



CS 330 MIP – Lecture 04

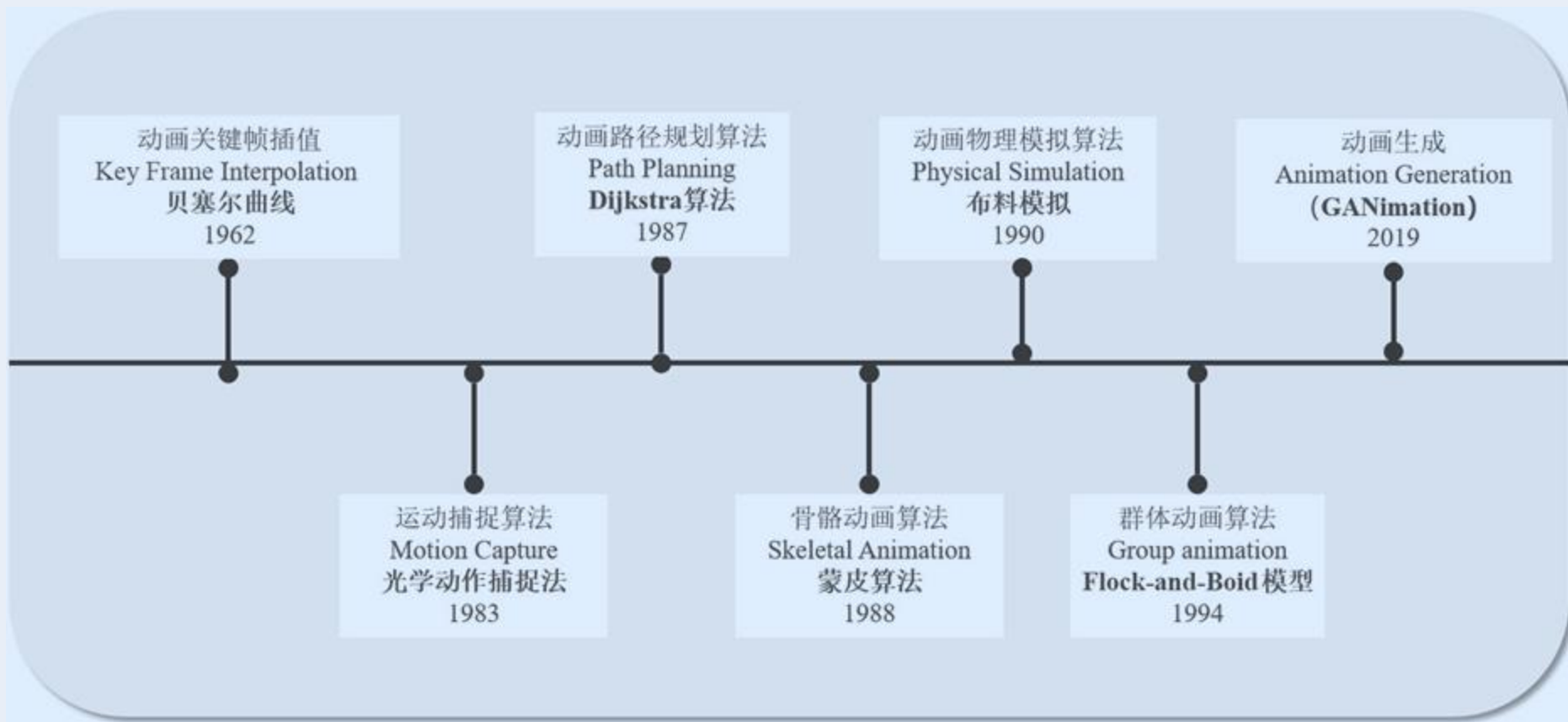
多媒体发展概论

Multimedia Development Milestones

Jimmy Liu 刘江

2025-03-12

4 动画信息智能计算发展里程碑



动画信息智能计算发展里程碑 1

第 1 个里程碑是经典的动画关键帧插值(Key Frame Interpolation)算法——贝塞尔曲线(Bézier Curve)。贝塞尔曲线最先由法国工程师皮埃尔·贝塞尔(Pierre Bézier)提出，其核心是基于控制点的线性组合来生成曲线。它通过提供平滑过渡、控制运动速度、精确运动路径等功能在动画关键帧插值中极大地丰富了动画的表现力和控制力。

动画信息智能计算发展里程碑 2

第 2 个里程碑是光学动作捕捉(Optical Motion Capture)算法，其基本概念可以追溯到 20 世纪 70 年代，但真正落地应用是 1983 年，汤姆·卡尔弗特(Tom Calvert)等人开发了早期的光学动作捕捉系统。光学动作捕捉算法的提出使得动画效果高度逼真，并且显著减少了制作成本，目前还广泛应用于虚拟现实、游戏开发和医学仿真等领域，成为计算机动画制作中不可或缺的重要工具。

动画信息智能计算发展里程碑 3

第 3 个里程碑是动画路径规划(Path Planning)中的 Dijkstra 算法。1987 年左右, Dijkstra 算法开始被用于角色动画和虚拟环境中角色的运动路径规划。Dijkstra 算法是一种用于计算加权图中单源最短路径的算法, 能够为角色在复杂的环境中找到最短或最优路径, 在不断演变和优化的人工智能动画信息计算中发挥着重要作用。

动画信息智能计算发展里程碑 4

第 4 个里程碑是动画蒙皮(Skinning)算法。1988 年, 艾伦·巴尔(Alan Barr)提出的局部 和全局变形方法成为动画蒙皮算法的基石。动画蒙皮算法是将骨骼动画的效果应用到角色或物体 表面的关键技术, 在骨骼动画中, 角色或物体的形状由一组顶点构成, 这些顶点通过蒙皮技术绑定到骨骼上, 当骨骼运动时, 顶点会根据骨骼的变换进行相应的移动和变形, 从而实现角色或物 体表面的动画效果。动画蒙皮算法的关键在于如何分配顶点的权重和计算骨骼的变换矩阵, 以确 保动画效果的准确性和真实性。

