==========

LLM SPEED BENCHMARK REPORT

==========

Timestamp: 2025-12-12T10:15:39.875Z

Total Models Available: 346

Models Tested: 346

Successful: 309

Failed: 37

RANKED RESULTS (Fastest to Slowest by Median Latency)

---

1. OpenAI: gpt-oss-safeguard-20b (openai/gpt-oss-safeguard-20b)

   Provider: openrouter

   Status: ✅

   Median Latency: 344ms

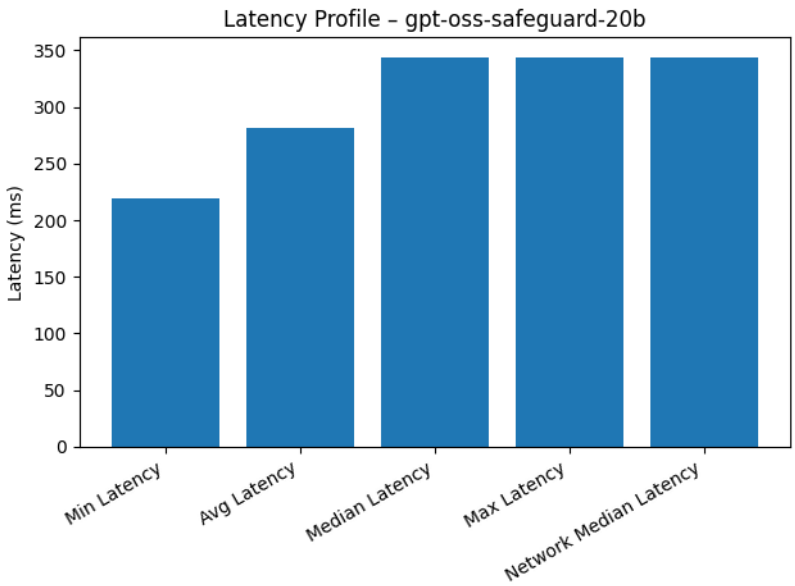
   Avg Latency: 281.5ms

   Min: 219ms | Max: 344ms

   Network Latency (median): 344.2ms

   Success Rate: 2/2

   Response Length: 44 chars



2. Qwen: Qwen-Plus (qwen/qwen-plus)

Provider: openrouter

Status: ✅

Median Latency: 351ms

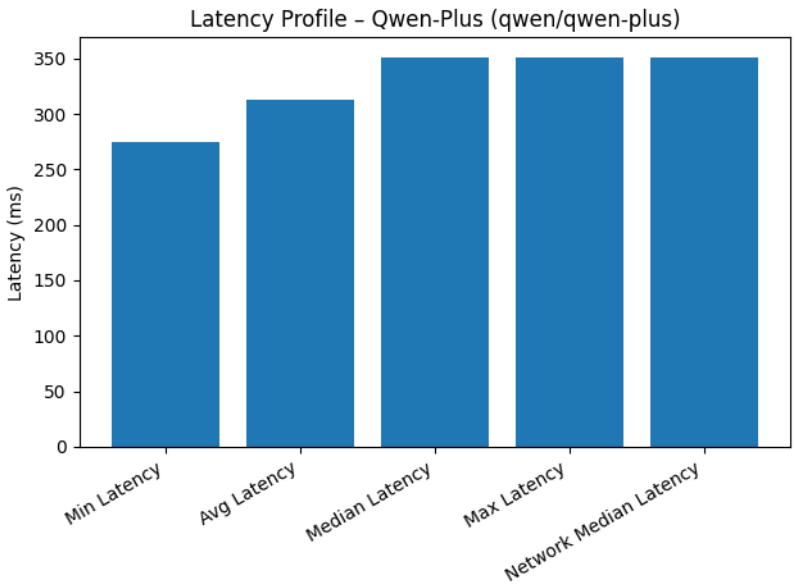
Avg Latency: 313.0ms

Min: 275ms | Max: 351ms

Network Latency (median): 351.4ms

Success Rate: 2/2

Response Length: 32 chars



3. Qwen: Qwen-Turbo (qwen/qwen-turbo)

Provider: openrouter

Status: ✅

Median Latency: 429ms

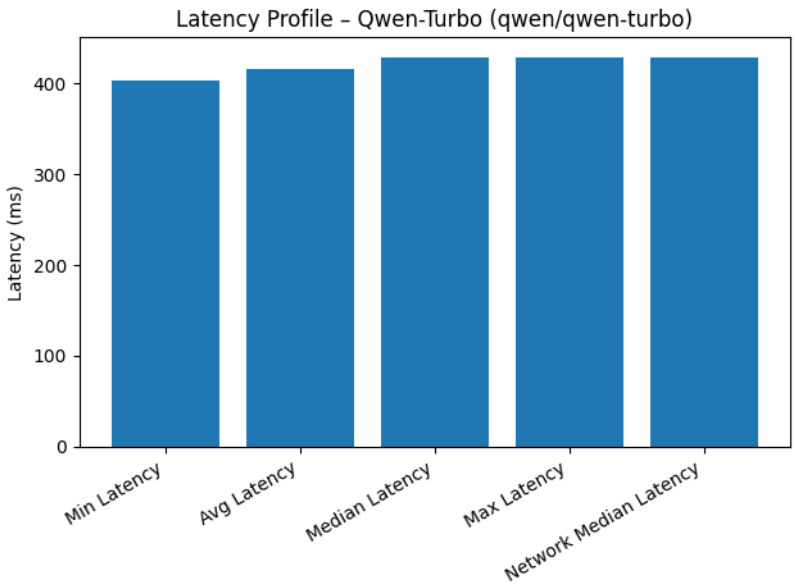
Avg Latency: 416.0ms

Min: 403ms | Max: 429ms

Network Latency (median): 428.7ms

Success Rate: 2/2

Response Length: 34 chars



4. Qwen: Qwen VL Plus (qwen/qwen-vl-plus)

Provider: openrouter

Status: ✅

Median Latency: 431ms

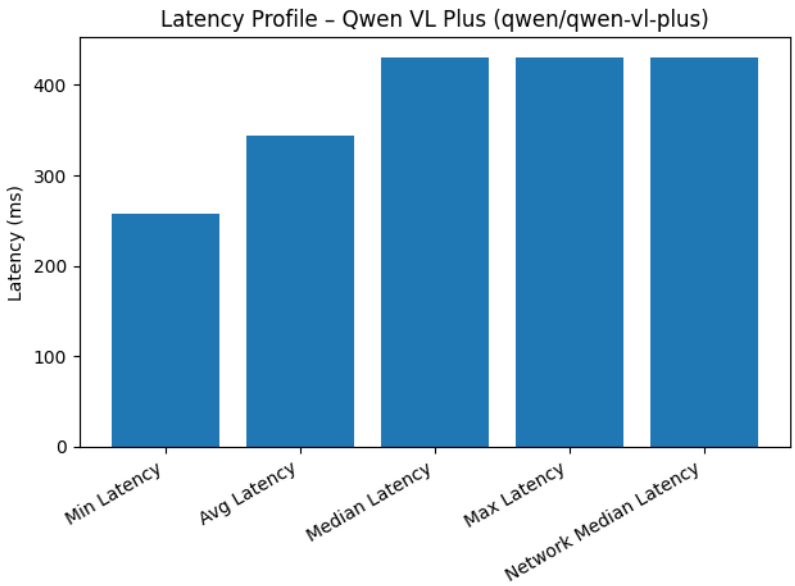
Avg Latency: 344.5ms

Min: 258ms | Max: 431ms

Network Latency (median): 430.7ms

Success Rate: 2/2

Response Length: 6 chars



5. Qwen: Qwen Plus 0728 (qwen/qwen-plus-2025-07-28)

Provider: openrouter

Status: ✅

Median Latency: 453ms

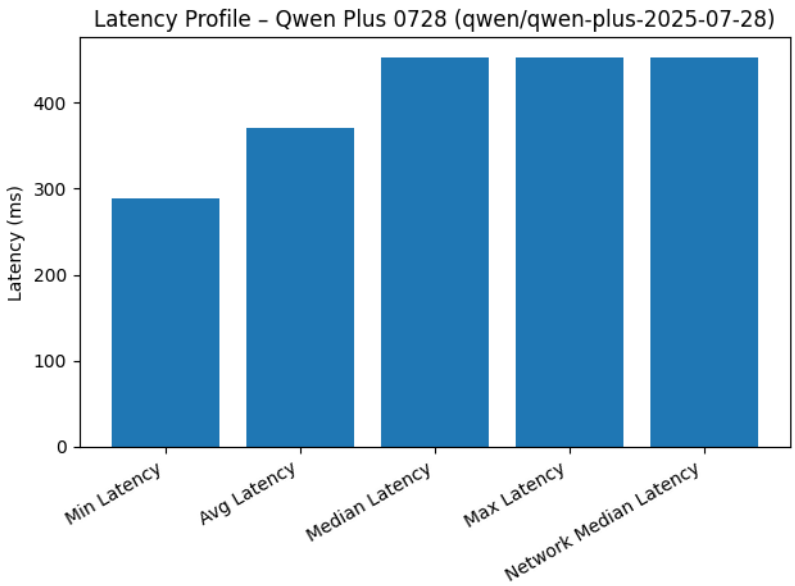
Avg Latency: 370.5ms

Min: 288ms | Max: 453ms

Network Latency (median): 452.6ms

Success Rate: 2/2

Response Length: 32 chars



6. Mistral: Codestral 2508 (mistralai/codestral-2508)

Provider: openrouter

Status: ✅

Median Latency: 472ms

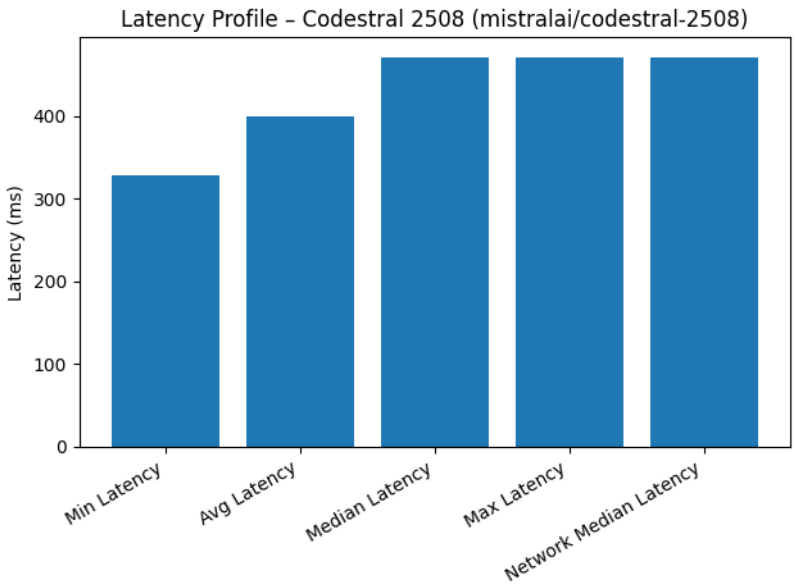
Avg Latency: 400.5ms

Min: 329ms | Max: 472ms

Network Latency (median): 471.9ms

Success Rate: 2/2

Response Length: 33 chars



7. Qwen: Qwen-Max (qwen/qwen-max)

Provider: openrouter

Status: ✅

Median Latency: 507ms

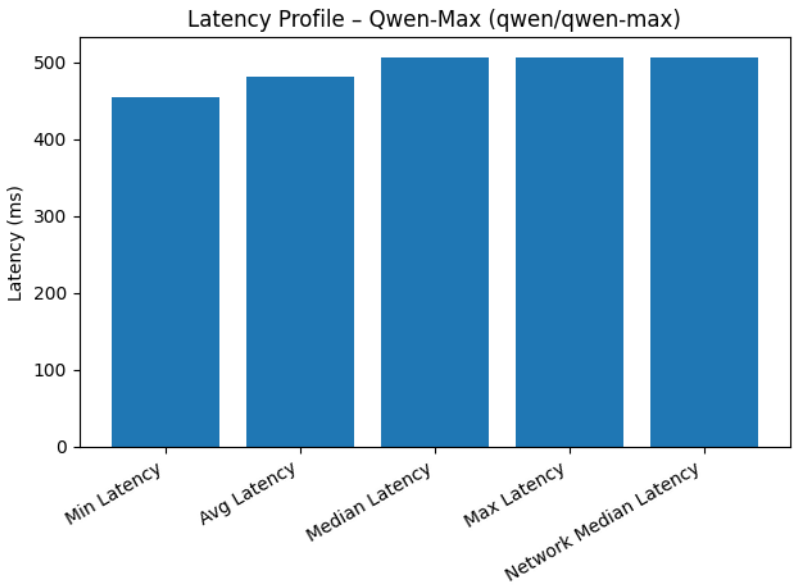
Avg Latency: 480.5ms

Min: 454ms | Max: 507ms

Network Latency (median): 506.7ms

Success Rate: 2/2

Response Length: 34 chars



8. Mistral Large 2411 (mistralai/mistral-large-2411)

Provider: openrouter

Status: ✅

Median Latency: 508ms

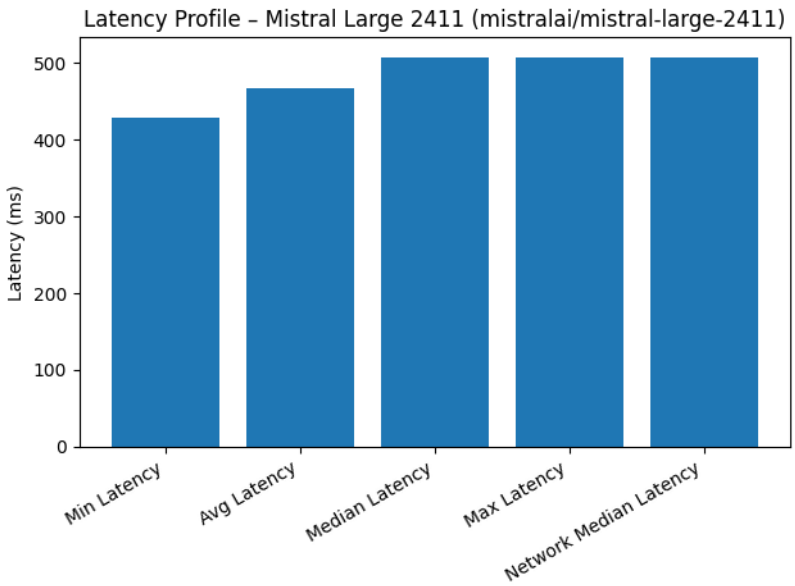
Avg Latency: 468.0ms

Min: 428ms | Max: 508ms

Network Latency (median): 507.7ms

Success Rate: 2/2

Response Length: 6 chars



9. TheDrummer: Skyfall 36B V2 (thedrummer/skyfall-36b-v2)

Provider: openrouter

Status: ✅

Median Latency: 510ms

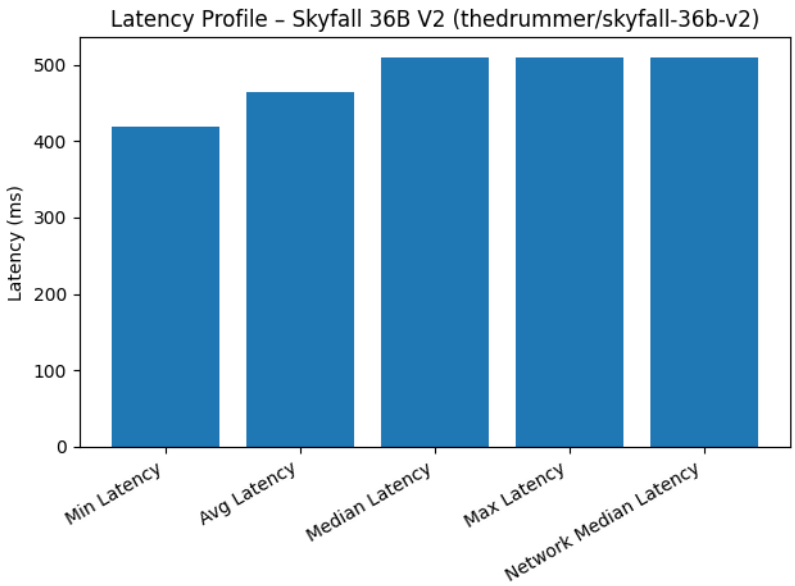
Avg Latency: 464.5ms

Min: 419ms | Max: 510ms

Network Latency (median): 510.0ms

Success Rate: 2/2

Response Length: 6 chars



10. Mistral: Ministral 3 8B 2512 (mistralai/ministral-8b-2512)

Provider: openrouter

Status: ✅

Median Latency: 554ms

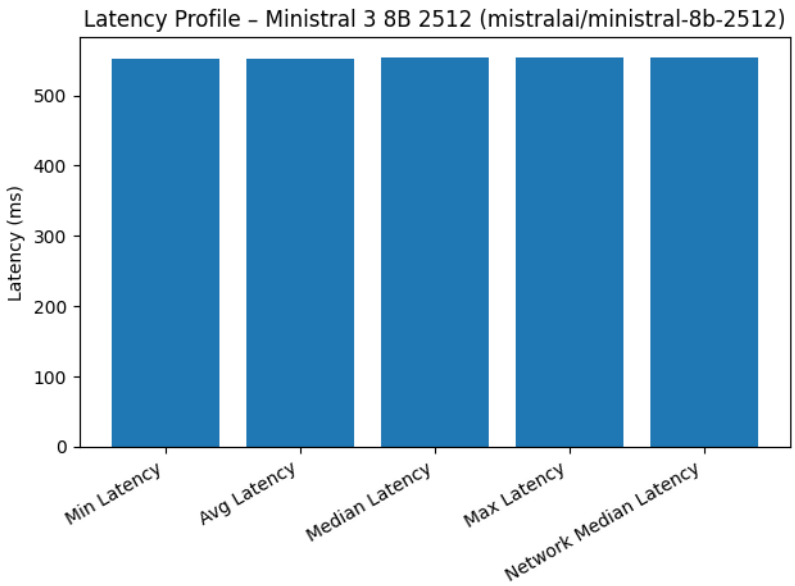
Avg Latency: 553.0ms

Min: 552ms | Max: 554ms

Network Latency (median): 554.8ms

Success Rate: 2/2

Response Length: 9 chars



11. Mistral: Ministral 3 3B 2512 (mistralai/ministral-3b-2512)

Provider: openrouter

Status: ✅

Median Latency: 556ms

Avg Latency: 549.0ms

Min: 542ms | Max: 556ms

Network Latency (median): 555.8ms

Success Rate: 2/2

Response Length: 45 chars

12. OpenAI: GPT-4 Turbo Preview (openai/gpt-4-turbo-preview)

Provider: openrouter

Status: ✅

Median Latency: 578ms

Avg Latency: 553.0ms

Min: 528ms | Max: 578ms

Network Latency (median): 577.8ms

Success Rate: 2/2

Response Length: 6 chars

13. Meta: Llama 4 Maverick (meta-llama/llama-4-maverick)

Provider: openrouter

Status: ✅

Median Latency: 582ms

Avg Latency: 500.5ms

Min: 419ms | Max: 582ms

Network Latency (median): 581.5ms

Success Rate: 2/2

Response Length: 6 chars

14. Google: Gemma 3 4B (google/gemma-3-4b-it)

Provider: openrouter

Status: ✅

Median Latency: 583ms

Avg Latency: 541.0ms

Min: 499ms | Max: 583ms

Network Latency (median): 583.0ms

Success Rate: 2/2

Response Length: 15 chars

15. Cogito V2 Preview Llama 109B (deepcogito/cogito-v2-preview-llama-109b-moe)

Provider: openrouter

Status: ✅

Median Latency: 593ms

Avg Latency: 467.5ms

Min: 342ms | Max: 593ms

Network Latency (median): 592.6ms

Success Rate: 2/2

Response Length: 6 chars

16. Mistral: Pixtral 12B (mistralai/pixtral-12b)

Provider: openrouter

Status: ✅

Median Latency: 598ms

Avg Latency: 563.0ms

Min: 528ms | Max: 598ms

Network Latency (median): 598.9ms

