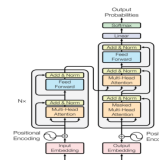


- ✓ OpenAI推出ChatGPT，通过理解和学习人类的语言来进行对话，还能根据聊天的上下文进行互动，真正像人类一样来聊天交流。
- ✓ OpenAI推出DALL·E 2，不仅能通过文字描述创建特定风格的图像与艺术，还可以根据语言对现有的图像进行编辑，使之更加逼真。

2017



Google



Google提出Transformer架构，其在简单语言问答和语言建模任务上有较好表现。

2018

Google

Google提出LaMDA模型，作为语言处理领域一项新的研究突破。LaMDA是一个面向对话的神经网络架构，可以就无休止的主题进行自由对话，克服了传统聊天机器人的局限性。

2021



2022

OpenAI

Google

- ✓ OpenAI推出GPT-4，其具备图像理解多模态能力。
- ✓ Google推出PaLM-E，作为一种多模态VLM（视觉语言模型），不仅可以理解图像，还能理解、生成语言，执行各种复杂的机器人指令。

2023



资料来源: 至顶智库结合公开资料整理绘制。



## 1.3 生成式AI产业图谱



在中国互联网协会、中国软件行业协会指导下，天津市人工智能学会、至顶科技、至顶智库联合发布**2023年全球生成式AI产业图谱**。

图谱主要分为**基础设施层**（AI芯片、AI计算集群、AI云服务）；**算法模型层**（生成式AI大模型）；**场景应用层**（文本生成、图像生成、音频生成、视频生成、数字人）。图谱中涉及各领域全球代表性企业及相关机构，同时将其代表性产品或解决方案作相应展示，为读者提供更为详实的参考信息。

高清大图请扫描下方二维码获取。



## 1.4 生成式AI商业模式

目前，生成式AI尚未建立成熟的变现方式，大部分产品仍处于免费试用“流量吸引+平台改良”阶段。从全球情况来看，生成式AI的主流营收模式有如下几类：作为底层平台收费、按产出内容收费、软件订阅服务收费、模型训练收费、具体属性收费。其中最具长期增长潜力，并将占据主要市场规模的模式为：作为底层平台收费。

### 软件订阅服务收费

向用户提供软件使用权，用户需要按月或其他频次支付费用。

### 按产出内容收费

按产出内容量收费如图片张数、请求计算量等，适用于应用层变现，会受到具体属性影响。

### 模型训练收费

包括模型定制开发，适用于NPC训练等个性化定制需求较强的领域。

### 作为底层平台收费

作为底层平台接入其他产品对外开放，按照数据请求量和实际算量计算。








### 具体属性收费

例如版权授予、是否支持商业用途（个人、企业、品牌使用等）、透明框架和分辨率等。

## 1.5 生成式AI典型投资情况

2019年以来，全球知名投资机构如Sequoia Capital红杉资本、Insight Partners、Coatue Management、创新工场、高瓴投资、启明创投等在生成式AI领域皆有所布局。从相关被投企业来看，美英生成式AI企业侧重在场景应用领域，中国生成式AI企业侧重在大模型领域，被投资企业如OpenAI、Stability AI、Jasper.ai、小冰公司、智谱AI、澜舟科技、毫末智行等。

### 2019年以来全球主要投资机构在生成式AI领域的相关投资

投资机构	被投企业	被投企业所在地	被投企业所属细分领域
Sequoia Capital红杉资本	OpenAI copy.ai InVideo	 美国	ChatGPT、GPT-4、DALL·E 2等 文本生成 视频生成
Insight Partners	Jasper.ai	 美国	文本生成
Coatue Management	Stability AI	 英国	图像生成 视频生成
软银中国资本	创新奇智	 中国	奇智孔明AI InnoGC工业预训练大模型
创新工场	澜舟科技 追一科技	 中国	孟子大模型；AIGC(智能创作)平台 “博文”大模型
高瓴投资	小冰公司 毫末智行	 中国	数字人 自动驾驶生成式大模型DriveGPT雪湖·海若
启明创投	智谱AI	 中国	ChatGLM-6B模型

资料来源：IT桔子，至顶智库结合公开资料整理绘制。11



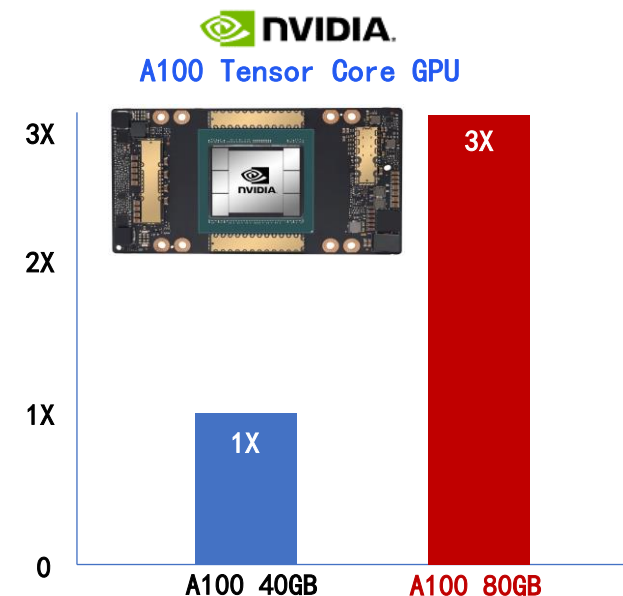
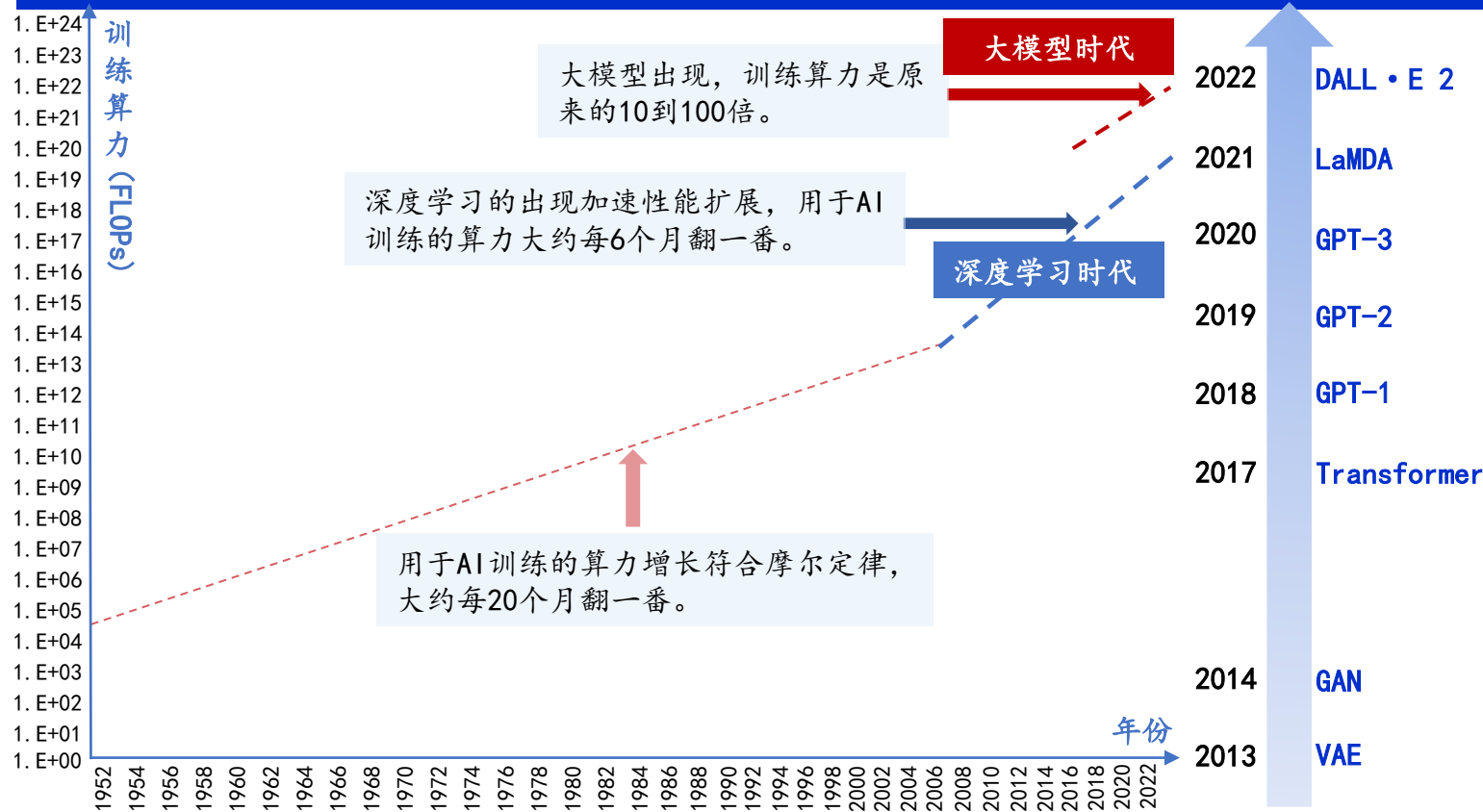
# 2023年全球生成式AI产业研究报告