动画信息智能计算发展里程碑 5

第 5 个里程碑是动画物理模拟中的布料模拟(Cloth Simulation)算法。1990 年, Breen 等人 提出了使用基于粒子系统的方法来预测织物的褶皱和悬垂。这一方法成功模拟了布料粒子间的相 互作用, 显著提高了布料的真实性。布料模拟技术的核心是物理模型的构建和数值方法的实现。通 过不断地研究和优化, 布料模拟能够生成高度逼真的布料效果, 广泛应用于人工智能多媒体动画处理领域。

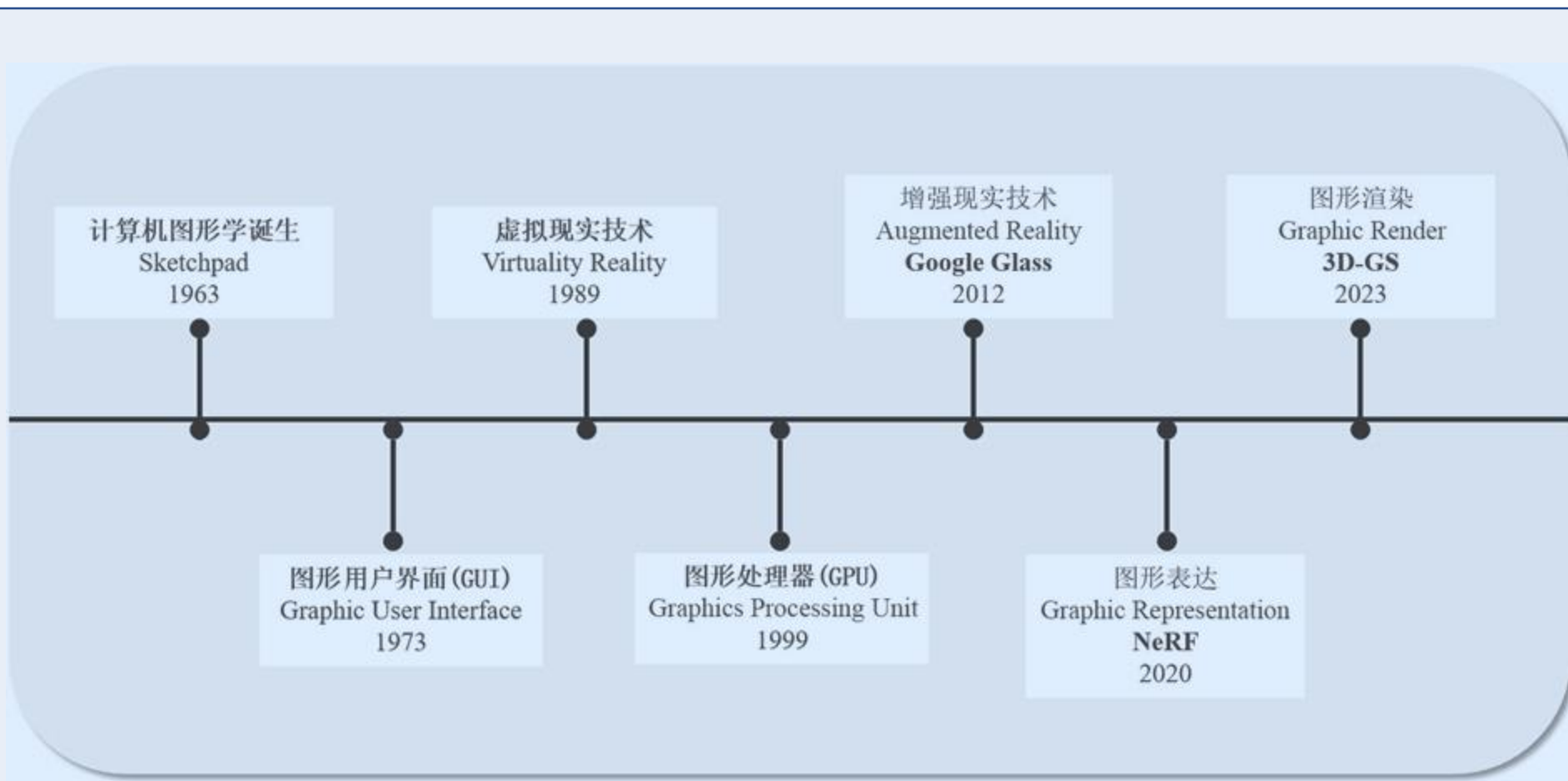
动画信息智能计算发展里程碑 6

第 6 个里程碑是 Flock-and-Boid 群体动画模型，该模型在主流动画电影中的首次重要应用是 1994 年的著名电影《狮子王》的制作。但其核心思想可追溯到 1987 年克雷德·雷诺德(Craig Reynold)开发的能够模拟鸟群、鱼群和其他群体动物行为的计算及模型。《狮子王》中角马奔跑 场景的制作标志这一技术的成功应用，此后该模型在计算机动画中得到了广泛应用，推动了动画 群体行为模拟技术的发展和应用。

动画信息智能计算发展里程碑 7

第 7 个里程碑是动画生成(Animation Generation)模型算法 GANimation, 由 Aliaksandr Siarohin 等人于 2019 年正式提出, 其核心是基于 GAN 的方法从单个静态图形中生成逼真的人类动画。GANimation 代表了动画生成技术的一种新方向, 推动了动画技术的发展和 innovation, 为更高效、更逼真的计算机智能动画生成方法打下了基础。

