

## 一、爬升知识金字塔：将数字资产转化为产业智能<sup>2</sup>

数字资产向产业智能的转化是一个系统化、分阶段的过程<sup>1</sup>，可分为三个阶段：数据转化为信息，信息转化为知识，最终知识转化为智能。

### 1、数据到信息<sup>4</sup>

在这一阶段，通过对数据的筛选、清理和预处理，将原始数据转化为“有用信息”。这些信息包括行业知识、流程知识以及隐性的专家经验等，为后续的知识提炼奠定基础。<sup>2</sup>

### 2、信息到知识<sup>1</sup>

信息转化为知识的过程包括以下几个关键步骤：<sup>9</sup>

- 提炼事实性知识：从信息中提取反映“是什么”的事实性知识，涵盖重点、规则以及其相互关系。<sup>5</sup>
- 推导程序性知识：基于事实性知识，推理生成“怎么用”的知识，如业务处理 workflows 和算法等。<sup>7</sup>
- 构建知识图谱：将事实性知识和程序性知识以知识图谱的形式组织起来，形成结构化的知识体系。<sup>6</sup>
- 生成长思维链数据：将事实性知识和程序性知识与具体任务结合，形成涵盖问题、推理过程（即如何使用知识逐步解决问题并得出答案）以及最终答案的长思维链数据。<sup>4</sup>

### 3、知识到智能<sup>3</sup>

通过使用长思维链知识训练模型，包括微调模型和模型蒸馏，最终实现智能化决策与推理。<sup>3</sup>

这一过程的核心科学问题是“如何借助大模型的能力，通过系统化方法爬升知识金字塔”。这一问题的解决对江西人工智能产业的发展至关重要。江西的特色在于丰富的产业数据，而如何将这些数字资产转化为产业智能，正是江西推动人工智能发展的关键所在。

目前，江西的产业数据大多处于“数据”阶段，部分已进入“信息”阶段，但尚未达到“长思维链知识”阶段。因此，依赖原始数据训练模型难以满足智能化需求。当前的主要挑战在于如何将数据转化为长思维链数据。

## 二、解决方案<sup>1</sup>

我们在“知识爬升”领域具备扎实的前期基础，能够有效应对上述挑战。

### 1、信息与知识的动态转化<sup>9</sup>

信息是静态的，而知识是动态的。同一组信息在不同需求下可提炼出不同知识。我们开发的 UniGraph 平台能够利用人工智能技术，根据用户需求对海量信息进行浅层学习、深层学习和迁移学习，自动提炼事实性知识、推导程序性知识，并高效构建知识图谱。

### 2、长思维链数据生成<sup>2</sup>

长思维链数据本质是一个“解题过程”，其输入为问题和

数据，输出为答案。由于数据量庞大，人工制作长思维链数据缺乏现实可行性。而直接依赖大模生成也存在“幻觉”问题，无法保障数据可靠性。<sup>3</sup>

我们开发的 Sapper 平台通过软件工程方法论，能够自动对输入的问题和数据进行需求分析，并针对分析结果进行系统设计、实现与测试，最终生成可靠的过程性解决方案，即真正合格的长思维链数据。<sup>2</sup>

### 3、小结<sup>1</sup>

基于这些前期技术积累，针对科学问题，我们提出解决方案：“通过应用软件工程方法论中的宏观设计原则、微观实现原则、数据结构以及相关工具，引导大模型实现知识金字塔爬升。”我们完全有能力将江西省的数字资产转化为产业智能，推动人工智能产业的进一步发展。<sup>1</sup>