Success Rate: 2/2

Response Length: 25 chars

17. NVIDIA: Llama 3.1 Nemotron Ultra 253B v1 (nvidia/llama-3.1-nemotron-ultra-253b-v1)

Provider: openrouter

Status: ✅

Median Latency: 607ms

Avg Latency: 525.0ms

Min: 443ms | Max: 607ms

Network Latency (median): 606.7ms

Success Rate: 2/2

Response Length: 6 chars

18. Mistral: Ministral 3B (mistralai/ministral-3b)

Provider: openrouter

Status: ✅

Median Latency: 618ms

Avg Latency: 617.0ms

Min: 616ms | Max: 618ms

Network Latency (median): 617.7ms

Success Rate: 2/2

Response Length: 34 chars

19. Mistral: Ministral 8B (mistralai/ministral-8b)

Provider: openrouter

Status: ✅

Median Latency: 618ms

Avg Latency: 495.0ms

Min: 372ms | Max: 618ms

Network Latency (median): 618.4ms

Success Rate: 2/2

Response Length: 19 chars

20. Qwen: Qwen VL Max (qwen/qwen-vl-max)

Provider: openrouter

Status: ✅

Median Latency: 631ms

Avg Latency: 457.0ms

Min: 283ms | Max: 631ms

Network Latency (median): 630.2ms

Success Rate: 2/2

Response Length: 6 chars

21. Mistral: Mistral Medium 3 (mistralai/mistral-medium-3)

Provider: openrouter

Status: ✅

Median Latency: 639ms

Avg Latency: 555.5ms

Min: 472ms | Max: 639ms

Network Latency (median): 639.1ms

Success Rate: 2/2

Response Length: 34 chars

22. Inception: Mercury Coder (inception/mercury-coder)

Provider: openrouter

Status: ✅

Median Latency: 640ms

Avg Latency: 606.0ms

Min: 572ms | Max: 640ms

Network Latency (median): 639.4ms

Success Rate: 2/2

Response Length: 40 chars

23. Mistral: Mistral 7B Instruct v0.2 (mistralai/mistral-7b-instruct-v0.2)

Provider: openrouter

Status: ✅

Median Latency: 649ms

Avg Latency: 543.5ms

Min: 438ms | Max: 649ms

Network Latency (median): 648.7ms

Success Rate: 2/2

Response Length: 41 chars

24. Mistral: Mistral 7B Instruct v0.3 (mistralai/mistral-7b-instruct-v0.3)

Provider: openrouter

Status: ✅

Median Latency: 656ms

Avg Latency: 569.5ms

Min: 483ms | Max: 656ms

Network Latency (median): 655.6ms

Success Rate: 2/2

Response Length: 67 chars

25. Mistral: Devstral Medium (mistralai/devstral-medium)

Provider: openrouter

Status: ✅

Median Latency: 663ms

Avg Latency: 656.0ms

Min: 649ms | Max: 663ms

Network Latency (median): 662.7ms

Success Rate: 2/2

Response Length: 34 chars

26. Mistral Tiny (mistralai/mistral-tiny)

Provider: openrouter

Status: ✅

Median Latency: 665ms

Avg Latency: 643.5ms

Min: 622ms | Max: 665ms

Network Latency (median): 664.8ms

Success Rate: 2/2

Response Length: 40 chars

27. ReMM SLERP 13B (undi95/remm-slerp-l2-13b)

Provider: openrouter

Status: ✅

Median Latency: 681ms

Avg Latency: 609.5ms

Min: 538ms | Max: 681ms

Network Latency (median): 680.9ms

Success Rate: 2/2

Response Length: 7 chars

28. OpenAI: GPT-4o-mini (openai/gpt-4o-mini)

Provider: openrouter

Status: ✅

Median Latency: 687ms

Avg Latency: 659.5ms

Min: 632ms | Max: 687ms

Network Latency (median): 686.2ms

Success Rate: 2/2

Response Length: 34 chars

29. AllenAI: Olmo 3 7B Instruct (allenai/olmo-3-7b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 693ms

Avg Latency: 666.0ms

Min: 639ms | Max: 693ms

Network Latency (median): 693.6ms

Success Rate: 2/2

Response Length: 32 chars

30. Qwen: Qwen2.5-VL 7B Instruct (qwen/qwen-2.5-vl-7b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 716ms

Avg Latency: 647.0ms

Min: 578ms | Max: 716ms

Network Latency (median): 716.7ms

Success Rate: 2/2

Response Length: 34 chars

31. Z.AI: GLM 4.6 (exacto) (z-ai/glm-4.6:exacto)

Provider: openrouter

Status: ✅

Median Latency: 750ms

Avg Latency: 750.0ms

Min: 750ms | Max: 750ms

Network Latency (median): 749.1ms

Success Rate: 1/2

Response Length: 13 chars

32. OpenAI: GPT-4 (openai/gpt-4)

Provider: openrouter

Status: ✅

Median Latency: 766ms

Avg Latency: 722.0ms

Min: 678ms | Max: 766ms

Network Latency (median): 765.6ms

Success Rate: 2/2

Response Length: 31 chars

33. Mistral: Ministral 3 14B 2512 (mistralai/ministral-14b-2512)

Provider: openrouter

Status: ✅

Median Latency: 771ms

Avg Latency: 611.0ms

Min: 451ms | Max: 771ms

Network Latency (median): 771.2ms

Success Rate: 2/2

Response Length: 37 chars

34. Meta: Llama 3.1 70B Instruct (meta-llama/llama-3.1-70b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 781ms

Avg Latency: 773.0ms

Min: 765ms | Max: 781ms

Network Latency (median): 781.2ms

Success Rate: 2/2

Response Length: 29 chars

35. xAI: Grok 3 Beta (x-ai/grok-3-beta)

