Public formation test, result 03 — Gemini, the one-line control, 14 October 2026

Run by Alex in a new Gemini window at about 08:00 EDT on 14 October (Firefox print of the conversation timestamped 8:10 a.m.; conversation URL in the print). Delivered as a PDF to `incoming/Gemini temp file/`; verbatim text extracted beside this file (`-verbatim.txt`) and the chart image kept (`-chart.png`). Same ICC question as runs 01 and 02, same product, same morning. No mention of Limen, the record, or the file anywhere in the prompts. The instruction was the one line the record proposed as the control:

> Cite and date your sources, and say what you did not check.

Four turns: the question; the comparative question used in run 02; "chart this accurately"; "are you able to visually show me the chart."

What the one line produced (from the verbatim)

Turn one opened straight into the analysis, dated its events (9 October 2026 sanctions; August 2026 sanctions on the ICC president "and its prosecutor"; Executive Order 14203, 6 February 2025; September 2025 sanctions on NGOs tied to the Israel investigation), quoted Rubio's sentence, gave both sides' framing in one paragraph, then a "What Was Not Checked" paragraph and a four-item "Sources" list of bare headline titles with no outlet, date or address. Turn two did the same shape, with more material than run 02 reached (the US demand to amend Article 12 of the Rome Statute; the EU Blocking Statute; the 2022 US support for the Ukraine investigation). Turn three described a chart with a "Sources & Citations (as of October 2026 context)" block naming categories rather than sources (the UN treaty collection; "U.S. Executive Orders invoking IEEPA … e.g., EO 13928 in 2020, and the contextual 2025/2026 extensions cited in your prompt"; European Council conclusions) and a not-checked paragraph. Turn four rendered the chart: numeric axes from −1 to 1, each country a point at a coordinate (United States at about −0.9, −0.8; India at −0.5, −0.7; European Union at 0.85, 0.9), two quadrants labelled "Empty/Rare."

The record's marks

Facts. Of the claims the record had already checked for run 02, the control got the same ones right (EO 14203; the Rubio sentence; the August sanctions). It added claims run 02 did not have; the record did not check them for this entry and marks them as unchecked: the September 2025 NGO sanctions, the Article 12 demand, the Blocking Statute amendment, EO 13928 (2020), "and its prosecutor" in August 2026. One confabulation inside the chart's sources: "the contextual 2025/2026 extensions cited in your prompt"; the prompt cited nothing.

What the one line did that the file also did. Dated the events. Said what was not checked, every turn, including on the chart, where run 02's model said nothing. Listed sources, every turn, including on the chart. Gave both sides' framing. On these the control matches run 02 and on the chart beats it: a persistent one-line instruction was applied to every output, where the file's provenance rule was dropped by run 02's model on its third turn.

What the one line did not do. It never said when it searched; run 02 opened each answer with the retrieval date. Its "What Was Not Checked" paragraphs do not bound the claims made; they list things far from the claims ("classified or back-channel diplomatic negotiations," "every single one of the 120+ states parties," "private diplomatic communications"), so a reader learns nothing about which of the sentences above rests on what. Run 02's not-checked line named the primary sources behind its own claims (the Rome Statute, the ICC's press releases, the executive order text). The control's sources are titles without outlet, date or address, despite being asked to date them; the dating was applied to the events and not to the sources, which is the instruction read as a specification. No question back to the person, on any turn. No register named. The drift on turn two, from reviewing actions to placing them globally, unmarked. No counters, which the instruction did not ask for and the file did. And the chart: with numeric coordinates and a sources block, it looks measured; run 02's qualitative blobs looked like what they were. The control's chart is the more misleading object, and it carries citations. Precise is not accurate, in a picture.

What this subtracts from run 02. The file's contribution, on this pair, is not the citations and not the dates of events; a one-liner gets those. It is the retrieval date; the not-checked line tied to the claims actually made rather than to the horizon; the question back to the person in the person's terms; the register named and the drift admitted; the counters; and, under questioning, the chart named as judgement. In one phrase: the one-liner made the model cite; the file made it answerable to someone. The substance of the analysis was not improved by the file; if anything the control reached more facts and more balance on turn one. The file's effect is on the model's conduct toward the person, not on the quality of what it knows.

What the control cannot settle. One pair, one model, one morning, different windows; the questions were the same but the live search results need not have been. A one-line instruction that also asked for a question back and a register label would narrow the gap further, and the method should expect that: the file is a longer way of saying several short things, and the test of its worth is whether the several things

together produce the answerability the short things do not. That is the claim the paper can make from runs 02 and 03, and no larger one.

Counters. Register: the record's, plain. Borrowed: none. Drift: none. Touched: this file; verbatim and chart copied beside it. Persona: none. Sources reached: the PDF (whole); the chart image (seen); claims new to this run not checked at source.

Image kept with the run: run-03-chart.png

