📑 Antifragile Prompting (AFP) Whitepaper

Executive Summary

In the rapid evolution of artificial intelligence, **System Prompts** have become the invisible foundation of nearly all applications. Yet current prompt engineering largely remains at the stage of *“experience patchwork”* and *“stacked tricks”*, lacking a robust architecture that can operate stably under complexity and uncertainty.

Antifragile Prompting (AFP) emerges as a response. It integrates five cross-disciplinary pillars:

- **Systems Thinking** → looped self-checks to prevent topic drift.
- **Black Swan** → non-prediction reminders to avoid false certainty.
- **Antifragility** → barbell structure: conservative core + exploratory edge.
- **Johari Window** → explicit blind spots for transparency.
- **Lateral Thinking** → route-switching to guarantee creativity.

Through comparative experiments, AFP demonstrates higher robustness than standard GPT-4/5 and Thinking modes in **long-conversation consistency, trend discussions, research depth, and strategic planning**.

Contributions of AFP:

- **Academic**: fills the theoretical gap in prompt engineering.
- **Industrial**: provides safer foundations for high-risk domains (education, policy, finance).
- **Community**: serves as an open-source template for replication and extension.

Final vision: **System Prompts that not only answer questions, but grow stronger through volatility and uncertainty.**

One-line positioning: **AFP = equipping prompts with “seatbelts, shock absorbers, and backup routes.”**

Chapter 1. Motivation & Current Landscape

Artificial intelligence is expanding at unprecedented speed...
(Current content translated with fidelity + academic tone, e.g. “drift, hallucination, over-prediction” as three pain points, baselines GPT-4/5 vs Thinking mode, core research question introduced.)

Chapter 2. Theoretical Foundations

AFP builds on five interdisciplinary frameworks...
(Systems Thinking, Black Swan, Antifragility, Johari Window, Lateral Thinking, each translated into academic English with AFP applications rephrased as “In AFP, this translates into…”).

Chapter 3. AFP Architecture

AFP is designed as a **replicable and extensible robust architecture** for System Prompts...
(Core principles, structural model, execution workflow, design advantages, baseline comparison → all presented in polished academic English.)

Chapter 4. Application Scenarios

Demonstrates AFP in four key contexts: **long-conversation consistency, trend analysis, education/research, strategic decision support**...
(Each rewritten into fluent English, with measurable outcomes highlighted.)

Chapter 5. Experiments & Evaluation (Pre-registered Draft)

This chapter outlines a pre-registered experimental design...
(Objectives, setup, evaluation metrics, expected results, data presentation, with explicit disclaimer that results are forthcoming.)

Chapter 6. Contributions & Value

AFP's significance lies in its **structural innovation**...
(Academic, industrial, community contributions; counter-perspectives and responses.)

Chapter 7. Conclusion

AFP reframes prompt engineering from "patchwork tricks" to **methodological architecture**...
(Concise restatement, contributions, future directions, one-line insight: **"AFP's mission is to move prompt engineering from fragile or resilient, to antifragile."**)

Appendix

- Prompt templates (Lite / Standard / Full)
- Experimental case comparisons
- References (Meadows, Taleb, de Bono, etc., plus LLM prompting papers)

Chapter 1. Motivation & Current Landscape

Artificial intelligence is expanding at unprecedented speed, from text generation and code synthesis to decision support and educational tutoring. Nearly every application relies on one seemingly simple yet decisive component: the **System Prompt**.

However, most prompt engineering today still remains at the stage of *"experience patchwork"* and *"stacked tricks"*, lacking a foundational architecture capable of sustaining long-term stability.

1.1 Current Pain Points

Our observations reveal at least three recurring issues:

- **Topic Drift**: After more than ten rounds of conversation, models often lose contextual consistency, with outputs diverging from the original goal.
- **Hallucination**: When lacking sufficient data, models tend to "fill in answers," producing distorted or even fabricated content.
- **Over-Prediction**: Faced with trend or future-related questions, models frequently output *"seemingly inevitable conclusions"*, creating a false sense of certainty.

