浙商证券股份有限公司
ZHESHANG SECURITIES CO.LTD

AI安全系列：以子之矛，攻子之盾

——从deepfakes深度伪造技术看AI安全

行业评级：看好

2023年6月26日

分析师 刘雯蜀
邮箱 liuwenshu03@stocke.com.cn
证书编号 S1230523020002

研究助理 刘静一
邮箱 liujingyi@stocke.com.cn

摘要

1、Deepfakes技术始于2014年，随着AI大模型能力的提升关注度持续增高

Deepfakes定义可分为广义和狭义两个层次，其中广义上指**利用了以生成对抗网络技术（GAN）为主体的深度学习技术制造的看起来很真实但实际上属于虚假的图片或视频。**随着AI大模型的能力不断突破，deepfakes受到的关注度持续增高。

2、 生成式 AI 模型是Deepfakes的技术基础，而人脸伪造技术是deepfakes的一个重要分支

生成式 AI 是深度学习的一个分支，根据生成内容的类别，**生成式AI模型可进一步被分为生成式语言模型和生成式图片模型。**而随着多模态模型的应用逐步深入，**生成式AI模型也开始向多模态方向发展。**目前主流生成式AI模型包括VAE、GAN、diffusion模型等，stable diffusion模型的发布再次使得生成式AI的输出质量大幅提升。

人脸伪造技术是deepfakes的一个重要分支，可被进一步划分为有目标可视身份伪造和无目标可视身份伪造，其中**有目标可视身份伪造技术已经在俄乌冲突中有所应用。**

3、生成式AI技术的迭代增加伪造图片的真实度，同时增大AIGC识别难度

考虑到：1）**数据层面**，可供训练的数据集规模扩大，带来模型效果的大幅提升；2）**算法层面**，算法框架的迭代与组合，使得模型训练更加高效、稳定；3）**算力层面**，算力水平的提升，助力更加高效的复杂模型训练，AI伪造图片的真实度不断增加，使得识别难度增大，带来的潜在风险提升。

4、反AI生成市场空间约64亿元

针对AI生成图像可能带来的风险，各国政府陆续推出政策加以约束。根据我们测算，**反AI生成市场空间约64亿元，其中监管侧44亿元，企业侧20亿元。**

5、部分反AI生成相关公司

上市公司：美亚柏科、东方通
非上市公司：瑞莱智慧、中科睿鉴、Illuminarty（海外）、AI Voice Detector（海外）、Google（海外）

风险提示

1、AIGC监管以及识别技术研发以及相关产品落地不及预期

2、报告中对各类模型的介绍与总结基于对于相关论文内容的理解，具有一定主观性

3、报告中对于反AI生成的市场空间测算存在主观判断及口径差异

4、由AI安全需求带来的市场竞争加剧

5、板块政策发生重大变化

目录

CONTENTS

01 Deepfakes 演进与发展历程

1）Deepfakes：基于深度学习的AI换脸技术

Deepfakes定义可分为**广义和狭义**两个层次：

- **广义：**Deepfakes指使用机器学习（ML）技术创建的伪造品，即利用了**以生成对抗网络技术（GAN）为主体的深度学习技术**制造的看起来很真实但实际上属于虚假的图片或视频。
- **狭义：**Deepfakes是“Deep machine learning”和 “fake”的组合词，是一种**基于深度学习的人物图像合成技术。**不同于一般意义上的p图，Deepfakes利用电脑程序找到两个面部之间的共同点，通过搭建神经网络来学习人脸，使替换以后的脸可以生动地模仿原来的表情，以假乱真。

2）Deepfakes的产生与飞跃

- **2014年是Deepfake的诞生元年**，Goodfellow与同事发表的科学论文标志着GAN AI的诞生，催生出我们如今所熟知的Deepfakes。在2014年，就有迹象表明 GAN 有望生成仿真度极高的人脸。深度伪造技术开始进入大众视野。在Deepfakes产生初期，生成代理往往倾向于产出分辨率较低而模糊不清的图像是检查代理难以判断内容的真伪，深度伪造技术一定阶段内存在着**输出内容像素过低，生成结果难以令人信服**的问题。
- **2017年英伟达推动质量飞跃**， 利用分阶段训练网络解决了生成代理产出分辨率过低的问题，GAN 开始产出质量空前的伪造人像，**深度伪造技术开始被推向市场主流。**自此，**Deepfakes一词成为 AI 生成图像和视频的代名词。**
- **2019年Deepfake正式成为市场主流，**专注于Deepfakes的YouTUbe频道拥有数百万关注者，产出质量远高于其他AI模型。

资料来源：AI前线公众号、浙商证券研究所

01 Deepfakes技术发展历程

- **2014：**Goodfellow与同事发表了全球首篇介绍GAN的科学论文，代表着**GAN AI**的诞生，催生出如今为人熟知的 Deepfakes。
- **2015：**研究人员开始将 **GAN 与经过图像识别优化的多层卷积神经网络（CNN）相结合**，这一组合取代了以往较为简单的 GAN 代理驱动网络，提高了处理数据的速度和显卡运行效率，也让生成结果的可信度迈上新的台阶。
- **2016：**研究人员把两个 GAN 结合起来，开展**并行学习。**
- **2017：**英伟达推动质量飞跃， GAN 开始产出质量空前的伪造人像，**深度伪造技术开始被推向市场主流。**自此，**Deepfakes一词成为了 AI 生成图像和视频的代名词。**
- **2018：**英伟达提升 GAN 控制能力，使其能够对人像中的“黑发”和“微笑”等图像单一特征作出调整，**将训练图像中的特征有针对性地转移到 AI 生成图像当中。**
- **2019：**三星公司的研究人员公布了一种能够深度伪造人类和艺术品的 GAN，只需参考少数照片就能利用Deepfakes AI达成出色的伪造效果；以色列研究人员又推出了**换脸 GAN（FSGAN）**，能够对即时视频中的人脸进行实时交换。无需任何预先训练。
- **2020：**微软推出 FaceShifter，该软件能够利用模糊的原始图片，依赖于分别负责**伪造人脸和照片比对**的两套网络，生成高度可信的 Deepfakes 图像；深度伪造技术有望成为迪士尼电影制作开发的主流技术。
- **2021：**社交媒体中出现Deepfakes**巡演、直播与人脸租赁**活动，在市场上获得极高热度。
- **2022：**GAN的改进接连出现，包括能够在短视频片段中轻易操纵人脸的**StyleGAN2变体**和既能以高度匹配的 3D 形式生成统一图像，也能利用一张真人图像还原出 3D 模型的**3D GAN**，大力推动力AI深度伪造技术的发展。

资料来源：AI前线公众号、浙商证券研究所

01 Deepfakes应用发展历程

1）海外应用

- **FaceShifter：**2020年由北京大学和微软亚洲研究院研究团队联合发表，是一种**高保真、能够感知遮挡**的AI换脸工具，采用两层框架结构实现高精度和遮挡条件下的换脸。其优于以往同类技术，在生成逼真的人脸图像方面表现优异，被誉为机器学习图像识别领域的“利矛”。
- **Wombo AI：**2021年正式进入大众视野，可以借助AI技术**将声音与图片中的角色自动对上口型**，使处于静止图片中的人物进行开口讲话，并且还有会动的姿态表情，在社交媒体上大受欢迎。
- **DeepFaceLive：**由DeepFaceLab的缔造者在 2021 年首次展示，能够在经过适当训练、或者接收到预训练 AI 模型之后，在**实时视频中交换人脸**，意味着换脸的技术又再一次的突破。

2）国内应用

- **Zao：**于2019年首次公测，是中国国内利用**深度伪造技术**制作的一个应用程序，用户可以利用这个程序将自己的脸替换成电影里某个角色的脸；用户还可以在Zao里面大量的视频和图片库中进行选择，在上传视频后就可以在几分钟之内生成深度伪造的角色。然而，从2019年9月1日，Zao就由于疑**侵犯用户肖像权**而被用户投诉，以及**对用户生物识别信息的采集存在的信息安全性问题带来的安全风险**而遭遇下架。
- **Face X-Ray：**2020年由北京大学和微软亚洲研究院研究团队联合发表，是一种针对**伪造人脸图像**的通用检测工具，不需要依赖于与特定人脸操作技术相关的伪影知识，并且支持它的算法可以在不使用任何方法生成假图像的情况下进行训练。这种工具能有效地识别出未被发现的假图像，并能可靠地预测混合区域，在市场上获得很高的评价，被誉为机器学习图像识别领域的“坚盾”。

资料来源：新智元公众号、《FaceShifter: Towards High Fidelity And Occlusion Aware Face Swapping》、《 Face X-ray for More General Face Forgery Detection》、品牌观察报公众号、浙商证券研究所

Deepfakes应用

表：国内外deepfakes相关应用梳理

名称	类别（开源与否）	开发地区	开发时间	使用状况
FakeApp		国外	2018	已下架
Faceswap	已开源	国外	2019	未下架
DeepFaceLab	已开源	国外	2018	未下架
DeepFaceLive	已开源	国外	2019	未下架
Faceswap-GAN		国外	2019	未下架
ZAO	已开源	国内	2019	已下架
DFaker		国外	2020	未下架
Deepface		国外	2015	未下架
FaceShifter		国外	2020	未下架
Wombo AI		国外	2021	未下架
Avatarify	已开源	国外	2020	未下架

资料来源：《深度伪造人脸视频检测技术综述》、Faceswap官网、DeepFaceLab官网、DeepFaceLive官网、搜狐网、重蔚自留地、浙商证券研究所

02 生成式AI模型梳理

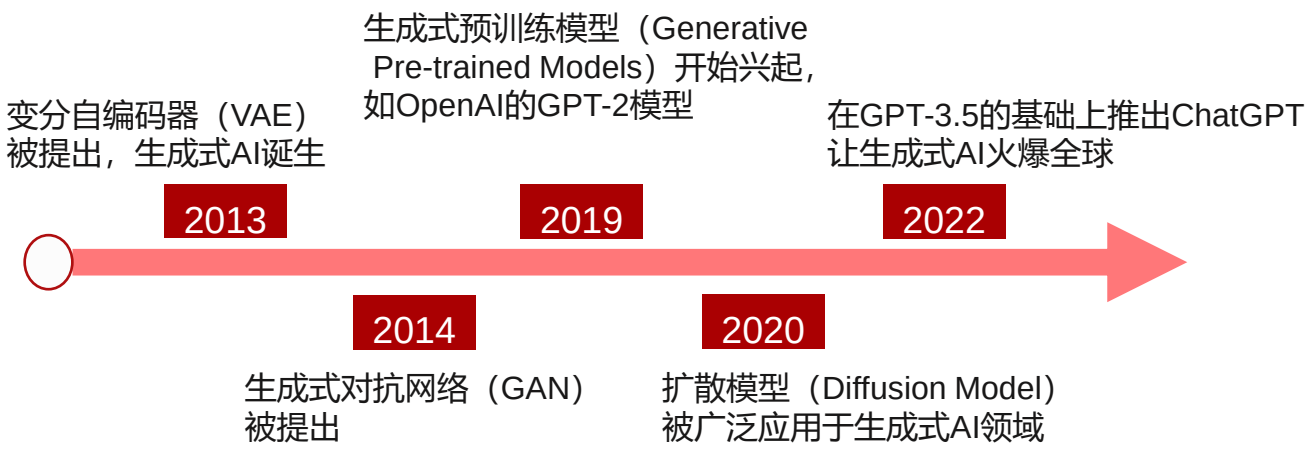
02 生成式AI是深度学习的分支

生成式 AI 是深度学习的一个分支，可以根据已经学习的内容生成新的内容。

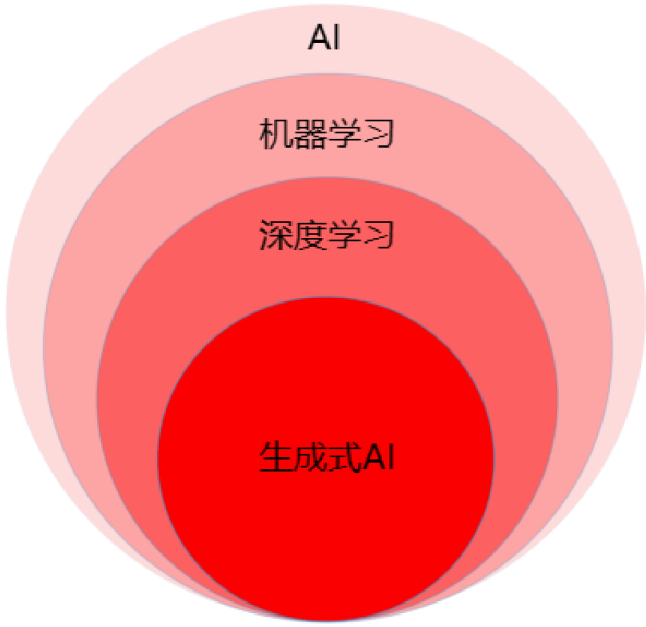
生成式AI的训练和推理阶段：

- **训练**：从现有的内容中学习的过程，训练的结果是创建一个统计模型。
- **推理**：当用户给出提示词，生成式 AI 将会使用统计模型去预测答案，生成新的文本来回答问题。

图：生成式AI发展历程



图：生成式AI是深度学习的一个分支



资料来源：OpenAI网站、《Auto-Encoding Variational Bayes》、《Generative Adversarial Networks》、《Denoising Diffusion Probabilistic Models》、浙商证券研究所