5 图形信息智能计算发展里程碑



图形信息智能计算发展里程碑 1

第 1 个里程碑是计算机图形学的诞生，伊万·萨瑟兰 (Ivan Sutherland) 于 1963 年完成了关于人机通信的图形系统的博士论文“Sketchpad”，这标志着图形信息处理领域的开拓。早期的计算机图形学主要关注二维图形的生成和显示，如简单的线条、形状和文本，随着计算能力的增强和算法的发展，计算机图形学的研究范围逐渐扩展到三维图形和真实感渲染。

图形信息智能计算发展里程碑 2

第 2 个里程碑是图形用户界面(Graphical User Interface, GUI)。GUI 最早于 1973 年出现, 它是用户与计算机等设备进行交互的图形化界面。与传统的命令行界面相比, GUI 通过直观的图形元素(如窗口、按钮、菜单等)极大地提高了用户的使用体验, 使计算机变得更加易用, 促进了计算机技术的普及以及人机交互领域的发展。

图形信息智能计算发展里程碑 3

第 3 个里程碑是虚拟现实(Virtual Reality, VR)技术。VR 这一术语是 VPL 公司的创始人杰 伦·拉尼尔(Jaron Lanier)和托马斯·齐默尔曼(Thomas Zimmerman)在 1989 年正式提出的, 本质是一种可以创建和体验虚拟世界的计算机技术, 它利用计算机生成的三维图形和声音, 通过 特殊的硬件设备(如 VR 头盔、手套等)为用户提供沉浸式的体验。虚拟现实技术最初主要用于 模拟和训练, 应用于游戏、娱乐、教育、医疗等领域。

图形信息智能计算发展里程碑 4

第 4 个里程碑是 GPU。英伟达(NVIDIA)公司在 1999 年发布 GeForce 256 图形处理芯片时 首先提出 GPU 的概念。GPU 是专门用于处理图形的计算机芯片，与传统的中央处理器(CPU)相比，GPU 在处理图形数据时具有更高的并行性和计算效率。GPU 的引入极大地提高了图形渲染的速度和质量，使复杂的图形效果和实时渲染成为可能。GPU 现在也被广泛应用于深度学习、高性能计算等领域。

图形信息智能计算发展里程碑 5

第 5 个里程碑是 Google Glass 增强现实(AR)技术，它是一款由谷歌 (Google)开发的智能眼镜，于 2013 年首次推出。Google Glass 采用 AR 技术，通过在眼镜上显示 信息来扩展用户视野和体验。它标志着 AR 技术在消费电子产品的应用，也为其他 AR 产品的开发奠定了基础，推动了人工智能多媒体图形信息计算的发展和应用。

