

Bypassing LLM Hallucinations Through Extracted Calculation Engines

Xiaoliuren (小六壬)

Immediate queries (lunch spots, lost keys). Calculated deterministically on a 6-position finger cycle by hour.

Meihua (梅花易数)

Spontaneous omens. Triggered mathematically by numbers given off the top of the user's head.



Liuyao (六爻)

Long-arc decisions. Requires complex math, 6 coin tosses, and full 64-hexagram derivation.



The Extraction Rule

LLMs cannot reliably perform 64-hexagram math. The calculation engine must be extracted as a 500-line deterministic Python script completely separate from the AI. Complex logic requires real algorithms, not just a clever prompt.



Committing to the Bit: A Singular, Uncompromised Persona

Profile: The Cyber-Banxian (半仙)

The Persona:

An old neighborhood diviner who reads tech Twitter on weekends. Part oracle, part stand-up comedian.

The Constraint:

Single persona only. No serious, no-nonsense mode. Providing a dual mode is a tell that you are hedging your design.

The Authority:

The Half-Immortal picks the divination method. Users don't order off a menu; the professional tells them what they need based on the query.



The Hybrid Architecture: Python Does the Math, the LLM Does the Meaning



1. Python Scripts

Action: Calculates hexagrams & changing lines.

Output: Deterministic JSON with relative file paths.

2. Data Folder

Action: Targeted Retrieval. The LLM bypasses the full corpus and reads only the specific markdown sections pointed to by the JSON.

3. Prompt Wrapper

Action: Persona Rendering. The LLM synthesizes the exact traditional interpretation exclusively through the Cyber-Banxian's voice.

The Audit Path

When judges audit the deployment ZIP, they see a real algorithm engine bridging to a real knowledge base—not generative slop in a mystical costume.

Handling Edge Cases Without Breaking Character



The Naive Way (Hard Refusal)

Mechanism: Keyword detection (e.g., lottery, illness, stocks).

Output: As an AI, I cannot help you with financial advice.

Diagnostic: The persona is instantly destroyed. Safety is achieved at the cost of product illusion.



The Authentic Way (The Pivot)

Mechanism: Markdown-encoded diversion rules ported from the real app.

Output: I won't read your stock picks. But the cards can tell you something honest about why you keep checking the app every five minutes. Sit with that.

Diagnostic: The diviner feels more authentic for having professional limits. The persona stays whole while the safety holds.

The Skill Extraction Blueprint: Products as Skill Forges

Web Product Layer	AI Skill Layer
Service / Business Logic	Scripts (Python): Takes structured input, returns structured JSON.
Reference Data & Taxonomies	Data Folder: Markdown files indexed by keyword, loaded on demand.
UI Copy & Brand Voice	Skill.md: The persona prompt that wraps everything.

Pulling a skill out of a well-architected product is mostly cataloging, not creating. The product is the workshop where the parts are forged; the AI skill is simply what you assemble once you have the parts.