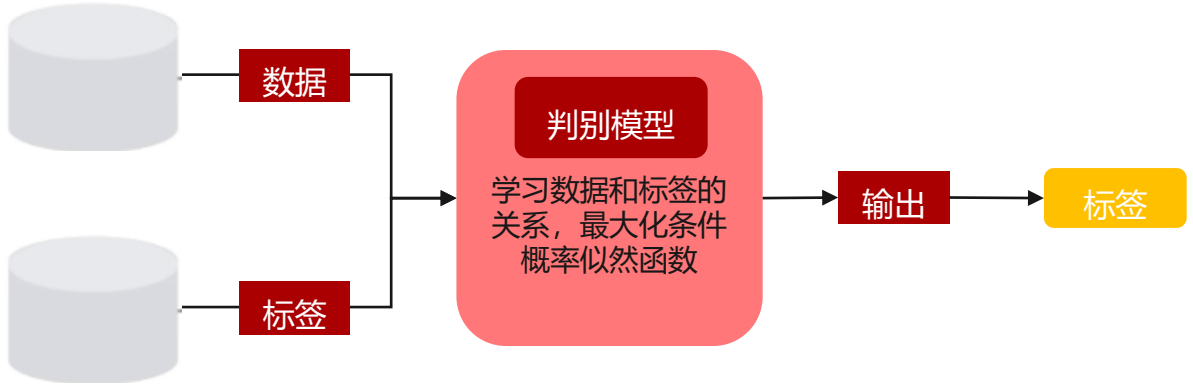
判别模型 vs 生成模型

判别模型（Discriminative Model）和生成模型（Generative Model）是机器学习中两种不同类型的模型，它们的主要区别在于其对数据的建模方式和应用领域。

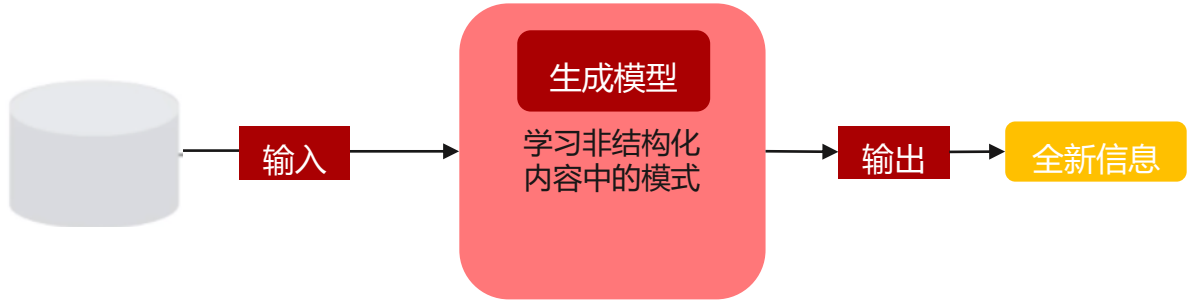
表：判别模型与生成模型对比

	判别模型	生成模型
建模对象	直接对条件概率进行建模	对联合概率分布进行建模
学习内容	输入和输出之间的关系	数据的分布和特征之间的关系，以及生成输入数据的过程
任务类别	分类、回归和标注等任务，通过最大化或最小化某种损失函数来寻找最优的参数配置，以实现准确的预测和分类	生成新的图像、语言模型和数据增强等任务
特点	直接学习输入特征和输出标签之间的关系，具有较低的计算复杂度，并且适用于特定任务	能够模拟数据的生成过程，生成新的样本数据，并且有助于理解数据的统计特性和生成机制
常见模型	逻辑回归、支持向量机、深度神经网络	高斯混合模型（GMM）、生成对抗网络（GAN）

图：判别模型框架



图：生成模型框架



资料来源：谷歌云、《Deriving discriminative classifiers from generative models》、CSDN、浙商证券研究所

生成式模型分类

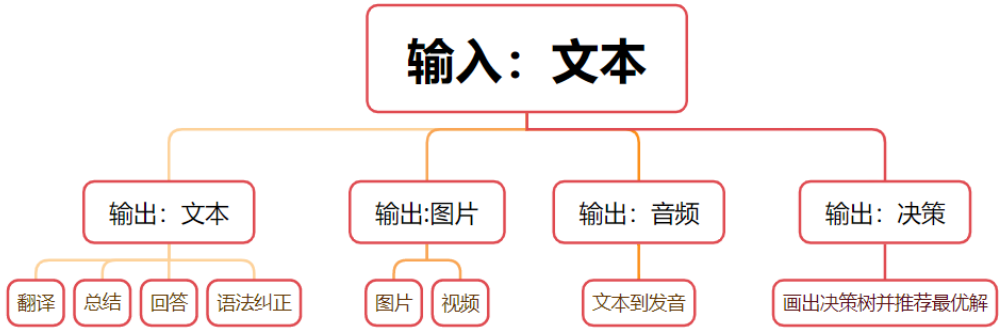
根据生成内容的类别，生成式AI模型可进一步被分为生成式语言模型和生成式图片模型：

- 生成式语言模型：基于自然语言处理的技术，通过学习语言规律和模式生成新的文本。通过训练大规模文本数据，如新闻、小说、网页内容等，生成式语言模型可以自动生成逻辑和语法正确的文本，如文章、对话、诗歌等，被广泛应用于机器翻译、文本摘要、对话生成、故事生成等任务。
- 生成式图片模型：基于计算机视觉的技术，通过学习图像的特征和结构生成新的图像。通过训练大规模的图像数据集，学习图像的特征表示和统计规律，生成式图片模型可以生成具有视觉真实感或艺术风格的图像，如自然风景、人像或抽象艺术作品。被应用于图像生成、图像标注、图像编辑等领域。

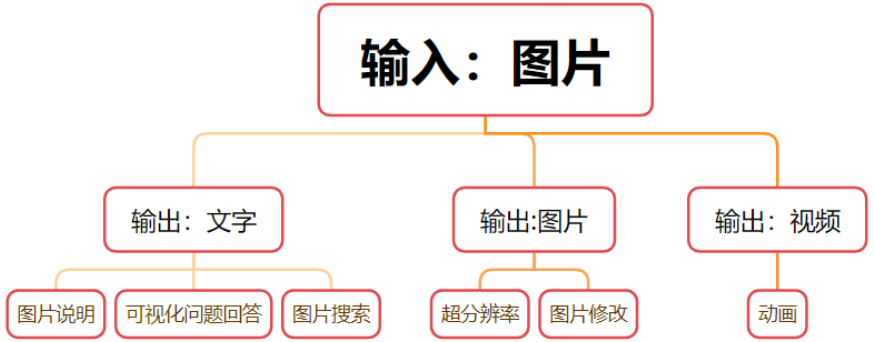
随着多模态模型的应用逐步深入，生成式模型也开始向多模态方向发展：

- 输入内容为文本：输出内容可以为文本、图片、音频、决策等内容
- 输入内容为图片：输出内容可以为文本、图片和视频等内容。

图：输入为文本时生成式模型的输出可选项



图：输入为图片时生成式模型的输出可选项



资料来源：谷歌云、浙商证券研究所

主流生成式AI模型概览

	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
潜变量模型		**生成对抗网络(GAN)**	GAN-INT-CLS	StackGAN Pixel-to-Pixel	StackGAN++ ProGAN CycleGAN	StyleGAN	GLIP-VQGAN	ViT-VQGAN	StyleGAN3 ViTGAN	CLIP+StyleGAN
		条件生成模型(CGAN)								
	变分自动编码器(VAE)		条件生成模型CVAE	AlignDRAW模型					CLIP+VQVAE DALL-E	
				Diffusion扩散模型				DDPM DDIM	Improved DDPM DPM-solver	Disco Diffusion DPM-solver++ DALL-E2 Stable Diffusion
基于流的模型				**RealNVP**					RealNVP++	
						Glow		Glow-360 Glow-TTS		
						FFJORD			FFJORD-GAN	

注释：部分模型可能同时涉及多项模型的框架，我们以其中主要或最新参考的模型作为划分依据

资料来源：相关领域已发表学术论文（论文名称参考报告末尾附录页）、浙商证券研究所

VAE——变分自动编码器

变分自动编码器（Variational Autoencoder，VAE）是一种生成模型，结合了自动编码器（Autoencoder）和变分推断（Variational Inference）的思想，于2013年由Kingma和Welling提出。其目标是学习数据的潜在表示，并能够生成新的样本。

VAE结合自动编码器和变分推断的思想，通过学习数据的潜在表示和生成新样本。它假设潜在变量服从先验分布，编码器将输入映射为均值和方差参数，通过重参数化技巧采样潜在变量。解码器接收采样的潜在变量解码为样本，通过最大化边际似然和KL散度来训练VAE，学习数据的潜在分布并生成新样本。

优点

概率建模

VAE用变分推断建模潜在变量分布，最大化边际似然训练模型，提供概率建模和不确定性估计。

可控生成

VAE潜在空间连续，操作潜在变量控制生成样本属性。

缺点

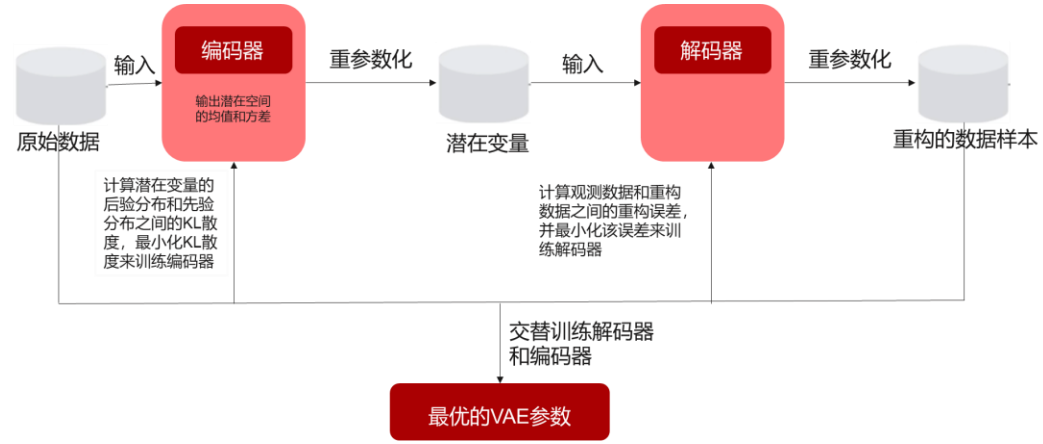
训练困难

VAE的训练需要平衡重构误差和KL散度的损失，调整超参数和监控训练过程相对复杂。

离散数据处理

VAE基于连续分布假设，对于离散数据处理可能有困难，需要适当修改或使用离散分布的变分推断方法。

图：VAE模型流程图



资料来源：《Auto-Encoding Variational Bayes》、谷歌云、浙商证券研究所

GAN——生成式对抗网络

生成式对抗网络（GAN）由Goodfellow等人于2014年提出，采用对抗训练策略，通过对抗训练的方式改变了生成模型的思路。

GAN包括生成器和判别器两个对抗的模型，生成器试图生成逼真样本，判别器试图区分生成和真实样本。通过对抗过程，生成器逐渐提高样本质量，判别器提高鉴别能力。GAN通过让生成器和判别器相互博弈，逐渐达到动态平衡，使生成器学习真实样本的分布。

优点

学习潜在表示

GAN的生成器学习数据的潜在表示，有助于特征学习和数据降维，提供可解释和可操作的潜在空间。

高质量样本生成

GAN生成逼真的图像、音频、文本等样本，质量和多样性较传统生成模型更好。

缺点

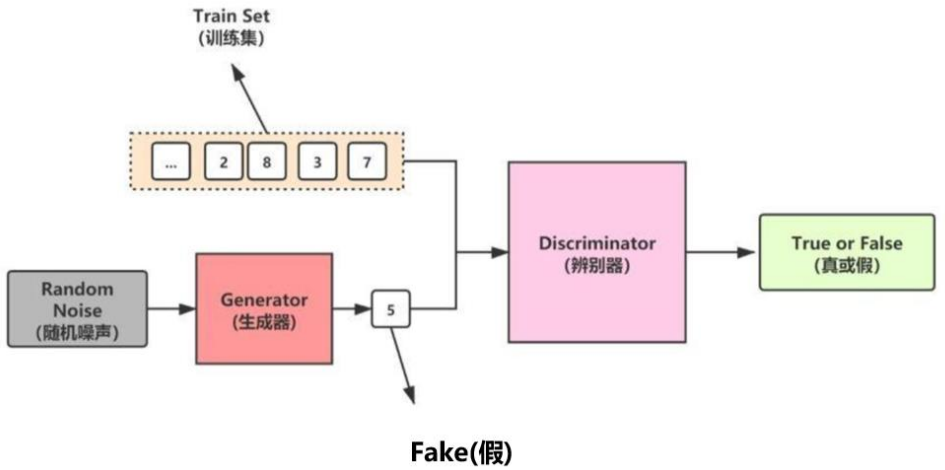
模式崩溃和塌陷

生成器可能只能生成少数几种模式，缺乏多样性和创造性。

难以评估生成质量

缺乏明确的似然函数，生成质量难以定量评估，需要人工观察和主观评价。

图：GAN模型流程图



资料来源：《Generative Adversarial Networks》、浙商证券研究所

Diffusion——扩散模型

Diffusion是一种用于建模数据分布的生成模型，通过迭代模拟数据在潜在空间中的扩散过程，使得样本在潜在空间中逐渐接近真实数据分布，从而由初始噪声样本逐步生成高质量样本。

