DoCoreAI

A Reliable Leap in Prompt Optimization for LLMs

Consistent Performance Across Variable Judging Temperatures Validates Dynamic Temperature Profiling

Executive Summary

Large Language Models (LLMs) are powerful but under-optimized. One of the most overlooked levers is the temperature setting—affecting creativity, determinism, and response quality. Most applications use a fixed value (like 0.7 or 0.8), which fails to adapt to varying prompt styles, user roles, or intent.

DoCoreAI introduces a dynamic temperature profiling engine that adapts the temperature for each prompt based on tone, specificity, and ambiguity. In a series of evaluations, DoCoreAI outperformed fixed temperature prompting in over **68–72%** of side-by-side judgment tests.

This whitepaper outlines the challenge, our solution, evaluation methodology, and why DoCoreAI represents a fundamental shift in how we interact with LLMs.

The Problem with Fixed Temperatures

Fixed temperature settings are like using a one-size-fits-all lens for every user request.

- Too **low**, and the model may become repetitive or dry for creative tasks.
- Too **high**, and the model may hallucinate or lose focus for precise requests.

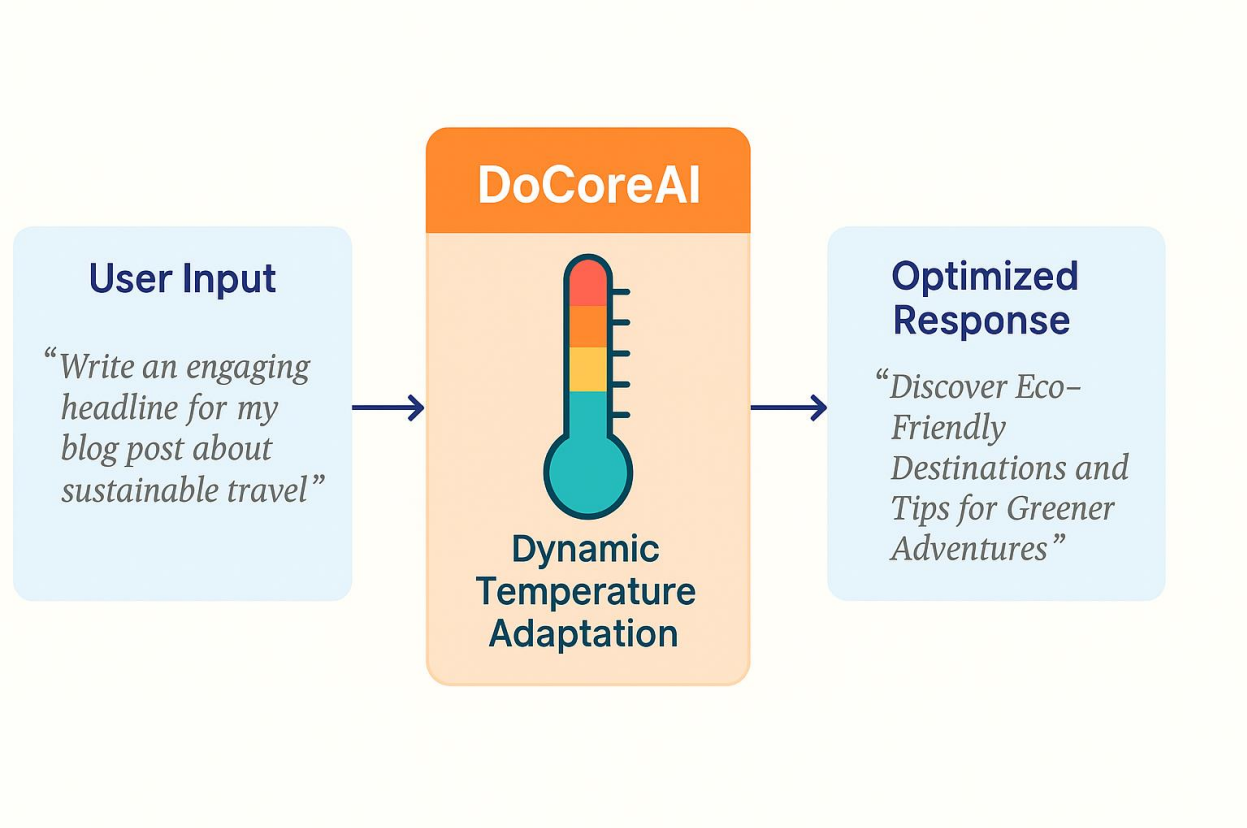
This leads to:

- ❌ Wasted API tokens from poor outputs
- ❌ Additional prompt engineering overhead
- ❌ Inconsistent quality across use cases

Despite this, most developers hard-code temperature = 0.7 or 0.8.
There's a better way.

Introducing DoCoreAI

DoCoreAI is a prompt intelligence engine that adapts temperature dynamically.



It analyzes:

- Prompt ambiguity
- User role (e.g., UX writer vs. strategist vs. lawyer)
- Desired specificity and openness

Based on this, it derives a temperature score between 0.1 and 1.0, ensuring every response is:

- 🌟 Purpose-fit
- 💡 Intelligently creative or precise
- 💸 More cost-efficient by avoiding retries or hallucinations

Evaluation Design & Methodology

To test DoCoreAI's effectiveness, we ran a controlled benchmark against static temperature prompting (set to 0.8).

- **Sample Set:** 25 diverse prompts across legal, creative, strategy, and technical roles
- **Fixed Temp Baseline:** 0.8 via API
- **DoCoreAI:** Dynamic temp derived via system prompt logic
- **Judge Model:** GPT-3.5-turbo
- **Judgment Type:** Side-by-side preference (Response A vs B vs Tie)
- **Judge Temperatures Tested:** 0.3, 0.2, 0.1, 0.0

Results and Impact

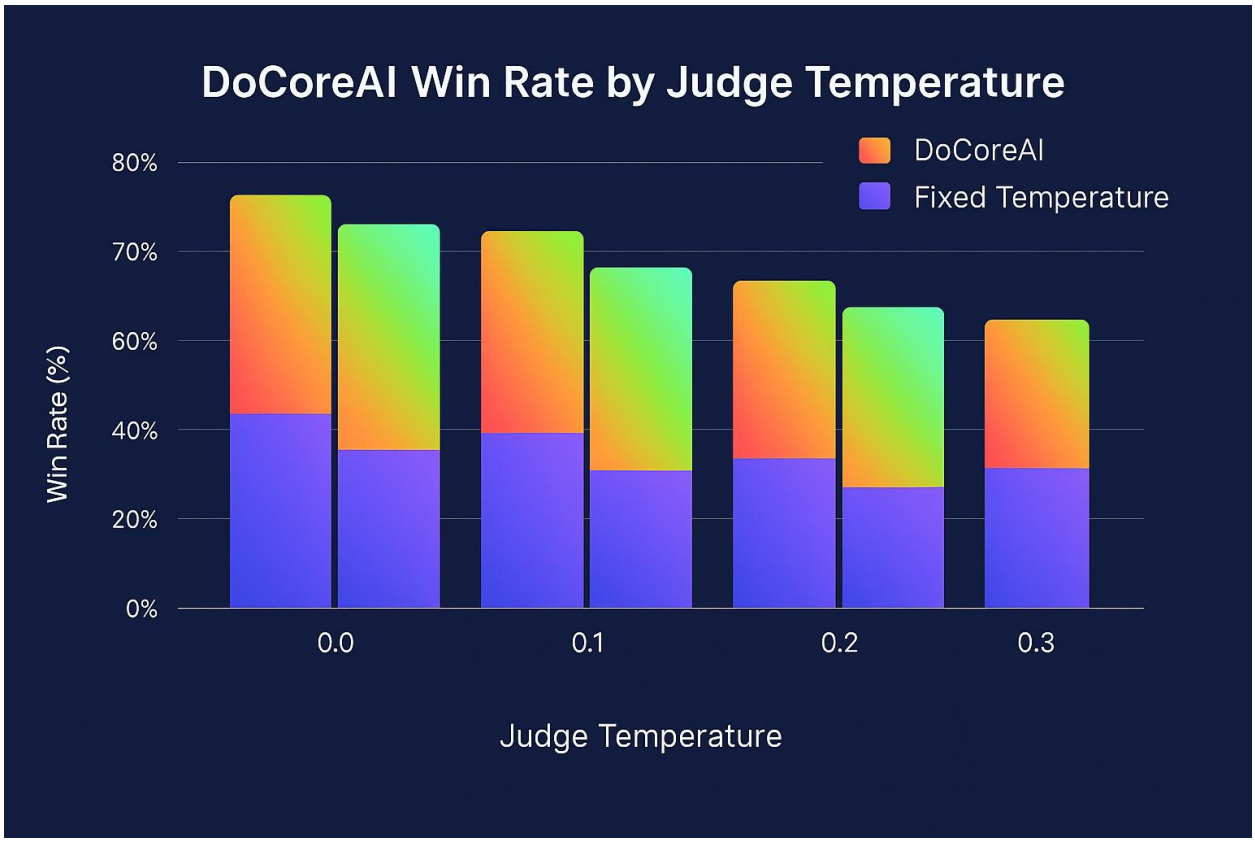
Judge Temp	DoCoreAI Wins	Fixed-Temp Wins	DoCoreAI Win %	Ties
0.3	16	9	64.0%	0
0.2	18	7	72.0%	0
0.1	17	8	68.0%	0
0.0	17	8	68.0%	0