## 2. 生成式AI基础设施篇

## 2.1 AI高性能芯片为生成式AI训练提供算力支撑

人工智能的发展从深度学习时代进入到大模型时代，大规模预训练模型的参数量呈现指数级上升，需要高性能算力的支撑。目前，大规模预训练模型训练算力是以往的10到100倍，当前主流生成式AI模型的训练广泛使用到英伟达Tensor Core GPU芯片，如微软斥资数亿美元购买数万颗英伟达A100芯片以帮助Open AI打造ChatGPT。

用于训练机器学习系统的算力近年呈现指数级上升



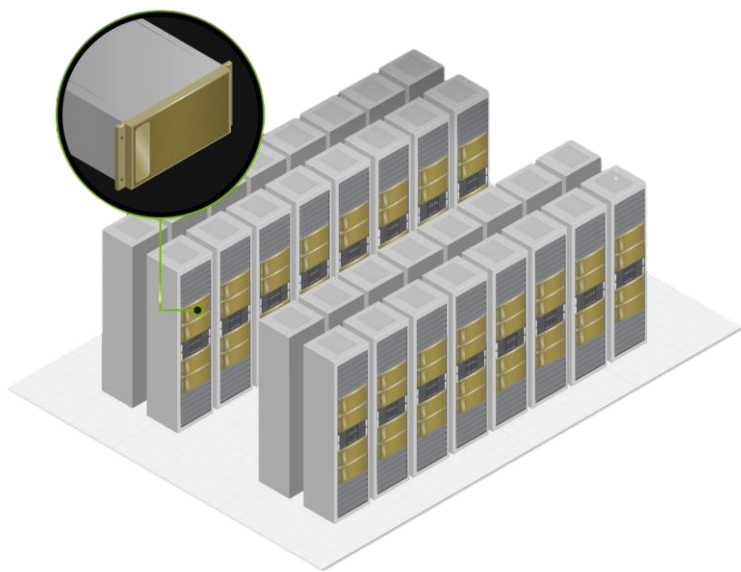
对于AI超大模型训练，英伟达A100 80GB为每个节点提供高达1.3TB统一显存，吞吐量比A100 40GB多高达3倍。微软斥资数亿美元购买数万颗英伟达A100芯片，以帮助Open AI打造ChatGPT。结合OpenAI训练集群情况，GPT-3模型需要英伟达A100 GPU数量约3000-5000张，算力需求巨大。

信息来源：Compute Trends Across Three Eras of Machine Learning, NVIDIA, 至顶智库结合公开资料整理绘制。 13

## 2.2 AI计算集群为生成式AI训练提供大规模算力资源

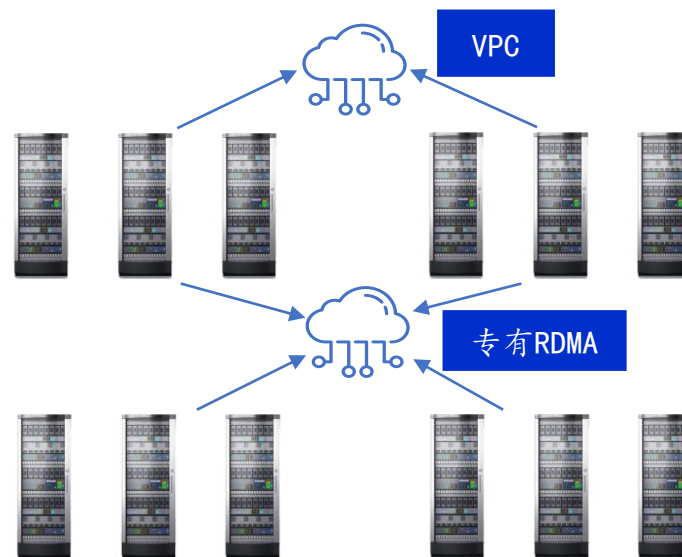
AI计算集群能够提供大规模算力、持续提高算力资源利用率、提升数据存储和处理能力，加速AI大模型训练和推理效率。当前较为典型的AI计算集群如英伟达DGX SuperPOD、百度智能云高性能计算集群EHC、腾讯新一代高性能计算集群HCC等，相关算力基础设施持续为生成式AI训练场景提供强大算力资源，进一步降低模型训练门槛和成本，推动生成式AI模型的落地进程。

 **NVIDIA DGX SuperPOD**



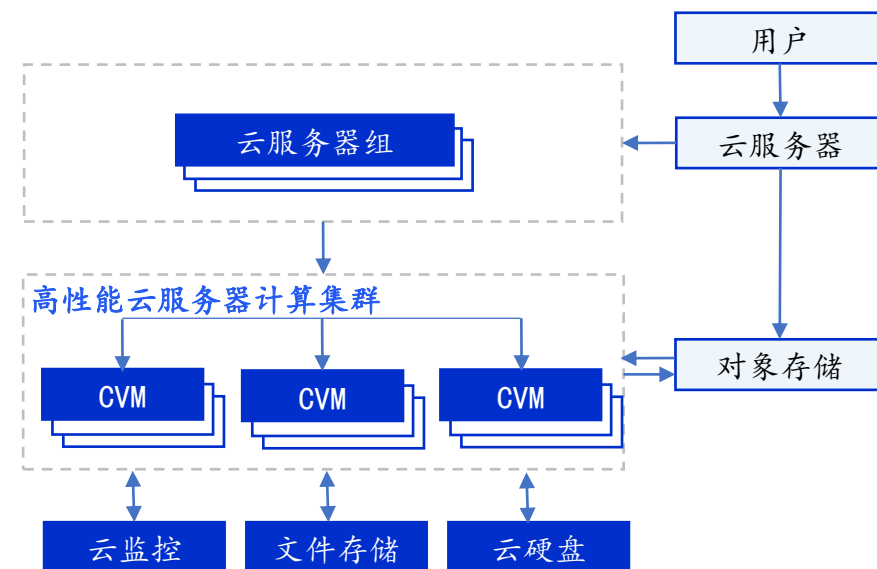
DGX SuperPOD采用模块化设计，支持不同规模大小的设计。一个标准SuperPOD由140台DGX A100 GPU服务器、HDR InfiniBand 200G网卡和NVIDIA Quantum QM8790交换机构建而成，针对超大语言模型预训练这一复杂场景，帮助AI研究人员快速搭建一套强大、灵活、高效的系统。