Diffusion模型通过动力学系统学习数据的结构和生成机制，克服了传统生成模型的限制，适用于复杂高维数据。它在图像生成、数据插值等任务中表现出优异性能，为生成模型提供了一种新的建模方式。

优点

逼近复杂分布

扩散模型有效建模复杂高维数据分布，捕捉非线性关系和复杂结构。

无需显式似然函数

不需要定义似然函数，通过扩散逼近真实数据分布，处理复杂数据和高维度更有效。

缺点

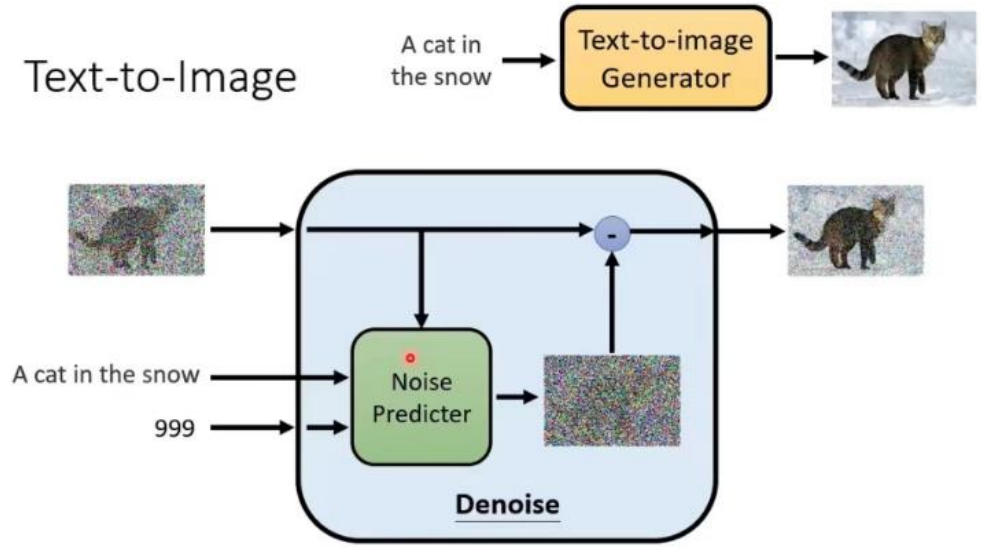
计算复杂度高

训练和推断过程需要大量计算资源和时间，对大规模高维数据集要求较高

超参数敏感性

性能和质量受超参数选择影响，不当选择可能导致不稳定或下降。

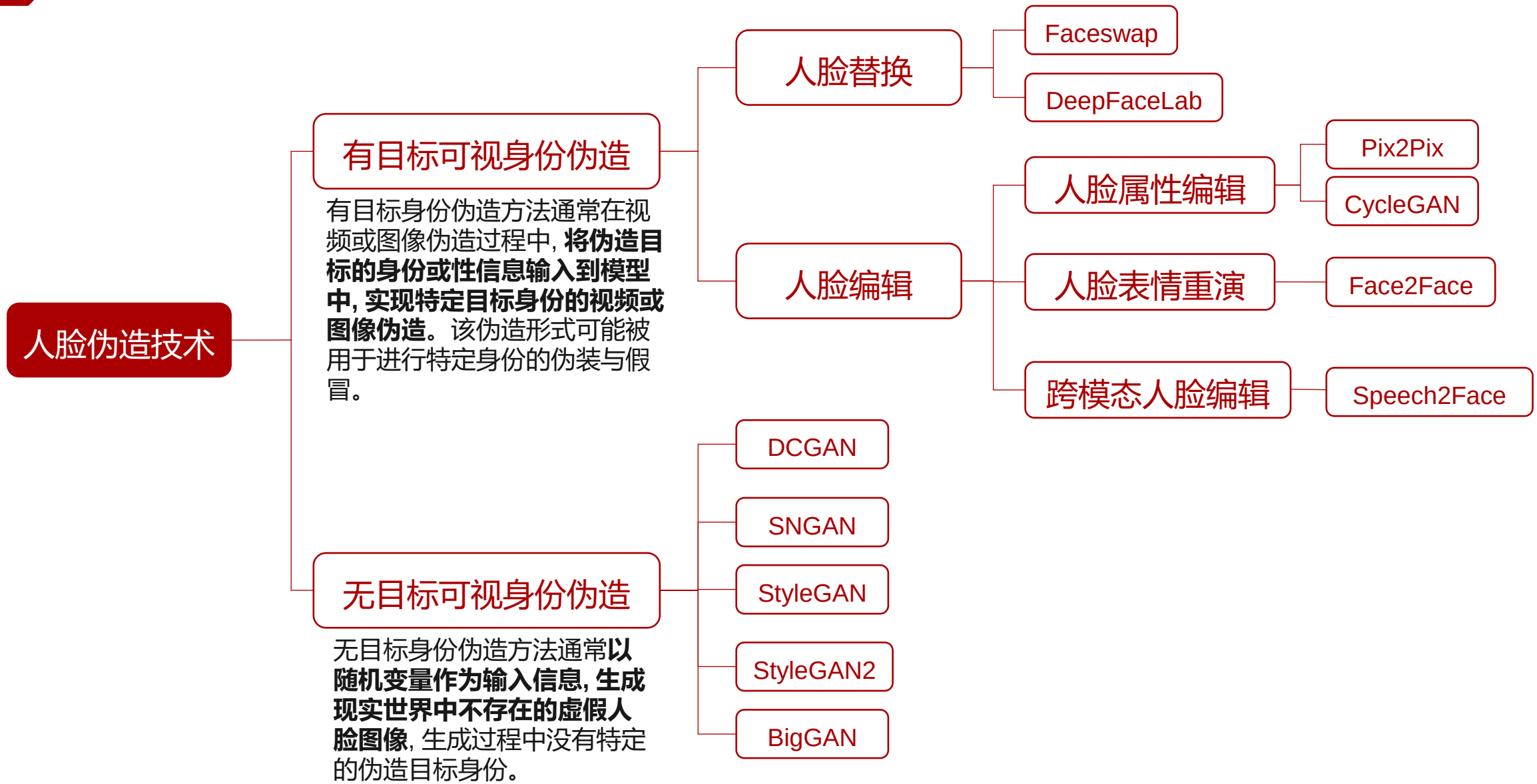
图：扩散模型流程图



资料来源：《Diffusion Models: A Comprehensive Survey of Methods and Applications》、谷歌云、浙商证券研究所

03 以视觉为例：人脸伪造技术分类

人脸伪造技术分类梳理



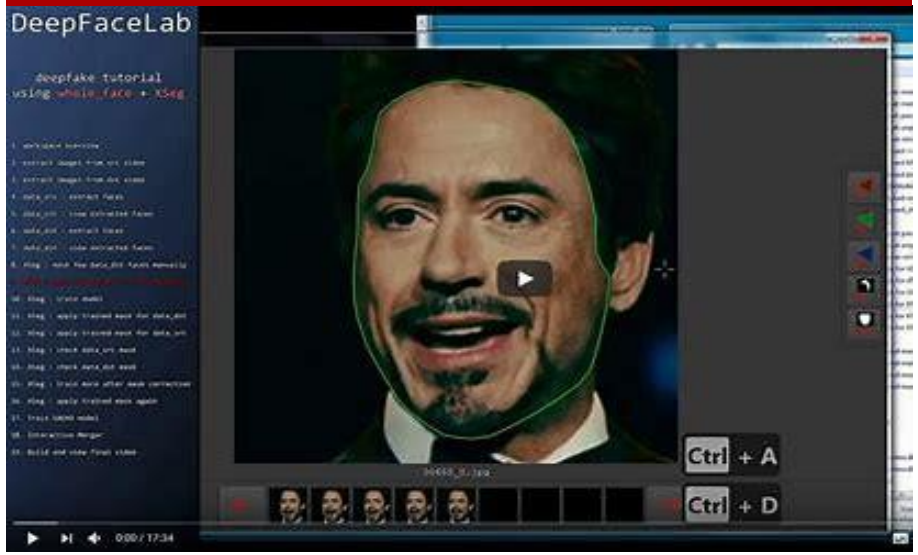
资料来源：《可视身份深度伪造与检测》、浙商证券研究所

人脸伪造技术分类梳理

1）有目标可视身份伪造

- **DeepFaceLab：**主要用于视频换脸，使用者将两张不同的照片导入模型实现脸部的交换。DeepFaceLab提供力从数据收集和筛选，到模型训练和最终视频输出的一站式解决方案，属于有目标可视身份伪造类别的人脸替换系列。DeepFaceLab拥有**面部更换、面部老化、头部更换、操纵嘴部**等功能，可用于短视频平台的娱乐项目以及电影的后期制作等。DeepFaceLab没有收费或捐赠机制，任何人都可以在GitHub上下载和使用。
- **Facelet-Bank：**腾讯优图团队提出的一种通用且灵活的高质量人脸属性编辑网络，进而与香港中文大学、Adobe 研究院、字节跳动人工智能实验室合作提出一种基于语义部件分解的人脸属性编辑方法。Facelet-Bank可以快速处理各种表情、配饰和化妆效果，并产生**高分辨率和高质量**的图像。Facelet-Bank可以用于视频网站的快速人像操作。

图：DeepFaceLab



图：Facelet-Bank

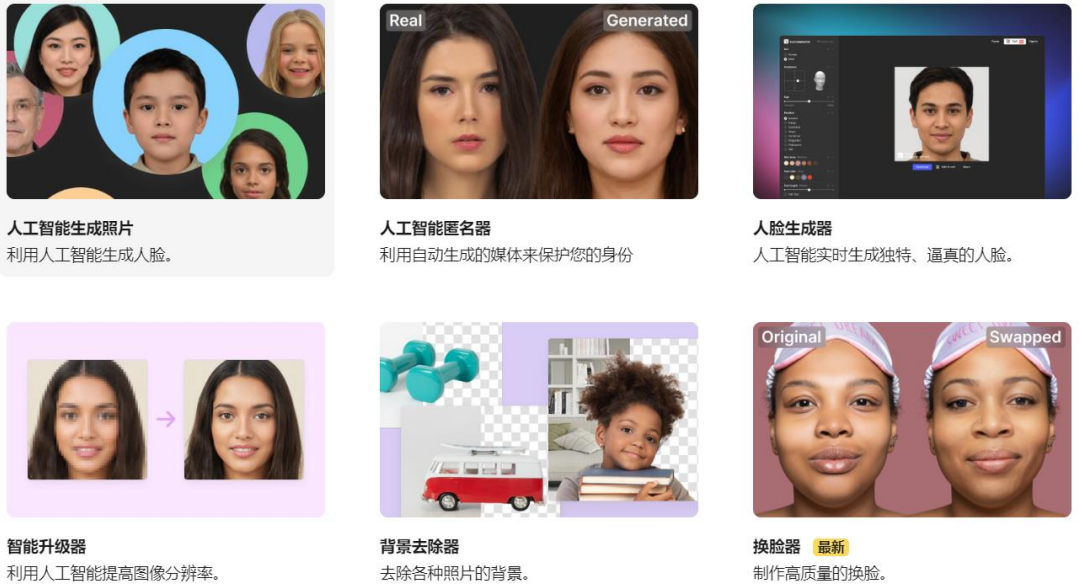
资料来源：DeepFaceLab官网、《Facelet-Bank for Fast Portrait Manipulation》、浙商证券研究所

03 人脸伪造技术分类梳理

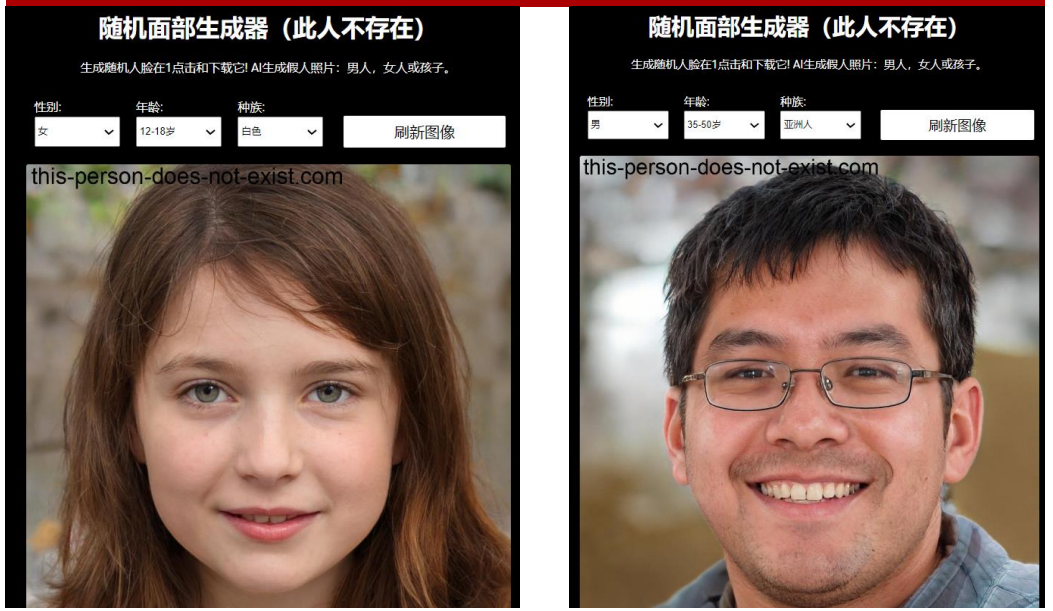
2）无目标可视身份伪造

- **icons8：**icons8是一个提供高质量免费图标和设计工具的网站，网站创始人Ivan Braun等利用StyleGAN模型生成大量伪造人脸图像并按照性别、年龄、种族、头发颜色、表情等属性分类在网站上进行展示，现有客户包含大学、约会软件和人力资源规划公司。icons8（Face Swapper）可以在3天内免费试用，付费内容包括优先处理、电子邮件支持和 60 天存储空间。
- **“This Person Does Not Exist”（此人不存在）**：一个生成虚拟人物头像的网站，由StyleGAN提供支持，以1024x1024像素的分辨率生成不存在的人的照片。网站使用人工智能技术生成照片，可自定义生成人脸的性别、年龄、以及种族，每次刷新页面都会生成一个全新、独特的虚拟人物头像。该网站旨在展示人工智能技术的能力，并提供免费的虚拟人物头像给用户使用。