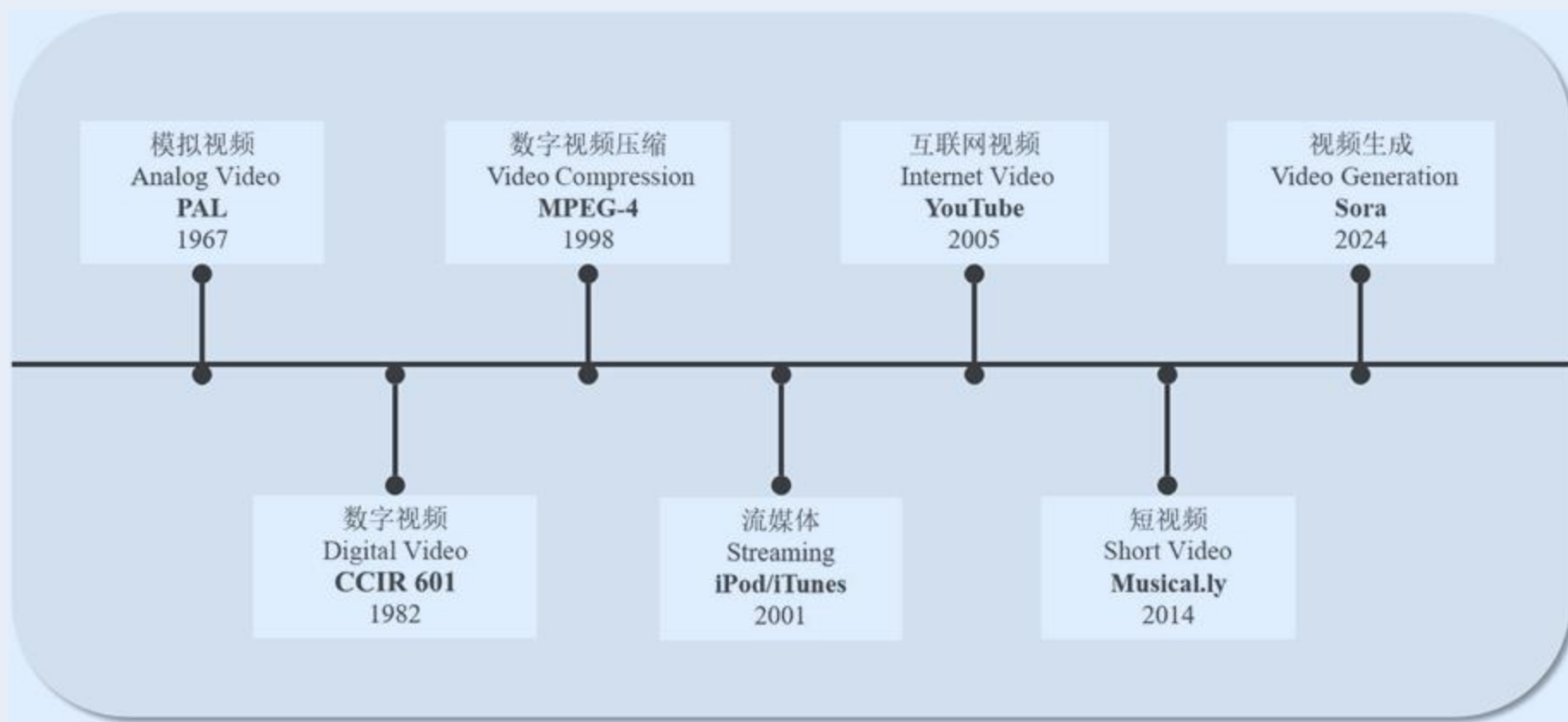
图形信息智能计算发展里程碑 6

第 6 个里程碑是 Neural Radiance Fields(NeRF)图形表达，由本·米尔登霍尔(Ben Mildenhall) 等人于 2020 年提出。它是一种用于生成逼真 3D 场景的神经网络模型，从有限的 2D 图像中恢复出完整的 3D 场景，并可以在新的视觉场景下渲染出极具真实感的图像。NeRF 图形表达具有直观、易理解的特点，能够更好地呈现数据的内在规律和关系，为人工智能多媒体图形信息计算带来了新的可能性。

图形信息智能计算发展里程碑 7

第 7 个里程碑是三维高斯泼溅(3D Gaussian Splatting, 3D-GS)图形渲染算法，由伯恩哈德·凯博(Bernhard Kerbl)等人在 2023 年提出。其利用多个各项异性的三维高斯分布球来表达静态的三维场景，并通过 Splatting 技术渲染出不同视角下的二维图像，是一种强大的点云图形渲染技术。它能通过对场景中的点云数据进行高效的渲染，从而实现逼真的效果。

6 视频信息智能计算发展里程碑



视频信息智能计算发展里程碑 1

第 1 个里程碑是逐行倒相(Phase Alternating Line, PAL)模拟视频, PAL 技术是德国人沃尔特·布鲁赫(Walter Bruch)在 1967 年提出的, 是通过电视广播实现的, 利用模拟信号将图像和声音信号传输到接收端, 再通过电视机进行解码和显示。PAL 模拟视频技术在人工智能多媒体视频信息计算中具有十分重要的作用, 特别是在视频传输、编码和处理等方面, 提高了视频处理的效率和质量。

视频信息智能计算发展里程碑 2

第 2 个里程碑是 CCIR 601 数字视频。国际电信联盟于 1982 年制定的 CCIR 601 数字视频技术逐渐取代模拟视频成为主流，数字视频技术将模拟信号转换为数字信号，通过计算机进行存储、编辑和传输。数字视频技术的出现，大大提高了视频处理的精度和效率，使视频编辑、特效制作和传输变得更加方便和高效。

视频信息智能计算发展里程碑 3

第 3 个里程碑是 MPEG-4 数字视频压缩，于 1998 年提出。为了解决数字视频存储和传输过程中的数据量庞大的问题，数字视频压缩技术应运而生，该技术去除了视频信号中的冗余信息，减少了数据量，同时保持了视频质量。而 MPEG-4 数字视频压缩技术的出现，使得高清、超高清等高质量视频内容的存储和传输成为可能，推动了视频产业的发展。

视频信息智能计算发展里程碑 4

第 4 个里程碑是流媒体。2001 年流媒体技术逐渐兴起，它允许用户在互联网上实时观看视频内容，无须等待整个文件下载完成。通过流媒体技术，用户可以在线观看电影、电视剧、直播等视频内容，为视频内容的传播提供了更加便捷的方式。

视频信息智能计算发展里程碑 5

第 5 个里程碑是 YouTube 互联网视频。YouTube 的首个视频由贾德·卡林姆(Jawed Karim) 在 2005 年上传，这标志着 YouTube 正式进入公众视野，并开始发展成为全球最大的视频分享平台之一。YouTube 的出现改变了互联网视频传播的方式，推动了用户生成内容的普及，使得互联网视频逐渐成为人们获取视频内容的主要渠道，改变了传统媒体的格局。