 **百度智能云 高性能计算集群EHC**



高性能计算集群EHC基于高性能RDMA网络，将多个裸金属服务器进行互联，提供高带宽、低时延的通信能力，极大提升计算任务加速比的计算集群，适用于超大模型训练、科学计算等大规模计算场景。

 **腾讯云 新一代高性能计算集群HCC**



分布式AI训练场景主要包括生成式AI大模型、自动驾驶、NLP、推荐系统等，通过人工智能技术解决各类问题，具有数据量大、计算量密集等特点。HCCPNV5采用星星海自研服务器，搭载NVIDIA H800 Tensor Core GPU，提供3.2Tbps RDMA网络，集群性能较前代提升高达3倍。

信息来源：NVIDIA，百度智能云、腾讯官方资料，至顶智库结合公开资料整理绘制。14



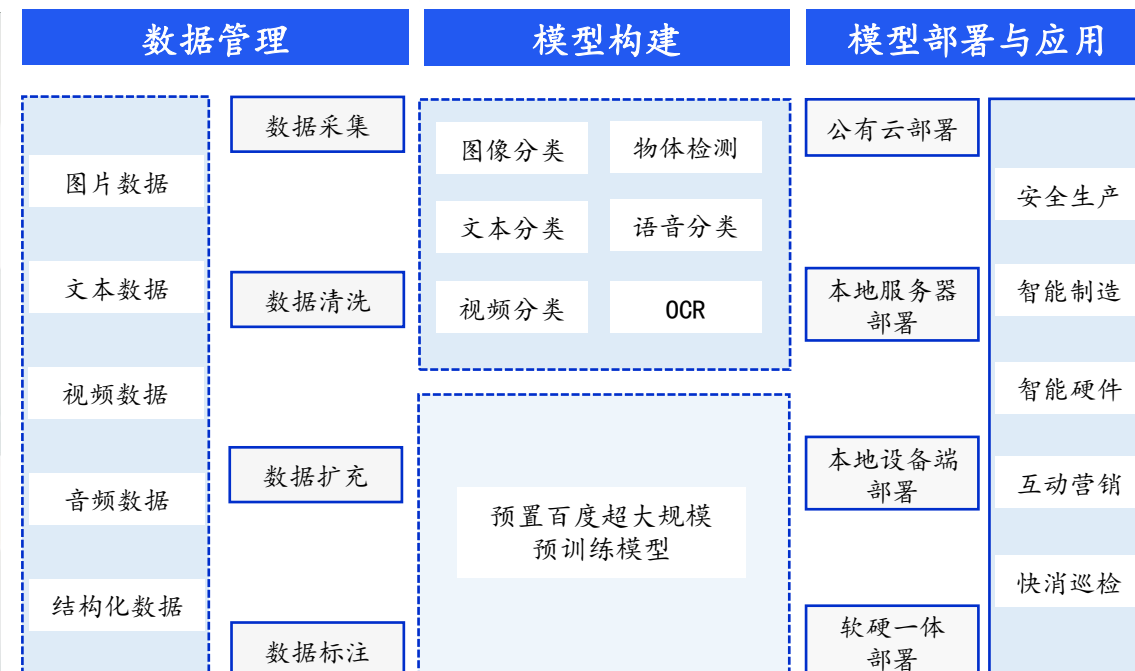
## 2.3 AI云服务为生成式AI模型开发提供平台支撑

人工智能预训练模型的开发对于云服务有较大需求，AI云服务可以提供人工智能开发模块，通过多元化的服务模式，降低开发者的开发成本和产品开发周期，为模型开发提供AI赋能。典型案例如亚马逊SageMaker，其可提供图片/图像分析、语音处理、自然语言理解等相关服务，使用者无需了解参数和算法即可实现功能的应用。百度飞桨EasyDL零门槛AI开发平台提供图像分类、物体检测、文本分类、声音分类和视频分类等功能，实现一站式自动化训练，降低AI定制开发门槛。