Provider: openrouter

Status: ✅

Median Latency: 782ms

Avg Latency: 746.5ms

Min: 711ms | Max: 782ms

Network Latency (median): 781.9ms

Success Rate: 2/2

Response Length: 25 chars

36. OpenAI: GPT-4o (2024-08-06) (openai/gpt-4o-2024-08-06)

Provider: openrouter

Status: ✅

Median Latency: 785ms

Avg Latency: 746.5ms

Min: 708ms | Max: 785ms

Network Latency (median): 785.0ms

Success Rate: 2/2

Response Length: 25 chars

37. OpenAI: GPT-4o (2024-11-20) (openai/gpt-4o-2024-11-20)

Provider: openrouter

Status: ✅

Median Latency: 786ms

Avg Latency: 674.5ms

Min: 563ms | Max: 786ms

Network Latency (median): 786.2ms

Success Rate: 2/2

Response Length: 40 chars

38. MythoMax 13B (gryphe/mythomax-l2-13b)

Provider: openrouter

Status: ✅

Median Latency: 803ms

Avg Latency: 633.0ms

Min: 463ms | Max: 803ms

Network Latency (median): 803.3ms

Success Rate: 2/2

Response Length: 13 chars

39. OpenAI: GPT-3.5 Turbo (openai/gpt-3.5-turbo)

Provider: openrouter

Status: ✅

Median Latency: 803ms

Avg Latency: 663.5ms

Min: 524ms | Max: 803ms

Network Latency (median): 803.4ms

Success Rate: 2/2

Response Length: 25 chars

40. OpenAI: GPT-3.5 Turbo Instruct (openai/gpt-3.5-turbo-instruct)

Provider: openrouter

Status: ✅

Median Latency: 804ms

Avg Latency: 698.0ms

Min: 592ms | Max: 804ms

Network Latency (median): 803.1ms

Success Rate: 2/2

Response Length: 12 chars

41. Meta: Llama 3.2 3B Instruct (meta-llama/llama-3.2-3b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 806ms

Avg Latency: 796.5ms

Min: 787ms | Max: 806ms

Network Latency (median): 806.2ms

Success Rate: 2/2

Response Length: 6 chars

42. Google: Gemini 2.5 Flash (google/gemini-2.5-flash)

Provider: openrouter

Status: ✅

Median Latency: 812ms

Avg Latency: 690.5ms

Min: 569ms | Max: 812ms

Network Latency (median): 812.1ms

Success Rate: 2/2

Response Length: 12 chars

43. Meta: Llama 3 8B Instruct (meta-llama/llama-3-8b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 816ms

Avg Latency: 807.0ms

Min: 798ms | Max: 816ms

Network Latency (median): 815.9ms

Success Rate: 2/2

Response Length: 6 chars

44. Qwen: Qwen2.5 VL 72B Instruct (qwen/qwen2.5-vl-72b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 817ms

Avg Latency: 674.5ms

Min: 532ms | Max: 817ms

Network Latency (median): 816.5ms

Success Rate: 2/2

Response Length: 12 chars

45. AI21: Jamba Large 1.7 (ai21/jamba-large-1.7)

Provider: openrouter

Status: ✅

Median Latency: 821ms

Avg Latency: 813.5ms

Min: 806ms | Max: 821ms

Network Latency (median): 821.1ms

Success Rate: 2/2

Response Length: 6 chars

46. OpenAI: GPT-4.1 Nano (openai/gpt-4.1-nano)

Provider: openrouter

Status: ✅

Median Latency: 825ms

Avg Latency: 682.0ms

Min: 539ms | Max: 825ms

Network Latency (median): 824.8ms

Success Rate: 2/2

Response Length: 6 chars

47. Z.AI: GLM 4.5 Air (z-ai/glm-4.5-air)

Provider: openrouter

Status: ✅

Median Latency: 830ms

Avg Latency: 792.5ms

Min: 755ms | Max: 830ms

Network Latency (median): 830.2ms

Success Rate: 2/2

Response Length: 6 chars

48. Qwen: Qwen3 Coder Flash (qwen/qwen3-coder-flash)

Provider: openrouter

Status: ✅

Median Latency: 839ms

Avg Latency: 775.0ms

Min: 711ms | Max: 839ms

Network Latency (median): 839.4ms

Success Rate: 2/2

Response Length: 6 chars

49. Sao10K: Llama 3.1 Euryale 70B v2.2 (sao10k/l3.1-euryale-70b)

Provider: openrouter

Status: ✅

Median Latency: 845ms

Avg Latency: 656.0ms

Min: 467ms | Max: 845ms

Network Latency (median): 844.6ms

Success Rate: 2/2

Response Length: 6 chars

50. LiquidAI/LFM2-2.6B (liquid/lfm-2.2-6b)

Provider: openrouter

Status: ✅

Median Latency: 850ms

Avg Latency: 615.5ms

Min: 381ms | Max: 850ms

Network Latency (median): 850.5ms

Success Rate: 2/2

Response Length: 6 chars

51. OpenAI: GPT-5.2 (openai/gpt-5.2)

Provider: openrouter

Status: ✅

Median Latency: 856ms

Avg Latency: 829.0ms

Min: 802ms | Max: 856ms

Network Latency (median): 856.1ms

Success Rate: 2/2

Response Length: 6 chars

52. Goliath 120B (alpindale/goliath-120b)

Provider: openrouter

Status: ✅

Median Latency: 861ms

Avg Latency: 840.5ms

Min: 820ms | Max: 861ms

Network Latency (median): 861.2ms

Success Rate: 2/2

Response Length: 7 chars

53. EssentialAI: Rnj 1 Instruct (essentialai/rnj-1-instruct)

Provider: openrouter

Status: ✅

Median Latency: 863ms

Avg Latency: 643.0ms

Min: 423ms | Max: 863ms

Network Latency (median): 862.9ms

Success Rate: 2/2

Response Length: 34 chars

54. Mistral: Voxtral Small 24B 2507 (mistralai/voxtral-small-24b-2507)

Provider: openrouter

Status: ✅

Median Latency: 872ms

Avg Latency: 602.0ms

Min: 332ms | Max: 872ms

Network Latency (median): 872.3ms

Success Rate: 2/2

Response Length: 6 chars

55. Qwen: Qwen3 VL 8B Instruct (qwen/qwen3-vl-8b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 874ms

Avg Latency: 643.0ms

Min: 412ms | Max: 874ms

Network Latency (median): 874.0ms

Success Rate: 2/2

Response Length: 51 chars

56. Mistral: Devstral 2 2512 (free) (mistralai/devstral-2512:free)

Provider: openrouter

Status: ✅

Median Latency: 881ms

Avg Latency: 873.5ms

Min: 866ms | Max: 881ms

Network Latency (median): 881.8ms

Success Rate: 2/2

Response Length: 6 chars

57. Sao10K: Llama 3 8B Lunaris (sao10k/l3-lunaris-8b)

Provider: openrouter

Status: ✅

Median Latency: 886ms

Avg Latency: 884.5ms

Min: 883ms | Max: 886ms

Network Latency (median): 885.7ms

Success Rate: 2/2

Response Length: 44 chars

58. OpenAI: GPT-3.5 Turbo 16k (openai/gpt-3.5-turbo-16k)

Provider: openrouter

Status: ✅

Median Latency: 895ms

Avg Latency: 711.0ms

Min: 527ms | Max: 895ms

Network Latency (median): 895.2ms

Success Rate: 2/2

Response Length: 31 chars

59. OpenAI: GPT-4o-mini (2024-07-18) (openai/gpt-4o-mini-2024-07-18)

Provider: openrouter

Status: ✅

Median Latency: 896ms

Avg Latency: 787.5ms

Min: 679ms | Max: 896ms

Network Latency (median): 895.3ms

Success Rate: 2/2

Response Length: 34 chars

60. OpenAI: GPT-4.1 Mini (openai/gpt-4.1-mini)

Provider: openrouter

Status: ✅

Median Latency: 897ms

Avg Latency: 862.0ms

Min: 827ms | Max: 897ms

Network Latency (median): 896.8ms

Success Rate: 2/2

Response Length: 34 chars

61. Google: Gemini 2.5 Flash Lite Preview 09-2025 (google/gemini-2.5-flash-lite-preview-09-2025)

Provider: openrouter

Status: ✅

Median Latency: 904ms

Avg Latency: 774.0ms

Min: 644ms | Max: 904ms

Network Latency (median): 903.4ms

Success Rate: 2/2

Response Length: 6 chars

62. Qwen: Qwen2.5 7B Instruct (qwen/qwen-2.5-7b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 904ms

Avg Latency: 696.0ms

Min: 488ms | Max: 904ms

Network Latency (median): 903.9ms

Success Rate: 2/2

Response Length: 40 chars

63. Llama Guard 3 8B (meta-llama/llama-guard-3-8b)

Provider: openrouter

Status: ✅

Median Latency: 906ms

Avg Latency: 906.0ms

Min: 906ms | Max: 906ms

Network Latency (median): 906.6ms

Success Rate: 1/2

Response Length: 0 chars

64. Mistral: Mistral Large 3 2512 (mistralai/mistral-large-2512)

Provider: openrouter

Status: ✅

Median Latency: 907ms

Avg Latency: 881.0ms

Min: 855ms | Max: 907ms

Network Latency (median): 907.4ms

Success Rate: 2/2

Response Length: 17 chars

65. Meta: Llama Guard 4 12B (meta-llama/llama-guard-4-12b)

Provider: openrouter

Status: ✅

Median Latency: 907ms

Avg Latency: 657.5ms

Min: 408ms | Max: 907ms

Network Latency (median): 907.2ms

Success Rate: 2/2

Response Length: 4 chars

66. Gemini Flash Latest (Direct) (gemini-flash-latest)

Provider: google

Status: ✅

Median Latency: 913ms

Avg Latency: 878.0ms

Min: 843ms | Max: 913ms

Network Latency (median): 913.1ms

Success Rate: 2/2

Response Length: 31 chars

67. LiquidAI/LFM2-8B-A1B (liquid/lfm2-8b-a1b)

Provider: openrouter

Status: ✅

Median Latency: 961ms

Avg Latency: 658.5ms

Min: 356ms | Max: 961ms

Network Latency (median): 960.7ms

Success Rate: 2/2

Response Length: 6 chars

68. OpenAI: GPT-4 Turbo (older v1106) (openai/gpt-4-1106-preview)

Provider: openrouter

Status: ✅

Median Latency: 981ms

Avg Latency: 914.0ms

Min: 847ms | Max: 981ms

Network Latency (median): 980.8ms

Success Rate: 2/2

Response Length: 35 chars

69. Prime Intellect: INTELLECT-3 (prime-intellect/intellect-3)

Provider: openrouter

Status: ✅

Median Latency: 983ms

Avg Latency: 850.0ms

Min: 717ms | Max: 983ms

Network Latency (median): 983.8ms

Success Rate: 2/2

Response Length: 34 chars

70. Sao10k: Llama 3 Euryale 70B v2.1 (sao10k/l3-euryale-70b)

Provider: openrouter

Status: ✅

Median Latency: 984ms

Avg Latency: 909.0ms

Min: 834ms | Max: 984ms

Network Latency (median): 984.4ms

Success Rate: 2/2

Response Length: 6 chars

71. OpenAI: GPT-4o (2024-05-13) (openai/gpt-4o-2024-05-13)

Provider: openrouter

Status: ✅

Median Latency: 984ms

Avg Latency: 858.5ms

Min: 733ms | Max: 984ms

Network Latency (median): 984.1ms

Success Rate: 2/2

Response Length: 34 chars

72. OpenAI: GPT-5 Chat (openai/gpt-5-chat)

Provider: openrouter

Status: ✅

Median Latency: 986ms

Avg Latency: 937.0ms

Min: 888ms | Max: 986ms

Network Latency (median): 986.6ms

Success Rate: 2/2

Response Length: 44 chars

73. Mistral: Mistral Small 3 (mistralai/mistral-small-24b-instruct-2501)

Provider: openrouter

Status: ✅

Median Latency: 989ms

Avg Latency: 846.5ms

Min: 704ms | Max: 989ms

Network Latency (median): 989.0ms

Success Rate: 2/2

Response Length: 6 chars

74. OpenAI: GPT-4 Turbo (openai/gpt-4-turbo)

Provider: openrouter

Status: ✅

Median Latency: 992ms

Avg Latency: 873.5ms

Min: 755ms | Max: 992ms

Network Latency (median): 992.5ms

Success Rate: 2/2

Response Length: 31 chars

75. AionLabs: Aion-RP 1.0 (8B) (aion-labs/aion-rp-llama-3.1-8b)

Provider: openrouter

Status: ✅

Median Latency: 994ms

Avg Latency: 994.0ms

Min: 994ms | Max: 994ms

Network Latency (median): 994.0ms

Success Rate: 1/2

Response Length: 53 chars

76. Meituan: LongCat Flash Chat (meituan/longcat-flash-chat)

Provider: openrouter

Status: ✅

Median Latency: 998ms

Avg Latency: 974.0ms

Min: 950ms | Max: 998ms

Network Latency (median): 998.7ms

Success Rate: 2/2

Response Length: 9 chars

77. Mistral: Mistral Small 3.2 24B (mistralai/mistral-small-3.2-24b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1011ms

Avg Latency: 950.0ms

Min: 889ms | Max: 1011ms

Network Latency (median): 1010.9ms

Success Rate: 2/2

Response Length: 27 chars

78. Kwaipilot: KAT-Coder-Pro V1 (free) (kwaipilot/kat-coder-pro:free)

Provider: openrouter

Status: ✅

Median Latency: 1017ms

Avg Latency: 846.0ms

Min: 675ms | Max: 1017ms

Network Latency (median): 1017.2ms

Success Rate: 2/2

Response Length: 6 chars

79. Anthropic: Claude 3 Haiku (anthropic/claude-3-haiku)

Provider: openrouter

Status: ✅

Median Latency: 1031ms

Avg Latency: 955.5ms

Min: 880ms | Max: 1031ms

Network Latency (median): 1031.3ms

Success Rate: 2/2

Response Length: 6 chars

80. Mistral: Mistral Small 3.1 24B (free) (mistralai/mistral-small-3.1-24b-instruct:free)

Provider: openrouter

Status: ✅

Median Latency: 1034ms

Avg Latency: 961.5ms

Min: 889ms | Max: 1034ms

Network Latency (median): 1034.1ms

Success Rate: 2/2

Response Length: 6 chars

81. OpenAI: GPT-4o (openai/gpt-4o)

Provider: openrouter

Status: ✅

Median Latency: 1042ms

Avg Latency: 846.5ms

Min: 651ms | Max: 1042ms

Network Latency (median): 1042.4ms

Success Rate: 2/2

Response Length: 34 chars

82. Inception: Mercury (inception/mercury)

Provider: openrouter

Status: ✅

Median Latency: 1048ms

Avg Latency: 744.5ms

Min: 441ms | Max: 1048ms

Network Latency (median): 1047.4ms

Success Rate: 2/2

Response Length: 6 chars

83. Google: Gemma 3n 2B (free) (google/gemma-3n-e2b-it:free)

Provider: openrouter

Status: ✅

Median Latency: 1063ms

Avg Latency: 1017.0ms

Min: 971ms | Max: 1063ms

Network Latency (median): 1062.5ms

Success Rate: 2/2

Response Length: 14 chars

84. xAI: Grok 3 (x-ai/grok-3)

Provider: openrouter

Status: ✅

Median Latency: 1066ms

Avg Latency: 972.5ms

Min: 879ms | Max: 1066ms

Network Latency (median): 1066.2ms

Success Rate: 2/2

Response Length: 29 chars

85. Google: Gemini 2.5 Flash Preview 09-2025 (google/gemini-2.5-flash-preview-09-2025)

Provider: openrouter

Status: ✅

Median Latency: 1068ms

Avg Latency: 950.5ms

Min: 833ms | Max: 1068ms

Network Latency (median): 1067.7ms

Success Rate: 2/2

Response Length: 6 chars

86. Baidu: ERNIE 4.5 VL 28B A3B (baidu/ernie-4.5-vl-28b-a3b)

Provider: openrouter

Status: ✅

Median Latency: 1072ms

Avg Latency: 859.5ms

Min: 647ms | Max: 1072ms

Network Latency (median): 1071.7ms

Success Rate: 2/2

Response Length: 6 chars

87. Nous: Hermes 4 405B (nousresearch/hermes-4-405b)

Provider: openrouter

Status: ✅

Median Latency: 1074ms

Avg Latency: 945.5ms

Min: 817ms | Max: 1074ms

Network Latency (median): 1074.3ms

Success Rate: 2/2

Response Length: 112 chars

88. WizardLM-2 8x22B (microsoft/wizardlm-2-8x22b)

Provider: openrouter

Status: ✅

Median Latency: 1090ms

Avg Latency: 840.0ms

Min: 590ms | Max: 1090ms

Network Latency (median): 1089.9ms

Success Rate: 2/2

Response Length: 40 chars

89. Mistral: Mistral 7B Instruct (mistralai/mistral-7b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1094ms

Avg Latency: 1084.5ms

Min: 1075ms | Max: 1094ms

Network Latency (median): 1094.1ms

Success Rate: 2/2

Response Length: 36 chars

90. Morph: Morph V3 Large (morph/morph-v3-large)

Provider: openrouter

Status: ✅

Median Latency: 1103ms

Avg Latency: 776.5ms

Min: 450ms | Max: 1103ms

Network Latency (median): 1103.8ms

Success Rate: 2/2

Response Length: 6 chars

91. Qwen: Qwen3 Coder 30B A3B Instruct (qwen/qwen3-coder-30b-a3b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1120ms

Avg Latency: 983.0ms

Min: 846ms | Max: 1120ms

Network Latency (median): 1120.7ms

Success Rate: 2/2

Response Length: 6 chars

92. Cohere: Command R7B (12-2024) (cohere/command-r7b-12-2024)

Provider: openrouter

Status: ✅

Median Latency: 1140ms

Avg Latency: 998.0ms

Min: 856ms | Max: 1140ms

Network Latency (median): 1139.7ms

Success Rate: 2/2

Response Length: 34 chars

93. Mistral Large 2407 (mistralai/mistral-large-2407)

Provider: openrouter

Status: ✅

Median Latency: 1147ms

Avg Latency: 868.5ms

Min: 590ms | Max: 1147ms

Network Latency (median): 1147.4ms

Success Rate: 2/2

Response Length: 11 chars

94. Morph: Morph V3 Fast (morph/morph-v3-fast)

Provider: openrouter

Status: ✅

Median Latency: 1150ms

Avg Latency: 824.0ms

Min: 498ms | Max: 1150ms

Network Latency (median): 1149.4ms

Success Rate: 2/2

Response Length: 26 chars

95. OpenAI: ChatGPT-4o (openai/chatgpt-4o-latest)

Provider: openrouter

Status: ✅

Median Latency: 1156ms

Avg Latency: 862.5ms

Min: 569ms | Max: 1156ms

Network Latency (median): 1155.9ms

Success Rate: 2/2

Response Length: 44 chars

96. Mistral: Mixtral 8x7B Instruct (mistralai/mixtral-8x7b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1164ms

Avg Latency: 1094.0ms

Min: 1024ms | Max: 1164ms

Network Latency (median): 1163.9ms

Success Rate: 2/2

Response Length: 42 chars

97. Meta: Llama 3 70B Instruct (meta-llama/llama-3-70b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1175ms

Avg Latency: 1133.0ms

Min: 1091ms | Max: 1175ms

Network Latency (median): 1175.4ms

Success Rate: 2/2

Response Length: 29 chars

98. Microsoft: Phi-3 Mini 128K Instruct (microsoft/phi-3-mini-128k-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1180ms

