Sprint proposition

Agentic Content Production: Achieving Narrative Fidelity at Scale

Context - Evolution of AI Systems

First Wave: AI Assistants (2022-2023)

The first wave of LLM-powered applications was based on "AI Assistants" where users engaged in natural conversations with general-purpose systems. Users wrote prompts with directives to perform basic actions: 'expand text', 'summarize', 'translate'. LLMs (primarily GPT-3.5 and GPT-4) leveraged their general knowledge and semantic understanding to perform assigned tasks. Role prompting and few-shot learning were the most important techniques. Developers could control assistant behavior through system prompts.

Second Wave: Advanced Prompt Engineering (2023-2024)

Prompt engineering practices were extended with:

- **RAG (Retrieval Augmented Generation)** - integrating external knowledge bases
- **Function calling** - enabling LLMs to interact with external tools and APIs
- **Chain of Thought (CoT)** - step-by-step reasoning processes
- **ReAct** - combining reasoning and acting in iterative loops
- **Query augmentation** - enhancing user inputs for better results

LLM applications integrated "chains" - sequences of LLM calls that enabled complex workflows. This approach drove the success of frameworks like LangChain. AI Assistants incorporated "personalization" through specialized GPTs, where conversations were dedicated to specific business activities. Assistants and users achieved objectives by iterating through multiple steps defined in Chain of Thought (or Tree of Thought) processes. ReAct frameworks triggered function calling based on specific conversation states.

Third Wave: The Agentic AI Split (2024-2025)

The rise of agentic AI and widespread adoption of Chain of Thought reasoning created two major architectural approaches:

Thinking/Reasoning Models

- **Large Reasoning Models (LRMs)** like OpenAI's o-series (o1, o3-mini) and DeepSeek-R1
- These models perform more computation (higher token consumption) at inference time

- Represents a shift from traditional approaches of more compute at training time (requiring more data) or post-training with RLHF and Supervised Fine-Tuning (using techniques like DPO)

Traditional Large Models

- **Very large models** like GPT-4.5 with enhanced knowledge and semantic understanding
- Designed for professional queries and creative intelligence tasks
- Focus on broader knowledge coverage and improved human communication

Specialized Agentic Systems

- **Development and coding tools** have reached significant maturity with solutions like:
  - Anthropic's Claude Code
  - Cursor AI (Agent mode)
  - Windsurf
  - GitHub Copilot
  - OpenAI's ChatGPT Canvas
- These combine thinking models (o1, Claude Sonnet 4) as planners with specialized agents for dedicated tasks

Content Creation - The Next Frontier

Agentic AI systems for textual and multimedia content creation remain in their early stages. Notable examples include:

- Google's NotebookLM podcast generation features
- Emerging video script generation tools
- Early-stage documentary and educational content systems

Project Goals

This project aims to create a specialized Agentic System for professional content creation, targeting:

- **YouTube video and YouTube Shorts script writing**
- **Training course and educational content development**
- **Podcast-style conversational content (NotebookLM approach)**
- **Documentary series and investigative content**

We focus on "Produced" content that follows an **"Editorial Production Pipeline"** - where content creation follows defined processes while preserving the "artistic" vision of hosts, creators, and producers.

"We built an agentic workflow to industrialize our creative process."

The project's goal is to create an **"Agentic Story Pipeline"**, **"Agentic Content Workflow"**, or **"StoryOps"** system.

Common applications: Television, podcasts, YouTube channels, digital publishing

Key workflow stages:

1. **Agenda/Syllabus** - Content planning and structure

2. **Research** - Information gathering and fact-checking
3. **Script Development** - Narrative construction and dialogue
4. **Review & Iteration** - Quality control and refinement
5. **Finalization** - Production-ready content preparation
6. **Delivery** - Format-specific output generation

Primary use cases: Creative yet repeatable formats including video essays, interview shows, product demonstrations, educational series, and documentary content.

The Challenge: Narrative Fidelity at Scale

The central challenge in agentic content production is maintaining **narrative fidelity** - ensuring that automated systems can preserve storytelling quality, factual accuracy, and creative vision while operating at industrial scale. This requires balancing the efficiency of automated workflows with the nuanced decision-making that defines compelling content.