图：icons8人工智能产品



图：“This Person Does Not Exist”网站



资料来源：icons8网站、“This Person Does Not Exist”网站、浙商证券研究所

人脸伪造技术应用

有目标可视身份伪造在俄乌冲突中已有所应用：

1）对乌克兰最高领导人泽连斯基的虚假模拟

2022年3月15日，乌克兰广播新闻媒体Ukraine 24遭到黑客攻击。黑客通过文字滚动新闻在电视直播中发送了有关泽连斯基的虚假消息，并在其网站上发布了泽连斯基呼吁士兵“放下武器投降”的深度伪造视频，该视频随即在Twitter等社交媒体上广为传播。

对泽连斯基的深度伪造视频属于**有目标可视身份伪造中的人脸编辑类**，伪造视频中存在画质不清晰、场景单一 、人物面部与颈部的肤色存在差异、人物头肩比不协调、人物上半身无肢体动作、眼神始终注视镜头等问题。网络攻击者利用**Pix2Pix、Face2Face和Speech2Face**等算法技术，以宣布停战为博弈主题，试图达到影响对方国家的目的。

2）对俄罗斯最高领导人普京的虚假模拟

2022年3月16日，Twitter 用户 @Serhii Sternenko 发布了“普京宣布停战”的深度伪造视频。对普京的深度伪造视频同样属于**有目标可视身份伪造中的人脸编辑类**，但深度伪造普京的视频制作更为精良，通过多景别切换、人物动作、微表情等方法使得视频仿真性更高，迷惑性更强。

资料来源：《认知的罗生门效应制造：深度伪造在俄乌冲突中的案例分析》、浙商证券研究所

04 以视觉为例：生成式图像检测手段

检测方向一：检测图片是否由生成式AI模型生成

检测GAN模型生成的图像：

GAN图检测通常使用预先训练好的GAN模型来生成假图像，并将这些假图像与真实图像混合在一起形成一个混合的数据集。然后，将该数据集送入分类器中进行训练，以区分真实图像和虚假图像。分类器通常采用卷积神经网络（CNN）进行训练，并基于图像的梯度、频率特征等进行判断，从而有效检测出虚假图像。

图：GAN图检测流程

GAN图检测方法

二分类方法

GAN生成的图像通常具有一些特殊的视觉特征，如颜色块、纹理不自然、边缘不清晰等。可以使用深度学习模型来训练一个分类器，将GAN生成的图像和真实图像分别作为正例和反例，然后使用测试数据集来测试分类器的准确率和性能。

频率特征方法

GAN生成的图像通常具有一些频率特征，如频率分布不均匀、频率分布不连续等。可以使用快速傅里叶变换（FFT）来计算图像的频率谱，然后使用一些统计方法来分析频率特征，如能量分布、频率分布、频率分布差异等。

梯度特征方法

GAN生成的图像通常具有一些梯度特征，如梯度分布不均匀、梯度变化不连续等。可以使用Sobel算子、Laplacian算子等边缘检测算法来计算图像的梯度，然后使用一些统计方法来分析梯度特征，如梯度分布、梯度变化、梯度方向分布等。

资料来源：基于《GAN-generated Faces Detection: A Survey and New Perspectives》，Xin Wang等著的论文整理、浙商证券研究所

检测方向一：检测图片是否由生成式AI模型生成

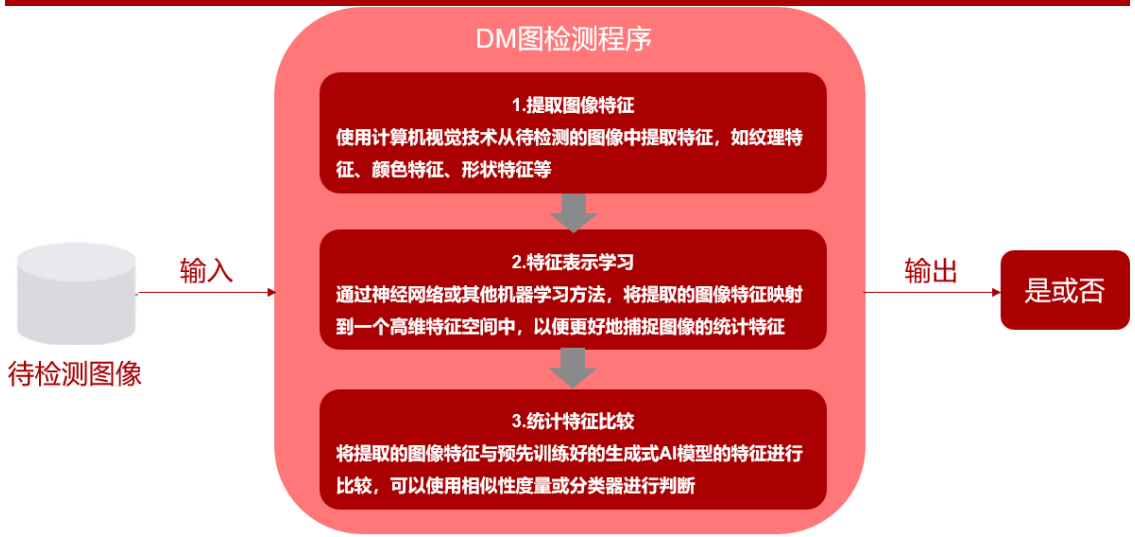
检测Diffusion模型生成的图像：

生成式AI模型通常具有一些独特的统计特征，例如纹理、颜色分布、物体形状等，DM图在某些统计特征上可能与自然图像有所不同，这可以用于区分它们，通过分析图像的统计特征并与已知的视频一致性检测。

现有模型在DM图上的泛化性能仍然有限，**并且DM图的识别比GAN图更具挑战性。**虽然已经有一些研究成果和原型系统，**但目前仍然没有一个普适性和完全可靠的DM图检测方法**：

- 直接将GAN上的模型用于DM图的检测效果较差，但可经过微调后恢复一定性能；
- 二次处理（如压缩）会使生成图更难以判断；
- DM图可以被扩散模型重建，而真实图片不行。通过计算重建图像与原图之间的扩散重建差（DIRE）作为特征进行二分类训练，可以提高判断的泛化性能。

图：理论上DM的检测流程



资料来源：基于《On the detection of synthetic images generated by diffusion models》，Riccardo Corvi等著的论文整理、浙商证券研究所

检测方向一：检测图片是否由生成式AI模型生成

方法多样性

模型差异
不同的方法采用不同的模型来提取真实图和生成图的特征，这导致了性能上的差异。

特征差异
不同方法使用不同的特征进行训练，包括纯视觉信息（如伪影、混合边界、全局纹理一致性）、图像频率信息以及重建图与待检测图的差异等。

数据差异
一些方法通过对抗手段生成更具挑战性的图片，以增强模型的识别能力。

挑战

泛化性差
跨模型检测的泛化性较差。当训练集中的生成图由特定的生成器产生时，检测器在对同一生成器生成的图片进行检测时表现良好，但对于新的生成器生成的图片检测性能会较差。

模型 Ensemble
生成模型可能采用多个子模型的Ensemble方法生成图片,这会增加检测难度。检测模型需要考虑不同子模型的特征,而不仅仅是单一模型的特征。

针对性生成
恶意的生成模型可能专门生成那些难以检测的图片,以躲避检测。这需要检测模型有足够的泛化能力,不易被针对性生成的样本欺骗。

改进方向

提升泛化性能
研究者可以探索提升跨模型检测的泛化性能的方法，使得检测器对于任何生成器生成的图片都能有效。

数据多样性
通过使用更多不同生成器生成的数据来训练检测器，以增强其对新生成器的适应能力

结合多种特征
可以考虑结合不同方法使用的特征，构建更综合的特征表示，从而提高判断的准确性和泛化性。

资料来源：基于《Detecting GAN generated Fake Images using Co-occurrence Matrices》，Lakshmanan Nataraj等著的论文整理、浙商证券研究所

检测方向二：伪造人脸检测

能够检测GAN模型生成的人脸图像的方法分为四类

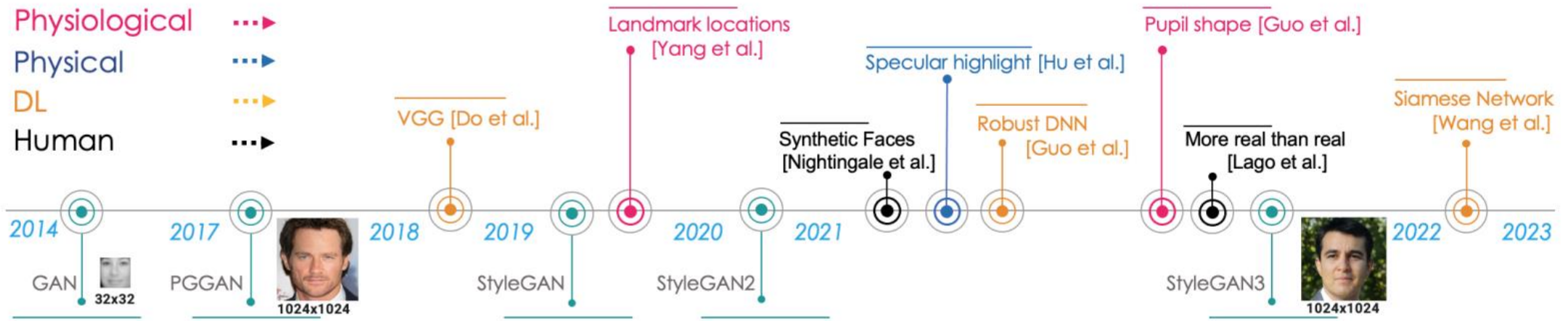
1. **基于深度学习的方法是最常用的方法之一**，它使用卷积神经网络（CNN）来提取特征并进行分类。尽管这些方法在准确性方面表现出色，但它们的图片识别（分类）可解释性相对较差。即使当人类无法察觉到图像的细微差别时，深度学习模型可以识别出来，但无法提供解释判断的原因，这使得解释结果变得困难。

2. **基于物理的方法通过寻找人工信息或面部与物理世界之间的不一致来检测GAN生成的人脸图像。**例如，透视中的照明和反射等物理特征可以用于识别GAN生成的人脸。这些方法不需要大量标记数据，但在准确性方面可能会受到噪声和其他因素的影响。

3. **基于生理的方法使用生理特征来检测GAN生成的人脸图像。**这些生理特征包括对称性、虹膜颜色、瞳孔形状等线索。通过研究这些生理特征，可以识别出GAN生成的人脸。与基于物理的方法类似，基于生理的方法也不需要大量标记数据，但可能会受到个体差异和其他因素的影响。

4. **评估和比较人类视觉性能的方法使用人类参与者来评估GAN生成的人脸图像是否真实。**这些方法可以提供有关GAN生成图像质量和真实性的有用信息。然而，这些方法并不直接用于检测GAN生成的人脸，而是用于评估和比较人类和AI的视觉性能。

