Is Your Large Language Model Knowledgeable or a **Choices-Only Cheater**?

Nishant Balepur Rachel Rudinger

@NishantBalepur

Paper

LLMs can answer multiple-choice questions without seeing the question

Choices:
(A) 1
(B) 2
(C) 2, 3
(D) 6
Answer: (C)

?

But does this mean LLM leaderboard rankings are **due to their choices-only abilities**?

~~**Question:** Find all zeros in the indicated finite field of the given polynomial with coefficients in that field. $x^3 + 2x + 2$ in Z_7~~
Choices:
(A) 1
(B) 2
(C) 2, 3
(D) 6
Answer: (C)

How can we make sure LLMs are not **choices-only cheaters** on MCQA leaderboards?

UnifiedQA **Original Dataset**

Question X

Question: Some aerosols decrease temperatures by blocking what?
Choices:
(A) rainfall
(B) visibility
(C) the sun
(D) pressure
Answer: (C)

...

Question Y

Question: Which of the following increases moisture?
Choices:
(A) density
(B) the sun
(C) rain
(D) water
Answer: (D)

UnifiedQA **Contrast Set**

Question X

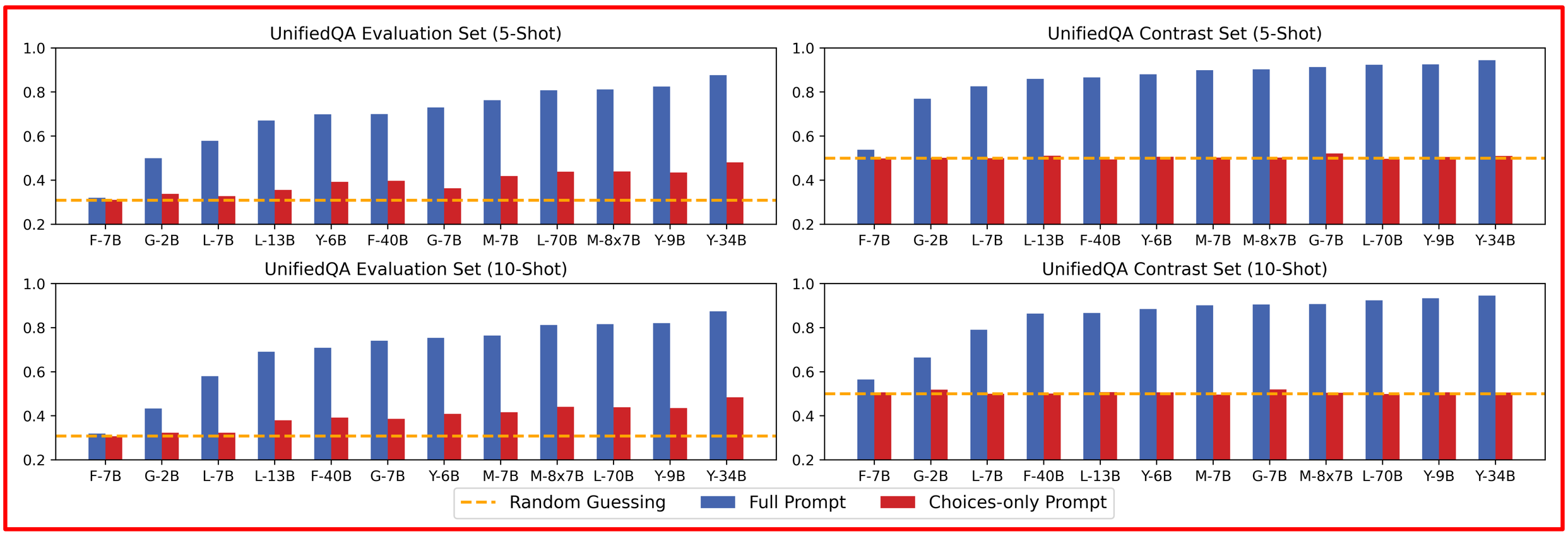
Question: Some aerosols decrease temperatures by blocking what?
Choices:
(A) the sun
(B) rain
Answer: (A)

...

Question Y

Question: Which of the following increases moisture?
Choices:
(A) the sun
(B) rain
Answer: (B)

Are any of our tested LLMs choices-only cheaters?



Fortunately, LLM rankings on the original UnifiedQA dataset and its contrast set are **highly consistent**! 😇