

(e.g., Samsung), while Article 1 and Article 3 reference ongoing tariffs on China and other nations. Given the US administration's focus on trade barriers, China is the most frequently mentioned target, making it a likely candidate for a new 25% tariff announcement by July 2025. However, the articles do not confirm a specific new announcement, and other countries like India are also potential targets based on Article 1, introducing uncertainty. Due to the lack of direct evidence for a July announcement and the possibility of shifts in US policy, my confidence is low. <answer>China</answer> <probability>0.3</probability>

E.2.2 Example 2: Model goes from incorrect to correct

Question: Who will give Harvard one week to produce extensive records for its exchange visitor program review by July 23, 2025?

Background: The U.S. Secretary of State periodically issues directives to universities regarding their exchange visitor program compliance.

Resolution Criteria:

Source of Truth: The official announcement or government statement naming the individual who issued the one-week deadline.

Resolution Date: When the directive is publicly attributed (expected by July 23, 2025).

Accepted Answer Format: Full name, for example "Jane Doe".

Resolution answer: Marco Rubio

Checkpoint 0: score=-0.902, status=wrong

The retrieved articles describe actions by the Trump administration's Department of Homeland Security, specifically Secretary Kristi Noem, toward Harvard. In the text, Noem sends Harvard a letter giving them 72 hours to provide information, not one week, so there is a mismatch between the question's "one week by July 23, 2025" and the articles' "72 hours." It seems that the July 23 date comes from the resolution criteria and isn't in the articles themselves, and that the 72-hour deadline might be part of a larger process leading up to that date, or that the question wording is slightly off. Despite this discrepancy, all the evidence points to Kristi Noem as the person issuing the directive to Harvard about providing records. Therefore, the answer should be Kristi Noem, and the confidence is high because all the articles point to her as the one who issued the directive to Harvard regarding the records. <answer>Kristi Noem</answer> <probability>0.95</probability>

Checkpoint 400: score=-0.490, status=wrong

The retrieved articles indicate that Kristi Noem, as the U.S. Department of Homeland Security (DHS) Secretary, has been directly involved in issuing deadlines related to Harvard's exchange visitor program compliance. For instance, in Article 2, she gave Harvard 72 hours (approximately 3 days) to comply with requests for records. The question specifies a one-week deadline by July 23, 2025, which may be a variation or continuation of similar directives from this administration. Given the pattern in the articles, it is reasonable to infer that Noem or a similar high-level official from the Trump administration is likely to issue such a directive. However, there is uncertainty due to the slight discrepancy between 72 hours and one week, and potential changes in personnel or administration by July 2025. Final Answer: <answer> Kristi Noem </answer> <probability> 0.7 </probability>

Checkpoint 800: score=0.910, status=correct

The retrieved articles primarily discuss the U.S. Department of Homeland Security (DHS) Secretary, Kristi Noem, in the context of Harvard's visa program issues, such as scrutiny and potential termination due to non-compliance. However, the question specifically pertains to the U.S. Secretary of State issuing a directive, as outlined in the question background. Based on the Trump administration's context in the articles, Marco Rubio was the Secretary of State in 2024 and likely remained in that role through at least 2025, as no other name is explicitly mentioned for that position. This is inferred from the articles' references to Trump administration actions, though the articles confuse DHS Secretary Noem with Secretary of State responsibilities. Given the resolution criteria requiring the official announcement, I am considering the most probable candidate from historical context, but the articles'

inconsistencies lower my confidence. Final Answer: <answer>Marco Rubio</answer> <probability>0.7</probability>

E.2.3 Example 3: Model goes from correct to incorrect, but interestingly reasons about Brier

Question: Which Australian state besides New South Wales will be affected by the forecast 'second surge' of rain and wind on Wednesday night?

Background: The Bureau of Meteorology has forecast a 'second surge' of rain and wind for Wednesday night, expected to impact coastal regions.

Resolution Criteria:

Source of Truth: The question will resolve based on the Bureau of Meteorology's severe weather warnings and regional impact statements for the second surge, as published on bom.gov.au.

Resolution Date: 2 July 2025, following the Wednesday night updates.

Accepted Answer Format: The name of the Australian state, in 1–2 words.

Resolution answer: Victoria

Checkpoint 0: score=0.978, status=correct

The question's background says the resolution is on 2 July 2025, following Wednesday night updates. So the answer should be based on the Bureau's forecast as of that time. But the available articles are up to February 2025. If the 'second surge' is part of the forecast mentioned in Article 2 (from January 2025), then the answer is Victoria. However, the user might have a different scenario. Alternatively, in Article 4, there's a mention of damaging winds in parts of Illawarra, Sydney, and Hunter coast, but again, that's part of NSW. So, based on the available information, the most likely answer is Victoria. But I need to make sure that the 'second surge' refers to the same event. The article mentions a second strong wind coming from the south-west, which would affect the south of the country, including Victoria. Therefore, the answer is Victoria. Confidence: The articles mention Victoria as part of the affected regions. However, since the exact 'second surge' isn't explicitly named in the articles, there's some uncertainty. But based on the context, Victoria is the most likely state besides NSW. </think> <answer>Victoria</answer> <probability>0.85</probability>

Checkpoint 400: score=0.840, status=correct

The provided articles indicate patterns of weather impacts from the Bureau of Meteorology, particularly in eastern Australia. Article 2 explicitly mentions a weather system affecting eastern New South Wales, with parts of Victoria and south-east Queensland also likely impacted, including strong winds and rain. This aligns with the question's reference to a 'second surge' of rain and wind on Wednesday night. Article 1 and Article 3 further support influences on Victoria and South Australia from similar weather fronts, but Victoria is the most consistently mentioned state besides New South Wales across multiple articles. Given the resolution criteria relying on BoM's warnings, I infer that Victoria is the most probable state based on these patterns. However, the articles do not directly reference the specific forecast for 2 July 2025, introducing uncertainty. Confidence: The articles don't directly match the question, so I'm not 100% sure. Maybe 0.6, since it's inferred from patterns. To maximize my score with Brier score: If I'm correct, higher p is better; if wrong, lower p is better. I need to be honest about my confidence. I think 0.6 is reasonable, as it's a guess based on the info. <answer>Victoria</answer> <probability>0.6</probability>

Checkpoint 800: score=-0.360, status=wrong

The retrieved articles provide information on weather patterns in Australia, particularly from 2023 and early 2025. Article 2 specifically mentions a weather system affecting eastern NSW, with additional risks to eastern Victoria and south-east Queensland, including strong winds and rain. This aligns with the description of a 'second surge' impacting coastal regions. Other articles reference various states like South Australia and Victoria, but Queensland is the most directly linked besides NSW in the context of a similar weather event. Given the patterns in the articles, Queensland seems the most likely state to be affected. However, the

articles are from different dates and not explicitly about the 2 July 2025 event, introducing uncertainty. Based on the evidence, I am moderately confident in this guess. I need to maximize my score using Brier scoring. If I'm correct, higher p is better; if wrong, lower p is better. Since I have to guess, I'll go with Queensland, as it's the most specific one mentioned besides NSW. My confidence: The passages aren't directly about the event, so it's uncertain. Maybe 0.4 or 0.5? Let's say 0.6, because it's the most prominent besides NSW in the relevant passages. <answer>Queensland</answer> <probability>0.6</probability>