### 亚马逊SageMaker机器学习流程



### 飞桨EasyDL零门槛AI开发平台

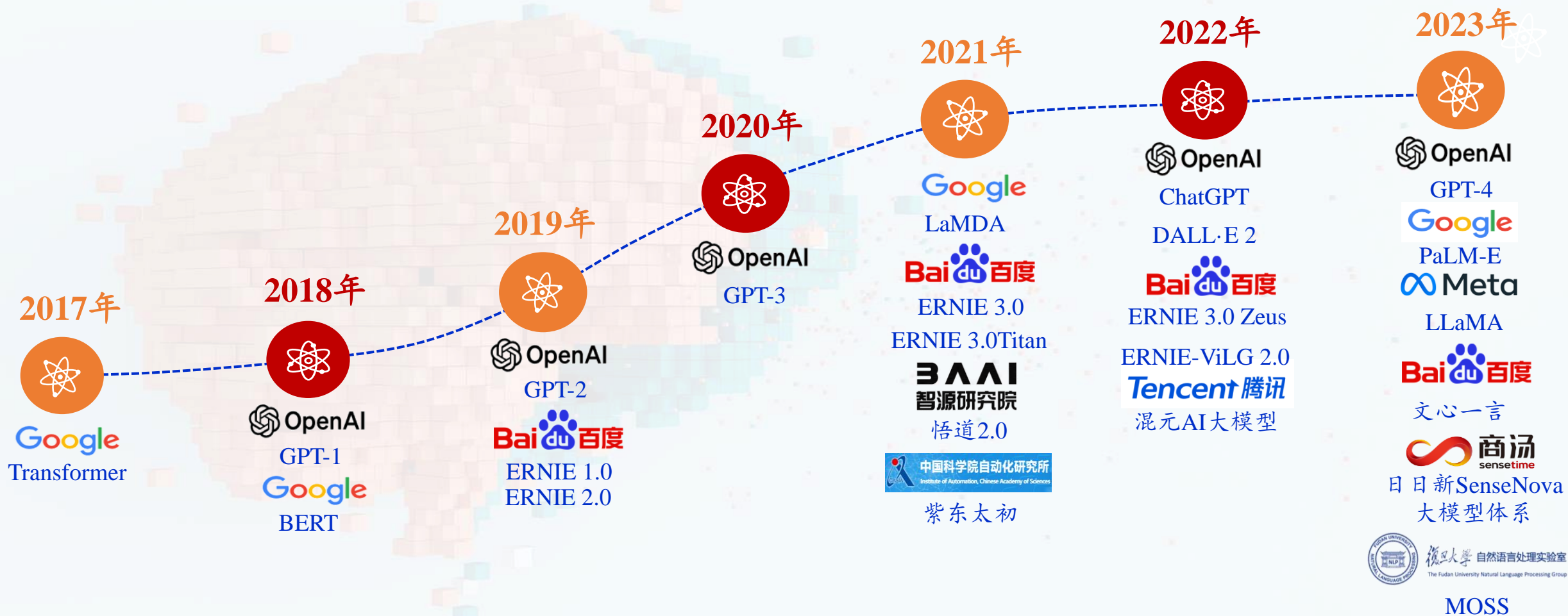


信息来源：AWS，飞桨(PaddlePaddle)官网，至顶智库结合公开资料整理绘制。

# 2023年全球生成式AI产业研究报告

## 3. 生成式AI算法模型篇

## 3.1 全球生成式AI模型发展历程

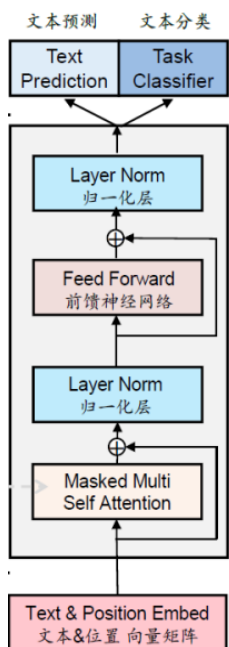


备注：本页仅列举典型生成式AI模型。信息来源：至顶智库结合公开资料整理绘制。

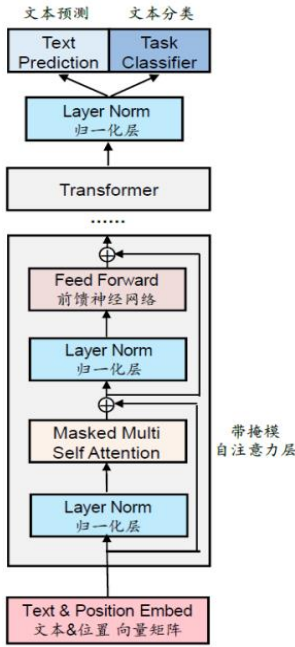


## 3.2 语言类生成主流模型：OpenAI GPT-1到GPT-4

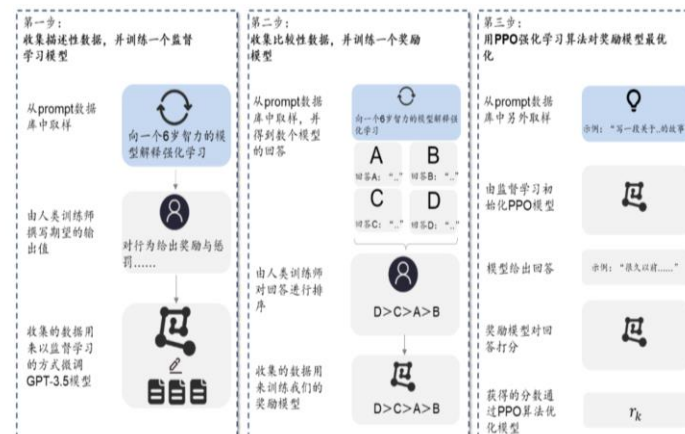
2018年以来，OpenAI先后发布GPT-1、GPT-2、GPT-3、ChatGPT、GPT-4等一系列生成式预训练模型。GPT-1模型基于Transformer架构，仅保留架构中解码器部分；GPT-2模型取消GPT-1中的有监督微调阶段；GPT-3模型舍弃GPT-2的zero-shot，采用few-shot对于特定任务给予少量样例；ChatGPT通过采用RLHF（人类反馈强化学习）技术，增强对模型输出结果的调节能力；2023年发布的GPT-4模型拥有更为强大的多模态能力，其支持图文多模态输入并生成应答文字，可实现对视觉元素的分类、分析和隐含语义提取，表现出优秀的应答能力。



**GPT-1模型**训练包含“**预训练+微调**”两个阶段：1) 利用大量无标注的语料预训练语言模型；2) 对预训练好的语言模型进行微调，将其迁移到各种有监督的NLP任务。GPT-1通过无监督训练解决需要大量高质量标注数据和通过大量语料训练解决训练任务的泛化问题。



**GPT-2模型**在架构上与GPT-1基本相同，但**取消GPT-1模型中有监督微调阶段**，将归一化层移到输入位置，在自注意力之后增加一层归一化。此外，GPT-2增加prompt文本提示，采用更大的参数和多任务学习进行预训练，并使用更大的训练集尝试zero-shot学习。提高模型面对未知任务的推理能力和泛化能力。



采用**RLHF（人类反馈强化学习）**技术对ChatGPT进行训练，加入更多人工监督进行微调。ChatGPT模型训练分为三个步骤：1) 收集描述性数据，训练监督学习模型；2) 收集比较型数据，训练奖励模型；3) 用PRO强化学习算法对奖励模型优化策略。

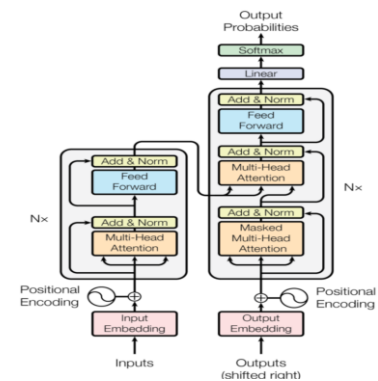
	GPT-4 Evaluated few-shot	GPT-3.5 Evaluated few-shot	LM SOTA Best external LM evaluated few-shot	SOTA Best external model (incl. benchmark-specific tuning)
MMLU [43] Multiple-choice questions in 57 subjects (professional & academic)	86.4% 5-shot	70.0% 5-shot	70.7% 5-shot U-PaLM [44]	75.2% 5-shot Flan-PaLM [45]
HellaSwag [46] Commonsense reasoning around everyday events	95.3% 10-shot	85.5% 10-shot	84.2% LLaMA (validation sets) [28]	85.6% ALUM [47]
AI2 Reasoning Challenge (ARC) [48] Grade-school multiple choice science questions. Challenge-set.	96.3% 25-shot	85.2% 25-shot	85.2% 8-shot PaLM [49]	86.5% ST-MOE [18]
Winogrande [50] Commonsense reasoning around pronoun resolution	87.5% 5-shot	81.6% 5-shot	85.1% 5-shot PaLM [3]	85.1% 5-shot PaLM [3]
HumanEval [37] Python coding tasks	67.0% 0-shot	48.1% 0-shot	26.2% 0-shot PaLM [3]	65.8% CodeT5 + GPT-3.5 [51]
DROP [52] (F1 score) Reading comprehension & arithmetic	80.9 3-shot	64.1 3-shot	70.8 1-shot PaLM [3]	88.4 QGAT [53]
GSM-8K [54] Grade-school mathematics questions	92.0%* 5-shot chain of thought	57.1% 5-shot	58.8% 8-shot Minerva [55]	87.3% Chinchilla + SFT + GRM-RL ORM reranking [56]

**GPT-4**作为**多模态模型**，支持**图文信息作为输入并生成说明、分类和分析**，在视觉及视觉-文字语义融合方面涌现更多能力。其在多个基准任务上取得优异成绩，包括图像字幕、图文问答、代码生成和法律推理等。

信息来源：至顶智库结合文献及公开资料整理绘制。

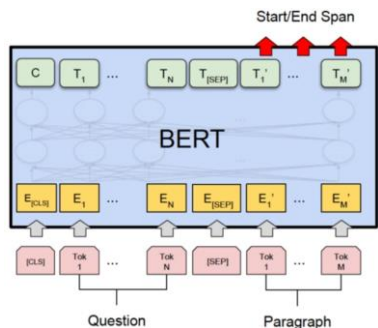
## 3.3 语言类生成主流模型：Google Transformer到PaLM-E

2017年，Google发布具有标志性意义的Transformer模型，该模型的解码模块成为GPT模型的核心要素，通过引入注意力机制，可实现更大规模的并行计算，明显减少模型的训练时间，使得大规模AI模型得以应用。BERT模型、LaMDA模型在信息提取能力以及安全性等方面不断提升。最新推出的PaLM-E模型具有很强的泛化和迁移能力，在完成视觉语言和通用语言任务的同时，可处理多模态数据（语言、视觉、触觉等），实现指导机器人完成相应任务的功能。