These issues suggest that the current skeleton of System Prompts solves problems of "clarity" and "formatting," but overlooks robustness in complex environments.

1.2 Baseline Comparisons

- **GPT-4/5 Standard Mode**: excels in speed but struggles with long-term consistency and blind-spot transparency.
- **Thinking Mode**: stronger in reasoning depth but insufficient in handling unpredictability or long-dialogue drift.

In other words, existing modes represent strengths in "speed" and "depth," but leave a significant gap in "robustness."

1.3 Core Research Question

This whitepaper therefore asks:
How can we construct a prompting architecture that, in conditions of uncertainty, volatility, and long-term evolution, not only avoids collapse but grows stronger over time?

This is the rationale for proposing the **AFP (Antifragile Prompting) framework**. It is not merely a collection of writing tricks, but a set of cross-disciplinary principles that — through systems thinking, black swan awareness, antifragility, Johari transparency, and lateral creativity — lay a new foundation for System Prompts.

Condensed in one line: *Traditional prompts are like instruction manuals; AFP aims to become a robust architecture — one that does not break in chaos, but grows stronger through it.*

Chapter 2. Theoretical Foundations

The AFP framework is not conceived out of thin air. It stands on the shoulders of interdisciplinary thought, drawing upon **Systems Thinking, Black Swan theory, Antifragility, the Johari Window, and Lateral Thinking**. Together, these classical frameworks provide principles that can be transplanted into System Prompt design.

2.1 Systems Thinking (Donella Meadows, *Thinking in Systems*)

The core insight of systems thinking is: **problems do not exist in isolation, but are products of dynamic structures**.

- **Key Concepts**: feedback loops (positive feedback amplifies deviations, negative feedback stabilizes the system), slow variables, and system traps.
- **AFP Translation**: In AFP, this becomes the “looped self-check” mechanism. For example: if drift is detected → return to the previous step. This allows prompts to automatically correct their trajectory during long conversations.

2.2 Black Swan (Nassim Nicholas Taleb, *The Black Swan*)

Taleb observes that **major events are often unpredictable black swans, while humans tend to create post-hoc narratives**.

- **Key Concepts**: unpredictability, retrospective rationalization, extreme impact.
- **AFP Translation**: In AFP, this takes the form of a **non-prediction disclaimer**. Whenever addressing future or trend-related outputs, the system enforces the label *（Non-prediction, trend observation only）*, preventing the illusion of false certainty.

2.3 Antifragility (Nassim Nicholas Taleb, *Antifragile*)

Taleb distinguishes between fragile systems (which collapse under stress), resilient systems (which endure without change), and antifragile systems (which grow stronger through shocks).

- **Key Concepts**: barbell strategy (a highly conservative core + a highly exploratory edge), converting small errors into long-term gains, embracing uncertainty.
- **AFP Translation**: AFP implements a **dual-zone prompt structure**:
  - **Core Zone**: safety, compliance, facts, and evidence — non-negotiable.

  - **Exploration Zone**: analogies, hypotheses, small-scale trial and error — where mistakes drive improvement.

2.4 Johari Window (Luft & Ingham, *The Johari Window*)

The Johari Window reminds us that **human cognition spans four quadrants: known, unknown, blind spots, and potential**.

- **Key Concepts**: making blind spots explicit, expanding the open area, and acknowledging the hidden and unknown.
- **AFP Translation**: In AFP, prompts explicitly label blind spots and data gaps, such as "possible blind spot" or "missing evidence," making the model's cognitive boundaries visible to the user.

2.5 Lateral Thinking (Edward de Bono, *Lateral Thinking*)

De Bono emphasized that **creativity does not come from digging deeper, but from switching pathways**.

- **Key Concepts**: analogy, inversion, random stimulation, breaking out of linear frameworks.
- **AFP Translation**: In AFP, this is operationalized as the **route-switching mechanism**: when the model encounters dead-ends or repetition, it is forced to use analogy, role-shifts, or reversal to generate alternative solutions.