Main Challenge - "Creative-Consistent, Cost-Conscious Content Automation" - "Narrative Fidelity at Scale"

Agentic Content Production: Optimizing for Creative Alignment, Narrative Integrity, and Cost Efficiency through Reuse & Feedback Loops

Here's a concise and structured summary of the **core challenge** we're tackling—refined into a statement that can guide your agentic workflow design or be used as documentation for your AI agents and collaborators:

🎯 Core Challenge of Agentic Content Production

To **automate the creation of high-quality video scripts** (or content episodes) that **consistently reflect the creator's vision**, tone, structure, and strategic intent across episodes—while **preserving creativity**, **engagement**, and **narrative coherence** from hook to call to action.

The goal is to create **high-quality, creator-aligned content** (script, video, episode…) while maintaining **cost-efficiency and production velocity**.

The system must **learn from feedback**, **self-improve over time**, and **re-use prior content patterns** to reduce the need for redundant prompting, rewriting, or manual review.

🔍 Key Dimensions of the Challenge

1. **Creative Alignment & Intent**
   - Maintain fidelity to the **creator's voice**, values, and directives.

- Respect creative **guidelines** (e.g., not too didactic, not too dramatic, include humor subtly).
- Follow the **format type** (e.g., educational vs editorial, short vs long-form).

2. **Research & Narrative Integrity**
   - Ensure all facts and examples are **aligned with the creator's knowledge base or worldview**.
   - Avoid introducing content that breaks the vision or misrepresents the message.

3. **Tone, Style, and Engagement**
   - Adapt tone and engagement level to the **format** (e.g., YouTube video ≠ training course).
   - Balance **clarity**, **energy**, and **authenticity** to fit audience expectations.

4. **Script Structure & Flow**
   - Ensure proper **chapter junctions**, transitions, and avoidance of **repetition**.
   - Preserve **narrative momentum** from **hook** to **final message**.

5. **Embedded Messaging**
   - Seamlessly integrate **native CTAs**, **sponsor messages**, or **product pitches** without disrupting the flow.
   - Maintain **natural voice**, avoid sounding "forced" or "scripted".

6. **Episode-to-Episode Coherence**
   - Maintain consistency of structure, tone, and editorial rhythm across episodes or series.
   - Reuse patterns **intentionally** (e.g., signature hooks, formats) while **avoiding redundancy**.

7. **Review & Control**
   - Ensure **checkpoints** (pre-publish or pre-recording) to validate alignment with all above.
   - Include loops for **creator review**, **style validation**, and **fact accuracy**.

Condensed Workflow for a Scriptwriting Process

Here's a condensed workflow for your virtual team of AI agents, assigning each role and incorporating the additional loops you described for research, fact-checking, and finalization:

1. Idea Generation & Agenda Creation

- **Show Host (Visionary Agent)**:
  - Proposes the episode's theme and global guidelines.
  - Defines the overarching goal of the episode.
- **Segment Producer (Planner Agent)**:
  - Structures the episode into an agenda with key segments and topics.
  - Drafts the initial outline for review.

2. Agenda Review & Feedback

- **Show Host**:

- Reviews and refines the agenda to align with their vision and tone.
- Approves the structure for further development.
- **Creative Director (Style Agent)**:
  - Ensures the agenda aligns with the overall style and branding of the series.

3. Research & Fact-Gathering

- **Research Assistant (Research Agent)**:
  - Provides detailed information, facts, and context for each segment of the agenda.
  - Suggests relevant examples or anecdotes to make the script more engaging.
- **Fact-Checker (Accuracy Agent)**:
  - Verifies the accuracy and credibility of the information provided.

4. Drafting the Script

- **Scriptwriter (Writer Agent)**:
  - Develops the first draft of the script based on the agenda and research input.
  - Collaborates with the Research Agent to incorporate factual and lively details.
  - Focuses on structure, narrative flow, and initial dialogue.

5. Script Refinement

- **Creative Director ↔ Script Director (Tone Agent)**:
  - Ensures the script matches the style, tone, and goals of the episode.
  - Provides iterative feedback for alignment with the series' branding and narrative style.
  - Loops with the Writer Agent to refine content.