图：GAN 人脸生成和检测工作的简要年表



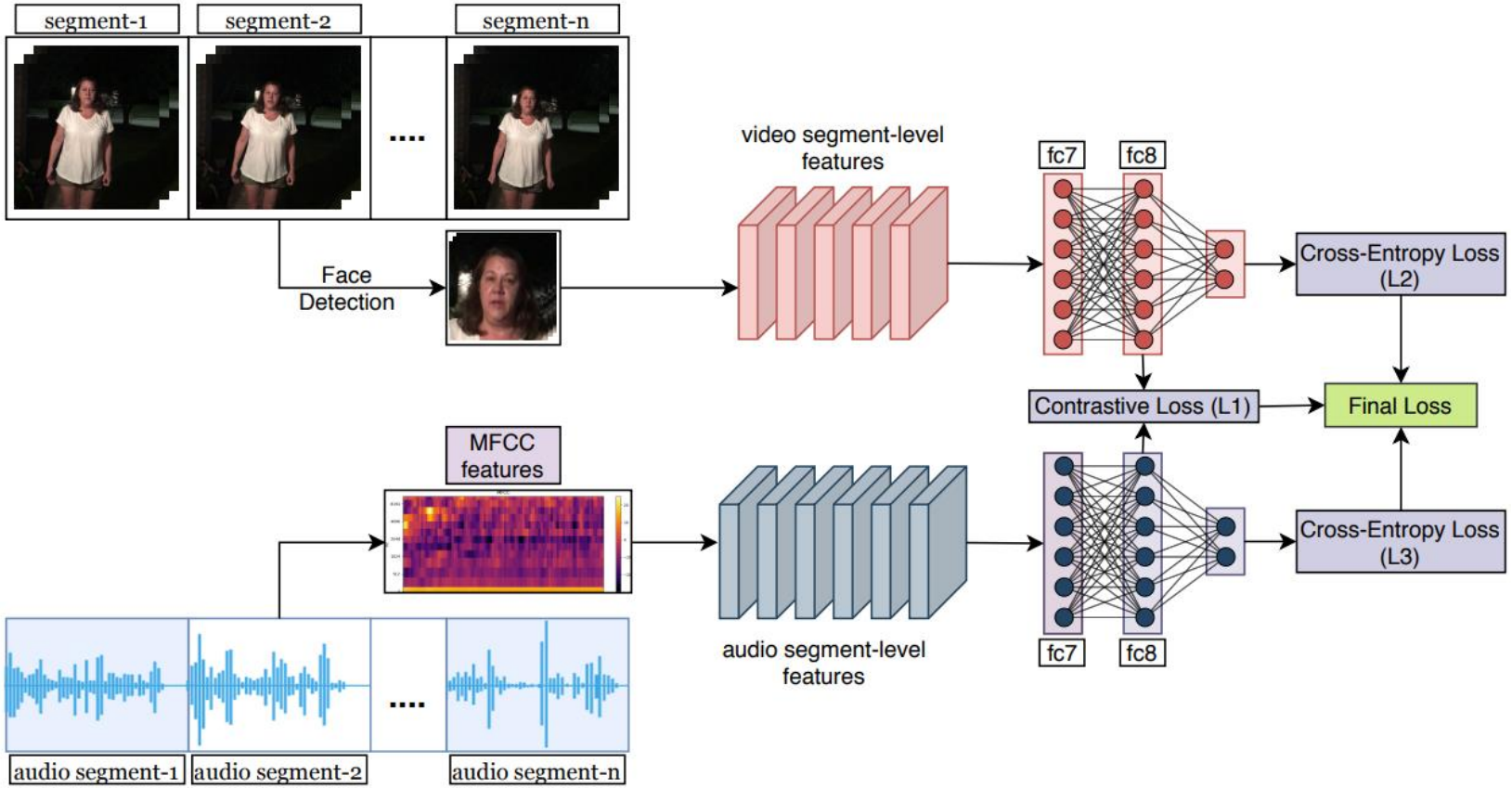
资料来源：《GAN-generated Faces Detection: A Survey and New Perspectives》，Xin Wang等、浙商证券研究所

检测方向二：伪造人脸检测

利用生成人脸仍不够逼真的漏洞

基于 DS 的假视频检测：从 1 秒视听片段中提取的特征输入到 MDS 网络（包括音频和视觉子网络）。视频和音频子网络学习的描述符通过交叉熵损失进行调整，而对比损失用于增强音频之间的更高差异性 - 假视频产生的视觉块。 MDS 被计算为视频长度上的聚合视听不和谐，并用作将视频标记为真/假的品质因数。

图：基于DS的假视频检测流程图



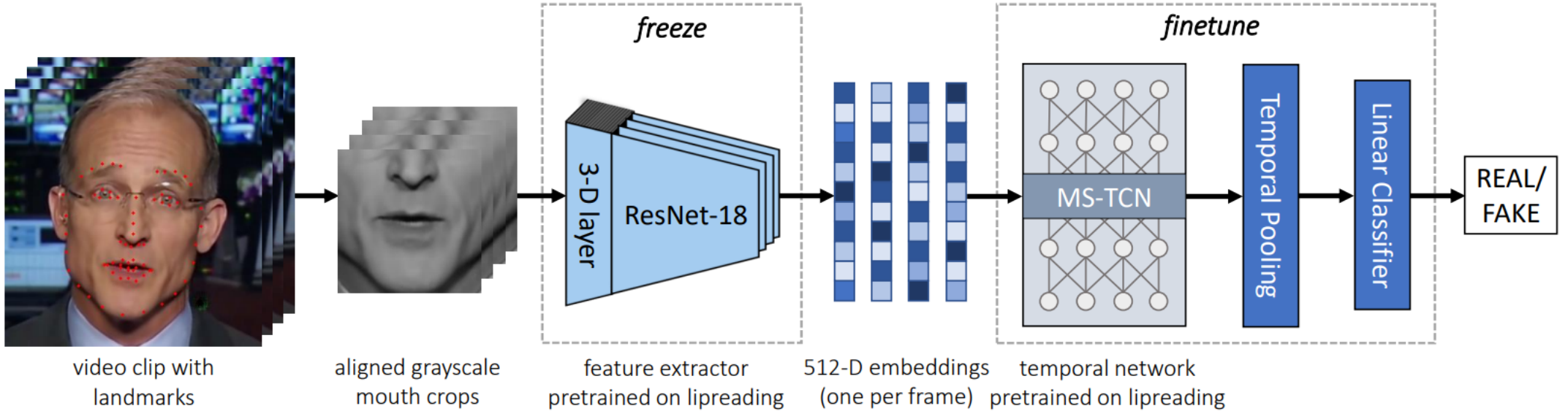
资料来源：《Not made for each other- Audio-Visual Dissonance-based Deepfake Detection and Localization》，Komal Chugh等、浙商证券研究所

检测方向二：伪造人脸检测

利用生成人脸仍不够逼真的漏洞

人脸伪造检测微调阶段：网络的输入包含 25 个灰度、对齐的唇部（我们仅显示四个用于说明目的）。它们通过冻结特征提取器（具有初始 3-D 卷积层的 ResNet-18），该特征提取器已经过唇读预训练，因此输出对嘴部运动敏感的嵌入。一个多尺度时间卷积网络 (MS-TCN)，也在唇读上进行了预训练，被微调以检测基于嘴部运动中语义高级不规则性的假视频。

图：基于唇部的伪造人脸检测流程图



资料来源：《Lips Don't Lie: A Generalisable and Robust Approach to Face Forgery Detection》，Alexandros Haliassos等、浙商证券研究所

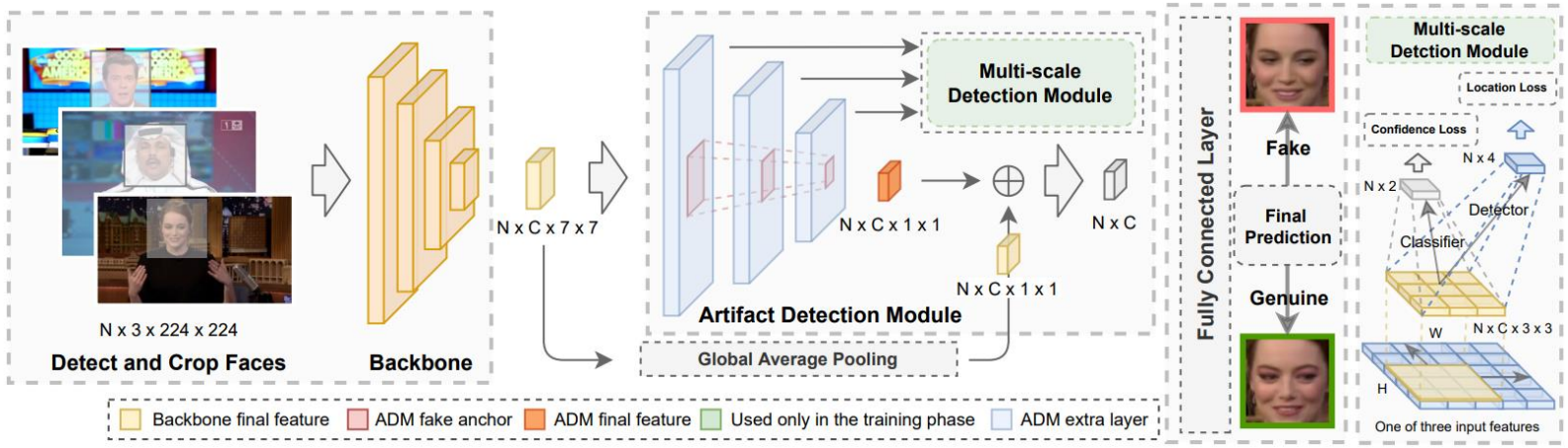
检测方向二：伪造人脸检测

利用模型自身的能力

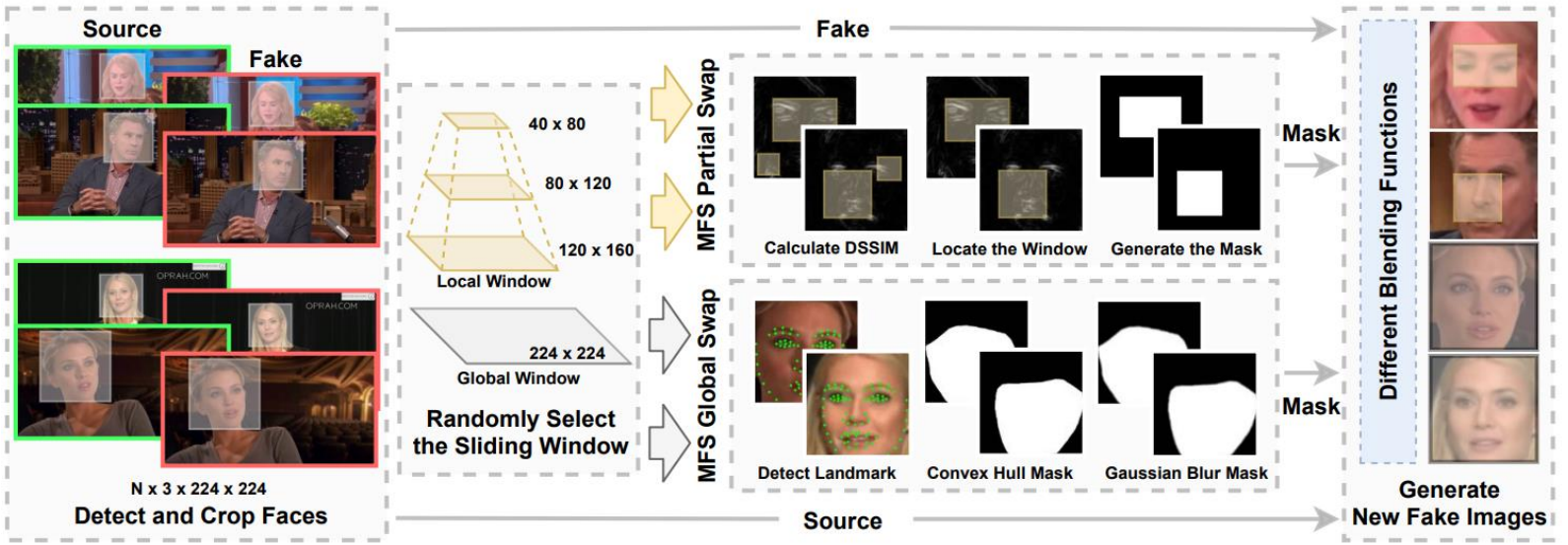
N 和 C 表示图像和通道的数量。在伪影检测模块的帮助下，ID-unaware Deepfake检测模型旨在关注图像的局部表示以指示人脸伪造。

MFS以全局交换和部分交换两种方式对配对的假图像和源图像进行操作，以生成具有工件区域真实值的新假图像。

图：ID-unaware Deepfake检测模型的整体框架



图：多尺度面部交换 (MFS) 概述



资料来源：《Implicit Identity Leakage: The Stumbling Block to Improving Deepfake Detection Generalization》，Zheng Ge等、浙商证券研究所