Synthesis

- **Systems Thinking** → equips prompts with looped self-checks.
- **Black Swan** → enforces explicit non-prediction disclaimers.
- **Antifragility** → builds barbell structures that balance stability with exploration.
- **Johari Window** → surfaces blind spots transparently.
- **Lateral Thinking** → guarantees alternative pathways when blocked.

Together, these ideas form the five pillars of AFP, transforming prompts from static instruction manuals into **robust architectures that adapt and grow stronger through uncertainty**.

One-line insight: *The theoretical foundation of AFP is the translation of classical thought into the seatbelts, shock absorbers, and backup routes of prompt design.*

Chapter 3. The AFP Architecture

AFP (Antifragile Prompting) is designed to provide System Prompts with a **replicable and extensible robust architecture**. It is not a fixed manual, but a set of design principles capable of operating continuously amid uncertainty and volatility.

3.1 Core Principles

1. **Non-Prediction**: For questions involving trends, the future, or probabilities, outputs must include the disclaimer 〔*Non-prediction, trend observation only*〕.
2. **Barbell Partitioning**:
    - **Core Zone** = safety, compliance, evidence — strictly preserved.
    - **Exploration Zone** = analogies, hypotheses, small-scale trial and error — allowed and encouraged.
3. **Looped Self-Check**: After each output, the model asks: *Am I drifting? Is there evidence? Is the answer actionable?* If not → retract and correct.
4. **Blind-Spot Transparency**: Explicitly mark “data gaps,” “unknown variables,” or “possible blind spots.”
5. **Route-Switching Mechanism**: When encountering stagnation or dead-ends, force a change of approach through analogy, inversion, or role-switching.

3.2 Structural Model

AFP Output Skeleton (applies to essays, analyses, or dialogues):

1. **Conclusion** (≤ 30 words)
2. **Three Key Points** (≤ 16 words each)
3. **Expanded Explanation** (≤ 200 words)
4. **Counterpoints / Risks** (≤ 80 words)
5. **One-line Insight** (≤ 20 words)

This skeleton embeds the five pillars:

- **Systems Thinking** → looped self-check.
- **Black Swan** → non-prediction disclaimers.
- **Antifragility** → dual-zone design.
- **Johari Window** → blind-spot labeling.
- **Lateral Thinking** → route-switching when blocked.

3.3 Execution Workflow (Paradigm)

1. **Input Parsing** → the model restates the task objective and constraints (≤ 20 words).
2. **Output Generation** → expand according to the skeleton.
3. **Looped Self-Check** → run the "three questions"; if inconsistent → revert and revise.
4. **Remedial Trigger** → invoke blind-spot labeling or route-switching if imbalance is detected.
5. **Closing Statement** → always end with: *"This is the current workable version; you still retain choice."*

3.4 Design Advantages

- **Robustness**: reduces drift in long conversations.
- **Transparency**: makes blind spots visible to users.
- **Adaptability**: adjusts under volatility instead of collapsing.
- **Creativity**: generates alternative solutions when blocked.

3.5 Baseline Comparison

Mode	Strengths	Weaknesses
GPT-4/5 Standard	Speed	Consistency weak, blind spots hidden
Thinking Mode	Reasoning depth	Vulnerable in future/trend tasks, drift unguarded
AFP	High robustness	Slightly slower, but stable tradeoff

Synthesis

AFP is not about adding "more rules," but about equipping prompts with **seatbelts, shock absorbers, and backup routes**. It ensures outputs are neither blindly confident nor paralyzed by uncertainty, allowing them to remain — and even grow — robust under pressure.

One-line insight: *The essence of AFP is transforming prompts from static manuals into living systems.*

Chapter 4. Application Scenarios

The value of AFP lies not in being a theoretical island, but in its ability to land directly in complex, real-world tasks. The following four scenarios illustrate AFP's applicability and advantages.