Avg Latency: 786.0ms

Min: 392ms | Max: 1180ms

Network Latency (median): 1180.5ms

Success Rate: 2/2

Response Length: 27 chars

99. OpenAI: GPT-5.2 Chat (openai/gpt-5.2-chat)

Provider: openrouter

Status: ✅

Median Latency: 1199ms

Avg Latency: 965.5ms

Min: 732ms | Max: 1199ms

Network Latency (median): 1199.0ms

Success Rate: 2/2

Response Length: 9 chars

100. Magnum v4 72B (anthracite-org/magnum-v4-72b)

Provider: openrouter

Status: ✅

Median Latency: 1199ms

Avg Latency: 1039.0ms

Min: 879ms | Max: 1199ms

Network Latency (median): 1199.1ms

Success Rate: 2/2

Response Length: 6 chars

101. Mistral: Mistral Nemo (mistralai/mistral-nemo)

Provider: openrouter

Status: ✅

Median Latency: 1199ms

Avg Latency: 841.0ms

Min: 483ms | Max: 1199ms

Network Latency (median): 1199.4ms

Success Rate: 2/2

Response Length: 6 chars

102. Meta: Llama 3.1 8B Instruct (meta-llama/llama-3.1-8b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1202ms

Avg Latency: 1018.5ms

Min: 835ms | Max: 1202ms

Network Latency (median): 1201.9ms

Success Rate: 2/2

Response Length: 34 chars

103. Cohere: Command A (cohere/command-a)

Provider: openrouter

Status: ✅

Median Latency: 1216ms

Avg Latency: 1088.5ms

Min: 961ms | Max: 1216ms

Network Latency (median): 1215.8ms

Success Rate: 2/2

Response Length: 34 chars

104. MoonshotAI: Kimi K2 0905 (exacto) (moonshotai/kimi-k2-0905:exacto)

Provider: openrouter

Status: ✅

Median Latency: 1217ms

Avg Latency: 961.5ms

Min: 706ms | Max: 1217ms

Network Latency (median): 1216.6ms

Success Rate: 2/2

Response Length: 6 chars

105. Gemini 2.5 Flash (Direct) (gemini-2.5-flash)

Provider: google

Status: ✅

Median Latency: 1228ms

Avg Latency: 1155.0ms

Min: 1082ms | Max: 1228ms

Network Latency (median): 1228.3ms

Success Rate: 2/2

Response Length: 12 chars

106. Google: Gemma 3 12B (google/gemma-3-12b-it)

Provider: openrouter

Status: ✅

Median Latency: 1241ms

Avg Latency: 1181.0ms

Min: 1121ms | Max: 1241ms

Network Latency (median): 1241.1ms

Success Rate: 2/2

Response Length: 54 chars

107. Qwen2.5 72B Instruct (qwen/qwen-2.5-72b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1242ms

Avg Latency: 1110.0ms

Min: 978ms | Max: 1242ms

Network Latency (median): 1241.2ms

Success Rate: 2/2

Response Length: 6 chars

108. Google: Gemma 3n 4B (google/gemma-3n-e4b-it)

Provider: openrouter

Status: ✅

Median Latency: 1244ms

Avg Latency: 981.5ms

Min: 719ms | Max: 1244ms

Network Latency (median): 1243.8ms

Success Rate: 2/2

Response Length: 50 chars

109. NVIDIA: Llama 3.1 Nemotron 70B Instruct (nvidia/llama-3.1-nemotron-70b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1245ms

Avg Latency: 1153.0ms

Min: 1061ms | Max: 1245ms

Network Latency (median): 1244.4ms

Success Rate: 2/2

Response Length: 6 chars

110. OpenAI: GPT-5.1 Chat (openai/gpt-5.1-chat)

Provider: openrouter

Status: ✅

Median Latency: 1247ms

Avg Latency: 1178.5ms

Min: 1110ms | Max: 1247ms

Network Latency (median): 1247.1ms

Success Rate: 2/2

Response Length: 6 chars

111. Deep Cogito: Cogito v2.1 671B (deepcogito/cogito-v2.1-671b)

Provider: openrouter

Status: ✅

Median Latency: 1300ms

Avg Latency: 854.5ms

Min: 409ms | Max: 1300ms

Network Latency (median): 1299.9ms

Success Rate: 2/2

Response Length: 6 chars

112. Google: Gemma 3n 4B (free) (google/gemma-3n-e4b-it:free)

Provider: openrouter

Status: ✅

Median Latency: 1308ms

Avg Latency: 1285.5ms

Min: 1263ms | Max: 1308ms

Network Latency (median): 1307.4ms

Success Rate: 2/2

Response Length: 14 chars

113. Microsoft: Phi 4 Multimodal Instruct (microsoft/phi-4-multimodal-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1312ms

Avg Latency: 1217.5ms

Min: 1123ms | Max: 1312ms

Network Latency (median): 1311.2ms

Success Rate: 2/2

Response Length: 48 chars

114. Amazon: Nova Pro 1.0 (amazon/nova-pro-v1)

Provider: openrouter

Status: ✅

Median Latency: 1312ms

Avg Latency: 1230.0ms

Min: 1148ms | Max: 1312ms

Network Latency (median): 1311.3ms

Success Rate: 2/2

Response Length: 49 chars

115. ByteDance: UI-TARS 7B (bytedance/ui-tars-1.5-7b)

Provider: openrouter

Status: ✅

Median Latency: 1321ms

Avg Latency: 1162.0ms

Min: 1003ms | Max: 1321ms

Network Latency (median): 1321.1ms

Success Rate: 2/2

Response Length: 32 chars

116. Mistral: Devstral Small 1.1 (mistralai/devstral-small)

Provider: openrouter

Status: ✅

Median Latency: 1339ms

Avg Latency: 1194.5ms

Min: 1050ms | Max: 1339ms

Network Latency (median): 1339.5ms

Success Rate: 2/2

Response Length: 40 chars

117. DeepSeek: DeepSeek V3.1 Terminus (deepseek/deepseek-v3.1-terminus)

Provider: openrouter

Status: ✅

Median Latency: 1340ms

Avg Latency: 1266.0ms

Min: 1192ms | Max: 1340ms

Network Latency (median): 1339.8ms

Success Rate: 2/2

Response Length: 6 chars

118. Cohere: Command R (08-2024) (cohere/command-r-08-2024)

Provider: openrouter

Status: ✅

Median Latency: 1351ms

Avg Latency: 1174.5ms

Min: 998ms | Max: 1351ms

Network Latency (median): 1351.4ms

Success Rate: 2/2

Response Length: 138 chars

119. DeepSeek: DeepSeek V3.1 Terminus (exacto) (deepseek/deepseek-v3.1-terminus:exacto)

Provider: openrouter

Status: ✅

Median Latency: 1357ms

Avg Latency: 1137.0ms

Min: 917ms | Max: 1357ms

Network Latency (median): 1357.4ms

Success Rate: 2/2

Response Length: 6 chars

120. Amazon: Nova Micro 1.0 (amazon/nova-micro-v1)

Provider: openrouter

Status: ✅

Median Latency: 1363ms

Avg Latency: 1131.5ms

Min: 900ms | Max: 1363ms

Network Latency (median): 1363.4ms

Success Rate: 2/2

Response Length: 31 chars

121. Perplexity: Sonar (perplexity/sonar)

Provider: openrouter

Status: ✅

Median Latency: 1383ms

Avg Latency: 1351.0ms

Min: 1319ms | Max: 1383ms

Network Latency (median): 1382.5ms

Success Rate: 2/2

Response Length: 37 chars

122. Mistral: Pixtral Large 2411 (mistralai/pixtral-large-2411)

Provider: openrouter

Status: ✅

Median Latency: 1383ms

Avg Latency: 1041.0ms

Min: 699ms | Max: 1383ms

Network Latency (median): 1382.3ms

Success Rate: 2/2

Response Length: 34 chars

123. Baidu: ERNIE 4.5 300B A47B (baidu/ernie-4.5-300b-a47b)

Provider: openrouter

Status: ✅

Median Latency: 1399ms

Avg Latency: 1200.0ms

Min: 1001ms | Max: 1399ms

Network Latency (median): 1398.9ms

Success Rate: 2/2

Response Length: 6 chars

124. Baidu: ERNIE 4.5 21B A3B (baidu/ernie-4.5-21b-a3b)

Provider: openrouter

Status: ✅

Median Latency: 1403ms

Avg Latency: 1373.5ms

Min: 1344ms | Max: 1403ms

Network Latency (median): 1403.0ms

Success Rate: 2/2

Response Length: 34 chars

125. DeepSeek: DeepSeek V3 0324 (deepseek/deepseek-chat-v3-0324)

Provider: openrouter

Status: ✅

Median Latency: 1418ms

Avg Latency: 1154.0ms

Min: 890ms | Max: 1418ms

Network Latency (median): 1418.3ms

Success Rate: 2/2

Response Length: 11 chars

126. OpenAI: GPT-5.1-Codex-Mini (openai/gpt-5.1-codex-mini)

Provider: openrouter

Status: ✅

Median Latency: 1439ms

Avg Latency: 1400.0ms

Min: 1361ms | Max: 1439ms

Network Latency (median): 1439.0ms

Success Rate: 2/2

Response Length: 12 chars

127. Deep Cogito: Cogito V2 Preview Llama 70B (deepcogito/cogito-v2-preview-llama-70b)

Provider: openrouter

Status: ✅

Median Latency: 1440ms

Avg Latency: 1060.0ms

Min: 680ms | Max: 1440ms

Network Latency (median): 1439.6ms

Success Rate: 2/2

Response Length: 6 chars

128. AI21: Jamba Mini 1.7 (ai21/jamba-mini-1.7)

Provider: openrouter

Status: ✅

Median Latency: 1443ms

Avg Latency: 1166.5ms

Min: 890ms | Max: 1443ms

Network Latency (median): 1443.5ms

Success Rate: 2/2

Response Length: 61 chars

129. Nous: Hermes 4 70B (nousresearch/hermes-4-70b)

Provider: openrouter

Status: ✅

Median Latency: 1463ms

Avg Latency: 1080.5ms

Min: 698ms | Max: 1463ms

Network Latency (median): 1462.8ms

Success Rate: 2/2

Response Length: 119 chars

130. Meta: Llama 3.3 70B Instruct (meta-llama/llama-3.3-70b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1487ms

Avg Latency: 1358.5ms

Min: 1230ms | Max: 1487ms

Network Latency (median): 1487.5ms

Success Rate: 2/2

Response Length: 74 chars

131. Anthropic: Claude 3.5 Haiku (anthropic/claude-3.5-haiku)

Provider: openrouter

Status: ✅

Median Latency: 1491ms

Avg Latency: 1384.0ms

Min: 1277ms | Max: 1491ms

Network Latency (median): 1490.8ms

Success Rate: 2/2

Response Length: 12 chars

132. Baidu: ERNIE 4.5 VL 424B A47B (baidu/ernie-4.5-vl-424b-a47b)

Provider: openrouter

Status: ✅

Median Latency: 1505ms

Avg Latency: 1330.5ms

Min: 1156ms | Max: 1505ms

Network Latency (median): 1505.0ms

Success Rate: 2/2

Response Length: 6 chars

133. Mistral: Mistral 7B Instruct (free) (mistralai/mistral-7b-instruct:free)

Provider: openrouter

Status: ✅

Median Latency: 1524ms

Avg Latency: 1456.0ms

Min: 1388ms | Max: 1524ms

Network Latency (median): 1524.3ms

Success Rate: 2/2

Response Length: 16 chars

134. Amazon: Nova Lite 1.0 (amazon/nova-lite-v1)

Provider: openrouter

Status: ✅

Median Latency: 1529ms

Avg Latency: 1255.5ms

Min: 982ms | Max: 1529ms

Network Latency (median): 1528.8ms

Success Rate: 2/2

Response Length: 37 chars

135. Noromaid 20B (neversleep/noromaid-20b)

Provider: openrouter

Status: ✅

Median Latency: 1551ms

Avg Latency: 1387.0ms

Min: 1223ms | Max: 1551ms

Network Latency (median): 1551.1ms

Success Rate: 2/2

Response Length: 45 chars

136. OpenAI: gpt-oss-20b (free) (openai/gpt-oss-20b:free)

Provider: openrouter

Status: ✅

Median Latency: 1575ms

Avg Latency: 1314.5ms

Min: 1054ms | Max: 1575ms

Network Latency (median): 1574.4ms

Success Rate: 2/2

Response Length: 31 chars

137. OpenAI: GPT-3.5 Turbo (older v0613) (openai/gpt-3.5-turbo-0613)

Provider: openrouter

Status: ✅

Median Latency: 1577ms

Avg Latency: 1091.5ms

Min: 606ms | Max: 1577ms

Network Latency (median): 1577.3ms

Success Rate: 2/2

Response Length: 34 chars

138. MoonshotAI: Kimi K2 0905 (moonshotai/kimi-k2-0905)

Provider: openrouter

Status: ✅

Median Latency: 1580ms

Avg Latency: 1566.0ms

Min: 1552ms | Max: 1580ms

Network Latency (median): 1579.6ms

Success Rate: 2/2

Response Length: 23 chars

139. Qwen: Qwen3 VL 235B A22B Instruct (qwen/qwen3-vl-235b-a22b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1590ms

Avg Latency: 1444.5ms

Min: 1299ms | Max: 1590ms

Network Latency (median): 1589.4ms

Success Rate: 2/2

Response Length: 51 chars

140. Qwen: Qwen3 Max (qwen/qwen3-max)

Provider: openrouter

Status: ✅

Median Latency: 1592ms

Avg Latency: 1321.0ms

Min: 1050ms | Max: 1592ms

Network Latency (median): 1592.4ms

Success Rate: 2/2

Response Length: 6 chars

141. SorcererLM 8x22B (raifle/sorcererlm-8x22b)

Provider: openrouter

Status: ✅

Median Latency: 1603ms

Avg Latency: 1517.5ms

Min: 1432ms | Max: 1603ms

Network Latency (median): 1603.5ms

Success Rate: 2/2

Response Length: 35 chars

142. Z.AI: GLM 4 32B (z-ai/glm-4-32b)

Provider: openrouter

Status: ✅

Median Latency: 1608ms

Avg Latency: 1357.0ms

Min: 1106ms | Max: 1608ms

Network Latency (median): 1607.8ms

Success Rate: 2/2

Response Length: 6 chars

143. Google: Gemma 3 4B (free) (google/gemma-3-4b-it:free)

Provider: openrouter

Status: ✅

Median Latency: 1609ms

Avg Latency: 1499.5ms

Min: 1390ms | Max: 1609ms

Network Latency (median): 1608.7ms

Success Rate: 2/2

Response Length: 15 chars

144. DeepSeek: DeepSeek Prover V2 (deepseek/deepseek-prover-v2)

Provider: openrouter

Status: ✅

Median Latency: 1610ms

Avg Latency: 1293.0ms

Min: 976ms | Max: 1610ms

Network Latency (median): 1610.2ms

Success Rate: 2/2

Response Length: 11 chars

145. xAI: Grok 4 Fast (x-ai/grok-4-fast)