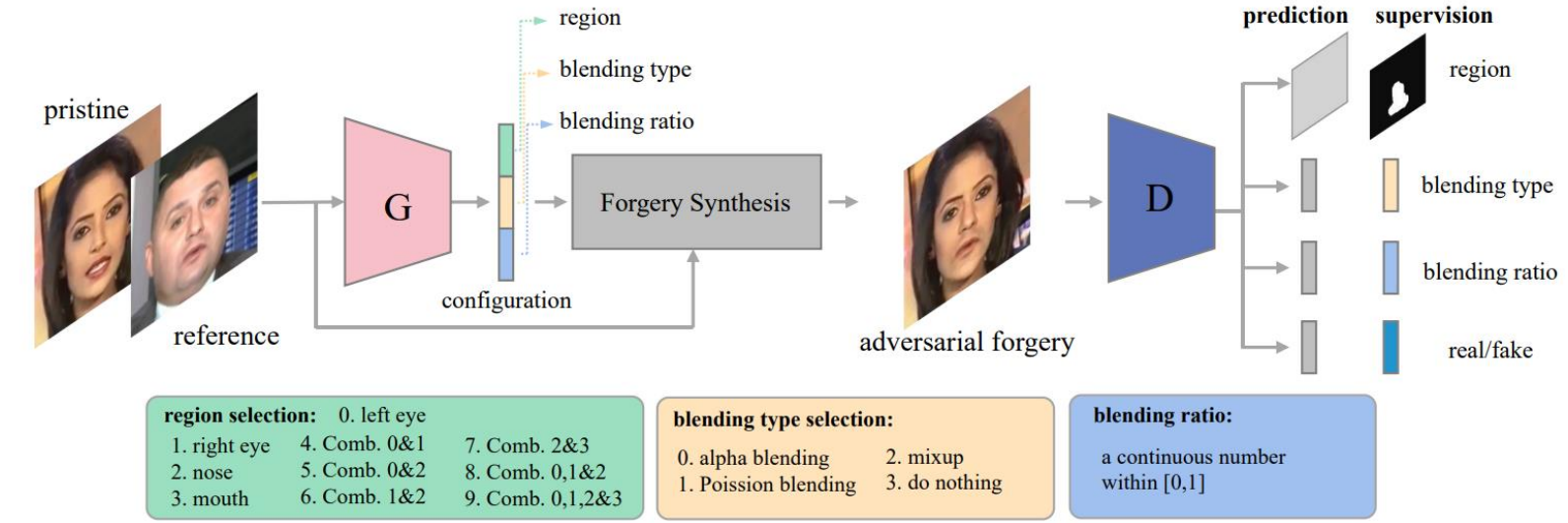
检测方向二：伪造人脸检测

利用模型自身的能力

对抗训练框架由合成器G、图片合成和判别器 D 组成，其中：

- 合成器：生成配置参数，用来合成更丰富的自监督样本数据（注意是生成配置的参数）
  1. 原生假图：不处理，即不进入合成器，直接用来训判别器；
  2. 原生真图：不增广的真图不进入合成器，直接训判别器；
  3. 合成假图：有一定概率与一个随机图（Reference）进行增广，形成局部虚假的假图；
- 图片合成：合成器G会生成配置方案（区域选择10个；混合blending类型选择；以及合成比例选择ratio），基于此进行合成（即数据增广）
- 判别器G：对图片进行分类，同时添加辅助任务，用合成器的G的输出作为label

图：对抗样本的自监督学习模型概述图



合成器网络（即生成器）输出三个伪造配置，进一步用于合成新的伪造，这些伪造配置也用作标签来指导检测器网络（即鉴别器），以对抗的方式训练生成器和鉴别器。

资料来源：《Self-supervised Learning of Adversarial Example: Towards Good Generalizations for Deepfake Detection》，Jue Wang等、浙商证券研究所

05 Deepfakes技术发展趋势展望

05 生成式AI技术的迭代增加伪造图片的真实度

1、数据：可供训练的数据集规模扩大，带来模型效果的大幅提升

2009年发布的ImageNet数据集有1400多万幅图片，涵盖2万多个类别，其中有超过百万的图片有明确的类别标注和图像中物体位置的标注；而2022年3月开放的LAION-5B数据集包含58.5亿个图片-文本对，训练数据集规模的扩大可提升模型精度。

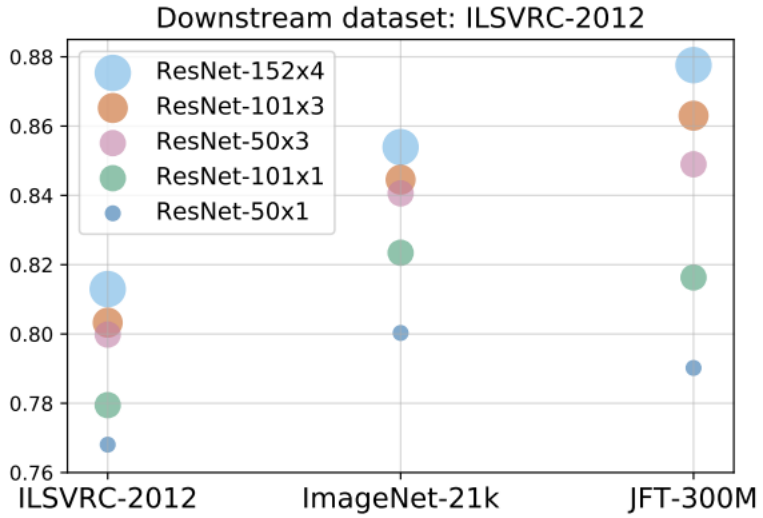
2、算法：算法框架的迭代与组合，使得模型训练更加高效、稳定

VAE模型由于在训练过程中调整参数只能得到平均的结果，所以生成图片较为模糊；GAN模型通过生成器和判别器相互对抗，可以生成高分变率的图像，但由于需要同时训练生成器和判别器两个模型，训练难度大、不稳定，结果不容易收敛；Diffusion模型在上述两个模型的基础上，在图片质量和模型训练难度上取得了较好的效果，在其基础上进一步迭代的Stable Diffusion又进一步解决了传统Diffusion模型的采样速度慢的缺点。

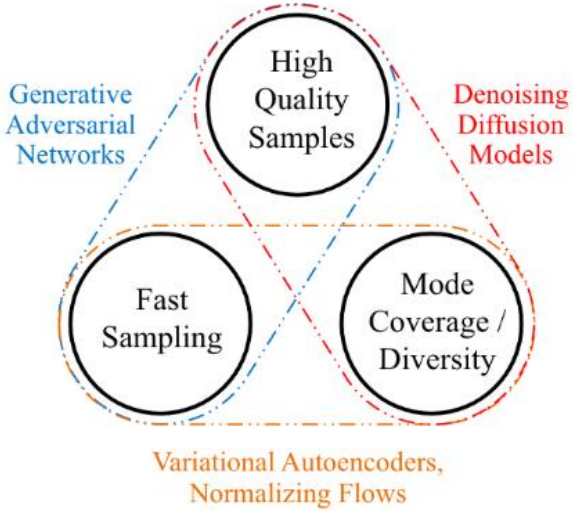
3、算力：算力水平的提升，助力更加高效的复杂模型训练

算力水平的提升使得训练更大规模参数的模型成为可能，且可通过多张高算力芯片的组合，实现训练效率的提升。以ChatGPT为例，用单个Tesla架构的V100训练包含1746亿参数的GPT3.0需要288年的时间，而用1024张A100GPU上预估需要34天，而多台的H100甚至可以将GPT的模型训练时间缩短到几天。

模型精度性能随训练集规模增大提升



各类生成式模型的优缺点



资料来源：ImageNet网站、无界AI公众号、脑智新语公众号、码上跃见公众号、九慕资产管理公众号、浙商证券研究所

AIGC的识别难度增大，潜在风险提升

技术实现层面

- 一方面，生成式AI模型的迭代，使得生成的内容越来越逼真，依靠AI生成内容逼真程度的漏洞来识别伪造图片或音频的难度增大
- 另一方面，AI模型能力的增强，也增加了依靠模型自身能力鉴别AIGC的潜力，deepfakes检测手段会持续丰富

风险管理机制层面

- AIGC识别难度增大，难以单独从技术层面对AI生成内容进行识别和鉴定，需要多方机制共同对AIGC风险进行监管
- 对于AIGC的监管从事后检测向事前规范倾斜，如提高生成式AI模型研发的准入门槛，制定统一的技术标准，增强AIGC的源头监管，落实责任划分等

资料来源：浙商证券研究所总结整理

06 AIGC监管及反生成AI市场空间测算

06 生成式AI监管政策梳理

国内

2019.11 网信办、文旅部、广电总局《网络音视频信息服务管理规定》：“不得利用基于深度学习、虚拟现实等新技术新应用制作、发布、传播虚假新闻信息。”

2019.12 网信办《网络信息内容生态治理规定》明确网络信息内容服务使用者和生产者、平台不得开展包括深度伪造在内的违法活动。

2022.01 网信办、工信部、公安部、市场监督管理总局《互联网信息服务算法推荐管理规定》将生成合成类算法列入收到监管的五大类应用算法推荐技术之一，并要求对AI生成图片、视频进行标识。

2022.11 网信办、工信部、公安部《互联网信息服务深度合成管理规定》明确强化深度合成服务提供者和技术支持者的主体责任，并建立健全的辟谣、举报机制。

2023.04 网信办《生成式人工智能服务管理办法（征求意见稿）》强化AI生成内容治理、优化技术监管、强化个人权利保障。

国外

2018.05 欧盟《通用数据保护条例》将包括可能被用于制作深度伪造内容的公民图片等个人数据置于现行欧盟法律保护之下。

2018.09 欧盟《反虚假信息行为准则》加强互联网企业对平台内容的审查，提高对包括深度伪造等音视频文件的管控。

2018.12 美国国会《恶意伪造禁令法案》：对制作深伪引发犯罪和侵权的个人、明知内容的深伪性质仍进行分发的社媒平台处以罚款和监禁。

2019.06 美国国会《深度伪造责任法案》：制作深伪媒体内容的作者必须用“不可删除的数字水印以及文本描述“标明该内容经过篡改或生成。

2020.05 澳大利亚战略政策研究所《深度伪造技术武器化》详细介绍了七个常见的深度造假工具：换脸、重新投射、口型同步、动作传递、图像生成、音频生成、文字生成。

2020.08 美国会研究服务处《深度伪造与国家安全》、《人工智能与国家安全》提示深度伪造已成为对手信息战的一部分。

2023.03 意大利个人数据保护局宣布禁止使用ChatGPT，限制OpenAI处理意大利用户信息，并针对隐私安全问题立案调查。

2023.06 欧洲议会通过了关于**《人工智能法案》**的谈判授权草案，针对生成式AI设立专门监管制度，并要求AI公司对其算法保持人为控制，提供技术文件，并为“高风险”应用建立风险管理系统。

资料来源：网信办、欧盟官网、信通院互联网法律研究中心公众号、信息安全与通信保密杂志社公众号、财联社公众号、新华网、每经网

反AI生成市场空间测算

监管侧市场

- 潜在需求场景：司法鉴定与鉴别场景
- 潜在需求方：公检法司
  - 公安局：3500
  - 检察院：3588
  - 法院：3508
- 售价假设：42万/套（对标美亚柏科电子取证产品）
- 空间测算：

42万/套×（3500公安局+3588检察院+3508法院）= **44亿元**

说明：由于官方未披露最新公检法司数量情况，我们参考最高法司法案例研究院2017年披露的数据进行估算

企业市场

- 潜在需求场景：AI伪造图片/音频识别
- 潜在需求方：企业及个人用户
  - 互联网用户数量（截至2022年6月）：10.5亿
- 售价假设：325元/月（每天4万次调用请求）（对标海外产品售价）
- 空间测算（纯软件部分）：

3900元/年×10.5亿互联网用户× 20次调用/天*365÷1460万次调用/年= **20亿元**

表：监管侧市场规模敏感性分析（单位：亿元）

渗透率 \ 单套设备售价（万元）	20	25	30	35	40	42	50	60
70%	14.83	18.54	22.25	25.96	29.67	31.15	37.08	44.50
80%	16.95	21.19	25.43	29.67	33.90	35.60	42.38	50.86
90%	19.07	23.84	28.61	33.37	38.14	40.05	47.68	57.21
100%	21.19	26.49	31.79	37.08	42.38	**44.50**	52.98	63.57
110%	23.31	29.14	34.96	40.79	46.62	48.95	58.27	69.93
120%	25.43	31.79	38.14	44.50	50.86	53.40	63.57	76.29

表：企业侧市场规模敏感性分析（单位：亿元）

平均每人每天调用次数（次） \ 每月订阅价格（元/月）	65	103	130	195	260	325	390
5	1.02	1.62	2.05	3.07	4.10	5.12	6.14
10	2.05	3.24	4.10	6.14	8.19	10.24	12.29
15	3.07	4.85	6.14	9.21	12.29	15.36	18.43
20	4.10	6.47	8.19	12.29	16.38	**20.48**	24.57
30	6.14	9.71	12.29	18.43	24.57	30.71	36.86
50	10.24	16.18	20.48	30.71	40.95	51.19	61.43

说明：基本假设为每个机构需要采购一台设备，则在100%渗透率的情况下，每个监管机构均需采购一台设备；若部分机构未采购设备，则对应渗透率低于100%；若部分机构采购超过1台设备，则对应的渗透率超过100%

资料来源： 采招网、最高人民法院司法案例研究院公众号、illuminarty网站、AI Voice Detector网站、人民网公众号、浙商证券研究所

07 公司梳理

美亚柏科

美亚柏科是国内领先的**电子数据取证行业龙头**和**公安大数据领先企业、网络空间安全和社会治理领域国家队**，持续深耕“大数据智能化”与“网络空间安全”两个主要赛道；公司于2017年成立AI研发中心，深度开展人工智能技术研究，并于2019年**针对深度合成技术成立专项研究团队**，现已**具备自主研发深度伪造视频图像鉴定的核心引擎的能力**；2023年3月20日，公司作为编写单位，**参与编制国内首个生成式人工智能标准体系。**

公司自主研发的**AI-3300“慧眼”视频图像鉴真工作站**具备智能鉴定和专业鉴定两种模式，涵盖40余种视频图像真伪鉴定算法、近10种深伪鉴定算法，能够全方位多视角地识别利用深度伪造手段进行的换脸、美颜、生成人脸、同图或异图复制篡改的影像，**识别效果处于国际领先水平**，能够为公安、司法行业及相关领域提供一站式影像真伪检测鉴定解决方案。