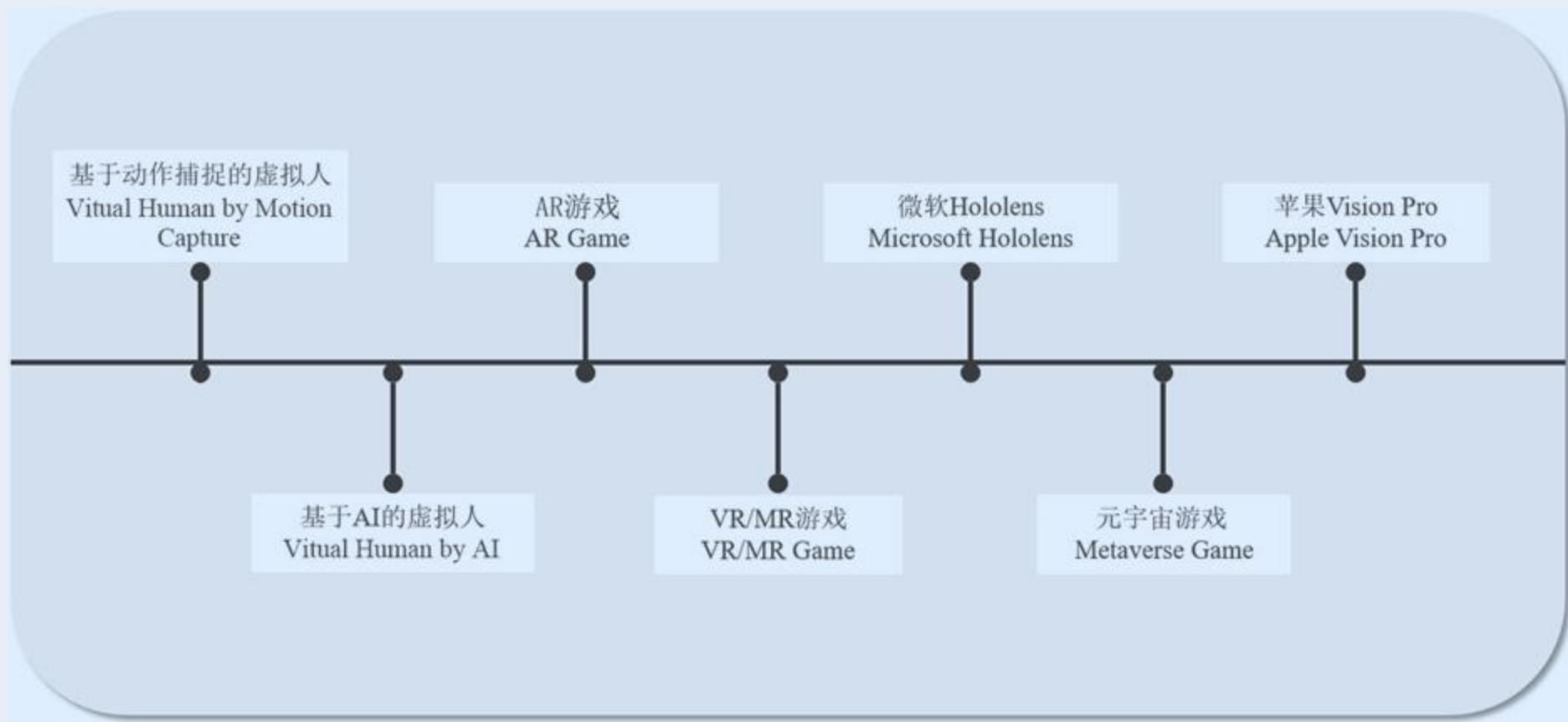
视频信息智能计算发展里程碑 6

第 6 个里程碑是 Musical.ly 短视频。2014 年 Musical.ly 在中国上海正式发布，它时间短、内容精炼、易于传播，逐渐成为人们记录生活、分享经验的重要方式。Musical.ly (阳陆育) 的创新和成功很大程度上塑造了现代短视频平台的特点和功能，为抖音、快手等其他短视频平台的兴起提供了基础，为短视频产业带来了新的机遇和挑战。

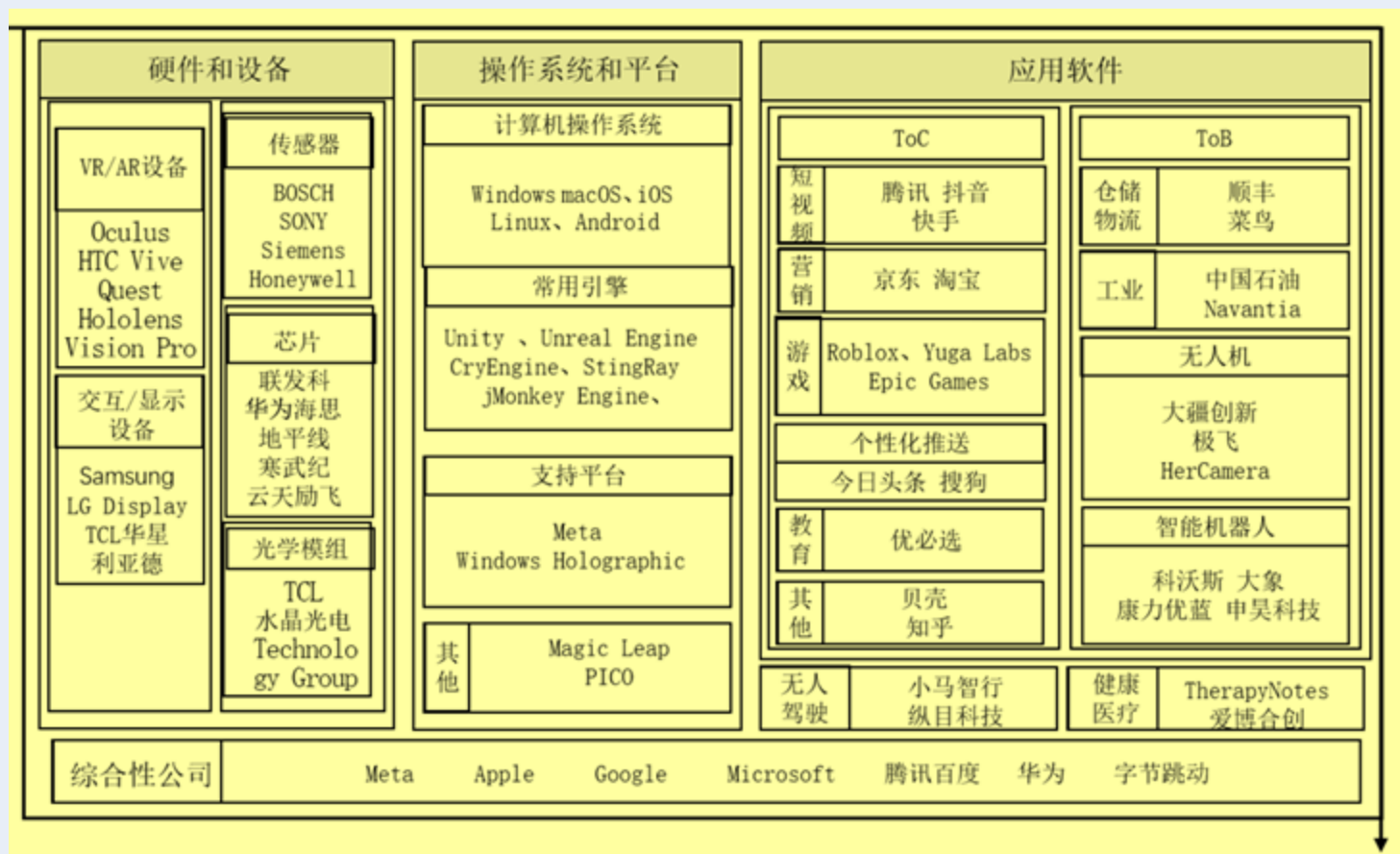
视频信息智能计算发展里程碑 7

第 7 个里程碑是 Sora 视频生成。2024 年发布的 Sora 是随着近年来生成式人工智能技术的发展而兴起的一种视频处理技术，它利用深度学习、多模态融合等算法，让计算机自动生成视频内容。视频生成技术的出现，为视频制作领域带来了革命性的变化，同时，也为用户提供了更加个性化、定制化的视频体验。

