FlowCabal

AI 半自动化写作辅助工具

田照涛

2026.04.14

贵州轻工职业大学

0.1 目录

0.2 上下文工程的兴起

- 2025 年 Karpathy 提出上下文工程概念
  - 工业级 LLM 应用的核心挑战：如何用恰当的信息填充上下文窗口
- 工业界验证:
  - Anthropic：上下文编辑减少 84% token 消耗
  - OpenAI Codex：25 小时运行消耗 1300 万 token
- AI 辅助写作面临类似挑战

0.3 长篇写作的核心问题

- 单次 LLM 调用受限于上下文窗口
- 如何在多次调用间维持一致性?
- 如何管理跨章节的记忆?
- 如何编排复杂的写作流程?

→ 需要系统性的上下文工程方案

0.4 FlowCabal 核心理念

工作流编排 + 记忆管理 + Agent 监控

三者结合，将长篇写作从高度依赖人类逐句干预的过程，变成半自动化的流水线

0.5 功能一：蓝图式工作流编排

将 LLM 调用封装为独立节点，节点间通过显式连线声明数据依赖

关键优势:

- DAG 拓扑序组织依赖关系
- 每个节点的查询仅包含其直接依赖节点的输出
- 从机制层面实现上下文的精确控制

0.6 工作流编排示意

角色辨析	→	引用下游两个节点
主要角色命运扩展		次要角色命运扩展

线性对话 → 节点 DAG: 上下文按需传递，避免冗余

0.7 功能二：内置记忆文件

拒绝 RAG 语义召回，采用纯文件系统的记忆架构

类比：代码库是记忆，grep 是检索

- `manuscripts/` 存放完整定稿章节
- `characters/`、`world/`、`voice.md` 是结构化摘要

三级渐进加载策略：

- L0：`index.md` 导航索引
- L1：各类记忆文件
- L2：完整原稿

0.8 功能二： 因果驱动检索

传统 RAG	FlowCabal
基于语义相似性召回	因果关系驱动
余弦相似度无法捕捉伏笔与回收	Agent 理解叙事上下文后自主决定加载什么
embedding 可能召回到无关段落	精确导航所需章节

核心洞察： 小说创作中， 因果链比语义相似性更可靠

Agent-inject 机制

节点的 prompt 中嵌入注入点（ hint ），执行时 Agent 自动从记忆系统中查询相关约束并注入

关键特性:

- 约束查询与上下文注入合二为一
- 增量构建模式：用户将待执行节点加入 Target Set
- 每个节点仅保留最近一次输出，不维护版本历史

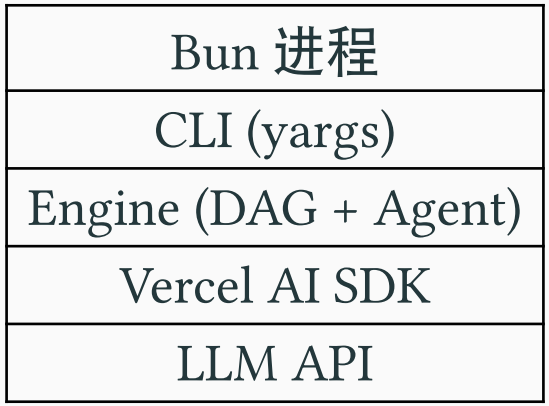
设计哲学：如果用户不重视生成的东西，那我们也认为它没有价值

0.10 技术栈概览

维度	选型
编程语言	TypeScript
运行时	Bun (`bun build --compile` 单二进制分发)
LLM 集成	Vercel AI SDK（统一 Provider 接口）
命令解析	yargs
数据校验	Zod Schema
存储方案	纯文件系统（JSON / Markdown）

零外部依赖 → 支持 Git 版本控制 → 降低部署门槛

0.11 架构分层



0.12 核心模块

模块	职责
types / schema	NodeDef、TextBlock、Workspace 等核心类型定义与 Zod 校验
llm	LLM Provider 构造、文本生成封装
workspace/core/node	Node 和 TextBlock 的 CRUD，自动维护 DAG
workspace/core/graph	Kahn 拓扑排序、Target Set / Stale Nodes 计算
workspace/core/runner	执行引擎：resolvePrompt、runSingle、runAll
agent	记忆加载、工具集、提示词模板

0.13 设计哲学：从完备到最小可用

早期设计追求全面性：

- 三角色 Agent
- 五种记忆侧写类型
- 向量数据库语义检索

转折：先实现能端到端跑通的最小内核，再在使用中发现真正需要的扩展点

每次简化都问：当前阶段的唯一用户是否真的需要这个抽象？

0.14 核心贡献

蓝图式工作流

DAG 拓扑序实现上下文精确控制

因果驱动记忆

三级渐进加载取代 RAG

Agent-inject

约束查询与上下文注入合一

FlowCabal 目前处于 Beta 阶段，核心功能均已可用

源代码: github.com/isirin1131/FlowCabal

0.15 未来方向

- **可视化编辑器**：节点拖拽和 DAG 连线体验
- **多 Agent 角色分拆**：以 Beta 验证后的真实需求为驱动
- **原生版本管理集成**：轻量级多版本存储
- **外部 Agent 深度集成**：Agent Gate 机制
- **工作流模板市场**：降低新用户上手门槛

始终让抽象服务于真实的创作需求，而非为复杂性而复杂性

FlowCabal

AI 半自动化写作辅助工具

田照涛

2026.04.14

贵州轻工职业大学