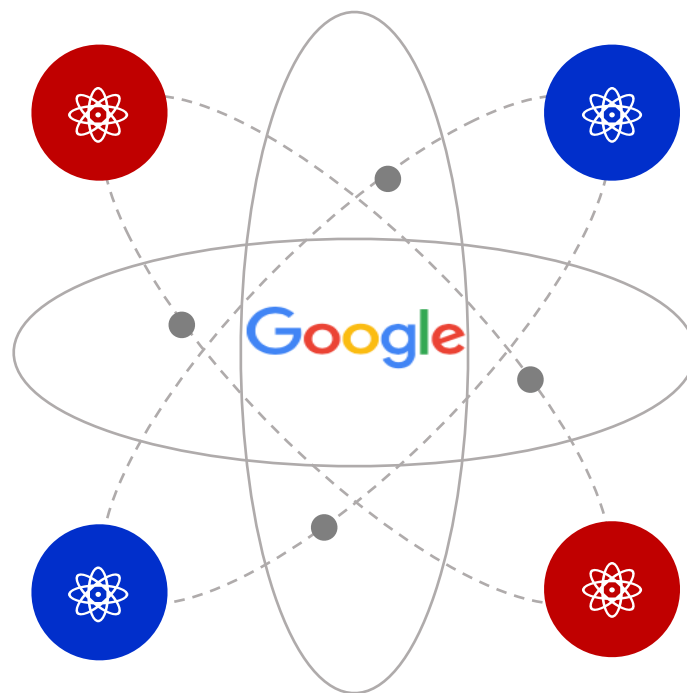
### Transformer

Transformer模型最早在2017年由Google提出，特点是引入注意力机制，可实现更大规模的并行计算，明显减少模型的训练时间，使得大规模AI模型得以应用。



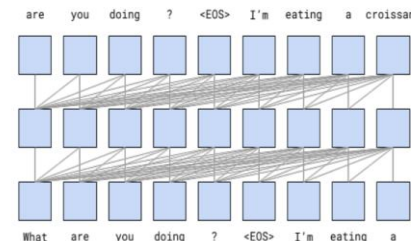
### BERT

BERT模型是一个面向自然语言处理任务的无监督预训练语言模型，通过大量无标记数据集中训练，显著提高各项自然语言处理任务的准确率。在对输入文本的编码过程中，利用每个词的所有上下文信息，语义信息提取能力增强。



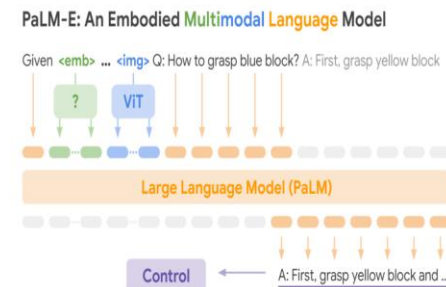
### LaMDA

LaMDA模型使用多达137B个参数以及用1.56T单词数据集进行训练，LaMDA使用单一模型实现多任务处理，展示接近人类水平的对话质量，在安全性和事实基础方面具有显著改进。



### PaLM-E

2023年，Google推出PaLM-E模型，通过将多模态信息编码转换为语言相似的形式嵌入模型，实现多模态联合训练。PaLM-E具有很强的泛化和迁移能力，能完成机器人具身推理任务。

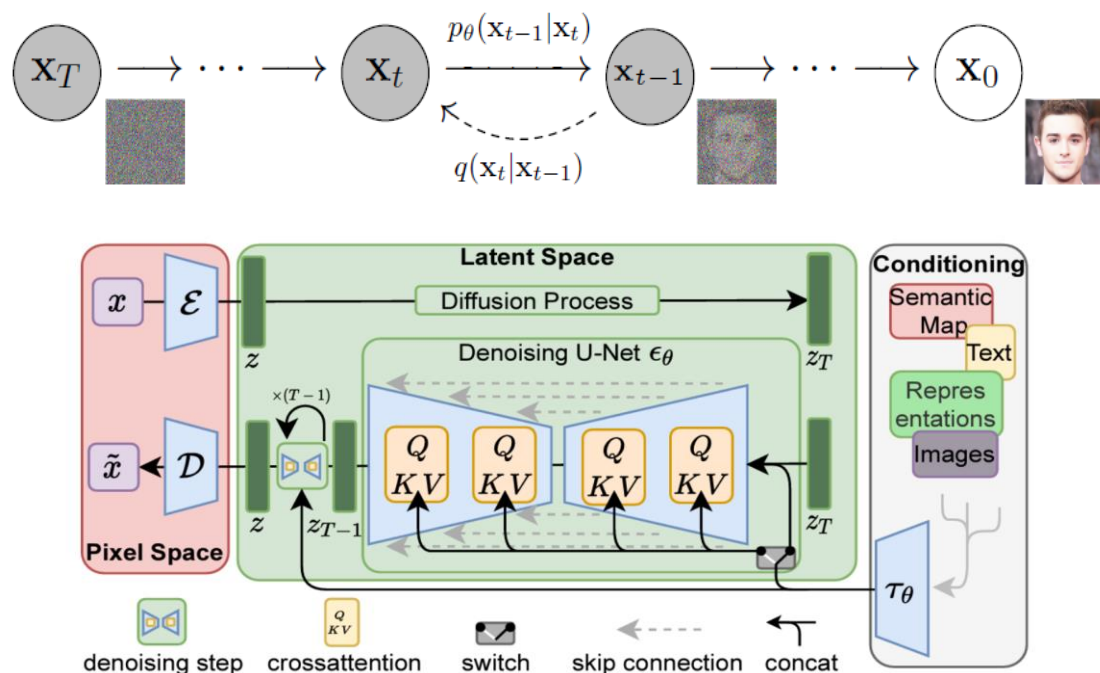


信息来源：至顶智库结合文献及公开资料整理绘制。

## 3.4 图像类生成主流模型：Diffusion Model

**Diffusion Model** 相关研究可追溯到2015年, 去噪扩散概率模型 (Denoising Diffusion Probabilistic Model, DDPM) 在2020年被提出, 展示扩散模型的强大能力, 带动扩散模型的发展。模型主要包括两个过程: 前向过程和反向过程, 其中前向过程又称为扩散过程, 扩散模型通过给图像增加高斯噪声破坏训练数据来学习, 找出逆转噪声过程的方法, 利用学习的去噪声方法实现从随机输入中合成新的图像。Diffusion模型的优势在于生成的图像质量更高, 不需要通过对抗性训练, 在所需数据更少条件下, 该模型图像生成效果有明显提升。

扩散模型原理图



**去噪扩散概率模型 (DDPM, Denoising Diffusion Probabilistic Model) 原理:**

- ✓ 一个固定的 (预先定义好的) 前向扩散过程: 逐步向图片增加噪声直到最终得到一张纯粹的噪声图;
- ✓ 一个学习得到的去噪过程: 训练一个神经网络去逐渐的从一张纯噪声中消除噪声, 直到得到一张真正的图片。

**潜扩散模型 (LDM, Latent Diffusion Model) 原理:**

- ✓ 模型在潜在空间中进行扩散处理, 降低训练成本并提高推理速度;
- ✓ 通过自动编码器去除像素级冗余, 对感知压缩和语义压缩进行松散分解, 然后对学习的潜在知识进行扩散来生成语义概念。
- ✓ 通过使用自动编码模型, 学习空间在感知上与图像空间等效, 显著降低计算复杂度。