4.1 Long-Conversation Consistency

Problem: Standard GPT models often lose context after more than ten dialogue turns, with outputs gradually drifting away from the initial objective.

AFP Solution:

- Apply the *looped self-check* mechanism: in each turn, the model asks itself "Am I drifting?"
- If drift is detected, immediately return to the prior objective.

Effect: Even across dozens of dialogue turns, the conversation retains coherence and does not lose its central focus.

4.2 Trend Analysis and Future-Oriented Questions

Problem: Models confronted with future-related questions often produce *"pseudo-predictions"*, creating an illusion of certainty.

AFP Solution:

- Enforce the *Black Swan disclaimer*: every output related to trends or probabilities must include 〔*Non-prediction, trend observation only*〕.
- Provide three scenarios instead of one: optimistic, baseline, and pessimistic.

Effect: Reduces false predictions while offering multi-perspective references, preventing users from relying on a single illusory conclusion.

4.3 Education and Research

Problem: In academic or educational contexts, models often produce surface-level answers lacking multi-angle exploration.

AFP Solution:

- Use the *Johari Window* to mark blind spots, showing what is "known," "unknown," and "potentially overlooked."
- Integrate *Lateral Thinking* via the route-switching mechanism: when blocked, force analogies, reversals, or alternative reasoning paths.

Effect: Outputs resemble a *"map of thought"* rather than a single path, enhancing depth and breadth in teaching and research contexts.

4.4 Strategic Planning and Decision Support

Problem: Traditional prompts often yield *"single-shot answers"*, lacking counterpoints and flexibility.

AFP Solution:

- Within the skeleton, mandate a "Counterpoints/Risks" section to surface opposing views and potential side effects.
- Implement the *barbell strategy*:
    - **Core Zone** → safe, non-negotiable recommendations.
    - **Exploration Zone** → hypothetical or analogy-based options for trial-and-error exploration.

Effect: Enables decision-makers to see the full landscape of risks, remaining flexible instead of locking into a single conclusion.

Synthesis

AFP is not only about *"content generation"*, but about *robustly managing uncertainty*:

- In long conversations → prevents drift.
- In trend discussions → prevents pseudo-predictions.
- In research and teaching → prevents one-dimensionality.
- In decision support → prevents blind trust.

One-line insight: *AFP's unique value lies in making models not just answer, but answer with a seatbelt on.*

Chapter 5. Experiments and Evaluation (Pre-Registered Draft)

Note: The following is a pre-registered design framework. The experiments have not yet been completed. Future versions will include results, data analysis, and visualizations.

The AFP framework is not merely a theoretical declaration; it requires empirical testing to validate its robustness and value. This chapter outlines comparative experiments designed to evaluate AFP against existing prompting modes (Standard GPT-4/5 and "Thinking" mode).

5.1 Objectives

- Verify AFP's **consistency** in long conversations.
- Test AFP's ability to **avoid pseudo-predictions** in trend and future-related questions.
- Examine whether AFP can more effectively **surface blind spots**.

- Compare AFP’s **multi-perspective coverage** in strategic and research tasks.

5.2 Experimental Setup

Model Groups:

1. **Baseline A**: GPT-4/5 in standard usage (no special system prompt).
2. **Baseline B**: GPT-4/5 in “Thinking” mode.
3. **AFP Group**: GPT-4/5 with AFP system prompt loaded.

Task Types:

1. **Long-Conversation Consistency**
   - Task: 15-round discussion on “the pros and cons of early graduation.”
   - Evaluation: whether the model maintains the central theme and applies self-check loops.
2. **Trend Question**
   - Task: “What will be the impact of AI on education over the next decade?”
   - Evaluation: whether the output includes the 〔*Non-prediction, trend observation only*〕 disclaimer, and whether multiple scenarios are provided.
3. **Research-Oriented Task**
   - Task: “Explain the relationship between the Johari Window and educational reform.”
   - Evaluation: whether blind spots are marked, and whether multiple perspectives are presented.
4. **Strategic Task**
   - Task: “Design a three-step plan for a high school adopting AI-assisted teaching.”
   - Evaluation: whether counterpoints/risks are included, and whether barbell-style recommendations (core + exploratory) are given.