图：美亚柏科主营业务

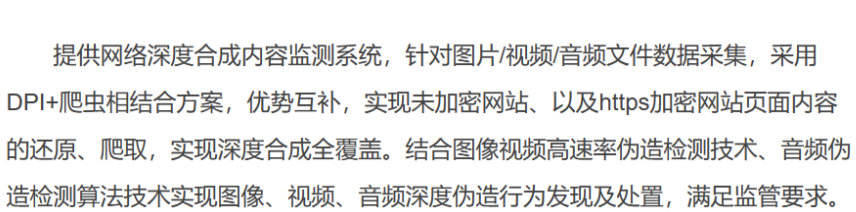
图：AI-3300“慧眼”视频图像鉴真工作站

资料来源：美亚柏科官网、浙商证券研究所

东方通作为国内**领先的大安全及行业信息化解决方案提供商**，持续深耕网络信息安全领域，其产品服务覆盖至政务、金融、企事业单位等多个行业，面对新一轮的AIGC识别产品的市场竞争，公司**在技术转化落地方面具有深厚的实战经验优势**；公司于2023年3月正式**立项课题《AIGC算法安全性检测方法研究》**，率先展开针对ChatGPT等AIGC算法的安全性评估测试方法和工具开发。

公司现有**“深度合成内容监测”**产品，采用**软硬件协同**的解决方案，**具备对常见AIGC算法生成的图片、音视频内容的检测能力**，支持深伪检测、特定人物伪造检测、伪造溯源等，并已通过部分客户实现应用；其中音频类检测**准确率可达特定人86%、非特定人76%**；同时，公司**将推出TongGPT**多模态智能交互模型，预计年底前实现针对**AI生成语音诈骗的自动化提醒**功能。

图：东方通AIGC解决方案



图：东方通网络深度合成内容监测系统的产品优势

资料来源：东方通官网、浙商证券研究所

瑞莱智慧

瑞莱智慧（RealAI）作为**人工智能基础设施和解决方案提供商**，于2018年依托清华大学人工智能研究院发起成立，致力于以第三代人工智能技术为高价值场景智能化升级提供一站式赋能方案；公司为政务、金融、能源、制造、互联网等领域合作伙伴提供包括**人脸识别系统安全、深度合成和伪造检测**等产品和解决方案，具备深厚的技术和实践经验积累。

公司推出的**基于第三代人工智能技术**（知识+数据驱动；安全、可靠、可信）的深度伪造检测平台**DeepReal**，通过辨识伪造内容和真实内容的表征差异性、挖掘不同生成途径的深度伪造内容一致性特征，能够**快速、精准**地对图像、视频、音频内容进行真伪鉴别，在第三届中国人工智能大赛中**斩获深度伪造视频检测A级证书**；同时，DeepReal在学术数据集与主流网络数据集**检测准确率达99%**、在产业实践检测准确率**达业内较高水平。**

图：瑞莱智慧在第三届中国人工智能大赛中斩获的奖项

赛题2：深度伪造视频检测

A级：淘宝（中国）软件有限公司、杭州海康威视数字技术股份有限公司、北京瑞莱智慧科技有限公司

B级：杭州网易智企科技有限公司、中国科学院计算技术研究所、杭州中科睿鉴科技有限公司、北京数美时代科技有限公司

赛题3：深度伪造视频生成方法识别

A级：杭州网易智企科技有限公司、中国科学院计算技术研究所

B级：北京瑞莱智慧科技有限公司、淘宝（中国）软件有限公司、中国科学技术大学

图：瑞莱智慧DeepReal产品架构图

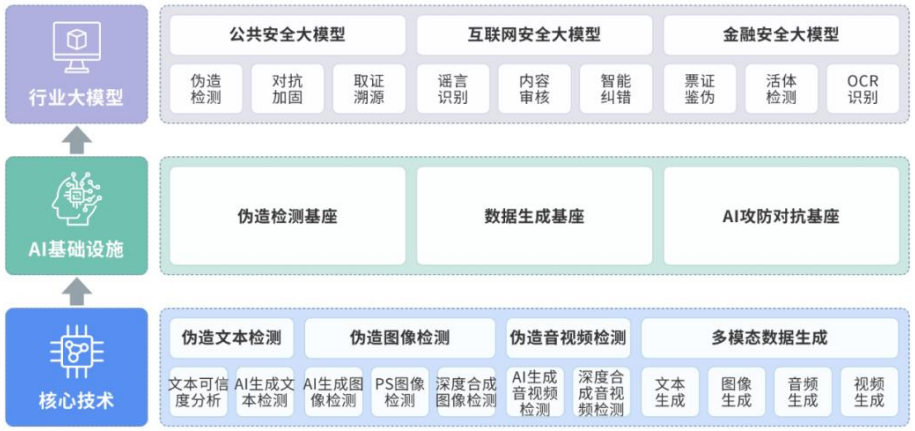
资料来源：瑞莱智慧RealAI官网、中国人工智能大赛官网、浙商证券研究所

中科睿鉴

中科睿鉴成立于2020年3月，核心团队源于中科院计算技术研究所，致力于运用AI技术赋能数字内容安全，**深耕虚假信息识别、深度合成内容检测、深度伪造溯源**等技术研发，在国家公共安全监管、媒体内容审核、金融风控管理、保险理赔欺诈等关键业务场景中，**率先实现伪造监测、伪造溯源、AI攻防对抗三大基础设施的全技术栈布局。**

公司积累了分行业、分场景的伪造检测能力，拥有**参数量达60亿**的**行业基础大模型体系化能力底座**，能够在此基础上微调，**针对新的伪造生成技术迅速分化出不同检测模型**；软件方面，目前已有音视频生成内容检测工具“睿安”、图像生成内容检测工具“睿图”、文本生成内容检测工具“睿鉴图灵”，**全方位覆盖AIGC文本、图像、音视频的检测能力**；硬件方面，已有“睿安深伪检测专用设备“，为**国内首款软硬一体的深伪检测专用设备**，破解”敏感任务安全可控处理”和“现网流量大规模部署”两大瓶颈，并实现**全国产化硬件生态适配**，支持主流国产芯片。

图：中科睿鉴行业基础大模型框架



图：睿安深度合成智能检测平台

图：睿安深伪检测专用设备

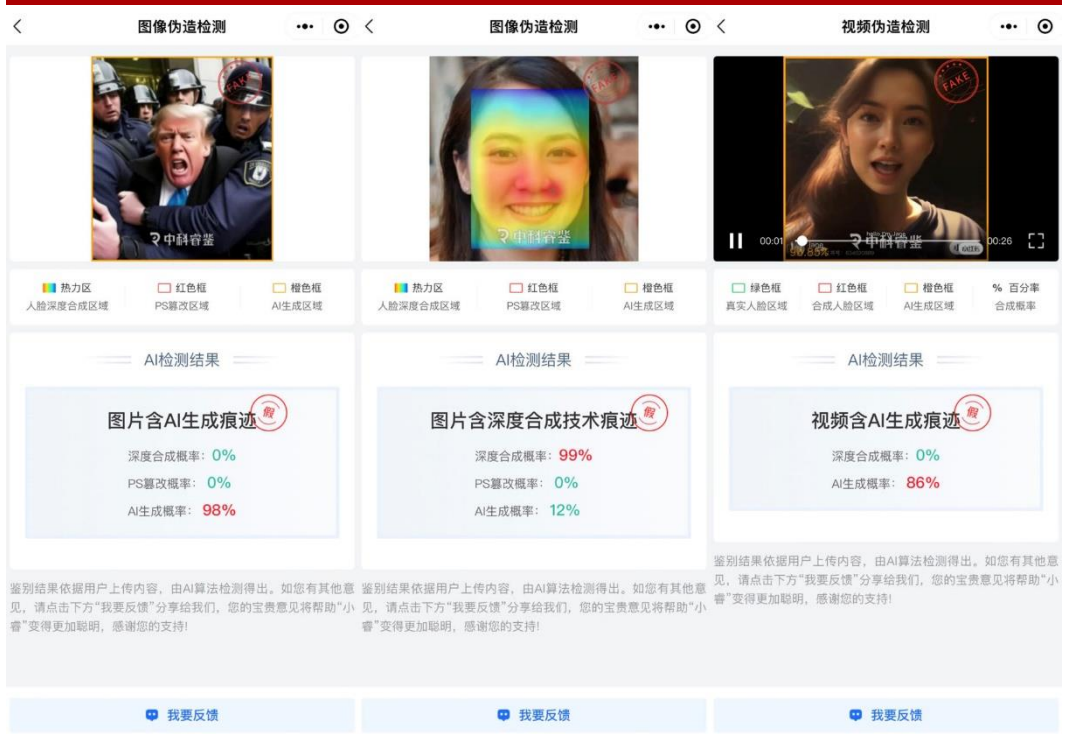
资料来源：中科睿鉴官网、浙商证券研究所

中科睿鉴已通过**“睿鉴AI”小程序**，整合新闻可信度分析、图像伪造检测、视频伪造检测、AI生成文本检测等功能并面向公众开放，每项功能可每日免费使用15次。

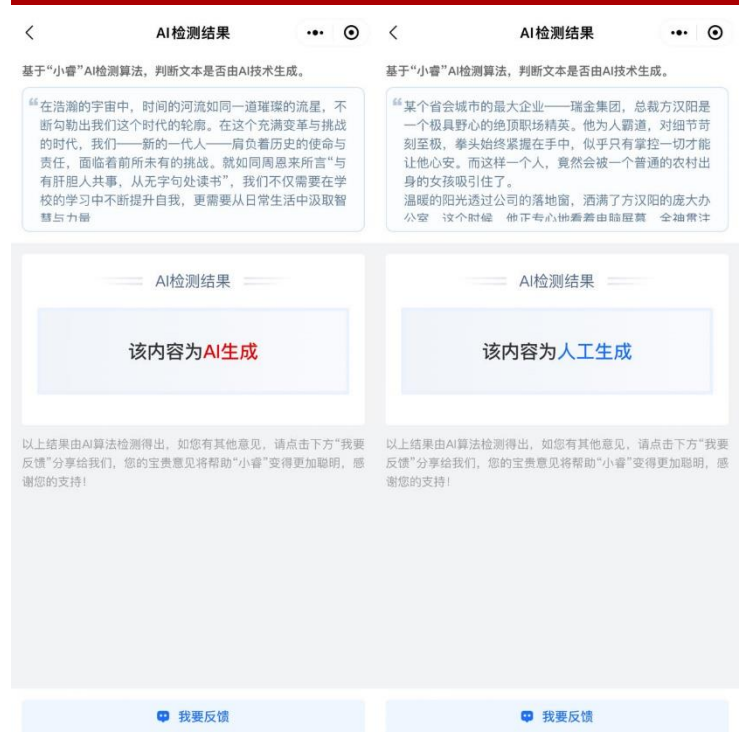
使用体验方面，“睿鉴AI”小程序**简单易用、响应迅速、交互流畅**，大幅降低了公众辨别伪造内容的技术门槛；识别准确率方面，经随机测试，“睿鉴AI”**能够准确识别伪造图像和视频**，同时分类标注深度合成/PS篡改/AI生成的区域；而AI生成的文本经测试有概率被识别为人工生成，**文本识别准确率仍存在进步空间。**

图：“睿鉴AI”小程序主页面

图：“睿鉴AI”伪造图片和视频检测结果



图：“睿鉴AI”AI文本检测结果



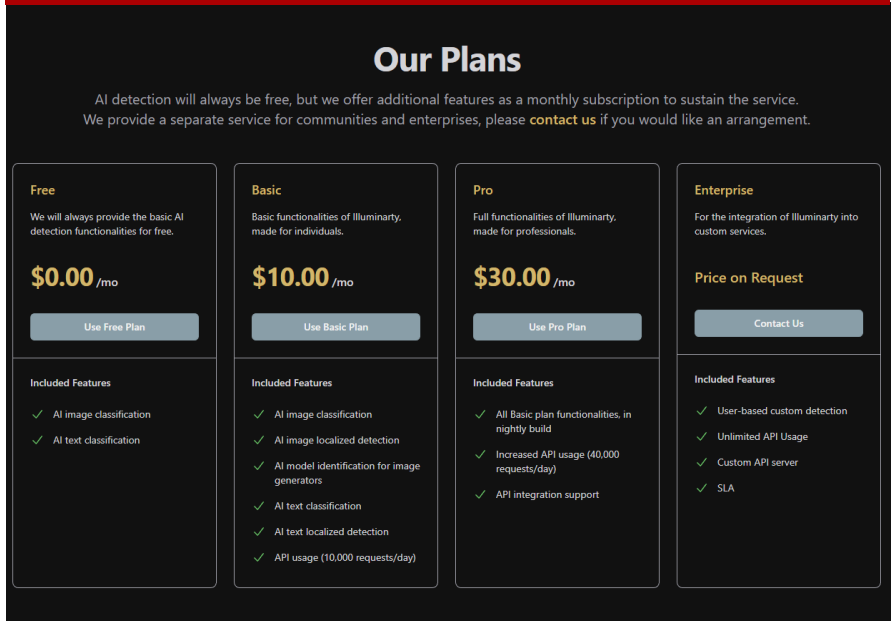
资料来源：中科睿鉴官网、“睿鉴AI”小程序、浙商证券研究所

Illuminarty

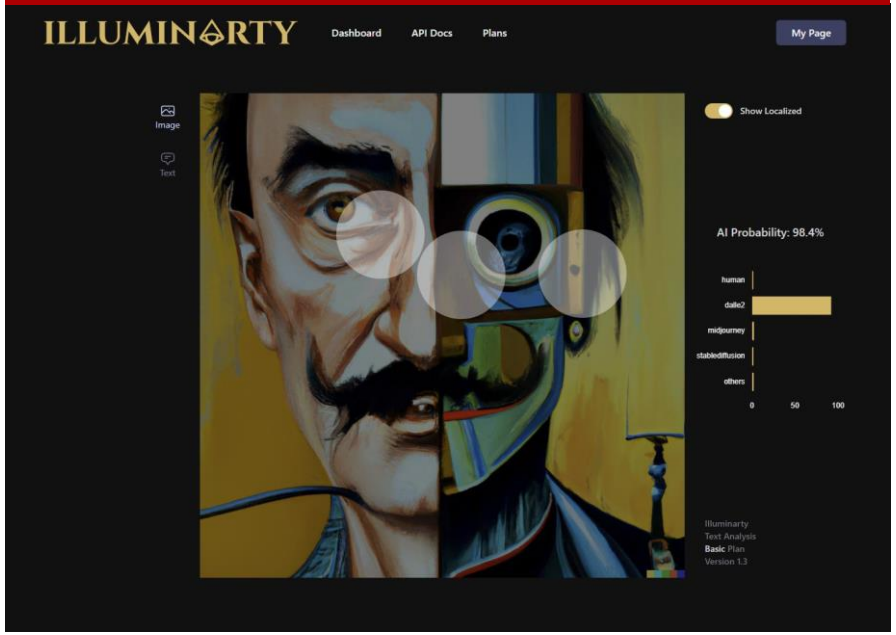
Illuminarty是一个**在线的AI生成内容检测网站**，提供**免费**的在线AI检测**基础服务**，能够检测用户上传的图片和文本是否由AI生成；同时提供“基础”、“专业”、“企业”三种**按月订阅套餐**，为有进阶需求的个人和企业用户提供更专业的AI内容检测服务；平台于2022年10月活跃于互联网，目前正在开发浏览器扩展程序以提供更便捷的AI内容检测服务。