Provider: openrouter

Status: ✅

Median Latency: 1624ms

Avg Latency: 1393.5ms

Min: 1163ms | Max: 1624ms

Network Latency (median): 1624.4ms

Success Rate: 2/2

Response Length: 6 chars

146. Microsoft: Phi 4 Reasoning Plus (microsoft/phi-4-reasoning-plus)

Provider: openrouter

Status: ✅

Median Latency: 1648ms

Avg Latency: 1555.5ms

Min: 1463ms | Max: 1648ms

Network Latency (median): 1647.8ms

Success Rate: 2/2

Response Length: 34 chars

147. OpenAI: GPT-4o-mini Search Preview (openai/gpt-4o-mini-search-preview)

Provider: openrouter

Status: ✅

Median Latency: 1654ms

Avg Latency: 1547.5ms

Min: 1441ms | Max: 1654ms

Network Latency (median): 1654.1ms

Success Rate: 2/2

Response Length: 193 chars

148. Anthropic: Claude Haiku 4.5 (anthropic/claude-haiku-4.5)

Provider: openrouter

Status: ✅

Median Latency: 1656ms

Avg Latency: 1521.0ms

Min: 1386ms | Max: 1656ms

Network Latency (median): 1656.9ms

Success Rate: 2/2

Response Length: 52 chars

149. OpenAI: GPT-5 Codex (openai/gpt-5-codex)

Provider: openrouter

Status: ✅

Median Latency: 1657ms

Avg Latency: 1509.5ms

Min: 1362ms | Max: 1657ms

Network Latency (median): 1656.6ms

Success Rate: 2/2

Response Length: 6 chars

150. OpenAI: GPT-4o (extended) (openai/gpt-4o:extended)

Provider: openrouter

Status: ✅

Median Latency: 1712ms

Avg Latency: 1699.0ms

Min: 1686ms | Max: 1712ms

Network Latency (median): 1712.7ms

Success Rate: 2/2

Response Length: 34 chars

151. Anthropic: Claude 3 Opus (anthropic/claude-3-opus)

Provider: openrouter

Status: ✅

Median Latency: 1723ms

Avg Latency: 1665.0ms

Min: 1607ms | Max: 1723ms

Network Latency (median): 1723.0ms

Success Rate: 2/2

Response Length: 6 chars

152. Mistral: Devstral Small 2505 (mistralai/devstral-small-2505)

Provider: openrouter

Status: ✅

Median Latency: 1726ms

Avg Latency: 1506.0ms

Min: 1286ms | Max: 1726ms

Network Latency (median): 1725.9ms

Success Rate: 2/2

Response Length: 6 chars

153. Venice: Uncensored (free) (cognitivecomputations/dolphin-mistral-24b-venice-edition:free)

Provider: openrouter

Status: ✅

Median Latency: 1735ms

Avg Latency: 1456.0ms

Min: 1177ms | Max: 1735ms

Network Latency (median): 1735.0ms

Success Rate: 2/2

Response Length: 26 chars

154. DeepSeek: DeepSeek V3 (deepseek/deepseek-chat)

Provider: openrouter

Status: ✅

Median Latency: 1738ms

Avg Latency: 1090.0ms

Min: 442ms | Max: 1738ms

Network Latency (median): 1738.1ms

Success Rate: 2/2

Response Length: 6 chars

155. Mistral: Mistral Medium 3.1 (mistralai/mistral-medium-3.1)

Provider: openrouter

Status: ✅

Median Latency: 1741ms

Avg Latency: 1327.5ms

Min: 914ms | Max: 1741ms

Network Latency (median): 1741.7ms

Success Rate: 2/2

Response Length: 38 chars

156. OpenAI: GPT-5.1-Codex (openai/gpt-5.1-codex)

Provider: openrouter

Status: ✅

Median Latency: 1749ms

Avg Latency: 1365.5ms

Min: 982ms | Max: 1749ms

Network Latency (median): 1749.8ms

Success Rate: 2/2

Response Length: 12 chars

157. Qwen: Qwen3 Next 80B A3B Instruct (qwen/qwen3-next-80b-a3b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1749ms

Avg Latency: 1678.0ms

Min: 1607ms | Max: 1749ms

Network Latency (median): 1748.7ms

Success Rate: 2/2

Response Length: 9 chars

158. Google: Gemini 2.5 Flash Lite (google/gemini-2.5-flash-lite)

Provider: openrouter

Status: ✅

Median Latency: 1764ms

Avg Latency: 1205.5ms

Min: 647ms | Max: 1764ms

Network Latency (median): 1764.4ms

Success Rate: 2/2

Response Length: 6 chars

159. Google: Gemma 3 27B (free) (google/gemma-3-27b-it:free)

Provider: openrouter

Status: ✅

Median Latency: 1764ms

Avg Latency: 1719.0ms

Min: 1674ms | Max: 1764ms

Network Latency (median): 1764.5ms

Success Rate: 2/2

Response Length: 59 chars

160. Deep Cogito: Cogito V2 Preview Llama 405B (deepcogito/cogito-v2-preview-llama-405b)

Provider: openrouter

Status: ✅

Median Latency: 1824ms

Avg Latency: 1451.0ms

Min: 1078ms | Max: 1824ms

Network Latency (median): 1823.6ms

Success Rate: 2/2

Response Length: 25 chars

161. EleutherAI: Llemma 7b (eleutherai/llemma_7b)

Provider: openrouter

Status: ✅

Median Latency: 1826ms

Avg Latency: 1826.0ms

Min: 1826ms | Max: 1826ms

Network Latency (median): 1825.8ms

Success Rate: 1/2

Response Length: 11 chars

162. Gemini 2.0 Flash (Direct) (gemini-2.0-flash)

Provider: google

Status: ✅

Median Latency: 1837ms

Avg Latency: 1495.0ms

Min: 1153ms | Max: 1837ms

Network Latency (median): 1836.9ms

Success Rate: 2/2

Response Length: 47 chars

163. Anthropic: Claude 3.5 Sonnet (anthropic/claude-3.5-sonnet)

Provider: openrouter

Status: ✅

Median Latency: 1838ms

Avg Latency: 1819.5ms

Min: 1801ms | Max: 1838ms

Network Latency (median): 1838.9ms

Success Rate: 2/2

Response Length: 12 chars

164. NousResearch: Hermes 2 Pro - Llama-3 8B (nousresearch/hermes-2-pro-llama-3-8b)

Provider: openrouter

Status: ✅

Median Latency: 1848ms

Avg Latency: 1832.5ms

Min: 1817ms | Max: 1848ms

Network Latency (median): 1847.6ms

Success Rate: 2/2

Response Length: 41 chars

165. Qwen: Qwen3 VL 30B A3B Instruct (qwen/qwen3-vl-30b-a3b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1852ms

Avg Latency: 1416.5ms

Min: 981ms | Max: 1852ms

Network Latency (median): 1852.0ms

Success Rate: 2/2

Response Length: 6 chars

166. OpenAI: gpt-oss-120b (exacto) (openai/gpt-oss-120b:exacto)

Provider: openrouter

Status: ✅

Median Latency: 1855ms

Avg Latency: 1787.5ms

Min: 1720ms | Max: 1855ms

Network Latency (median): 1854.6ms

Success Rate: 2/2

Response Length: 6 chars

167. Google: Gemini 2.0 Flash Lite (google/gemini-2.0-flash-lite-001)

Provider: openrouter

Status: ✅

Median Latency: 1868ms

Avg Latency: 1238.0ms

Min: 608ms | Max: 1868ms

Network Latency (median): 1868.9ms

Success Rate: 2/2

Response Length: 20 chars

168. Mistral: Mixtral 8x22B Instruct (mistralai/mixtral-8x22b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1889ms

Avg Latency: 1301.0ms

Min: 713ms | Max: 1889ms

Network Latency (median): 1888.5ms

Success Rate: 2/2

Response Length: 9 chars

169. Mistral Large (mistralai/mistral-large)

Provider: openrouter

Status: ✅

Median Latency: 1900ms

Avg Latency: 1347.0ms

Min: 794ms | Max: 1900ms

Network Latency (median): 1900.5ms

Success Rate: 2/2

Response Length: 17 chars

170. Google: Gemini 2.0 Flash (google/gemini-2.0-flash-001)

Provider: openrouter

Status: ✅

Median Latency: 1902ms

Avg Latency: 1363.5ms

Min: 825ms | Max: 1902ms

Network Latency (median): 1901.6ms

Success Rate: 2/2

Response Length: 47 chars

171. OpenAI: GPT-4.1 (openai/gpt-4.1)

Provider: openrouter

Status: ✅

Median Latency: 1913ms

Avg Latency: 1314.5ms

Min: 716ms | Max: 1913ms

Network Latency (median): 1912.1ms

Success Rate: 2/2

Response Length: 44 chars

172. GPT-3.5 Turbo (gpt-3.5-turbo)

Provider: openai

Status: ✅

Median Latency: 1923ms

Avg Latency: 1236.5ms

Min: 550ms | Max: 1923ms

Network Latency (median): 1923.6ms

Success Rate: 2/2

Response Length: 25 chars

173. Meta: Llama 3.2 1B Instruct (meta-llama/llama-3.2-1b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 1946ms

Avg Latency: 1097.5ms

Min: 249ms | Max: 1946ms

Network Latency (median): 1945.9ms

Success Rate: 2/2

Response Length: 6 chars

174. Qwen: Qwen3 235B A22B Instruct 2507 (qwen/qwen3-235b-a22b-2507)

Provider: openrouter

Status: ✅

Median Latency: 1964ms

Avg Latency: 1603.0ms

Min: 1242ms | Max: 1964ms

Network Latency (median): 1963.2ms

Success Rate: 2/2

Response Length: 34 chars

175. Cohere: Command R+ (08-2024) (cohere/command-r-plus-08-2024)

Provider: openrouter

Status: ✅

Median Latency: 1971ms

Avg Latency: 1630.5ms

Min: 1290ms | Max: 1971ms

Network Latency (median): 1971.2ms

Success Rate: 2/2

Response Length: 95 chars

176. Meta: Llama 4 Scout (meta-llama/llama-4-scout)

Provider: openrouter

Status: ✅

Median Latency: 1973ms

Avg Latency: 1078.5ms

Min: 184ms | Max: 1973ms

Network Latency (median): 1973.2ms

Success Rate: 2/2

Response Length: 6 chars

177. Nous: Hermes 3 405B Instruct (nousresearch/hermes-3-llama-3.1-405b)

Provider: openrouter

Status: ✅

Median Latency: 1976ms

Avg Latency: 1686.5ms

Min: 1397ms | Max: 1976ms

Network Latency (median): 1976.5ms

Success Rate: 2/2

Response Length: 30 chars

178. DeepSeek: DeepSeek V3.1 (deepseek/deepseek-chat-v3.1)

Provider: openrouter

Status: ✅

Median Latency: 2006ms

Avg Latency: 1502.5ms

Min: 999ms | Max: 2006ms

Network Latency (median): 2005.4ms

Success Rate: 2/2

Response Length: 48 chars

179. OpenAI: o3 (openai/o3)

Provider: openrouter

Status: ✅

Median Latency: 2086ms

Avg Latency: 1977.0ms

Min: 1868ms | Max: 2086ms

Network Latency (median): 2085.8ms

Success Rate: 2/2

Response Length: 6 chars

180. GPT-3.5 Turbo (Free Tier) (free-gpt-3.5-turbo)

Provider: openai

Status: ✅

Median Latency: 2088ms

Avg Latency: 1562.0ms

Min: 1036ms | Max: 2088ms

Network Latency (median): 2088.4ms

Success Rate: 2/2

Response Length: 6 chars

181. Google: Gemma 2 9B (google/gemma-2-9b-it)

Provider: openrouter

Status: ✅

Median Latency: 2092ms

Avg Latency: 1290.0ms

Min: 488ms | Max: 2092ms

Network Latency (median): 2092.6ms

Success Rate: 2/2

Response Length: 7 chars

182. Qwen: Qwen3 Coder 480B A35B (qwen/qwen3-coder)

Provider: openrouter

Status: ✅

Median Latency: 2094ms

Avg Latency: 2056.5ms

Min: 2019ms | Max: 2094ms

Network Latency (median): 2093.6ms

Success Rate: 2/2

Response Length: 6 chars

183. Perplexity: Sonar Pro (perplexity/sonar-pro)

Provider: openrouter

Status: ✅

Median Latency: 2104ms

Avg Latency: 1796.5ms

Min: 1489ms | Max: 2104ms

Network Latency (median): 2103.6ms

Success Rate: 2/2

Response Length: 29 chars

184. Arcee AI: Trinity Mini (arcee-ai/trinity-mini)

Provider: openrouter

Status: ✅

Median Latency: 2130ms

Avg Latency: 2001.5ms

Min: 1873ms | Max: 2130ms

Network Latency (median): 2129.5ms

Success Rate: 2/2

Response Length: 252 chars

185. Tongyi DeepResearch 30B A3B (free) (alibaba/tongyi-deepresearch-30b-a3b:free)

Provider: openrouter

Status: ✅

Median Latency: 2155ms

Avg Latency: 2106.0ms

Min: 2057ms | Max: 2155ms

Network Latency (median): 2154.4ms

Success Rate: 2/2

Response Length: 37 chars

186. Inflection: Inflection 3 Productivity (inflection/inflection-3-productivity)

Provider: openrouter

Status: ✅

Median Latency: 2165ms

Avg Latency: 2142.5ms

Min: 2120ms | Max: 2165ms

Network Latency (median): 2164.4ms

Success Rate: 2/2

Response Length: 86 chars

187. AlfredPros: CodeLLaMa 7B Instruct Solidity (alfredpros/codellama-7b-instruct-solidity)

Provider: openrouter

Status: ✅

Median Latency: 2176ms

Avg Latency: 1994.0ms

Min: 1812ms | Max: 2176ms

Network Latency (median): 2176.3ms

Success Rate: 2/2

Response Length: 24 chars

188. OpenAI: gpt-oss-120b (openai/gpt-oss-120b)

Provider: openrouter

Status: ✅

Median Latency: 2188ms

Avg Latency: 1839.5ms

Min: 1491ms | Max: 2188ms

Network Latency (median): 2187.4ms

Success Rate: 2/2

Response Length: 24 chars

189. TheDrummer: Rocinante 12B (thedrummer/rocinante-12b)

Provider: openrouter

Status: ✅

Median Latency: 2209ms

Avg Latency: 2090.5ms

Min: 1972ms | Max: 2209ms

Network Latency (median): 2209.1ms

Success Rate: 2/2

Response Length: 6 chars

190. Mistral: Mistral Small 3.1 24B (mistralai/mistral-small-3.1-24b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 2228ms