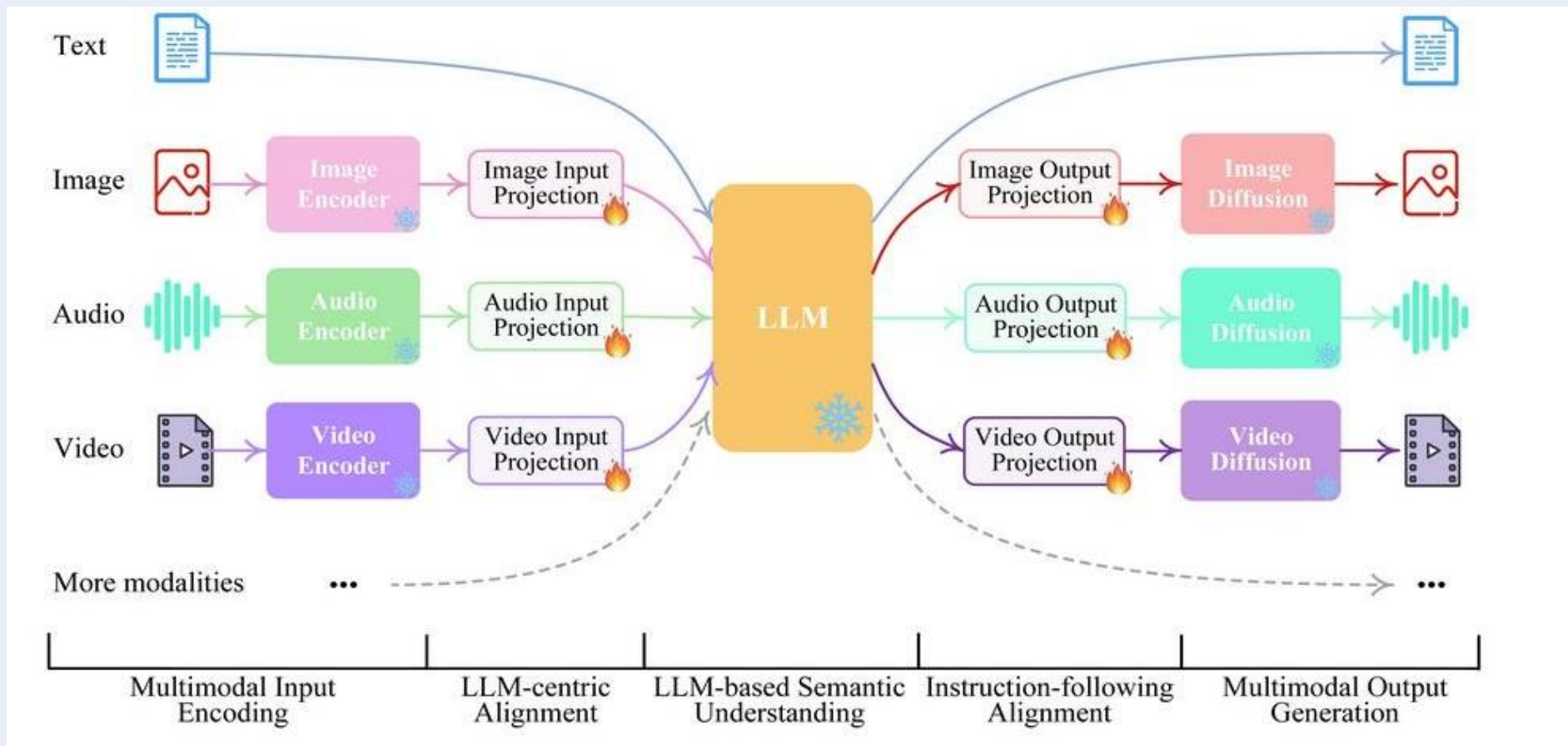
人工智能多媒体信息融合



人工智能多媒体计算的产业结构现状



Next- AI生成式多媒体



多技术交叉

高清晰度与沉浸式体验	未来的多媒体技术将追求更高清晰度和沉浸式体验，VR、AR和MR技术继续发展可以提升感觉效果，并在教育、医疗等领域发挥重要作用。新一代视频编解码标准也会更加提高压缩效率，为高速的信息传输提供可能。
人工智能与多媒体内容融合	人工智能技术将深刻改变多媒体内容的创作和处理方式，实现智能化生成、编辑和推荐，提高内容质量和用户体验。
多媒体与物联网融合	5G网络部署，提供高速低延迟的多媒体传输。物联网技术的普及将促进多媒体内容与智能设备的广泛连接，实现跨平台无缝传输和展示，拓展多媒体的应用场景。
交互性与社交性提升	未来的多媒体技术将更加注重用户交互体验和社交需求，通过引入交互技术和社交功能，使多媒体内容更加直观、易用、有趣，并促进信息共享和交流。同时多媒体内容的加密、水印和隐私保护技术的发展也为交互性提供了隐私保障。
云端化与边缘计算结合	云计算和边缘计算技术的发展将推动多媒体技术的云端和边缘端协同工作，提升处理速度、效率，降低存储和传输成本，为用户提供更加灵活和便捷的多媒体服务。

