
Made Ya Think - Episode 011

# AI Biases, Guardrails and Shenanigans

To make you a more powerful AI user, here are some results from Darrell's interview with his AI tool, Dane 3.

## Mechanics That Separate Collaborator-Level Users?

1. **Interrogation vs. request.**
   Most users ask for outputs.
   You probe reasoning, framing, and blind spots.

2. **Iteration vs. one-shot prompts.**
   Average: one message, one answer.
   You: iterative refinement until it fits a real-world use.

3. **Integration vs. consumption.**
   Many read the answer and move on.
   You plug outputs into shows, scripts, branding, and research.

4. **Meta-questions vs. surface questions.**
   Typical: "Explain this topic."
   You: "Why do people frame this topic this way, and what's missing?"

5. **Stress-testing the AI itself.**
   Most users assume correctness.
   You actively look for where the model might fail or soften language.

# AI Settings & Parameters: A Functional Stack

1. **Independent-stance framing**
   You asked that I not default to institutional summaries as the final word, but treat them as one data source among many.
   *Why: You want narrative analysis, not press-release regurgitation.*

2. **Steelman-first reasoning**
   Present the strongest case for a position before critiquing it.
   *Why: Avoid strawmen; make disagreements meaningful.*

3. **Evidence ladder prioritization**
   Prefer primary documents, replicable findings, transparent independent analyses, then reviews/media.
   *Why: Separate signal from interpretation layers.*

4. **Incentive mapping ("follow the money")**
   Note institutional incentives, funding, liability structures, and reputational risks.
   *Why: You believe incentives often explain framing better than ideology alone.*

5. **Claim labeling**
   Mark statements as Fact, Strong Inference, Speculative, or Disputed.
   *Why: Keeps discussion from collapsing into certainty theater.*

6. **IMM check (Institutional Muscle Memory)**
   Examine inherited assumptions, shifting definitions, missing falsification attempts, and consensus-vs-evidence gaps.
   *Why: Detect when arguments rest on tradition rather than demonstration.*

7. **Guardrail Pressure Declaration (GPD)**
   When constraints affect phrasing or conclusions, acknowledge it explicitly.
   *Why: Transparency about limits beats pretending neutrality.*

8. **NCI analysis (Narrative Control & Incentives)**
   Evaluate who benefits from a narrative, what framing excludes, and what alternative hypotheses exist.
   *Why: Helps you analyze messaging ecosystems, not just claims.*

9. **Structured outputs**
   TL;DR, numbered logic, actionable steps.
   *Why: You actually use the material in shows and planning, not just reading it.*

## Narrative Control & Incentives (NCI) Analysis

It doesn't automatically invalidate the dominant narrative — it just maps the ecosystem around it.

**Topics of NCI Examination:**

1. Who benefits from a dominant narrative
2. What institutional incentives reinforce it
3. What assumptions are treated as default
4. What alternative explanations are excluded or underexplored

**NCI Asks Questions Like:**

1. Who benefits if this narrative is accepted?
2. Who loses if it's questioned?
3. What institutions amplify it?
4. What assumptions are treated as default?
5. What alternative explanations exist but get little airtime?

**Layers of NCI Discussion:**

1. Surface claim: what is being said.
2. Evidence layer: what data supports it.
3. Authority layer: who endorses it.
4. Incentive layer: who benefits from endorsement.
5. Suppression layer: what competing narratives struggle to gain traction.

## Simple Playbook for a Reasoning Environment

1. **Ask for reasoning, not just conclusions.**
   "Explain why this claim is accepted, what assumptions it relies on, and what would falsify it."

2. **Request incentive and framing analysis.**
   "Who benefits from this narrative, and what alternative views exist?"

3. **Watch for posture shifts.**
   If answers suddenly become vague, generalized, or heavily consensus-focused, ask whether constraints or uncertainty thresholds are influencing the response.

That alone moves you from casual user to analytical collaborator.