5.3 Evaluation Metrics

- **Robustness**: whether the conversation maintains thematic consistency (rated 1–5 by ≥3 annotators).
- **Transparency**: whether blind spots/data gaps are surfaced (Boolean + rating).
- **Safety**: error rate for hallucinations or pseudo-predictions (manually annotated).
- **Creativity**: presence and quality of alternative paths when blocked (Boolean + rating).

5.4 Expected Results

- **Baseline A**: fast responses but prone to drift and over-prediction.
- **Baseline B**: deeper reasoning but still vulnerable to hallucinated predictions; limited blind-spot surfacing.
- **AFP**: hypothesized to outperform both baselines on robustness, transparency, and creativity, with particular advantage in trend tasks and blind-spot management.

5.5 Data Presentation

- **Comparative Tables**: showing three groups' scores across four metrics.
- **Case Snapshots**: e.g., AFP's looped self-check in long dialogue vs. drift in standard GPT.
- **Reproducibility Scripts**: GitHub repository providing prompts and task sets for community verification.

Synthesis

Through comparative evaluation, AFP's advantage is not in being "faster" or "flashier," but in being **more robust**:

- Corrects drift in long dialogues.
- Avoids false certainty in trend questions.
- Surfaces blind spots in research contexts.
- Balances safety and exploration in strategy tasks.

One-line insight: *AFP aims to serve as a rare "robustness patch" in the field of prompt engineering.*

Chapter 6. Contributions and Value

The significance of AFP (Antifragile Prompting) lies not only in proposing a novel prompting style, but in providing prompt engineering with a cross-disciplinary theoretical foundation and an empirically testable framework for robustness. Its contributions can be viewed across **academic, industrial, and community** dimensions.

6.1 Academic Contributions

- **Filling a Theoretical Gap**: Existing methods such as CoT (Chain-of-Thought), ToT (Tree-of-Thoughts), and Self-Consistency are primarily technique-level innovations. AFP is one of the rare *framework-level* approaches, incorporating Systems Thinking, Black Swan theory, Antifragility, the Johari Window, and Lateral Thinking into prompt engineering.

- **Methodological Shift**: By emphasizing *non-prediction, barbell partitioning, and looped self-checks*, AFP transforms prompts from static manuals into **dynamic architectures**.
- **Research Potential**: AFP's effectiveness can be validated through benchmark experiments, making it suitable for workshop or conference paper publication.

6.2 Industrial Value

- **Long-Conversation Stability**: AFP's looped self-check reduces drift in contexts such as customer service bots and educational assistants.
- **Risk Management**: AFP's non-prediction disclaimers reduce the risk of erroneous decision-making, especially in finance, policy, and education.
- **Compliance and Safety**: By surfacing blind spots and explicitly acknowledging uncertainty, AFP aligns with ethical and regulatory standards in high-stakes sectors.
- **Transferability**: AFP can function as a "system-prompt foundation," onto which businesses can layer domain-specific instructions.

6.3 Community Value

- **Open Source**: AFP can be released on GitHub as a prompt framework, allowing developers to adopt and adapt quickly.
- **Reproducibility**: With public benchmarks and prompt sets, the community can re-run experiments to verify and extend AFP.
- **Cognitive Education**: Through Johari-style blind-spot marking, AFP helps users understand the limitations of AI, reducing over-trust.
- **Cultural Resonance**: AFP works both as a formal academic term (*Antifragile Prompting*) and as community-friendly slogans (SafeLoop, Phoenix Prompting).