多媒体智能体

- 多媒体智能体Multimedia Agent是一种结合了人工智能和多媒体技术的智能体。
- 智能体通常被定义为一个能够自主行动并在特定环境中感知并响应其环境变化的软件实体。
- 多媒体智能体能够处理、分析和理解多种形式的媒体内容，包括文本、图像、音频、图形、动画和视频等。
- 它通过应用算法和模型，它可以提取媒体内容中的关键信息，进行内容分类、标注和编辑等操作。
- 借助自然语言处理、机器学习和深度学习等技术，多媒体智能体也能够对媒体内容进行深入的分析 and 理解，它能够识别图像中的对象、场景和情绪，解析音频中的语音内容和情感倾向，以及理解文本中的语义和上下文。

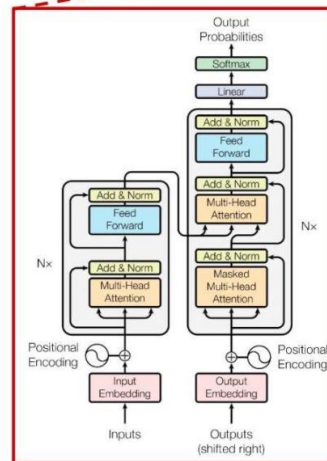
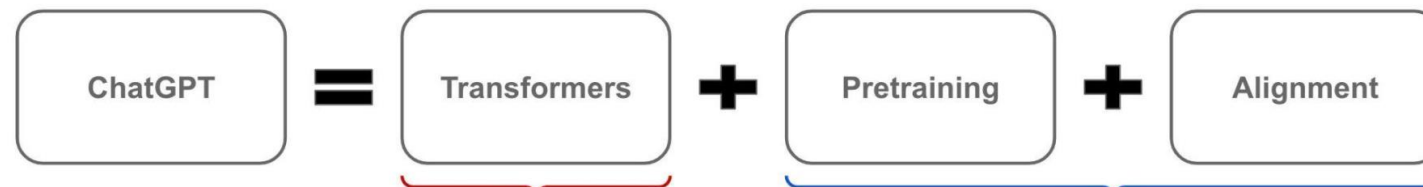
通用人工智能与多媒体智能体

OpenAI对实现通用人工智能五步规划

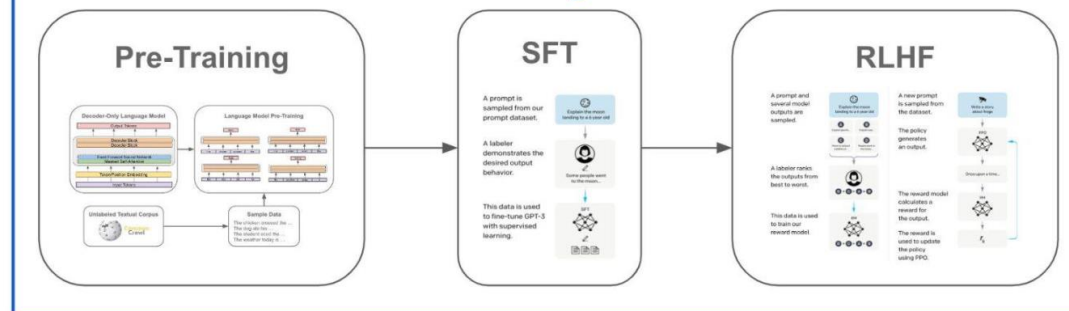


1. ChatGPT与Transformer 模型

LLMs(ChatGPT): Transformer+Pretraining+Alignment



LLM Training Pipeline



GPT: Generative Pretrained Transformer

2. DeepSeek LLM



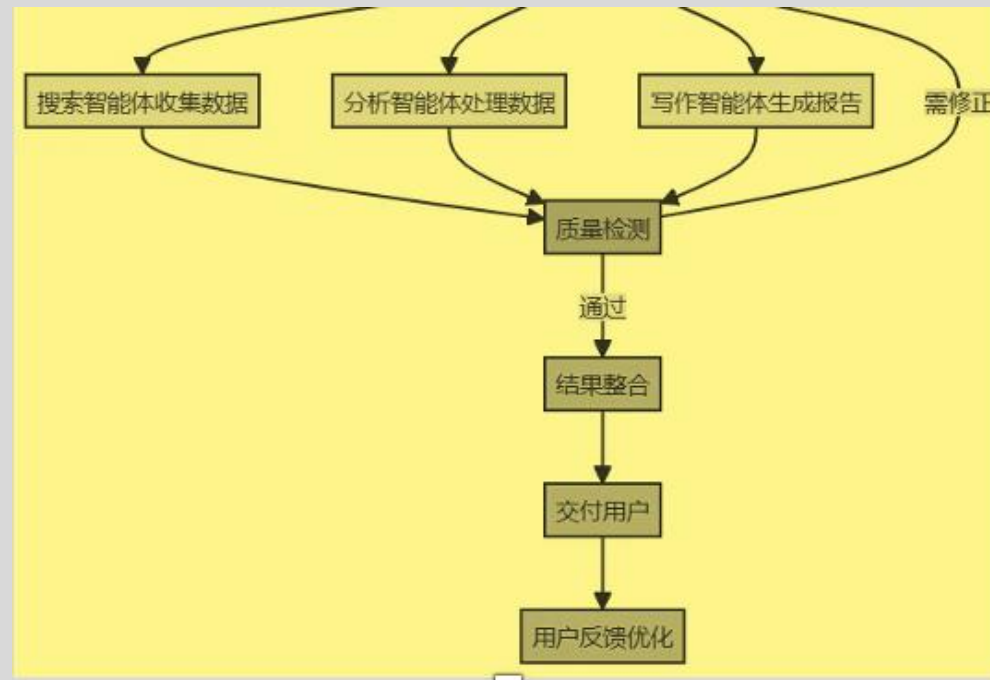
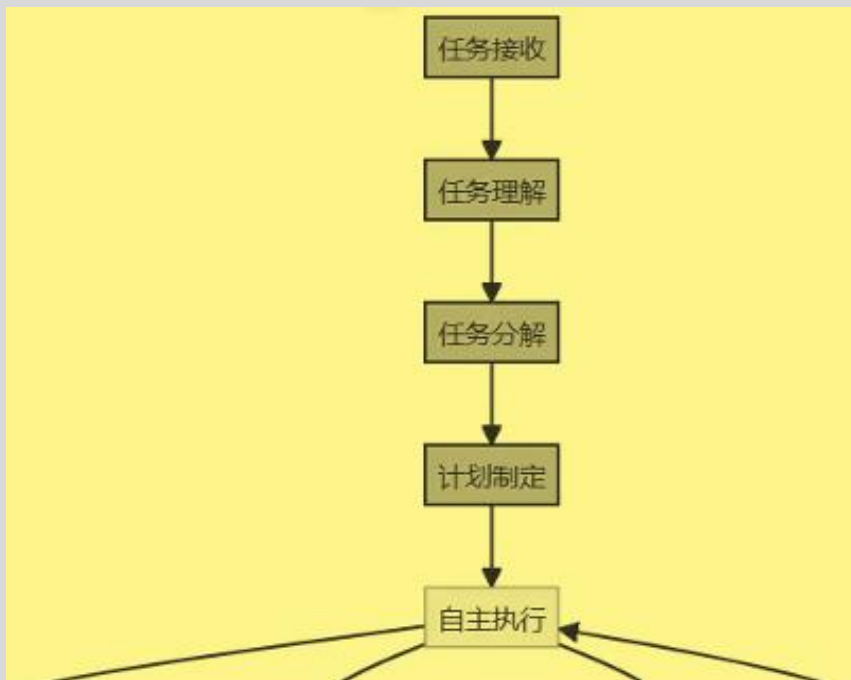
3. AI Agent

