□ Model Settings – 100 MCQs with Answers

Temperature

1.

Q: In language models, what does "temperature" primarily control?
A) Randomness of output □

2.

Q: A temperature value close to 0 makes output:
A) More deterministic □

3.

Q: A temperature value close to 1 makes output:
A) More diverse/random □

4.

Q: Setting temperature to 2.0 will:
A) Increase randomness drastically □

5.

Q: Best temperature for factual Q&A tasks is usually:
A) 0 – 0.3 □

6.

Q: Best temperature for creative writing is usually:
A) 0.7 – 1.0 □

7.

Q: What happens if temperature is set too high?
A) Output becomes chaotic/unreliable □

8.

Q: Temperature balances between:
A) Creativity and determinism □

9.

Q: Is temperature a hard filter or probability reweighting?
A) Probability reweighting

10.

Q: If model always repeats the same output, try:
A) Increasing temperature

Top-p (Nucleus Sampling)

11.

Q: What does top_p (nucleus sampling) control?
A) Cumulative probability mass

12.

Q: Top_p = 1 means:
A) No restriction (full distribution)

13.

Q: Top_p = 0.9 means model samples from:
A) Smallest set of tokens with 90% probability mass

14.

Q: Lowering top_p reduces:
A) Diversity

15.

Q: Raising top_p increases:
A) Diversity

16.

Q: For safe factual answers, set top_p to:
A) 0.8 – 0.9

17.

Q: For highly creative tasks, set top_p:
A) Near 1

18.

Q: Is top_p deterministic?
A) No, it still involves randomness

19.

Q: Top_p and temperature together:
A) Control randomness and diversity

20.

Q: Which is more commonly tuned, temperature or top_p?
A) Temperature

Top-k Sampling

21.

Q: Top-k restricts output to:
A) The top k most likely tokens

22.

Q: Top-k = 1 means:
A) Always choose the most likely token

23.

Q: Top-k = 50 means:
A) Sample from 50 most likely tokens

24.

Q: Smaller k values cause:
A) More deterministic results

25.

Q: Larger k values cause:
A) More diverse/random results

26.

Q: If k is too large, output may:
A) Become incoherent

27.

Q: If k is too small, output may:
A) Become repetitive

28.

Q: Which is more flexible: top-p or top-k?
A) Top-p

29.

Q: If both top-p and top-k are used, final selection is:
A) Intersection of both

30.

Q: Setting top-k to 0 usually means:
A) No restriction

Penalties

31.

Q: Frequency penalty discourages:
A) Repetition of words □

32.

Q: Presence penalty encourages:
A) Exploring new topics/tokens □

33.

Q: High frequency penalty makes text:
A) Less repetitive □

34.

Q: High presence penalty makes text:
A) More exploratory □

35.

Q: Which penalty reduces overuse of common words?
A) Frequency penalty □

36.

Q: Which penalty encourages mentioning new entities?
A) Presence penalty □

37.

Q: Both penalties together improve:
A) Output variety □

38.

Q: Penalties typically range from:
A) -2.0 to +2.0 □

39.

Q: Negative frequency penalty effect:
A) Encourages repetition □

40.

Q: Best penalty setting for essays requiring uniqueness:
A) Presence penalty > 0

Tool Choice

41.

Q: Tool choice decides:
A) Which tool the agent calls

42.

Q: Automatic tool_choice means:
A) Model picks tool itself

43.

Q: Manual tool_choice means:
A) User specifies tool explicitly

44.

Q: Disabling tool_choice prevents:
A) Model from calling any tool

45.

Q: Best setting when user wants strict control:
A) Manual tool_choice

46.

Q: Best setting for automation tasks:
A) Automatic tool_choice

47.

Q: Tool_choice parameter ensures:
A) Correct delegation to tools

48.

Q: Wrong tool_choice may cause:
A) Errors in execution

49.

Q: In exam systems, tool_choice helps in:
A) Directing AI to use calculators, search, etc.

50.