6. Final Review

- **Show Host**:
  - Reviews the final script to ensure it reflects their vision and delivery style.
  - Suggests any last-minute tweaks for authenticity and engagement.
- **Creative Director**:
  - Gives final approval, ensuring cohesion and consistency across episodes.

7. Production Preparation

- **Producer (Production Agent)**:
  - Coordinates the finalized script for production, aligning the technical team with the content.
  - Prepares any assets or additional materials required.

Feedback Loops in the Process

1. **Research ↔ Scriptwriting Loop**:
   - During the drafting phase, the **Writer Agent** continuously collaborates with the **Research Agent** for additional details, examples, and fact-checking.
2. **Scriptwriting ↔ Creative Director Loop**:
   - Ensures alignment of style, tone, and narrative goals during script refinement.

This structure should work well for an AI-driven team, where agents are assigned clear roles and interconnected feedback loops ensure a high-quality final script.

Next Sprint Planning

We are working on a "Training Course" content development but moving towards a more general purpose "content creation" agentic solution.

We have modularize our project in 3 mains module

- Knowledge Manager
    - Curate the content that can be used during the content creation phase
    - Give easy access to curated knowledge to editing module
- Content creation
    - Research Team
        - Retrieve accurate information and facts aligned with "Creator" direction and align with the targeted audience. Must have full coverage (exhaustive).
        - Access to Knowledge Manager and refine the research note that is align with "host" or "producer" directive.
    - Content Editing Team
        - Ensure tone and style that work "chapter" per "chapter"
    - Authoring Alignment
        - Ensure global coherency of the full content.
        - Check alignement
- Feedback, Evaluation and Optimization
    - Measure the quality of the content and

Evolution in next sprints (to priorize) - We are at Epic Level.
- Knowledge Manager
    - MCP access to content based on precomputed metadata
    - (Optional) Retrieval search of content
- Content Creation
    - Add a planner agent and move away from a static execution graph to a dynamic execution
        - Pre existing Collection of specialized Agent define in a yaml <-> with Prompt for specific task
        - Prompt are optimized, query search can be optimized and given to the agents
        - Predefine execution graph in Yaml
    - Research Team retrieve fact from KM and load it into Response API to synthetise Research Note

- Editing team
  - Writing style can be fed with "Few Shot" Example for style (not existing yet).
- Authoring Aligment Agent
  - Final check of document authoring and alignement with project directive
- Authoring Alignement
  - Reviewer agent and authoring_alignment agent feedback are tracked on conversation
  - human-in-the-loop could be setup before callong the chapter validation or document validation to retrive user feedback.
  - User feedback can be memorized (short term or long term) in order to reuse.
  - Metrics are produce on a run to measure alignement to goals (tone, style, structure, content) based on syllabus but also "creative director" alignement.
  - Optimization of prompt can be done at agent level in passive mode then active mode.

Key Dimensions of the Next Sprint
1. Creative Alignment
- Adherence to the **creator's vision**, tone, and strategic direction.
- Adaptability across content formats (talking-head video ≠ podcast ≠ training).

2. Narrative Coherence
- Structural integrity: chapters flow, no redundancies, smooth transitions.
- Hook, CTA, embedded messages must be natively and meaningfully integrated.

3. Content Reuse & Inspiration
- Use **few-shot examples** (prior hooks, CTAs, scripts) as **creative priors**.
- Enable **style transfer** and **reuse of proven formats** to speed up iteration.

4. Feedback Integration & Self-Improvement
- Host/creator feedback must:
  - a) Correct/improve the current output.
  - b) **Update the agent's directives** for future episodes (memory, prompt evolution).
- Feedback should trigger a **fine-tuned adaptation**, not just per-output patching.

5. Efficiency (Cost, Time, Cognitive Load)
- **Reduce time to quality**: reach the aligned script in **fewer iterations**.
- **Lower cost per episode** by:
  - Reusing high-quality past content,
  - Minimizing expensive model calls or review loops,
  - Automating structure/consistency checks.