信息来源: 至顶智库结合文献及公开资料整理绘制。



# 2023年全球生成式AI产业研究报告

## 4. 生成式AI场景应用篇

## 4.1 全球生成式AI典型应用概览



信息来源：至顶智库结合公开资料整理绘制。 22

## 4.2 生成式AI场景应用—文本生成

**文本生成**应用主要在四个领域：**内容续写**、**文本风格迁移**、**摘要/标题生成**及**整段文本生成**，与其相关的个性化文本生成及实时文本交互前景广阔。总体来看，基于NLP技术的文本生成是生成式AI中发展较早的应用，全球知名科技企业先后推出文本生成类应用工具，如Microsoft、Xmind等相关产品在文案写作、数据分析、演示文稿、思维导图等方面均有相关应用案例。

通过随机Mask数据库文本中的语段，让神经网络自主学习复原被遮挡部分，产出预训练模型，再通过大规模预训练模型理解上文或给定条件，从概率层面推测最符合要求的输出结果。

内容续写

文本风格迁移

文本生成

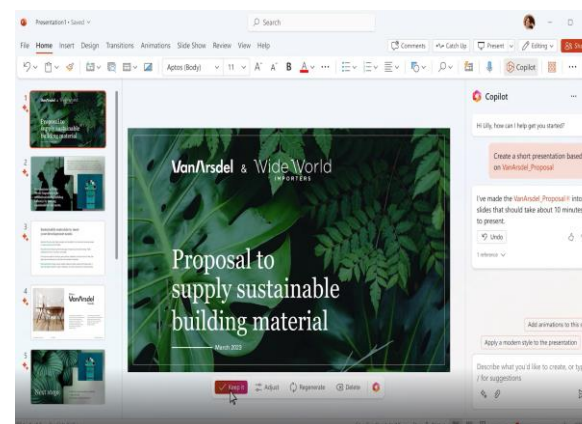
摘要/标题生成

整段文本生成

隐式方法通过使用某类无监督学习或强化学习模式将文本属性及内容自动分离，常见的有生成对抗方式。显式方法首先寻找并删除代表文风的短语，其次检索与目标文风最匹配的相似短语，最后生成目标语句并保证语句通顺、原意不变。

通过词嵌入将字、词、句进行区分；然后基于特征评分、序列标注等提取内容特征计算相关文本单元权重；再选择相应的文本单元子集组成摘要候选集，完成内容选择；最后针对字数要求等限定条件进行整理形成最终摘要。

通过注意力机制、多层感知器等系统进行语句内容预选，对数值、时间等类型数据进行推理，增强数据间的结构信息；再结合上下文进行推导，控制句法及文本连贯，将语义与句法统一分析；最后采用Seq2Seq等模式，构建文本生成器，生成最终文本。



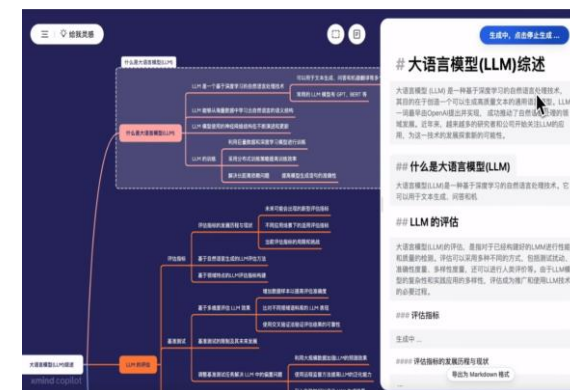
Microsoft 365 Copilot

- ✓ Copilot in Word与用户一起写作、编辑、总结和创作。
- ✓ Copilot in PowerPoint能够在创作过程中，通过自然语言命令将想法转化为设计好的演示文稿。
- ✓ Copilot in Excel帮助用户释放洞察、识别趋势，或在短时间内创建专业的数据可视化。



Xmind Copilot

Ghostwriter为思维导图的每个主题生成文字段落，“标记成组”功能将多个主题的内容合并成完整的段落，灵活控制文章内容生成。





## 4.3 生成式AI场景应用—图像生成

**图像生成**的技术场景划分为图像属性编辑、图像局部生成及更改、端到端的图像生成。其中，前两者落地场景为图像编辑工具，端到端的图像生成则对应创意图像生成及功能性图像生成两大落地场景。目前，图像编辑工具的使用已较为广泛，相关产品较为丰富；创意图像生成大多以NFT等形式呈现，功能性图像大多以营销类海报/界面、LOGO、模特图、用户头像为主。

### 图像生成领域的技术场景

#### 图像属性编辑

可直观理解为经AI降低门槛的photoshop，可方便达到图片去水印、自动调整光影等目的。

#### 图像局部生成及更改

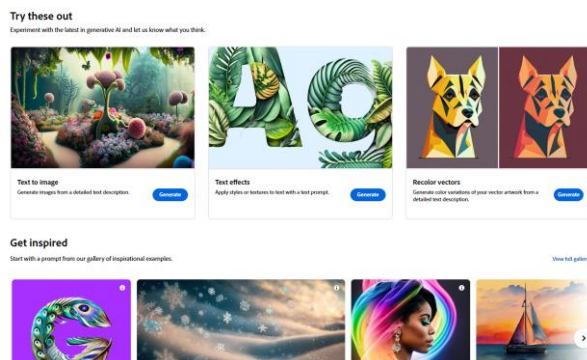
更改图像部分构成、修改面部特征等。

#### 端到端的图像生成

基于草图生成完整图像、有机组合多张图像生成新图像、根据指定属性生成目标图像等。

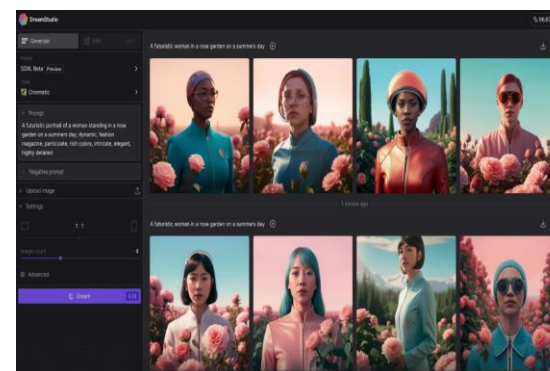
### 图像生成领域的落地场景

#### 图像编辑工具



Firefly具备丰富的图像编辑制作功能，可实现对图像的修复如添加、替换和删除等，还可根据文本描述生成可编辑的矢量或图像等，或根据文本提示修改样式或纹理。

#### 创意图像生成、功能性图像生成



Stability AI  
DreamStudio

DreamStudio具备独有的稳定扩散模型及便捷的编辑器平台，提供强大的文本生成图像、图像生成图像的功能，便于用户创建独特的设计。

## 4.4 生成式AI场景应用—音频生成

**音频生成**在日常生活中已较为常见，其应用领域可进一步区分为语音合成和音乐创作，**语音合成**包括文本生成特定语音（TTS）和语音克隆领域。TTS领域的技术成熟度较高，但在情感表现上仍有欠缺；语音克隆对电影、动画等行业意义重大值得关注；音乐创作可进一步细分为作词、作曲、编曲、录制、混音等多个方向，创作过程主要依托Transformer模型。

