# AI SAMPLING PARAMETERS MASTERY GUIDE

AIMCLEAR® MARKETING AGENCY

## TEMPERATURE

0.0→2.0 | The Creativity Controller | Default: 0.7-1.0

MECHANISM: Adjusts probability distribution for token selection. Low=predictable, High=creative

FORMULA: `P(token) = exp(logit/T) / Σexp(all_logits/T) — Higher T flattens distribution`

**PROBABILITY DISTRIBUTIONS:**


**ACTUAL OUTPUTS for "Write about the future of AI":**

| | | |
|---|---|---|
| **T=0.0** | **DETERMINISTIC** | "The future of AI involves continued advancement in machine learning a |
| *Same every time, Wikipedia-like, factual* | | |
| **T=0.5** | **PROFESSIONAL** | "The future of AI promises transformative changes across industries an |
| *Varied but safe, business appropriate* | | |
| **T=1.0** | **CREATIVE** | "The future of AI dances between silicon dreams and digital consciousn |
| *Metaphorical, engaging, marketing-ready* | | |
| **T=1.5+** | **EXPERIMENTAL** | "AI's future spirals through quantum possibilities birthing synthetic |
| *Artistic but may lose coherence* | | |

**OPTIMAL SETTINGS BY USE CASE:**

| 0.0-0.2 | 0.3-0.5 | 0.6-0.8 | 0.9-1.2 | 1.3-2.0 |
|---|---|---|---|---|
| Legal, Medical, Data | Support, Docs, FAQs | Blogs, Emails, Reports | Creative, Marketing, Ideas | Fiction, Poetry, Art |

OpenAI: 0-2 | Claude: 0-1 | Gemini: 0-1 | Images: CFG 1-30 | Video: Motion 0-255

## MAX TOKENS

1→128K | Output Length Limiter | 1 token ≈ 0.75 words

PURPOSE: Controls response length & cost. Model stops at limit OR natural completion.

CRITICAL: You pay per token! GPT-4=0.06/1*Koutput*, *GPT* − 3.5 = 0.0015/1K

**TOKEN TO CONTENT MAPPING:**

| | | | |
|---|---|---|---|
| 10 | **MICRO** | ~8 words | Response: "Yes" |
| 50 | **TWEET** | ~38 words | Single sentence answer |
| 150 | **PARAGRAPH** | ~110 words | Short explanation |
| 500 | **EMAIL** | ~375 words | Detailed response |
| 1500 | **ARTICLE** | ~1125 words | Full documentation |
| 4000 | **CHAPTER** | ~3000 words | Complete guide |
| 8000 | **BOOK SECTION** | ~6000 words | Long-form content |

**MODEL LIMITS:**

GPT-4 Turbo: 4K output / 128K context
Claude 3 Opus: 4K output / 200K context
Gemini 1.5 Pro: 8K output / 1M context
Context = input capacity ≠ output length!

**COST EXAMPLE:**

2000 token response:
GPT-4: $0.12
GPT-3.5: $0.003
Claude: Varies by tier

**PRO TIP: Set higher than needed—model stops naturally. Monitor usage to optimize costs.**

## TOP P (NUCLEUS)

Dynamic vocabulary via cumulative probability

**HOW P=0.9 WORKS:**

| | | | |
|---|---|---|---|
| **blue** | 40% | Σ=40% | ✓ INCLUDED |
| **azure** | 30% | Σ=70% | ✓ INCLUDED |
| **cyan** | 15% | Σ=85% | ✓ INCLUDED |
| **teal** | 8% | Σ=93% | ✗ EXCLUDED |
| **navy** | 4% | Σ=97% | ✗ EXCLUDED |
| **other** | 3% | Σ=100% | ✗ EXCLUDED |

**SETTINGS:**

| P=0.1: Top 10% only | P=0.5: Top 50% | P=0.9: Top 90% | P=1.0: Everything |
|---|---|---|---|
| Ultra focused | Balanced | Creative | Max variety |

## TOP K

Fixed vocabulary limit regardless of probability

**HOW K=5 WORKS:**

| | |
|---|---|
| 1. blue | ✓ INCLUDED |
| 2. azure | ✓ INCLUDED |
| 3. cyan | ✓ INCLUDED |
| 4. teal | ✓ INCLUDED |
| 5. navy | ✓ INCLUDED |
| — CUT OFF — | |
| 6. cobalt | ✗ EXCLUDED |
| 7. indigo | ✗ EXCLUDED |
| 8. sapphire | ✗ EXCLUDED |

**SETTINGS:**

K=1: Greedy | K=10: Limited | K=40: Natural | K=100: Wide
**OpenAI doesn't support K. Use P instead!**

## REPETITION PENALTIES

Prevent Redundant Output | 3 Types

WHY: AI tends to repeat. These penalties force vocabulary diversity.

**FREQUENCY**
OpenAI
-2 to 2
Counts occurrences
*"very very" → "extremely"*

**PRESENCE**
OpenAI
-2 to 2
Binary (used/not)
*Forces new vocabulary*

**REPETITION**
Claude/Gemini
0 to 2
Multiplier (1=off)
*Single parameter*

**EFFECT DEMONSTRATION:**

OFF: "The cat was very very very happy" — *Natural repetition allowed*

MED: "The feline was remarkably joyful" — *Clear diversity*

LOW: "The cat was extremely happy" — *Subtle vocabulary shift*

HIGH: "That feline seemed remarkably euphoric" — *May feel unnatural*

## IMAGE & VIDEO GENERATION PARAMETERS

### IMAGE GENERATION

| | | |
|---|---|---|
| **CFG SCALE** *→ Higher=stricter but artifacts >15* | 1-30 (7.5 default) | Prompt adherence strength |
| **STEPS** *→ Quality vs time tradeoff* | 10-150 (30-50 optimal) | Denoising iterations |
| **SEED** *→ Save for consistent style* | -1 or specific number | Reproducibility control |
| **SAMPLER** *→ Affects aesthetic style* | Euler a, DPM++, DDIM | Algorithm selection |

### VIDEO GENERATION

| | | |
|---|---|---|
| **MOTION** *→ 100=natural, 255=chaotic* | 0-255 | Movement intensity |
| **FPS** *→ 24=cinematic standard* | 8-60 | Frame rate |
| **DURATION** *→ Exponential compute cost* | 2-16 seconds | Clip length |
| **INIT STRENGTH** *→ For img2vid/vid2vid* | 0-1 | Source influence |

**QUICK START: Temp=0.7 | MaxTokens=1000 | TopP=0.9 | RepPenalty=1.1**

AIMCLEAR® MARKETING AGENCY

Integrated Marketing Superpower
**aimclear.com**