

《智元觉醒》：文生图与图生视频创作技术报告

项目名称：《智元觉醒》 作者：黄泽美，学号：20253889 日期：2025.11.8

摘要

本报告旨在阐述交互式人工智能短片《智元觉醒》的完整创作流程。项目以“科技”为核心主题，通过文生图与图生视频技术，构建了一个从概念诞生、城市崛起、人机融合到奇点降临的20秒叙事。报告将详细说明所使用的工具链、提示词的迭代与优化过程、以及在实现视觉连贯性与叙事一致性方面所遇到的挑战与解决方案。

使用工具链说明

在此次实验中，使用了deepseek，海豚实验室平台，即梦，Adobe Premiere Pro 以及 VScode：

- deepseek用来生成 文生图和图生视频 的提示词工作
- 海豚实验室平台用于文生图工作，~~因为全用即梦的话免费积分不够~~
- 即梦用于图生视频工作，质量很好
- Adobe Premiere Pro 用于剪辑合成总视频
- VScode 用于写markdown，进而转成PDF

提示词迭代过程

提示词的构建是本项目的核心。我的策略是：首先通过文生图确定每个模块的“关键帧”视觉，再通过图生视频提示词描述该关键帧中需要发生的运动。以【模块一：创世】为例：

文生图提示词的挑战：

初版提示词（文生图）：

a brain made of light in space (太空中的光构成的大脑)

问题：结果过于简单，缺乏科技感和细节，更像一个医学概念图。

迭代过程：逐步加入风格限定词和细节描述。

增加风格：cyberpunk, ethereal (赛博朋克，空灵)

增加构成元素：neural network, data particles, binary code (神经网络，数据粒子，二进制代码)

增加画质与渲染：cinematic lighting, 8K, Unreal Engine 5 (电影级光照，8K，虚幻引擎5渲染)

最终定稿提示词（文生图）：

A glowing, intricate neural network floating in the void of space, composed of luminous blue and white data particles. Complex mathematical formulas and binary code flow like stardust towards the center,

forming a bright core. Cyberpunk, ethereal, cosmic, dark background, hyper-detailed, 8K, cinematic lighting, Unreal Engine 5.

成果：成功生成了兼具神秘感与科技感的“创世”画面，为后续视频生成奠定了完美的视觉基础。

图生视频提示词的挑战：

目标：将静态的神经元网络变为“正在形成”的动态过程。

我尝试了不同的动作描述：

The neural network is forming (神经网络正在形成) - 动作过于模糊，效果不佳。

The particles are moving (粒子正在移动) - 运动随机，缺乏方向性。

最终方案：A slow zoom-in on the glowing neural network. The data particles pulse rhythmically with a soft light. The streams of binary code gently flow and converge towards the bright core...

策略：使用具体、可想象的动词（zoom-in, pulse, flow and converge）来明确指导AI的运动生成，并将运动强度(--s)设置为较低的30，以匹配“诞生”的缓慢节奏。

遇到的问题及解决方案

挑战一：视频片段间的视觉与叙事连贯性

问题描述：最初四个模块独立生成，导致风格、色调、甚至“宇宙规则”不统一。例如，模块一的粒子颜色与模块二的城市主色调差异巨大，破坏了整体感。

解决方案：

确立统一视觉语言：在项目开始前，预先定义了核心配色方案（霓虹蓝、科技紫、金属银）和视觉元素（粒子、线框、全息光）。在所有模块的文生图提示词中，都强制加入这些关键词。

“种子”图片参考法：在生成后续模块时，将前一个模块的成品图作为视觉参考，在Midjourney中使用/describe功能分析其风格，并融入新的提示词中，确保视觉基因的延续。

挑战二：图生视频运动控制的不确定性

问题描述：即梦对提示词的理解有时会产生偏差。例如，在模块四，希望实现“向内坍缩”，但AI可能生成“向外爆炸”的效果。

解决方案：

精确的动作描述：避免使用“变化”、“改变”等模糊词汇。使用物理学和电影语言中的精确动词，如implosion（内爆）、collapse into（坍缩进）、assemble from the ground up（从地面组装而起）。

迭代生成与筛选：对同一关键帧和提示词进行3-5次生成，从中挑选运动最符合预期的一版。

利用运动强度参数：深刻理解参数的影响。慢节奏、精细的动作使用低强度（25-40）；快节奏、宏大的动作使用高强度（50-70）。

技术选型的原因

- **deepseek, 海豚实验室平台**：整体上，学生创作足够用，对简单场景的分析力不错，易获得，上手难度低，免费，方便
 - **即梦**：比海豚实验室平台略微专业些，同样上手难度低，作品精度和画面渲染更强
 - **Adobe Premiere Pro**：为了实现对音画同步、色彩校正和转场效果的精细控制，专业桌面级非编软件是必不可少的。它确保了最终输出的最高质量。
 - **VScode**：强劲的轻量编译器，顺手，上手markdown的难度低
-

结论与未来展望

通过本次《智元觉醒》的创作，我对基于现有AIGC工具链制作高质量、强叙事性短片的可行性有了更深的感悟。**成功的核心在于**:前瞻性的整体视觉规划、精准且迭代优化的提示词工程、以及对不同工具特性的深刻理解与灵活运用。

未来，随着多模态模型（如Sora）的普及和控制技术的增强（如角色一致性、摄像机运动控制），我们可以挑战更复杂的叙事、更长的时间跨度和更生动的人物互动。AIGC正在极大地降低动态影像创作的门槛，并将成为未来影视、游戏和媒体行业中不可或缺的创意生产力工具。