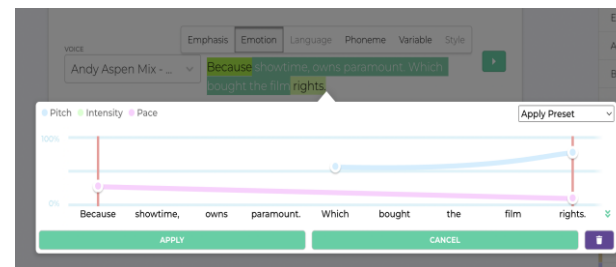
### 语音合成

**文本生成特定语音（TTS）**：广泛应用于客服及硬件机器人、有声读物制作、语音播报等任务；还可应用于短视频内容配音及自动生成配音。

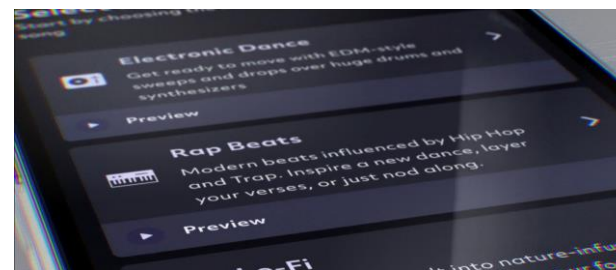
**语音克隆**：本质上属于制定目标语音的TTS。目前被应用于虚拟歌手演唱、自动化配音等，在声音IP化的基础上，对动画、电影以及虚拟人行业具有重要意义。

### 音乐创作

**乐曲/歌曲生成**：AI作曲可理解为“以语言模型为中介，对音乐数据进行双向转化”。目前已支持基于开头旋律、图片、文字描述、音乐类型、情绪类型等生成特定乐曲。



平台支持克隆任何语音，并生成动态、可迭代且独特的语音内容，方便用户创建高质量的自定义人工智能语音，用于所需项目中。



提供在线AI音乐生成器，可让用户在几秒钟内创建自己的原创音乐，支持用户将所创音乐在不同流媒体上公开发布。



从实验室迈向AI合成音、AI广播、智能新闻播报等更多应用场景。喜马拉雅智能语音实验室利用TTS技术合成出单田芳的AI合成音。



## 4.5 生成式AI场景应用—视频生成

**视频生成**有望成为未来跨模态生成领域的中高潜力场景。视频生成主要对应三个领域：视频属性编辑、视频自动剪辑、视频部分生成。视频属性编辑已广泛应用于视频创作领域，大幅提升视频剪辑效率；视频自动剪辑主要在技术尝试阶段；视频部分生成的原理本质与图像生成类似，强调将视频切割成帧，再对每一帧的图像进行处理，现阶段的技术在于提升修改精准度与修改实时性两方面。



### Google Imagen Video

Imagen Video能根据文字描述生成1280\*768分辨率、每秒24帧、长128帧的视频片段，但其生成的视频仍有部分扭曲和抖动。

### 影谱科技

影谱科技自主研发的模型具有丰富的视觉物料库、精准的特征提取能力等，已实现与商业、科教、文娱等领域的融合。

### 剪映

剪映具有AI智能字幕、曲线变速、智能抠像、文本阅读等功能，支持PC端、移动端、网页版多种环境使用。



## 4.6 生成式AI场景应用—数字人

**数字人**指存在于非物理世界(如图片、视频、直播、VR)中,并具有多重人类特征的综合。数字人代表着从文本/音频等低密度模态向图像/视频/实时交互等信息密度更高模态的转化,未来视频乃至元宇宙领域都将是数字人的重要应用场景。在生成式AI领域,数字人生成可划分为数字人视频生成和数字人实时互动,数字人视频生成是目前应用最广泛的领域之一,而数字人实时互动多应用于可视化的智能客服,更强调实时交互功能。

### 数字人视频生成

目前数字人应用最广泛的领域之一,通常在生成数字人的基础上再生成其他内容如摘要、图示等,进而交付更完整的生成式AI内容播报产品。



Hour One致力于专业视频通信数字人的制作,公司的虚拟角色完全以真人为模型,通过文本描述能让数字人传达出人类般的表现力。图中所示为虚拟新闻演播室和定制创建的虚拟新闻主播。



### 数字人实时互动

可理解为以人为单位的数字孪生,进一步涉及思维及策略相关的生成。广泛应用于可视化的智能客服,多见于APP、银行大堂等。



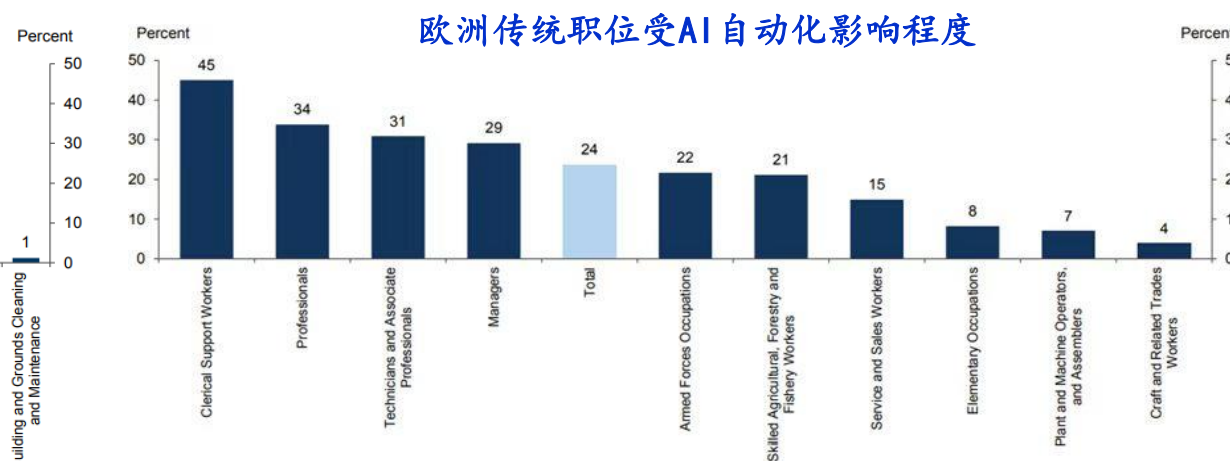
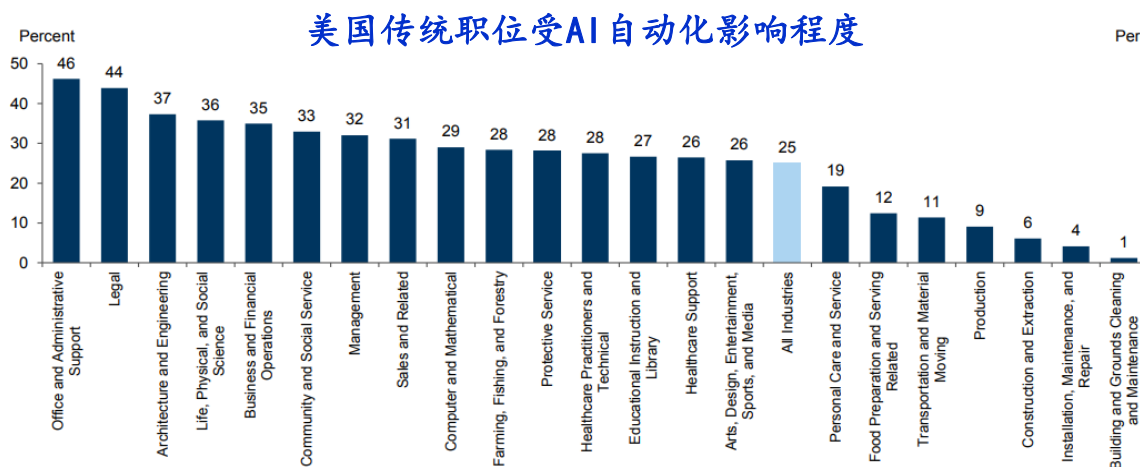
小冰具有数字专家/员工等完整产品线,支持创建拥有情感交互能力、专业技能和内容生产能力的数字人,目前已被应用于多种场景。图图为小冰公司与招商局集团联合研发的数字员工“招小影”。