Q: Can model settings override tool_choice?
A) No, tool_choice is explicit

Mixed Questions

51.

Q: Which controls randomness more directly: temperature or top_p?
A) Temperature

52.

Q: Which focuses on cumulative probability mass?
A) Top_p

53.

Q: Which focuses on fixed number of tokens?
A) Top-k

54.

Q: Which prevents repetition?
A) Frequency penalty

55.

Q: Which encourages new topics?
A) Presence penalty

56.

Q: Setting temperature=0, top_p=1 gives:
A) Fully deterministic results

57.

Q: Setting temperature=1, top_p=1 gives:
A) Maximum diversity

58.

Q: Setting top-k=1 is similar to:
A) Greedy decoding

59.

Q: Combining top-p=0.9 and temperature=0.7 results in:
A) Balanced diversity

60.

Q: Setting penalties to high values may cause:
A) Weird, unnatural sentences

Scenario Based

61.

Q: Best settings for math solving?
A) Temperature=0, top_p=1

62.

Q: Best settings for poetry writing?
A) Temperature=0.9, top_p=1

63.

Q: Best settings for chatbot?
A) Temperature=0.7, top_p=0.9

64.

Q: Best settings for legal document summary?
A) Temperature=0.2, top_p=0.8

65.

Q: Best settings for brainstorming startup ideas?
A) Temperature=1, top_p=1

66.

Q: Best settings for SQL query generation?
A) Temperature=0.1, top_p=0.8

67.

Q: Best settings for short story?
A) Temperature=0.8, top_p=0.95

68.

Q: Best settings for factual Q&A?
A) Temperature=0 – 0.2

69.

Q: Best settings for translation?
A) Temperature=0.3 – 0.5

70.

Q: Best settings for customer support bot?
A) Temperature=0.5, top_p=0.9

True/False

71. T/F: Temperature affects probability scaling. □ True

72. T/F: Top-p = 0 disables randomness completely. □ False

73. T/F: Top-k = 1 equals greedy decoding. □ True

74. T/F: Penalties can be negative. □ True

75. T/F: Presence penalty reduces repetition. □ False

76. T/F: Frequency penalty discourages repeating same words. □ True

77. T/F: Tool_choice can be auto or manual. □ True

78. T/F: Setting temperature high guarantees correctness. □ False

79. T/F: Top-p = 1 is equivalent to no filtering. □ True

80. T/F: Frequency penalty improves creativity. □ False

Advanced Understanding

81.

Q: Which parameter adjusts softmax distribution sharpness?
A) Temperature

82.

Q: Which is also called nucleus sampling?
A) Top-p

83.

Q: Which is also called truncated sampling?
A) Top-k

84.

Q: Which parameter helps against "hallucinations"?
A) Lower temperature

85.

Q: Which setting is critical for real-time safety filters?
A) Tool_choice

86.

Q: Overusing penalties can result in:
A) Nonsensical output

87.

Q: Frequency penalty modifies token:
A) Likelihood

88.

Q: Presence penalty modifies token:
A) Likelihood for unseen tokens

89.

Q: Temperature works at what stage?
A) During token probability scaling

90.

Q: Top-k works at what stage?
A) After sorting probabilities

Applied Cases

91.

Q: Which setting ensures chatbot doesn't repeat "hello"?
A) Frequency penalty

92.

Q: Which setting ensures chatbot tries new topics?
A) Presence penalty

93.

Q: Which setting ensures output is reliable for medical use?
A) Low temperature

94.

Q: Which setting ensures model experiments with wording?
A) High temperature

95.

Q: Which parameter ensures fairness between tool selection?
A) Tool_choice

96.

Q: Which parameter ensures model won’t hallucinate math answers?

A) Temperature=0

97.

Q: Which setting to avoid repetitive phrases in stories?
A) Frequency penalty

98.

Q: Which setting helps generate diverse ad slogans?
A) High temperature + high top_p

99.

Q: Which setting helps debugging structured output?
A) Low temperature

100.

Q: Which two parameters together best control randomness?
A) Temperature + Top-p