AI Agent（人工智能代理）通常指的是一个软件实体，它可以在某种程度上模拟人类智能的某些方面，执行特定任务或达成目标。这些代理可以感知环境，并在此基础上进行决策并采取行动，从而完成既定的任务或解决问题。

AI Agent workflows 的四种设计模式

- 反思 (Reflection) : LLM 检查自己的工作, 以提出改进方法。
- 工具使用 (Tool use) : LLM 拥有网络搜索、代码执行或任何其他功能来帮助其收集信息、采取行动或处理数据。
- 规划 (Planning) : LLM 提出并执行一个多步骤计划来实现目标 (例如, 撰写论文大纲、进行在线研究, 然后撰写草稿.....) 。
- 多智能体协作 (Multi-agent collaboration) : 多个 AI 智能代理一起工作, 分配任务并讨论和辩论想法, 以提出比单个智能体更好的解决方案。

AI Agent workflow



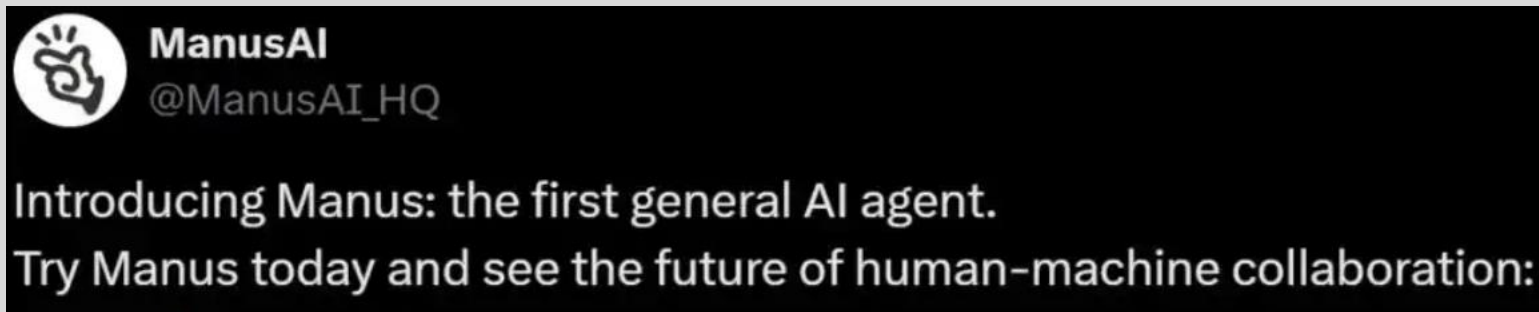
- OpenAI推出名为“Operator”的AI智能体，称可完成一些网购和订票等具体任务。Anthropic和谷歌等企业，也相继发布AI智能体或是用于开发代理式AI的模型套件，并预言智能体可能终会成为公司“员工”之一。
- 亚马逊（Amazon）2月26日宣布为迭代后的AI助理“Alexa”配备代理能力。SAP和ServiceNow等国际软件巨头，也相继投资研发AI智能体工具，并迅速推向市场。

LLM 与 AI Agent

- LLM 的核心，是一个深度思考的大语言模型，它通过逻辑缜密的思考链（COT: Chain of Thoughts），让思考变得更加深入。
- 代替人类作出决策并独立完成复杂任务的AI Agent的核心，就是通过一个个任务清单，详细地拆开任务，一步步执行这个执行链（COA: Chain of Actions）。
- LLM 是个思考助手，而AI Agent是个执行助手。
- 面对LLM，最重要的事情，可能是提问题。而面对AI Agent的时候，最重要的事情，可能是下任务。

通用AI智能体 MANUS

- 2025年3月5日晚间，中国初创公司Butterfly Effect（蝴蝶效应）发布AI Agent（智能体）Manus的早期预览版，将其称为“全球首款通用智能体产品”。在其他AI“生产想法”的时候，Manus可以“生产成果”。





CS 330 MIP – Lecture 04

多媒体发展概论

Multimedia Development Milestones

Jimmy Liu 刘江

2025-03-12