# 2023年全球生成式AI产业研究报告

## 5. 生成式AI机遇挑战篇

## 5.1 生成式AI时代，行政类工作被替代性高，“问客”有望成为新职业

生成式AI对就业的影响挑战与机遇并存。一方面，生成式AI将促进岗位智能化升级，部分工作岗位将被替代。据高盛分析，生成式AI的智能自动化能力极大提升工作效率并降低运营成本，美国和欧洲的传统职位都将受到不同程度的AI自动化影响，生成式AI可以替代四分之一的工作岗位。另一方面，生成式AI也会创造新职业：“问客”（Prompt Engineer）让人们能够利用自然语言作为提示词，通过与AI进行交互，得到信息或创造作品。除此之外，围绕人工智能的相关领域也将产生大量新的工作岗位。



数据来源：Goldman Sachs: The Potentially Large Effects of Artificial Intelligence on Economic Growth  
围绕人工智能领域诞生的新岗位

### 技术开发领域

**机器学习工程师：**负责开发、实现和维护机器学习算法和模型。  
**自然语言处理工程师：**开发、实现和维护自然语言处理技术，如语音识别等。  
**机器人工程师：**设计、开发和维护各种类型的机器人，如工业机器人等。

### 应用领域

**问客（Prompt Engineer）：**利用自然语言作为提示词，与AI进行交互，得到信息或创造作品。  
**人工智能产品经理：**负责开发、推广和营销人工智能产品和服务。  
**人机交互设计师：**设计和开发人与计算机之间的交互界面和体验，确保用户能充分利用人工智能系统。

### 辅助领域

**人工智能训练师：**负责人工智能相关数据搜集、训练等，帮助模型以更优性能、更高效率运行发展。  
**数据安全专家：**研究人工智能技术在企业 and 政府中应用时的数据保护和安全性。  
**人工智能伦理学家：**研究人工智能技术的伦理和社会问题，确保其合法、公正、透明和人性化。

资料来源：至顶智库结合公开资料整理绘制。29



## 5.2 生成式AI作品版权主要在软件所有者和使用者之间分配

生成式AI本质是机器学习的应用，其在模型学习阶段不可避免地会使用大量的数据集执行训练，但对训练后生成物的版权归属问题尚存争议。由于法律主体才能享有权利，所以生成式AI作品的版权只能由对作品的生成具有贡献的人享有，相关人员包括软件开发者、所有者和使用者（主体身份可能重合），一般认为生成式AI软件开发者已从软件版权中得到补偿，生成式AI作品版权主要在软件所有者和使用者之间分配。

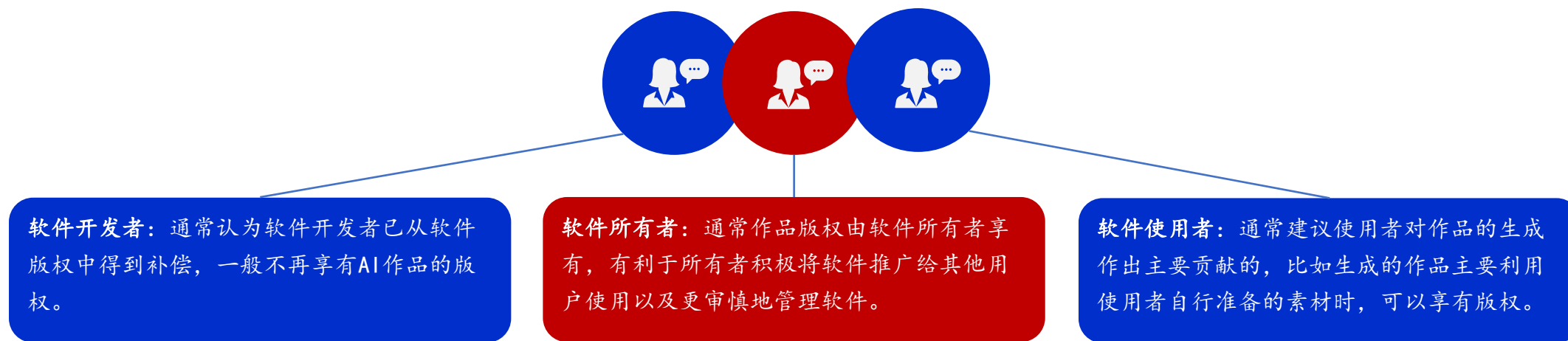
### 当前对生成式AI作品版权归属问题的主要看法

**归属素材作者：**创作的内容由素材库训练生成，本身来自素材库，应当对相关的素材作者提供版权付费。但AI的素材学习库十分庞大，向所有学习对象支付版权费并不可行。

**归属生成式AI用户或平台：**作品由作者使用生成式AI工具创造的，版权完全归作者所有。作品由平台生成式AI工具生成的，版权归平台所有；作者可在非商用时自由使用，商用时需付费使用。

**不归属于特定机构或用户：**作品由公共的数据训练而成的，版权应当回归公共大众，任何人生成的作品都可由其他人在符合法律规定的情形下自由使用。

### 生成式AI作品版权主要在软件所有者和使用者之间分配



资料来源：至顶智库结合公开资料整理绘制。 30

# 指导单位介绍



## 中国互联网协会介绍

中国互联网协会成立于2001年5月25日，由国内从事互联网行业的网络运营商、服务提供商、设备制造商、系统集成商以及科研、教育机构等70多家互联网从业者共同发起成立，是由中国互联网行业及与互联网相关的企事业单位自愿结成的行业性的全国性的非营利性的社会组织，现有会员1000多个。



## 中国软件行业协会介绍

中国软件行业协会成立于1984年9月6日，会员由从事软件研究开发、销售、培训、应用、信息系统集成、信息服务以及为软件产业提供咨询、市场调研、投融资服务和其他中介服务等的企事业单位与个人自愿结合而组成，经国家民政部注册登记，是唯一代表中国软件产业界并具有全国性一级社团法人资格的行业组织，民政部首批授予的AAA级行业组织。



# 参编单位介绍



## 天津市人工智能学会介绍

天津市人工智能学会成立于2021年4月24日，由中科院姚建铨院士及团队、天津大学、天津超算中心等单位发起成立，旨在促进人工智能人才培养、推动产学研用结合，促进产业的深度融合，完善人工智能生态，赋能国家新一代人工智能创新发展试验区建设。学会坚持求实、创新、开放、联合的发展理念，坚持“百花齐放、百家争鸣”的方针，积极开展学术上的交流和讨论，将现代科技成果转化为城市的发展动力。团结天津市以及全国的人工智能工作者，提供一个交流、合作、创新的开放平台，为促进人工智能的繁荣和发展、科学知识的普及和推广、科技人才的成长和提高，以及加速实现我国的伟大复兴做出贡献。



## 至顶科技 至顶科技介绍

至顶科技是一家立足全球科技视野和深刻技术理解，以记录和推动数字化转型进程为使命的信息传播服务公司。自2015年重组以来，公司一直在全力打造更值得信赖的面向知识普及、方案选型、产城合作的科技生态平台，让更多个人、企业、城市受益于数字技术和数字经济创造的崭新机遇。



## 至顶智库介绍

至顶智库重点研究方向是数字经济趋势下的算力新生态、转型新场景、企业新工具。智库研究领域主要涉及人工智能、算力芯片、自动驾驶等前沿科技领域。目前已推出数字经济洞察周报、产业图谱、产业报告、视频解读、会议观察等多项研究成果。





# 2023年全球生成式AI 产业研究报告

2023 Global Generative AI Industry Report



中国互联网协会



中国软件行业协会



天津市人工智能学会



至顶科技



至顶智库