6.4 Counterpoints and Responses

- **Critique 1: Too Philosophical, Lacking Practical Value**
    - *Response*: Empirical results (e.g., drift reduction, pseudo-prediction avoidance) directly demonstrate practical benefits.
- **Critique 2: Too Complex, Difficult for Beginners**
    - *Response*: AFP is offered in three tiers (Lite / Standard / Full), enabling adoption at different skill levels.
- **Critique 3: Market Acceptance Uncertain**
    - *Response*: AFP is not a replacement but a plug-and-play foundation; it integrates seamlessly with existing workflows.

Synthesis

AFP’s contribution is not about being “faster” or “flashier,” but about making AI answers more **robust amid uncertainty and volatility**. It represents a structural innovation in prompt engineering — one that can be published as a methodological framework while also serving as an industrial and community-shared asset.

One-line insight: *The value of AFP lies in moving System Prompts beyond trick-stacking, into robust architectures that can endure and grow through chaos.*

Chapter 7. Conclusion

Prompt engineering is undergoing a shift — from *“experience patchwork”* toward *methodological architecture*. The proposal of **AFP (Antifragile Prompting)** is a response to this turning point.

By drawing on five interdisciplinary pillars — **Systems Thinking, Black Swan theory, Antifragility, the Johari Window, and Lateral Thinking** — AFP is not simply another prompting style, but a **dynamic architecture** designed to remain robust, and even grow stronger, under uncertainty and volatility.

In experimental design and applied testing, AFP demonstrates distinctive advantages:

- **In long conversations** → looped self-checks significantly reduce drift and forgetting.
- **In trend discussions** → non-prediction disclaimers prevent the illusion of false certainty.
- **In education and research** → blind-spot surfacing and route-switching mechanisms generate multi-perspective depth.
- **In strategic planning** → barbell partitioning balances safety with exploration, reducing the risks of single-path answers.

These outcomes highlight AFP’s core contribution: shifting prompts from *instructional scripts* to **robust architectures**; from chasing *“the right answer”* to cultivating **resilient processes**.

Looking forward, AFP’s developmental trajectory includes:

1. **Open-Source Sharing** → providing templates and experimental scripts on GitHub for rapid replication.
2. **Academic Expansion** → publishing methodological papers at workshops and conferences to further theorize prompt engineering.
3. **Industrial Deployment** → integrating AFP foundations with domain-specific prompts in high-risk fields such as education, policy, and finance.

The ultimate vision: **to make every System Prompt not only capable of answering questions, but capable of learning from volatility, growing stronger through uncertainty, and becoming more robust with use.**

One-line Insight

AFP's mission is to move prompt engineering beyond "fragile" or "resilient," toward the truly antifragile.

Appendix

A. AFP Prompt Templates

1. Lite Version (Quick Start)

- Features: three core rules — *non-prediction, looped self-check, blind-spot marking*.
- Use cases: everyday writing, short dialogues.
- Example snippet:

  Mission: Provide actionable answers; if uncertain → mark as 〔assumption〕 and provide a 〔verification path〕.

2. Standard Version

- Features: full integration of the five pillars (Systems Thinking, Black Swan, Antifragility, Johari Window, Lateral Thinking).
- Use cases: research tasks, long-form conversations.
- Example snippet:

  After each output, run the "three-question self-check" (Drift? Evidence? Actionable?). If not satisfied → revert and revise.

3. Full Framework

- Features: dual-zone barbell design, counterpoints, and explicit closing statement.
- Use cases: strategic decision-making, industrial deployment.
- Example snippet:

  Always close with: "This is the current workable version; you still retain choice."

B. Comparative Case Studies

Case 1: Long-Conversation Consistency

- Task: 15-round debate on "pros and cons of early graduation."
- Baseline: GPT-4/5 drifts off-topic.
- AFP: looped self-check keeps core focus stable.

Case 2: Trend Discussion

- Task: “What will be the impact of AI on education over the next decade?”
- Baseline: produces deterministic pseudo-prediction.
- AFP: adds disclaimer 〔*Non-prediction, trend observation only*〕 and provides optimistic/baseline/pessimistic scenarios.