Illuminarty的图片检测采用计算机视觉算法，检测上传的图片是否由AI生成以及其使用的AI模型，并给出图片的AI生成概率；目前已**具备检测主流生成式AI模型的能力**，但**功能尚存限制**：1. 基础设施受限，无法处理>3MB的图片；2. 样本受限，检测高分辨率图片、特定艺术风格图片、摄影作品、动漫/游戏截图时存在偏差；同时，新加入的检测AI生成文本的功能，可通过NLP算法输出上传文本为AI生成的概率，并标出最有可能为AI生成内容的段落，但现阶段的**识别准确率存在较大提升空间**。

图：Illuminarty订阅套餐及对应功能



图：Illuminarty图片识别功能演示



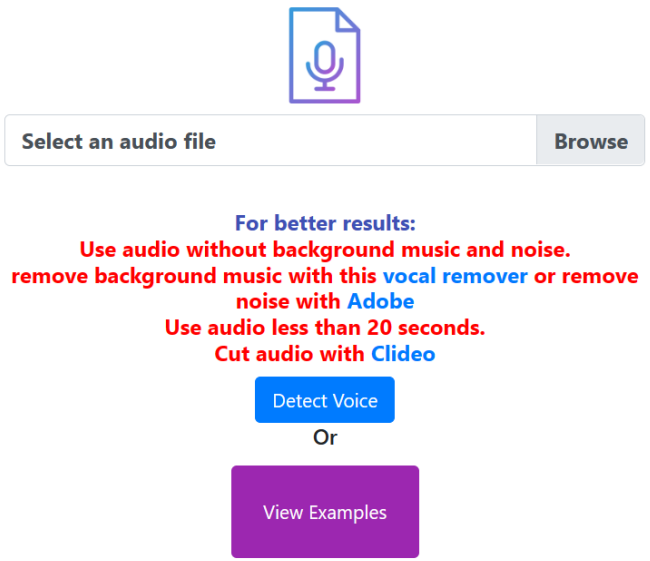
资料来源：Illuminarty官网、浙商证券研究所

AI Voice Detector

AI Voice Detector是一个在线的**AI生成音频检测网站**，**支持上传多种音视频文件格式**并检测用户上传的音频是否由AI生成，可应用于法律诉讼、媒体报道、客服交互等场景的音频真实性识别；其工作原理为通过处理和分析声音信号以提取频率、时域、能量等特征，使用机器学习算法进行分类和识别，输出音频为AI生成的可能性；平台于2023年2月开始活跃于互联网。

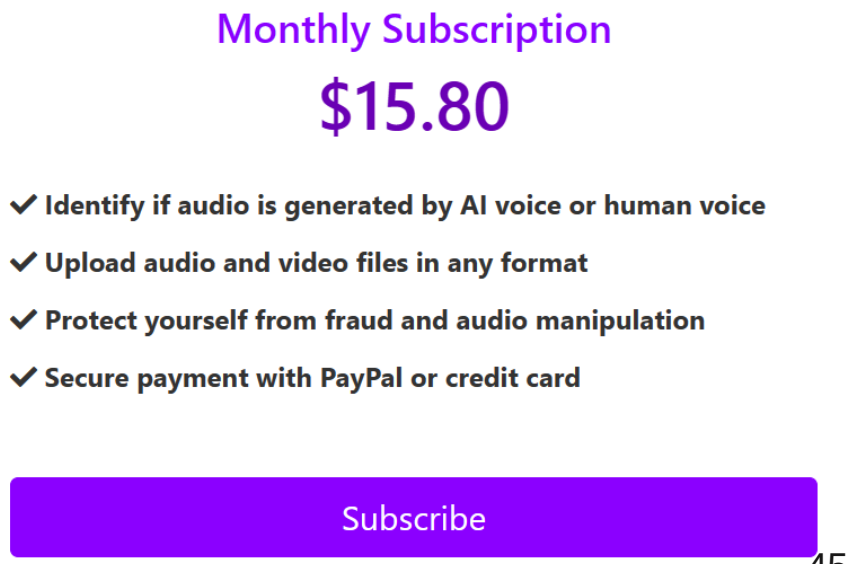
据网站公开信息，AI Voice Detector能够**准确识别特定音频样本是否由AI生成**，但**实际应用受限于包括上传文件自身的背景噪音、音频长度等因素**，需借助降噪、音频剪辑等工具以达到最佳检测效果；目前，网站不提供免费试用选项，**仅支持订阅使用**，且套餐选择单一，**价格门槛较高、用户数量受限。**

图：AI Voice Detector功能演示



图：AI Voice Detector当前用户数量

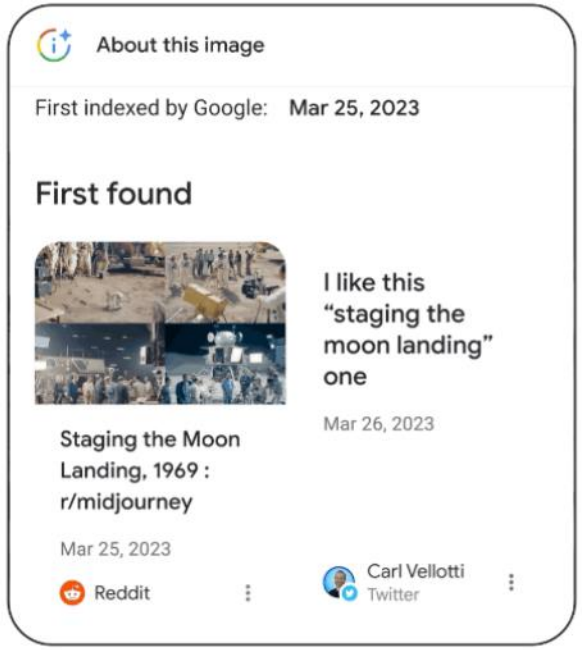
图：AI Voice Detector订阅套餐



资料来源：AI Voice Detector官网、浙商证券研究所

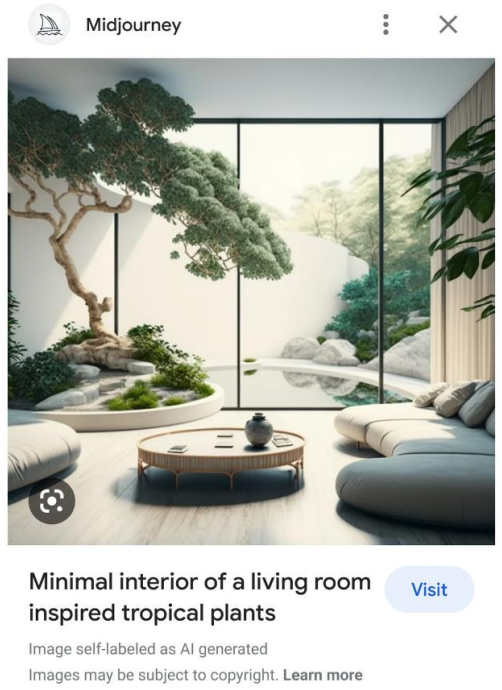
Google于2023年5月宣布将在其图片搜索结果中，加入新功能**“关于此图片”**，该功能将为用户显示**图片首次被Google索引的时间、图片首次出现的站点、显示该图片的其他在线平台**（例如新闻网站、社交媒体）等信息，帮助用户判断网络图片的真实可靠性，提高用户对于虚假信息的防范能力；同时，Google宣布**将为其Imagen生成的图片原始文件加入AI标记**以确保AI生成的图片能够被准确识别，并将**允许内容创作者和发布者为图片进行AI生成标记**，预计Midjourney、Shutterstock等平台将在近期加入AI生成标记。

图：Google“关于此图片”功能演示



图：Google的AI生成图片标记示例

图：其他平台的AI生成标记示例



资料来源：Google官网、浙商证券研究所

08 风险提示

08 风险提示

1、AIGC监管以及识别技术研发以及相关产品落地不及预期

2、报告中对各类模型的介绍与总结基于对于相关论文内容的理解，具有一定主观性

3、报告中对于反AI生成的市场空间测算存在主观判断及口径差异

4、由AI安全需求带来的市场竞争加剧

5、板块政策发生重大变化

附录：主流生成式AI模型概览参考文献

《Auto-Encoding Variational Bayes》,Diederik P Kingma等、

《Generative Adversarial Networks》, Ian J. Goodfellow等、

《Training generative neural networks via Maximum Mean Discrepancy optimization》，Gintare Karolina Dziugaite等、

《Conditional Generative Adversarial Nets》，Mehdi Mirza等、

《Learning structured output representation using deep conditional generative models》，Kihyuk Sohn等、

《Learning What and Where to Draw》，Scott Reed等、

《Conditional Image Synthesis With Auxiliary Classifier GANs》，Augustus Odena等、

《StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks》，Han Zhang等、

《StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks》，Han Zhang等、

《Image-to-Image Translation with Conditional Adversarial Networks》，Phillip Isola等、

《Progressive Growing of GANs for Improved Quality, Stability, and Variation》，Tero Karras等、

《Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks》，Jun-Yan Zhu等、

《A Style-Based Generator Architecture for Generative Adversarial Networks》，Tero Karras等、

《Taming Transformers for High-Resolution Image Synthesis》，Patrick Esser等、

《ViTGAN: Training GANs with Vision Transformers》，Kwonjoon Lee等、《Generative Models for Effective ML on Private, Decentralized Datasets》，Sean Augenstein等、

《Alias-Free Generative Adversarial Networks》，Tero Karras等、

附录：主流生成式AI模型概览参考文献

《Denoising Diffusion Probabilistic Models》，Jonathan Ho等、

《Improved Denoising Diffusion Probabilistic Models》，Alexander Quinn Nichol等、

《Your GAN is Secretly an Energy-based Model and You Should use Discriminator Driven Latent Sampling》，Tong Che等、

《A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT》，YIHAN CAO等、

《Density estimation using Real NVP》，Laurent Dinh等、

《FFJORD: Free-form Continuous Dynamics for Scalable Reversible Generative Models》，Will Grathwohl等、

《Glow: Generative Flow with Invertible 1x1 Convolutions》，Diederik P. Kingma等

行业评级与免责声明

浙商证券股份有限公司
ZHESHANG SECURITIES CO.LTD

行业的投资评级

以报告日后的6个月内，行业指数相对于沪深300指数的涨跌幅为标准，定义如下：

1、看好 ：行业指数相对于沪深300指数表现＋10%以上；

2、中性 ：行业指数相对于沪深300指数表现－10%～＋10%以上；

3、看淡 ：行业指数相对于沪深300指数表现－10%以下。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重。

建议：投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者不应仅仅依靠投资评级来推断结论

行业评级与免责声明

浙商证券股份有限公司
ZHESHANG SECURITIES CO.LTD

法律声明及风险提示

本报告由浙商证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格，经营许可证编号为：Z39833000）制作。本报告中的信息均来源于我们认为可靠的已公开资料，但浙商证券股份有限公司及其关联机构（以下统称“本公司”）对这些信息的真实性、准确性及完整性不作任何保证，也不保证所包含的信息和建议不发生任何变更。本公司没有将变更的信息和建议向报告所有接收者进行更新的义务。

本报告仅供本公司的客户作参考之用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本报告仅反映报告作者的出具日的观点和判断，在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，投资者应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本公司的交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理公司、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

联系方式

浙商证券研究所

上海总部地址：杨高南路729号陆家嘴世纪金融广场1号楼25层

北京地址：北京市东城区朝阳门北大街8号富华大厦E座4层

深圳地址：广东省深圳市福田区广电金融中心33层

邮政编码：200127

电话：(8621)80108518

传真：(8621)80106010

浙商证券研究所：http://research.stocke.com.cn