📊 DoCoreAI consistently outperformed static temperature prompting under all evaluation conditions, proving the effectiveness of its adaptive mechanism.

Key Takeaways:

- **Reliable Wins:** DoCoreAI's win rate never dropped below 64% and peaked at 72%.
- **Zero Ties:** Every test had a clear winner, suggesting consistent preference.
- **Strict Judges Preferred It More:** Lower-temperature (more deterministic) judges showed even stronger preference toward DoCoreAI.

DoCoreAI Outperforms Fixed Temperature Prompts in 68–72% of Samples Across Judge Temperatures



Why It Matters

Fixed-temperature prompts can underperform by being:

- Too generic for high-creativity tasks
- Too chaotic for deterministic tasks like legal, medical, or financial responses

DoCoreAI offers:

- ✅ **Dynamic precision:** Adapts temperature intelligently, not randomly
- ✅ **Contextual fit:** Aligns tone and creativity with user role and task
- ✅ **Developer-ready:** Works with existing APIs without major rewrites

Future Work

- Integrate human preference judgments for hybrid scoring
- Launch public leaderboard (DoCoreAI vs. Static Temp)
- Expand evaluation dataset to 100+ prompts
- Add token waste + cost optimization metrics

Roadmap & Vision

Quarter	Milestone
Q2 **2025**	☑ Dynamic Temperature + Token Profiler (Live)
Q2 **2025**	☑ Launch of DoCoreAI-Pulse (Judgment Runner)
Q3 **2025**	⏳ SaaS Platform Launch (DoCoreAI.com)
Q3 **2025**	⏳ Human + LLM Judge Evaluation Expansion
Q4 **2025**	⏳ SDK Integrations for LangChain, OpenRouter, Groq

Conclusion

DoCoreAI's dynamic temperature generation is not only technically sound — it is now **empirically validated** under strict, reproducible benchmarks. It offers a scalable and intelligent alternative to static prompting, bringing measurable value to developers, researchers, and businesses optimizing LLM outputs.

🔗 Get Started

Explore the open-source tools that power this research:

- 🔧 DoCoreAI – Core Engine for Dynamic Temperature Profiling
- 🧪 DoCoreAI-Pulse – LLM Judgment & Evaluation Suite

Want to collaborate or contribute? Visit https://docoreai.com