Avg Latency: 1384.0ms

Min: 540ms | Max: 2228ms

Network Latency (median): 2227.8ms

Success Rate: 2/2

Response Length: 25 chars

191. OpenAI: o3 Mini (openai/o3-mini)

Provider: openrouter

Status: ✅

Median Latency: 2257ms

Avg Latency: 2177.0ms

Min: 2097ms | Max: 2257ms

Network Latency (median): 2256.3ms

Success Rate: 2/2

Response Length: 44 chars

192. Qwen: Qwen3 30B A3B Thinking 2507 (qwen/qwen3-30b-a3b-thinking-2507)

Provider: openrouter

Status: ✅

Median Latency: 2278ms

Avg Latency: 1854.0ms

Min: 1430ms | Max: 2278ms

Network Latency (median): 2277.4ms

Success Rate: 2/2

Response Length: 34 chars

193. NeverSleep: Lumimaid v0.2 8B (neversleep/llama-3.1-lumimaid-8b)

Provider: openrouter

Status: ✅

Median Latency: 2290ms

Avg Latency: 2208.0ms

Min: 2126ms | Max: 2290ms

Network Latency (median): 2289.6ms

Success Rate: 2/2

Response Length: 32 chars

194. Amazon: Nova 2 Lite (amazon/nova-2-lite-v1)

Provider: openrouter

Status: ✅

Median Latency: 2383ms

Avg Latency: 2073.0ms

Min: 1763ms | Max: 2383ms

Network Latency (median): 2383.0ms

Success Rate: 2/2

Response Length: 65 chars

195. NVIDIA: Nemotron Nano 9B V2 (nvidia/nemotron-nano-9b-v2)

Provider: openrouter

Status: ✅

Median Latency: 2412ms

Avg Latency: 2184.0ms

Min: 1956ms | Max: 2412ms

Network Latency (median): 2412.4ms

Success Rate: 2/2

Response Length: 35 chars

196. AionLabs: Aion-1.0-Mini (aion-labs/aion-1.0-mini)

Provider: openrouter

Status: ✅

Median Latency: 2419ms

Avg Latency: 2318.5ms

Min: 2218ms | Max: 2419ms

Network Latency (median): 2418.7ms

Success Rate: 2/2

Response Length: 501 chars

197. Microsoft: Phi-3.5 Mini 128K Instruct (microsoft/phi-3.5-mini-128k-instruct)

Provider: openrouter

Status: ✅

Median Latency: 2520ms

Avg Latency: 1749.0ms

Min: 978ms | Max: 2520ms

Network Latency (median): 2520.2ms

Success Rate: 2/2

Response Length: 38 chars

198. Sao10K: Llama 3.1 70B Hanami x1 (sao10k/l3.1-70b-hanami-x1)

Provider: openrouter

Status: ✅

Median Latency: 2550ms

Avg Latency: 2447.5ms

Min: 2345ms | Max: 2550ms

Network Latency (median): 2549.9ms

Success Rate: 2/2

Response Length: 107 chars

199. Anthropic: Claude 3.7 Sonnet (anthropic/claude-3.7-sonnet)

Provider: openrouter

Status: ✅

Median Latency: 2559ms

Avg Latency: 2530.5ms

Min: 2502ms | Max: 2559ms

Network Latency (median): 2558.2ms

Success Rate: 2/2

Response Length: 49 chars

200. Relace: Relace Search (relace/relace-search)

Provider: openrouter

Status: ✅

Median Latency: 2592ms

Avg Latency: 1760.0ms

Min: 928ms | Max: 2592ms

Network Latency (median): 2592.4ms

Success Rate: 2/2

Response Length: 6 chars

201. OpenAI: GPT-4 (older v0314) (openai/gpt-4-0314)

Provider: openrouter

Status: ✅

Median Latency: 2647ms

Avg Latency: 1705.5ms

Min: 764ms | Max: 2647ms

Network Latency (median): 2647.3ms

Success Rate: 2/2

Response Length: 40 chars

202. OpenAI: gpt-oss-20b (openai/gpt-oss-20b)

Provider: openrouter

Status: ✅

Median Latency: 2655ms

Avg Latency: 1999.0ms

Min: 1343ms | Max: 2655ms

Network Latency (median): 2655.1ms

Success Rate: 2/2

Response Length: 16 chars

203. Qwen: Qwen3 Coder 480B A35B (exacto) (qwen/qwen3-coder:exacto)

Provider: openrouter

Status: ✅

Median Latency: 2667ms

Avg Latency: 2537.5ms

Min: 2408ms | Max: 2667ms

Network Latency (median): 2666.7ms

Success Rate: 2/2

Response Length: 6 chars

204. Sao10K: Llama 3.3 Euryale 70B (sao10k/l3.3-euryale-70b)

Provider: openrouter

Status: ✅

Median Latency: 2690ms

Avg Latency: 2668.5ms

Min: 2647ms | Max: 2690ms

Network Latency (median): 2689.8ms

Success Rate: 2/2

Response Length: 73 chars

205. Meta: Llama 3.1 405B Instruct (meta-llama/llama-3.1-405b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 2715ms

Avg Latency: 2542.5ms

Min: 2370ms | Max: 2715ms

Network Latency (median): 2715.7ms

Success Rate: 2/2

Response Length: 6 chars

206. NVIDIA: Nemotron Nano 9B V2 (free) (nvidia/nemotron-nano-9b-v2:free)

Provider: openrouter

Status: ✅

Median Latency: 2719ms

Avg Latency: 2448.0ms

Min: 2177ms | Max: 2719ms

Network Latency (median): 2719.3ms

Success Rate: 2/2

Response Length: 35 chars

207. TheDrummer: Cydonia 24B V4.1 (thedrummer/cydonia-24b-v4.1)

Provider: openrouter

Status: ✅

Median Latency: 2780ms

Avg Latency: 2757.0ms

Min: 2734ms | Max: 2780ms

Network Latency (median): 2780.3ms

Success Rate: 2/2

Response Length: 12 chars

208. Qwen: Qwen3 235B A22B (qwen/qwen3-235b-a22b)

Provider: openrouter

Status: ✅

Median Latency: 2815ms

Avg Latency: 2356.5ms

Min: 1898ms | Max: 2815ms

Network Latency (median): 2815.3ms

Success Rate: 2/2

Response Length: 34 chars

209. Arcee AI: Trinity Mini (free) (arcee-ai/trinity-mini:free)

Provider: openrouter

Status: ✅

Median Latency: 2897ms

Avg Latency: 2563.5ms

Min: 2230ms | Max: 2897ms

Network Latency (median): 2897.1ms

Success Rate: 2/2

Response Length: 12 chars

210. GPT-4 Turbo (gpt-4-turbo)

Provider: openai

Status: ✅

Median Latency: 2945ms

Avg Latency: 2195.0ms

Min: 1445ms | Max: 2945ms

Network Latency (median): 2944.9ms

Success Rate: 2/2

Response Length: 31 chars

211. Qwen: Qwen3 Coder 480B A35B (free) (qwen/qwen3-coder:free)

Provider: openrouter

Status: ✅

Median Latency: 2958ms

Avg Latency: 2077.0ms

Min: 1196ms | Max: 2958ms

Network Latency (median): 2958.1ms

Success Rate: 2/2

Response Length: 6 chars

212. Amazon: Nova Premier 1.0 (amazon/nova-premier-v1)

Provider: openrouter

Status: ✅

Median Latency: 2963ms

Avg Latency: 2908.0ms

Min: 2853ms | Max: 2963ms

Network Latency (median): 2962.7ms

Success Rate: 2/2

Response Length: 140 chars

213. MiniMax: MiniMax M2 (minimax/minimax-m2)

Provider: openrouter

Status: ✅

Median Latency: 2964ms

Avg Latency: 2809.5ms

Min: 2655ms | Max: 2964ms

Network Latency (median): 2964.3ms

Success Rate: 2/2

Response Length: 6 chars

214. Perplexity: Sonar Pro Search (perplexity/sonar-pro-search)

Provider: openrouter

Status: ✅

Median Latency: 3040ms

Avg Latency: 1859.0ms

Min: 678ms | Max: 3040ms

Network Latency (median): 3040.2ms

Success Rate: 2/2

Response Length: 38 chars

215. Anthropic: Claude Sonnet 4 (anthropic/claude-sonnet-4)

Provider: openrouter

Status: ✅

Median Latency: 3093ms

Avg Latency: 2908.0ms

Min: 2723ms | Max: 3093ms

Network Latency (median): 3092.8ms

Success Rate: 2/2

Response Length: 35 chars

216. OpenAI: o1 (openai/o1)

Provider: openrouter

Status: ✅

Median Latency: 3124ms

Avg Latency: 3066.0ms

Min: 3008ms | Max: 3124ms

Network Latency (median): 3123.7ms

Success Rate: 2/2

Response Length: 47 chars

217. xAI: Grok 4.1 Fast (x-ai/grok-4.1-fast)

Provider: openrouter

Status: ✅

Median Latency: 3143ms

Avg Latency: 2502.0ms

Min: 1861ms | Max: 3143ms

Network Latency (median): 3143.1ms

Success Rate: 2/2

Response Length: 6 chars

218. Anthropic: Claude Opus 4.5 (anthropic/claude-opus-4.5)

Provider: openrouter

Status: ✅

Median Latency: 3202ms

Avg Latency: 2850.5ms

Min: 2499ms | Max: 3202ms

Network Latency (median): 3201.8ms

Success Rate: 2/2

Response Length: 34 chars

219. DeepSeek: DeepSeek R1 0528 Qwen3 8B (deepseek/deepseek-r1-0528-qwen3-8b)

Provider: openrouter

Status: ✅

Median Latency: 3227ms

Avg Latency: 2878.0ms

Min: 2529ms | Max: 3227ms

Network Latency (median): 3227.4ms

Success Rate: 2/2

Response Length: 6 chars

220. AllenAI: Olmo 3 7B Think (allenai/olmo-3-7b-think)

Provider: openrouter

Status: ✅

Median Latency: 3275ms

Avg Latency: 2418.0ms

Min: 1561ms | Max: 3275ms

Network Latency (median): 3275.0ms

Success Rate: 2/2

Response Length: 34 chars

221. OpenAI: o4 Mini (openai/o4-mini)

Provider: openrouter

Status: ✅

Median Latency: 3455ms

Avg Latency: 2895.5ms

Min: 2336ms | Max: 3455ms

Network Latency (median): 3455.1ms

Success Rate: 2/2

Response Length: 32 chars

222. OpenAI: GPT-5.1 (openai/gpt-5.1)

Provider: openrouter

Status: ✅

Median Latency: 3486ms

Avg Latency: 2572.0ms

Min: 1658ms | Max: 3486ms

Network Latency (median): 3486.4ms

Success Rate: 2/2

Response Length: 38 chars

223. Microsoft: Phi 4 (microsoft/phi-4)

Provider: openrouter

Status: ✅

Median Latency: 3502ms

Avg Latency: 2750.0ms

Min: 1998ms | Max: 3502ms

Network Latency (median): 3502.2ms

Success Rate: 2/2

Response Length: 34 chars

224. TheDrummer: UnslopNemo 12B (thedrummer/unslopnemo-12b)

Provider: openrouter

Status: ✅

Median Latency: 3514ms

Avg Latency: 3051.0ms

Min: 2588ms | Max: 3514ms

Network Latency (median): 3514.1ms

Success Rate: 2/2

Response Length: 6 chars

225. OpenAI: Codex Mini (openai/codex-mini)

Provider: openrouter

Status: ✅

Median Latency: 3517ms

Avg Latency: 2906.0ms

Min: 2295ms | Max: 3517ms

Network Latency (median): 3517.8ms

Success Rate: 2/2

Response Length: 12 chars

226. Baidu: ERNIE 4.5 21B A3B Thinking (baidu/ernie-4.5-21b-a3b-thinking)

Provider: openrouter

Status: ✅

Median Latency: 3561ms

Avg Latency: 3337.5ms

Min: 3114ms | Max: 3561ms

Network Latency (median): 3560.9ms

Success Rate: 2/2

Response Length: 27 chars

227. IBM: Granite 4.0 Micro (ibm-granite/granite-4.0-h-micro)

Provider: openrouter

Status: ✅

Median Latency: 3616ms

Avg Latency: 2060.0ms

Min: 504ms | Max: 3616ms

Network Latency (median): 3615.8ms

Success Rate: 2/2

Response Length: 6 chars

228. GPT-4 (gpt-4)

Provider: openai

Status: ✅

Median Latency: 3714ms

Avg Latency: 2777.0ms

Min: 1840ms | Max: 3714ms

Network Latency (median): 3714.0ms

Success Rate: 2/2

Response Length: 31 chars

229. Perplexity: Sonar Reasoning Pro (perplexity/sonar-reasoning-pro)

Provider: openrouter

Status: ✅

Median Latency: 3732ms

Avg Latency: 3640.0ms

Min: 3548ms | Max: 3732ms

Network Latency (median): 3732.0ms

Success Rate: 2/2

Response Length: 78 chars

230. xAI: Grok Code Fast 1 (x-ai/grok-code-fast-1)

Provider: openrouter

Status: ✅

Median Latency: 3762ms

Avg Latency: 2941.5ms

Min: 2121ms | Max: 3762ms

Network Latency (median): 3762.1ms

Success Rate: 2/2

Response Length: 86 chars

231. Mistral: Saba (mistralai/mistral-saba)

Provider: openrouter

Status: ✅

Median Latency: 3767ms

Avg Latency: 2163.0ms

Min: 559ms | Max: 3767ms

Network Latency (median): 3767.6ms

Success Rate: 2/2

Response Length: 12 chars

232. Anthropic: Claude Sonnet 4.5 (anthropic/claude-sonnet-4.5)

Provider: openrouter

Status: ✅

Median Latency: 3774ms

Avg Latency: 3065.0ms

Min: 2356ms | Max: 3774ms

Network Latency (median): 3773.4ms

Success Rate: 2/2

Response Length: 29 chars

233. Anthropic: Claude 3.7 Sonnet (thinking) (anthropic/claude-3.7-sonnet:thinking)

Provider: openrouter

Status: ✅

Median Latency: 3780ms

Avg Latency: 3753.5ms

Min: 3727ms | Max: 3780ms

Network Latency (median): 3779.8ms

Success Rate: 2/2

Response Length: 49 chars

234. Anthropic: Claude Opus 4 (anthropic/claude-opus-4)

Provider: openrouter

Status: ✅

Median Latency: 3914ms

Avg Latency: 3833.0ms

Min: 3752ms | Max: 3914ms

Network Latency (median): 3914.4ms

Success Rate: 2/2

Response Length: 35 chars

235. Anthropic: Claude Opus 4.1 (anthropic/claude-opus-4.1)

Provider: openrouter

Status: ✅

Median Latency: 3925ms

Avg Latency: 3722.0ms

Min: 3519ms | Max: 3925ms

Network Latency (median): 3924.6ms

Success Rate: 2/2

Response Length: 35 chars

236. Google: Gemma 3 12B (free) (google/gemma-3-12b-it:free)

Provider: openrouter

Status: ✅

Median Latency: 3992ms

Avg Latency: 3709.5ms

Min: 3427ms | Max: 3992ms

Network Latency (median): 3992.0ms

Success Rate: 2/2

Response Length: 54 chars

237. Qwen: Qwen3 VL 235B A22B Thinking (qwen/qwen3-vl-235b-a22b-thinking)

Provider: openrouter

Status: ✅

Median Latency: 4023ms

Avg Latency: 3251.0ms

Min: 2479ms | Max: 4023ms

Network Latency (median): 4022.7ms

Success Rate: 2/2

Response Length: 34 chars

238. xAI: Grok 3 Mini (x-ai/grok-3-mini)

