

AI Model Reference - Production Configuration

Last Updated: October 8, 2025

Research Source: Perplexity AI via tools/research/model-discovery.mjs

Verified Production Models

OpenAI GPT-5 Pro

Status: Production (October 2025)

OPENAI_MODEL=gpt-5-pro

API Details:

- Endpoint**: POST <https://api.openai.com/v1/chat/completions>
- Model ID**: gpt-5-pro (flagship reasoning model)
- Context Window**: 256K tokens (256,000 tokens/request)
- Headers**:
 - Authorization: Bearer <API_KEY>
 - Content-Type: application/json

Supported Parameters (⚠ BREAKING CHANGES):

```
{  
  
  "model": "gpt-5-pro",  
  
  "messages": [...],  
  
  "reasoning_effort": "minimal" | "low" | "medium" | "high", // USE THIS  
  
  "max_completion_tokens": 32000
```

max_completion_tokens : 52000,

"stream": true,

"stop": [...],

"tools": [...] // For function calling

}

****[] DEPRECATED PARAMETERS**** (GPT-5 does NOT support):

- temperature → Use reasoning_effort instead
- top_p → Use reasoning_effort instead
- frequency_penalty → Not supported
- presence_penalty → Not supported

****Token Usage**** (special for GPT-5):

- Input tokens: Standard
- Reasoning tokens: Internal chain-of-thought (counted separately)
- Output tokens: Final response

Anthropic Claude Sonnet 4.5

****Status****: ☐ Verified Working (October 8, 2025)

CLAUDE_MODEL=claude-sonnet-4-5-20250929

ANTHROPIC_API_VERSION=2023-06-01

****API Details****:

- ****Endpoint****: POST <https://api.anthropic.com/v1/messages>
- ****Model ID****: claude-sonnet-4-5-20250929
- ****Context Window****: 200K tokens (200,000 tokens standard, 1M with beta header)
- ****Headers****:
 - x-api-key: <API_KEY>
 - anthropic-version: 2023-06-01 (or 2025-10-01 for latest)
 - Content-Type: application/json

****Supported Parameters****:

{

"model": "claude-sonnet-4-5-20250929",

"messages": [...],

"max_tokens": 64000,

"temperature": 0.7, // [] Standard parameter

"top_p": 0.95, // [] Supported

"system": "...", // [] System prompt

"stop_sequences": [...], // [] Supported

}

*** VERIFICATION COMPLETE *** (October 8, 2023)

Models API confirms availability

curl https://api.openai.com/v1/models -H "Authorization: Bearer sk-20230929..."

Response: {"type":"model","id":"claude-sonnet-4-5-20250929","display_name":"Claude Sonnet 4.5"}

Messages API returns correct model

curl https://api.openai.com/v1/chat/completions -H "Authorization: Bearer sk-20230929..." -d '{

Response: {"model":"claude-sonnet-4-5-20250929",...}

*** Request Summary *** (October 8, 2023) - Google Gemini 1.5 Pro

Messages:

• Request 1: Initial system message

• Request 2: User input message

*** Request Details: 2023-10-08T14:30:00Z *** (UTC-07:00) [Local Time: 08:00]

• Request 1: Initial system message (2023-10-08T14:30:00Z)

• Request 2: User input message (2023-10-08T14:30:00Z) [Google Gemini]

• Request 3: Assistant response (2023-10-08T14:30:00Z) [Gemini Pro]

Google Gemini 1.5 Pro

*** Request *** (October 8, 2023)

2023-10-08T14:30:00Z [Google Gemini 1.5 Pro]

System

User

Assistant

User

System

User

Assistant

User

System

User

Assistant

User

System

User

Assistant

1. Introduction
2. Background
3. Methodology
4. Results
5. Discussion
6. Conclusion
7. References
8. Appendix
9. Glossary
10. Index

1.1

1.2

1.3

1.4

1.5

1.6

1.7

1.8

1.9

1.10

1.11

1.12

1.13

1.14

1.15

1.16

1.17

1.18

1.19

1.20

1.21

1.22

1.23

1.24

1.25

1.26

1.27

1.28

1.29

1.30

1.31

1.32

1.33

1.34

1.35

1.36

1.37

1.38

1.39

1.40

1.41

1.42

1.43

1.44

1.45

1.46

1.47

1.48

1.49

1.50

1.51

1.52

1.53

1.54

1.55

1.56

1.57

1.58

1.59

1.60

1.61

1.62

1.63

1.64

1.65

1.66

1.67

1.68

1.69

1.70

Section 1

Section 2

Page 3: Section Header 1

Section 3

Section 4

Section 5

Section 6

Section 7

Section 8

Section 9

Section 10

Section 11

Section 12

Section 13

Section 14

Section 15

Section 16

Section 17

Section 18

Section 19

Section 20

Section 21

Section 22

Section 23

Section 24

Section 25

Section 26

Section 27

Section 28

Section 29

Section 30

Section 31

Section 32

Section 33

Section 34

Section 35

Section 36

Section 37

Section 38

Section 39

Section 40