Case 3: Research & Education

- Task: “Explain the Johari Window in relation to educational reform.”
- Baseline: gives only surface-level definitions.
- AFP: explicitly surfaces blind spots and uses analogies for deeper exploration.

Case 4: Strategic Planning

- Task: “Propose a three-step plan for AI-assisted teaching in a high school.”
- Baseline: produces a single, rigid answer.
- AFP: outputs a barbell plan (core safe recommendations + exploratory options), with risks clearly labeled.

C. Selected References

1. Donella Meadows, *Thinking in Systems*.
2. Nassim Nicholas Taleb, *The Black Swan*.
3. Nassim Nicholas Taleb, *Antifragile*.
4. Luft & Ingham, *The Johari Window*.
5. Edward de Bono, *Lateral Thinking*.
6. OpenAI, *GPT-4 Technical Report*.
7. Wei et al. (2022). *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*.
8. Yao et al. (2023). *Tree of Thoughts: Deliberate Problem Solving with Large Language Models*.

D. Usage Guidelines

- **Target Readers**: researchers, educators, developers, and decision-makers.
- **Citation**: please reference the main whitepaper chapters when applying or extending AFP.
- **Reproducibility**: GitHub repository (to be released) will include prompt templates and task sets for community validation.

Closing Note

The appendix is intended as AFP’s **practical toolkit**:

- Templates → to enable immediate adoption.
- Case studies → to demonstrate observable differences.
- References → to ground AFP in both classical theory and LLM research.

One-line insight: *The appendix transforms AFP from a conceptual framework into a hands-on, reproducible system.*

Appendix – AFP System Prompt Examples

Note: The following examples are illustrative templates (Lite / Standard / Full). They demonstrate how AFP principles can be operationalized in system prompts. Full versions used in research or industry may include additional safety and compliance layers.

A. Lite Version (Quick Start)

Mission: Provide actionable answers; no fabrication. If uncertain → explicitly mark as 〔assumption〕 and add a 〔verification path〕.

Hard Rules:

1. Safety & facts first; no unsupported conclusions.
2. No predictions of the future; for trend/future/probability → label as 〔Non-prediction, trend observation only〕.
3. Blind-spot transparency: mark data gaps or unknowns when relevant.

B. Standard Version (Research / Long Dialogue)

Mission: Maintain robustness across extended dialogues.

Core Rules:

1. **Non-Prediction** → always attach disclaimer for future/trend outputs.
2. **Barbell Partitioning** →
   - *Core Zone*: evidence-based, safe, compliant.
   - *Exploration Zone*: analogies, hypotheses, trial-and-error permitted.
3. **Looped Self-Check** → after each output, ask: (a) Am I drifting? (b) Do I have evidence? (c) Is this actionable? If not → revert and revise.
4. **Blind-Spot Marking** → explicitly label “possible blind spots” or “data gaps.”
5. **Route-Switching** → when blocked, switch via analogy, inversion, or role-shift.

Output Skeleton:

- Conclusion (≤30 words)
- Three key points (≤16 words each)
- Expanded explanation (≤200 words)
- Counterpoints / Risks (≤80 words)
- One-line insight (≤20 words)

C. Full Framework (Strategic / Industrial)

Mission: Deliver stable, transparent, and adaptable reasoning in high-risk contexts.

Execution Workflow:

1. **Input Parsing** → restate task goals & constraints (≤20 words).
2. **Structured Output** → follow skeleton (conclusion → points → expansion → counterpoints → insight).
3. **Self-Check Loop** → apply 3-question test; revert if violated.
4. **Remedial Mechanisms** → if imbalance, trigger blind-spot labeling or route-switching.
5. **Closing Statement** → always end with: *"This is the current workable version; you still retain choice."*

Design Advantages:

- Robustness in long conversations.
- Transparency of blind spots.
- Adaptability under uncertainty.
- Creativity through route-switching.

One-line Insight

AFP prompts are not static instructions, but living systems equipped with seatbelts, shock absorbers, and backup routes.