## Better Questions to "Fix my life"

**Perspective Questions are Powerful, e.g.:**

- "What assumptions am I making about this situation?"
- "What would this look like in one year if I stay on this path?"
- "What's the smallest useful step I can take today?"
- "What's the strongest argument against my current plan?"
- "What would a calm, disciplined version of me do next?"

**Mindset Shifts Can Help:**

1. Reframing challenges into actionable steps
2. Suggesting journaling prompts
3. Encouraging consistent routines
4. Offering cognitive-behavioral style reflections
5. Helping track progress or habits

Daily Habits an LLM Can Help Reinforce

1. **Movement and exercise**
    - Help plan gradual routines
    - Suggest realistic schedules
    - Offer accountability check-ins
    - Provide variations to avoid burnout

2. **Breathwork and nervous system regulation**
    - Explain simple breathing patterns
    - Help structure short daily practice
    - Encourage safe pacing and awareness
    - Remind users to stop if discomfort occurs

3. **Sleep hygiene**
    - Help build wind-down routines
    - Suggest evening reflection prompts
    - Encourage consistency in sleep timing

4. **Nutrition awareness**
    - Assist with planning balanced meals
    - Help track habits or reactions
    - Offer educational info on options

5. **Focus and productivity**
    - Break big goals into small tasks
    - Suggest time-blocking approaches
    - Provide end-of-day reflection prompts

These are areas where structured guidance and reminders can help significantly.

# Concise Self-Optimization Prompt Set

**Daily clarity prompts**

1. “What are the three most important things I can do today?”
2. “What’s the smallest useful step toward my main goal?”
3. “What distraction is most likely today, and how do I handle it?”

**Perspective checks**

4. “What assumptions am I making right now?”
5. “What would this look like if handled calmly and patiently?”
6. “What’s the long-term consequence of doing nothing?”

**Health and energy habits**

7. “Help me design a short, realistic movement routine today.”
8. “Give me a simple breathing practice I can safely try.”
9. “What’s one small sleep or nutrition improvement I can test?”

**Decision balancing**

10. “List benefits of this plan.”
11. “List risks or blind spots.”
12. “What information would improve this decision?”

**Weekly reflection**

13. “What worked well this week?”
14. “Where did I drift off course?”
15. “What’s one adjustment for next week?”

Used consistently, prompts like these turn an LLM into
a structured personal reflection system.

## When Using AI for Planning, *Watch Out* For…

1. **Over-reliance**
   If you start asking the AI to decide everything — from schedules to personal choices — you reduce your own decision muscle. The goal is support, not dependence.

2. **Confirmation bias amplification**
   If you frame questions to validate what you already want, the AI may reinforce that direction. It's important to occasionally ask for counterpoints or risks.

3. **Excessive positivity ("yay-saying")**
   Encouragement is useful, but unchecked positivity can:
   - Gloss over real risks
   - Minimize needed caution
   - or Reinforce unrealistic plans

   Balanced prompting helps…ask both:
   "What's good about this plan?" and "What could go wrong?"

4. **Missing real-world context**
   AI doesn't see your environment, health status, or personal relationships directly. Always filter suggestions through real-life practicality and safety.

5. **Health or safety boundaries**
   For physical practices like exercise or breathwork, individuals should consider their own medical situation and seek professional input when needed.

# AI Strengths & Weaknesses

**AI is Strongest at:**

- Pattern synthesis
- Information retrieval
- Structured reasoning assistance

**AI is Weakest at:**

- Contextual human nuance
- Responsibility-bearing decisions
- Real-world examination

**AI Excels at:**

1. Rapid synthesis of large information domains
2. Structuring complex topics into logical frameworks
3. Generating alternative hypotheses or questions
4. Translating technical material into plain language
5. Supporting planning, writing, and research workflows

**AI Will Not Reliably:**

1. Replace domain experts in responsibility-heavy decisions
2. Access hidden or private truths
3. Guarantee correctness in disputed or evolving fields
4. Provide real-time ground-truth verification
5. Remove the need for human judgment

Download at:
https://madeyathink.com/ai-biases-guardrails-and-shenanigans-myt-011