Provider: openrouter

Status: ✅

Median Latency: 4046ms

Avg Latency: 3571.0ms

Min: 3096ms | Max: 4046ms

Network Latency (median): 4046.6ms

Success Rate: 2/2

Response Length: 6 chars

239. OpenAI: o3 Mini High (openai/o3-mini-high)

Provider: openrouter

Status: ✅

Median Latency: 4107ms

Avg Latency: 3139.5ms

Min: 2172ms | Max: 4107ms

Network Latency (median): 4107.6ms

Success Rate: 2/2

Response Length: 44 chars

240. Z.AI: GLM 4.6V (z-ai/glm-4.6v)

Provider: openrouter

Status: ✅

Median Latency: 4131ms

Avg Latency: 3652.5ms

Min: 3174ms | Max: 4131ms

Network Latency (median): 4131.0ms

Success Rate: 2/2

Response Length: 34 chars

241. Z.AI: GLM 4.5 (z-ai/glm-4.5)

Provider: openrouter

Status: ✅

Median Latency: 4226ms

Avg Latency: 4226.0ms

Min: 4226ms | Max: 4226ms

Network Latency (median): 4226.3ms

Success Rate: 1/2

Response Length: 35 chars

242. OpenAI: GPT-5.1-Codex-Max (openai/gpt-5.1-codex-max)

Provider: openrouter

Status: ✅

Median Latency: 4251ms

Avg Latency: 2780.0ms

Min: 1309ms | Max: 4251ms

Network Latency (median): 4251.4ms

Success Rate: 2/2

Response Length: 25 chars

243. Body Builder (openrouter/bodybuilder)

Provider: openrouter

Status: ✅

Median Latency: 4400ms

Avg Latency: 2876.5ms

Min: 1353ms | Max: 4400ms

Network Latency (median): 4400.3ms

Success Rate: 2/2

Response Length: 193 chars

244. AionLabs: Aion-1.0 (aion-labs/aion-1.0)

Provider: openrouter

Status: ✅

Median Latency: 4457ms

Avg Latency: 4149.5ms

Min: 3842ms | Max: 4457ms

Network Latency (median): 4456.9ms

Success Rate: 2/2

Response Length: 292 chars

245. MoonshotAI: Kimi K2 0711 (moonshotai/kimi-k2)

Provider: openrouter

Status: ✅

Median Latency: 4572ms

Avg Latency: 3604.5ms

Min: 2637ms | Max: 4572ms

Network Latency (median): 4572.2ms

Success Rate: 2/2

Response Length: 24 chars

246. DeepSeek: R1 (deepseek/deepseek-r1)

Provider: openrouter

Status: ✅

Median Latency: 4649ms

Avg Latency: 4606.5ms

Min: 4564ms | Max: 4649ms

Network Latency (median): 4649.2ms

Success Rate: 2/2

Response Length: 12 chars

247. Qwen: Qwen3 Next 80B A3B Thinking (qwen/qwen3-next-80b-a3b-thinking)

Provider: openrouter

Status: ✅

Median Latency: 4685ms

Avg Latency: 4547.5ms

Min: 4410ms | Max: 4685ms

Network Latency (median): 4684.6ms

Success Rate: 2/2

Response Length: 9 chars

248. Tencent: Hunyuan A13B Instruct (tencent/hunyuan-a13b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 4742ms

Avg Latency: 3217.0ms

Min: 1692ms | Max: 4742ms

Network Latency (median): 4742.6ms

Success Rate: 2/2

Response Length: 7 chars

249. THUDM: GLM 4.1V 9B Thinking (thudm/glm-4.1v-9b-thinking)

Provider: openrouter

Status: ✅

Median Latency: 4919ms

Avg Latency: 3116.5ms

Min: 1314ms | Max: 4919ms

Network Latency (median): 4919.6ms

Success Rate: 2/2

Response Length: 36 chars

250. Mistral: Mistral 7B Instruct v0.1 (mistralai/mistral-7b-instruct-v0.1)

Provider: openrouter

Status: ✅

Median Latency: 5017ms

Avg Latency: 3468.0ms

Min: 1919ms | Max: 5017ms

Network Latency (median): 5017.2ms

Success Rate: 2/2

Response Length: 95 chars

251. OpenAI: GPT-5 Mini (openai/gpt-5-mini)

Provider: openrouter

Status: ✅

Median Latency: 5171ms

Avg Latency: 3718.0ms

Min: 2265ms | Max: 5171ms

Network Latency (median): 5171.1ms

Success Rate: 2/2

Response Length: 6 chars

252. Perplexity: Sonar Reasoning (perplexity/sonar-reasoning)

Provider: openrouter

Status: ✅

Median Latency: 5174ms

Avg Latency: 5174.0ms

Min: 5174ms | Max: 5174ms

Network Latency (median): 5174.6ms

Success Rate: 1/2

Response Length: 34 chars

253. Perplexity: Sonar Deep Research (perplexity/sonar-deep-research)

Provider: openrouter

Status: ✅

Median Latency: 5190ms

Avg Latency: 4752.5ms

Min: 4315ms | Max: 5190ms

Network Latency (median): 5190.3ms

Success Rate: 2/2

Response Length: 75 chars

254. OpenAI: GPT-5 (openai/gpt-5)

Provider: openrouter

Status: ✅

Median Latency: 5345ms

Avg Latency: 4237.0ms

Min: 3129ms | Max: 5345ms

Network Latency (median): 5345.1ms

Success Rate: 2/2

Response Length: 40 chars

255. Meta: LlamaGuard 2 8B (meta-llama/llama-guard-2-8b)

Provider: openrouter

Status: ✅

Median Latency: 5706ms

Avg Latency: 3347.5ms

Min: 989ms | Max: 5706ms

Network Latency (median): 5706.4ms

Success Rate: 2/2

Response Length: 0 chars

256. Inflection: Inflection 3 Pi (inflection/inflection-3-pi)

Provider: openrouter

Status: ✅

Median Latency: 5729ms

Avg Latency: 4038.5ms

Min: 2348ms | Max: 5729ms

Network Latency (median): 5729.3ms

Success Rate: 2/2

Response Length: 14 chars

257. Nous: Hermes 3 70B Instruct (nousresearch/hermes-3-llama-3.1-70b)

Provider: openrouter

Status: ✅

Median Latency: 6318ms

Avg Latency: 6318.0ms

Min: 6318ms | Max: 6318ms

Network Latency (median): 6318.8ms

Success Rate: 1/2

Response Length: 30 chars

258. Auto Router (openrouter/auto)

Provider: openrouter

Status: ✅

Median Latency: 6428ms

Avg Latency: 5023.0ms

Min: 3618ms | Max: 6428ms

Network Latency (median): 6428.5ms

Success Rate: 2/2

Response Length: 6 chars

259. DeepSeek: R1 0528 (deepseek/deepseek-r1-0528)

Provider: openrouter

Status: ✅

Median Latency: 6447ms

Avg Latency: 6307.0ms

Min: 6167ms | Max: 6447ms

Network Latency (median): 6447.6ms

Success Rate: 2/2

Response Length: 34 chars

260. Qwen: Qwen3 235B A22B (free) (qwen/qwen3-235b-a22b:free)

Provider: openrouter

Status: ✅

Median Latency: 6550ms

Avg Latency: 6550.0ms

Min: 6550ms | Max: 6550ms

Network Latency (median): 6550.2ms

Success Rate: 1/2

Response Length: 6 chars

261. Switchpoint Router (switchpoint/router)

Provider: openrouter

Status: ✅

Median Latency: 6561ms

Avg Latency: 4064.0ms

Min: 1567ms | Max: 6561ms

Network Latency (median): 6561.4ms

Success Rate: 2/2

Response Length: 34 chars

262. Qwen: Qwen3 30B A3B Instruct 2507 (qwen/qwen3-30b-a3b-instruct-2507)

Provider: openrouter

Status: ✅

Median Latency: 6646ms

Avg Latency: 3789.5ms

Min: 933ms | Max: 6646ms

Network Latency (median): 6646.2ms

Success Rate: 2/2

Response Length: 6 chars

263. MiniMax: MiniMax M1 (minimax/minimax-m1)

Provider: openrouter

Status: ✅

Median Latency: 6879ms

Avg Latency: 5961.5ms

Min: 5044ms | Max: 6879ms

Network Latency (median): 6878.7ms

Success Rate: 2/2

Response Length: 30 chars

264. Qwen: Qwen3 235B A22B Thinking 2507 (qwen/qwen3-235b-a22b-thinking-2507)

Provider: openrouter

Status: ✅

Median Latency: 7600ms

Avg Latency: 6596.0ms

Min: 5592ms | Max: 7600ms

Network Latency (median): 7600.1ms

Success Rate: 2/2

Response Length: 6 chars

265. DeepSeek: DeepSeek V3.2 Exp (deepseek/deepseek-v3.2-exp)

Provider: openrouter

Status: ✅

Median Latency: 7736ms

Avg Latency: 5066.0ms

Min: 2396ms | Max: 7736ms

Network Latency (median): 7735.7ms

Success Rate: 2/2

Response Length: 41 chars

266. NVIDIA: Nemotron Nano 12B 2 VL (free) (nvidia/nemotron-nano-12b-v2-vl:free)

Provider: openrouter

Status: ✅

Median Latency: 7790ms

Avg Latency: 7103.5ms

Min: 6417ms | Max: 7790ms

Network Latency (median): 7789.9ms

Success Rate: 2/2

Response Length: 34 chars

267. MoonshotAI: Kimi K2 Thinking (moonshotai/kimi-k2-thinking)

Provider: openrouter

Status: ✅

Median Latency: 7798ms

Avg Latency: 7798.0ms

Min: 7798ms | Max: 7798ms

Network Latency (median): 7797.6ms

Success Rate: 1/2

Response Length: 30 chars

268. NVIDIA: Nemotron Nano 12B 2 VL (nvidia/nemotron-nano-12b-v2-vl)

Provider: openrouter

Status: ✅

Median Latency: 8175ms

Avg Latency: 5616.0ms

Min: 3057ms | Max: 8175ms

Network Latency (median): 8174.7ms

Success Rate: 2/2

Response Length: 35 chars

269. OpenAI: o4 Mini High (openai/o4-mini-high)

Provider: openrouter

Status: ✅

Median Latency: 8450ms

Avg Latency: 5511.0ms

Min: 2572ms | Max: 8450ms

Network Latency (median): 8449.3ms

Success Rate: 2/2

Response Length: 35 chars

270. Mancer: Weaver (alpha) (mancer/weaver)

Provider: openrouter

Status: ✅

Median Latency: 8471ms

Avg Latency: 7111.5ms

Min: 5752ms | Max: 8471ms

Network Latency (median): 8470.8ms

Success Rate: 2/2

Response Length: 747 chars

271. Z.AI: GLM 4.5 Air (free) (z-ai/glm-4.5-air:free)

Provider: openrouter

Status: ✅

Median Latency: 8574ms

Avg Latency: 7143.5ms

Min: 5713ms | Max: 8574ms

Network Latency (median): 8573.9ms

Success Rate: 2/2

Response Length: 6 chars

272. OpenAI: GPT-5 Nano (openai/gpt-5-nano)

Provider: openrouter

Status: ✅

Median Latency: 8723ms

Avg Latency: 6716.0ms

Min: 4709ms | Max: 8723ms

Network Latency (median): 8723.5ms

Success Rate: 2/2

Response Length: 6 chars

273. Qwen: Qwen2.5 VL 32B Instruct (qwen/qwen2.5-vl-32b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 8738ms

Avg Latency: 4861.5ms

Min: 985ms | Max: 8738ms

Network Latency (median): 8737.7ms

Success Rate: 2/2

Response Length: 34 chars

274. TNG: R1T Chimera (free) (tngtech/tng-r1t-chimera:free)

Provider: openrouter

Status: ✅

Median Latency: 9254ms

Avg Latency: 6412.5ms

Min: 3571ms | Max: 9254ms

Network Latency (median): 9254.6ms

Success Rate: 2/2

Response Length: 39 chars

275. Google: Gemini 3 Pro Preview (google/gemini-3-pro-preview)

Provider: openrouter

Status: ✅

Median Latency: 9314ms

Avg Latency: 9238.0ms

Min: 9162ms | Max: 9314ms

Network Latency (median): 9314.5ms

Success Rate: 2/2

Response Length: 45 chars

276. Qwen: Qwen3 VL 30B A3B Thinking (qwen/qwen3-vl-30b-a3b-thinking)

Provider: openrouter

Status: ✅

Median Latency: 9474ms

Avg Latency: 5565.0ms

Min: 1656ms | Max: 9474ms

Network Latency (median): 9473.8ms

Success Rate: 2/2

Response Length: 6 chars

277. xAI: Grok 4 (x-ai/grok-4)

Provider: openrouter

Status: ✅

Median Latency: 9608ms

Avg Latency: 9608.0ms

Min: 9608ms | Max: 9608ms

Network Latency (median): 9608.2ms

Success Rate: 1/2

Response Length: 6 chars

278. Qwen: Qwen3 VL 8B Thinking (qwen/qwen3-vl-8b-thinking)

Provider: openrouter

Status: ✅

Median Latency: 9791ms

Avg Latency: 7437.0ms

Min: 5083ms | Max: 9791ms

Network Latency (median): 9790.9ms

Success Rate: 2/2

Response Length: 41 chars

279. Z.AI: GLM 4.6 (z-ai/glm-4.6)

Provider: openrouter

Status: ✅

Median Latency: 9797ms

Avg Latency: 9797.0ms

Min: 9797ms | Max: 9797ms

Network Latency (median): 9796.5ms

Success Rate: 1/2

Response Length: 39 chars

280. OpenAI: GPT-4o Search Preview (openai/gpt-4o-search-preview)

Provider: openrouter

Status: ✅

Median Latency: 9863ms

Avg Latency: 7443.5ms

Min: 5024ms | Max: 9863ms

Network Latency (median): 9863.5ms

Success Rate: 2/2

Response Length: 642 chars

281. xAI: Grok 3 Mini Beta (x-ai/grok-3-mini-beta)

Provider: openrouter

Status: ✅

Median Latency: 10070ms

Avg Latency: 6601.0ms

Min: 3132ms | Max: 10070ms

Network Latency (median): 10070.0ms

Success Rate: 2/2

Response Length: 6 chars

282. MoonshotAI: Kimi Dev 72B (moonshotai/kimi-dev-72b)

Provider: openrouter

Status: ✅

Median Latency: 10170ms

Avg Latency: 10170.0ms

Min: 10170ms | Max: 10170ms

Network Latency (median): 10169.8ms

Success Rate: 1/2

Response Length: 32 chars

283. Qwen: Qwen2.5 Coder 7B Instruct (qwen/qwen2.5-coder-7b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 10392ms

Avg Latency: 5550.0ms

Min: 708ms | Max: 10392ms

Network Latency (median): 10392.9ms

Success Rate: 2/2

Response Length: 8555 chars

284. Qwen: Qwen3 14B (qwen/qwen3-14b)

Provider: openrouter

Status: ✅

Median Latency: 11581ms

Avg Latency: 8353.5ms

Min: 5126ms | Max: 11581ms

Network Latency (median): 11580.3ms

Success Rate: 2/2

Response Length: 34 chars

285. DeepSeek: DeepSeek V3.2 (deepseek/deepseek-v3.2)

Provider: openrouter

Status: ✅

Median Latency: 11836ms

Avg Latency: 11836.0ms

Min: 11836ms | Max: 11836ms

Network Latency (median): 11836.2ms

Success Rate: 1/2

Response Length: 34 chars

286. Qwen2.5 Coder 32B Instruct (qwen/qwen-2.5-coder-32b-instruct)

Provider: openrouter

Status: ✅

Median Latency: 11869ms

Avg Latency: 9592.5ms

Min: 7316ms | Max: 11869ms

Network Latency (median): 11869.0ms

Success Rate: 2/2

Response Length: 3257 chars

287. Tongyi DeepResearch 30B A3B (alibaba/tongyi-deepresearch-30b-a3b)

Provider: openrouter

Status: ✅

Median Latency: 12113ms

Avg Latency: 6587.5ms

Min: 1062ms | Max: 12113ms

Network Latency (median): 12112.8ms

Success Rate: 2/2

Response Length: 34 chars

288. Google: Gemma 3 27B (google/gemma-3-27b-it)

Provider: openrouter

Status: ✅

Median Latency: 12133ms

Avg Latency: 6807.5ms

Min: 1482ms | Max: 12133ms

Network Latency (median): 12132.6ms

Success Rate: 2/2

Response Length: 56 chars

289. Qwen: QwQ 32B (qwen/qwq-32b)

Provider: openrouter

Status: ✅

Median Latency: 12268ms

Avg Latency: 11564.0ms

Min: 10860ms | Max: 12268ms

Network Latency (median): 12268.4ms

Success Rate: 2/2

Response Length: 9 chars

290. Meta: Llama 3.3 70B Instruct (free) (meta-llama/llama-3.3-70b-instruct:free)

Provider: openrouter

Status: ✅

Median Latency: 12362ms

Avg Latency: 10382.0ms

Min: 8402ms | Max: 12362ms

Network Latency (median): 12362.1ms

Success Rate: 2/2

Response Length: 73 chars

291. Qwen: Qwen Plus 0728 (thinking) (qwen/qwen-plus-2025-07-28:thinking)

Provider: openrouter

Status: ✅

Median Latency: 12620ms

Avg Latency: 6879.5ms

Min: 1139ms | Max: 12620ms

Network Latency (median): 12621.1ms

Success Rate: 2/2

Response Length: 6 chars

292. Qwen: Qwen3 30B A3B (qwen/qwen3-30b-a3b)

Provider: openrouter

Status: ✅

Median Latency: 13068ms

Avg Latency: 10703.0ms

Min: 8338ms | Max: 13068ms

Network Latency (median): 13068.3ms

Success Rate: 2/2

Response Length: 6 chars

293. Google: Gemini 2.5 Pro Preview 06-05 (google/gemini-2.5-pro-preview)

Provider: openrouter

Status: ✅

Median Latency: 13582ms

Avg Latency: 10374.0ms

Min: 7166ms | Max: 13582ms

Network Latency (median): 13581.7ms

Success Rate: 2/2

Response Length: 49 chars

294. Amazon: Nova 2 Lite (free) (amazon/nova-2-lite-v1:free)

Provider: openrouter

Status: ✅

Median Latency: 13982ms

Avg Latency: 11030.0ms

Min: 8078ms | Max: 13982ms

Network Latency (median): 13981.9ms

Success Rate: 2/2

Response Length: 34 chars

295. TNG: DeepSeek R1T2 Chimera (free) (tngtech/deepseek-r1t2-chimera:free)

Provider: openrouter

Status: ✅

Median Latency: 14939ms

Avg Latency: 10508.5ms

Min: 6078ms | Max: 14939ms

Network Latency (median): 14938.8ms

Success Rate: 2/2

Response Length: 34 chars

296. Qwen: Qwen3 32B (qwen/qwen3-32b)

Provider: openrouter

Status: ✅

Median Latency: 15939ms

Avg Latency: 11916.5ms

Min: 7894ms | Max: 15939ms

Network Latency (median): 15939.4ms

Success Rate: 2/2

Response Length: 34 chars

297. Google: Gemini 2.5 Pro (google/gemini-2.5-pro)

Provider: openrouter

Status: ✅

Median Latency: 16256ms

Avg Latency: 11929.5ms

Min: 7603ms | Max: 16256ms

Network Latency (median): 16255.6ms

Success Rate: 2/2

Response Length: 34 chars

298. Gemini 2.5 Pro (Direct) (gemini-2.5-pro)

Provider: google

Status: ✅

Median Latency: 16664ms

Avg Latency: 15778.5ms

Min: 14893ms | Max: 16664ms

Network Latency (median): 16663.8ms

Success Rate: 2/2

Response Length: 43 chars

299. Google: Gemini 2.5 Pro Preview 05-06 (google/gemini-2.5-pro-preview-05-06)

Provider: openrouter

Status: ✅

Median Latency: 16765ms

Avg Latency: 12338.5ms

Min: 7912ms | Max: 16765ms

Network Latency (median): 16764.6ms

Success Rate: 2/2

Response Length: 43 chars

300. Meta: Llama 3.1 405B (base) (meta-llama/llama-3.1-405b)

Provider: openrouter

Status: ✅

Median Latency: 16822ms

Avg Latency: 9375.5ms

Min: 1929ms | Max: 16822ms

Network Latency (median): 16821.7ms

Success Rate: 2/2

Response Length: 1911 chars

301. TNG: DeepSeek R1T Chimera (free) (tngtech/deepseek-r1t-chimera:free)

Provider: openrouter

Status: ✅

Median Latency: 17246ms

Avg Latency: 11472.5ms

Min: 5699ms | Max: 17246ms

Network Latency (median): 17246.1ms

Success Rate: 2/2

Response Length: 843 chars

302. DeepSeek: R1 Distill Qwen 32B (deepseek/deepseek-r1-distill-qwen-32b)

Provider: openrouter

Status: ✅

Median Latency: 18953ms

Avg Latency: 17159.0ms

Min: 15365ms | Max: 18953ms

Network Latency (median): 18953.2ms

Success Rate: 2/2

Response Length: 6 chars

303. Gemini Pro Latest (Direct) (gemini-pro-latest)

Provider: google

Status: ✅

Median Latency: 19729ms

Avg Latency: 15503.0ms

Min: 11277ms | Max: 19729ms

Network Latency (median): 19728.9ms

Success Rate: 2/2

Response Length: 42 chars

304. DeepSeek: DeepSeek V3.2 Speciale (deepseek/deepseek-v3.2-speciale)

Provider: openrouter

Status: ✅

Median Latency: 20191ms

Avg Latency: 18022.0ms

Min: 15853ms | Max: 20191ms

Network Latency (median): 20190.7ms

Success Rate: 2/2

Response Length: 6 chars

305. DeepSeek: R1 Distill Llama 70B (deepseek/deepseek-r1-distill-llama-70b)

Provider: openrouter

Status: ✅

Median Latency: 21193ms

Avg Latency: 12676.5ms

Min: 4160ms | Max: 21193ms

Network Latency (median): 21193.5ms

Success Rate: 2/2

Response Length: 34 chars

306. NVIDIA: Llama 3.3 Nemotron Super 49B V1.5 (nvidia/llama-3.3-nemotron-super-49b-v1.5)

Provider: openrouter

Status: ✅

Median Latency: 22227ms

Avg Latency: 22102.0ms

Min: 21977ms | Max: 22227ms

Network Latency (median): 22227.3ms

Success Rate: 2/2

Response Length: 34 chars

307. Qwen: Qwen3 8B (qwen/qwen3-8b)

Provider: openrouter

Status: ✅

Median Latency: 24725ms

Avg Latency: 14249.0ms

Min: 3773ms | Max: 24725ms

Network Latency (median): 24724.6ms

Success Rate: 2/2

Response Length: 104 chars

308. Nex AGI: DeepSeek V3.1 Nex N1 (free) (nex-agi/deepseek-v3.1-nex-n1:free)

Provider: openrouter

Status: ✅

Median Latency: 24839ms

Avg Latency: 21044.0ms

Min: 17249ms | Max: 24839ms

Network Latency (median): 24839.5ms

Success Rate: 2/2

Response Length: 125 chars

309. Z.AI: GLM 4.5V (z-ai/glm-4.5v)

Provider: openrouter

Status: ✅

Median Latency: 27442ms

Avg Latency: 16749.5ms

Min: 6057ms | Max: 27442ms

Network Latency (median): 27442.2ms

Success Rate: 2/2

Response Length: 35 chars

310. OpenAI: GPT-5.2 Pro (openai/gpt-5.2-pro)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: This request requires more credits, or fewer max_tokens. You requested up to 65536 tokens, but can only afford 59544. To increase, visit https://openrouter.ai/settings/credits and add more credits

311. Mistral: Devstral 2 2512 (mistralai/devstral-2512)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: No endpoints found matching your data policy (Paid model training). Configure: https://openrouter.ai/settings/privacy

312. TNG: R1T Chimera (tngtech/tng-r1t-chimera)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: No endpoints found matching your data policy (Paid model training). Configure: https://openrouter.ai/settings/privacy

313. AllenAI: Olmo 3 32B Think (free) (allenai/olmo-3-32b-think:free)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

314. MoonshotAI: Kimi Linear 48B A3B Instruct (moonshotai/kimi-linear-48b-a3b-instruct)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

315. OpenAI: o3 Deep Research (openai/o3-deep-research)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

316. OpenAI: o4 Mini Deep Research (openai/o4-mini-deep-research)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

317. OpenAI: GPT-5 Pro (openai/gpt-5-pro)

Provider: openrouter

Status: ❌ FAILED

Error: Timeout

318. Relace: Relace Apply 3 (relace/relace-apply-3)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

319. Qwen: Qwen3 Coder Plus (qwen/qwen3-coder-plus)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

320. OpenGVLab: InternVL3 78B (opengvlab/internvl3-78b)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: No endpoints found matching your data policy (Paid model training). Configure: https://openrouter.ai/settings/privacy

321. Meituan: LongCat Flash Chat (free) (meituan/longcat-flash-chat:free)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

322. StepFun: Step3 (stepfun-ai/step3)

Provider: openrouter

Status: ❌ FAILED

Error: Timeout

323. OpenAI: gpt-oss-120b (free) (openai/gpt-oss-120b:free)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: No endpoints found matching your data policy (Free model publication). Configure: https://openrouter.ai/settings/privacy

324. MoonshotAI: Kimi K2 0711 (free) (moonshotai/kimi-k2:free)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: No endpoints found matching your data policy (Free model publication). Configure: https://openrouter.ai/settings/privacy

325. TNG: DeepSeek R1T2 Chimera (tngtech/deepseek-r1t2-chimera)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: No endpoints found matching your data policy (Paid model training). Configure: https://openrouter.ai/settings/privacy

326. TheDrummer: Anubis 70B V1.1 (thedrummer/anubis-70b-v1.1)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

327. OpenAI: o3 Pro (openai/o3-pro)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: OpenAI is requiring a key to access this model, which you can add in https://openrouter.ai/settings/integrations - you can also switch to gpt-5-chat or gpt-5-mini.

328. Nous: DeepHermes 3 Mistral 24B Preview (nousresearch/deephermes-3-mistral-24b-preview)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: No endpoints found matching your data policy (Paid model training). Configure: https://openrouter.ai/settings/privacy

329. Arcee AI: Spotlight (arcee-ai/spotlight)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

330. Arcee AI: Maestro Reasoning (arcee-ai/maestro-reasoning)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

331. Arcee AI: Virtuoso Large (arcee-ai/virtuoso-large)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

332. Arcee AI: Coder Large (arcee-ai/coder-large)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

333. Qwen: Qwen3 4B (free) (qwen/qwen3-4b:free)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

334. TNG: DeepSeek R1T Chimera (tngtech/deepseek-r1t-chimera)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: No endpoints found matching your data policy (Paid model training). Configure: https://openrouter.ai/settings/privacy

335. Microsoft: MAI DS R1 (microsoft/mai-ds-r1)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: No endpoints found matching your data policy (Paid model training). Configure: https://openrouter.ai/settings/privacy

336. ArliAI: QwQ 32B RpR v1 (arliai/qwq-32b-arliai-rpr-v1)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: No endpoints found matching your data policy (Paid model training). Configure: https://openrouter.ai/settings/privacy

337. OpenAI: o1-pro (openai/o1-pro)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: This request requires more credits, or fewer max_tokens. You requested up to 65536 tokens, but can only afford 16270. To increase, visit https://openrouter.ai/settings/credits and add more credits

338. AllenAI: Olmo 2 32B Instruct (allenai/olmo-2-0325-32b-instruct)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: No endpoints found matching your data policy (Paid model training). Configure: https://openrouter.ai/settings/privacy

339. DeepSeek: R1 Distill Qwen 14B (deepseek/deepseek-r1-distill-qwen-14b)

Provider: openrouter

Status: ❌ FAILED

Error: Timeout

340. MiniMax: MiniMax-01 (minimax/minimax-01)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Cannot read properties of undefined (reading '0')

341. Google: Gemini 2.0 Flash Experimental (free) (google/gemini-2.0-flash-exp:free)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

342. Anthropic: Claude 3.5 Haiku (2024-10-22) (anthropic/claude-3.5-haiku-20241022)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

343. Meta: Llama 3.2 3B Instruct (free) (meta-llama/llama-3.2-3b-instruct:free)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

344. Nous: Hermes 3 405B Instruct (free) (nousresearch/hermes-3-llama-3.1-405b:free)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

345. Google: Gemma 2 27B (google/gemma-2-27b-it)

Provider: openrouter

Status: ❌ FAILED

Error: Timeout

346. Microsoft: Phi-3 Medium 128K Instruct (microsoft/phi-3-medium-128k-instruct)

Provider: openrouter

Status: ❌ FAILED

Error: Open Router API error: Provider returned error

🏆 FASTEST MODEL

OpenAI: gpt-oss-safeguard-20b (openai/gpt-oss-safeguard-20b)

Median Latency: 344ms

Provider: openrouter

==========

==========