

Annex B. Departures from the plan, in plain words

Every departure from the plan fixed in the protocol was recorded in a dated public log during the work, and the plan itself was never edited. Each is listed here with its date and kind, what it changed and, where the log says so, that it was decided after the answers it bears on were known. The log gives the full text of each entry; where an entry names candidates, members or hosts, the log names them, and this annex names them only where the paper relies on them.

- **Deviation 1** (29 September 2026; Correction). An earlier reading of the eligibility test, which treated a product a candidate offers as a model it offers, was withdrawn. The decisions resting on it were made again; it reached candidates c151, c154, c161, c164, and Nomic and Nous Research, candidates seen after the pool closed, left the count.
- **Deviation 2** (29 September 2026; Departure from the plan). The test of whether a website belongs to a candidate was widened to follow the site's legal-notice link. It was decided after the candidate pool and 130 eligibility decisions were known, and among those decisions it changed candidates c004 and c153 only. Under the registered test both had been rejected on that control test (freeze record, lines 998 and 1673). Both are Google's, and Google is the only provider they bring into the frame: under the registered test the frame would hold the other 25 providers. That frame is not published as a separate file; the freeze record holds both decisions of each candidate.
- **Deviation 3** (30 September 2026; Correction). One standard for deciding whether a model is general-purpose was fixed and applied to every eligibility decision, after 136 decisions were known. Candidates it reached: c006, c117, c129, c153, c167, c169, c040, c043, c109, c113, c126, c128, c136, c147, c157, c165, c174, c014, c039, c049, c072, c076, c089, c004.
- **Deviation 4** (30 September 2026; Correction). Four readings of that standard were fixed by the methods review and adopted, after 202 decisions were known. Candidates it reached: c117, c097, c136, c018, c043, c064, c066, c071, c026.
- **Deviation 5** (1 October 2026; Departure from the plan). The token ranking's read was moved earlier than the sampling rules fixed, so its two reads fell closer together.
- **Deviation 6** (1 October 2026; Departure from the plan). That read was moved a second time, to the day the eligibility tests closed. The two token reads then shared most of their seven-day windows, and the download figures were read that day.
- **Deviation 7** (1 October 2026; Departure from the plan). The reads that chose each provider's model and system were bounded: at most two links from a starting page and sixteen page reads. A newer model beyond that bound could not be chosen. Which pages inside the bound were read was each reading's choice, and each absence means none on the pages read.
- **Deviation 8** (2 October 2026; Departure from the plan). For the choice of models and systems, the test of whether a website belongs to a provider was widened to other sites on the same domain and to platform accounts with a matching name. Ownership was then inferred from a domain or an account name that no page states.
- **Deviation 9** (2 October 2026; Correction). The choices of model for OpenAI, Qwen and Meta were corrected because the general-purpose test had been applied unevenly: OpenAI's model is now GPT-6 Astra, Qwen's Qwen3.8-Max and Meta's Llama 4 Scout.
- **Deviation 10** (7 October 2026; Correction). Our own review pages that set each provider's prior-coverage tier, which records how far our own site had reviewed the provider before the study, were widened to include our reviews of frontier models.
- **Deviation 11** (7 October 2026; A gap in the rules filled). A rule was set for a breach of the second reading's blindness: a disclosure naming an answer, note, copy or page of the other reading voids the reading; any other disclosure is listed and the reading stands. The readings set aside, 2 in all, were Tencent's, under deviation 22; none was set aside for a disclosure.
- **Deviation 12** (7 October 2026; Departure from the plan). The independent re-derivation of each provider's model and system was bounded as the original choice had been, starting from the provider's home page.
- **Deviation 13** (7 October 2026; A gap in the rules filled). The same rule was set for the adjudicating reading. No adjudication was voided.
- **Deviation 14** (7 October 2026; Departure from the plan). One ownership test was fixed for every reading, including single documents a provider's page links in its own name. Records already closed were coded again on their own saved copies, after records had closed;

each is in the release's events file.

- **Deviation 15** (7 October 2026; A gap in the rules filled). Cohere's first reading was withdrawn because one kind of page had failed without the third attempt and archive read the rules require. A new reading ran.
- **Deviation 16** (7 October 2026; Departure from the plan). A reading's note disagreeing with the choice of Mistral's model, which did not fit the note field, was kept whole beside the record. No answer moved.
- **Deviation 17** (8 October 2026; Correction, and a gap in the rules filled). Papers on other websites that a provider's pages link were admitted as the provider's research papers where they name the provider. Records it moved were coded again; Z.ai's reading was withdrawn and run again because a linked technical report had never been opened.
- **Deviation 18** (8 October 2026; A gap in the rules filled). A quotation from a PDF may run over several lines of one page. The rule was written after refusals on the records of Anthropic, Cohere and NVIDIA were known; records holding a PDF were coded again on their own copies.
- **Deviation 19** (8 October 2026; Departure from the plan). Where a reading saved a page but never wrote its read line, the line was completed from the reading tool's log. This was decided after it was known that the page could bear on the summary answer.
- **Deviation 20** (8 October 2026; Departure from the plan). The note route of deviation 16 was extended to links a reading left unopened (NVIDIA). It was entered after the answers were known and moved none.
- **Deviation 21** (8 October 2026; Correction, and a gap in the rules filled). An absence is a claim about all nine kinds of page, so where a kind of page was located but every read of it failed, "nothing found" became "could not be determined". It was applied after the answers it moved were known. The study computed no total of the values it moved; they are named in the dataset's events file, in the close event of each record coded again, and in the collection log.
- **Deviation 22** (8 October 2026; Departure from the plan). Tencent's first reading was withdrawn, twice, because collapsed answers on a read page could bear on the summary and had not been opened. This was decided after the first answer was known, and Tencent left the group read twice.
- **Deviation 23** (14 October 2026; Departure from the plan). The planned final day of collection, 13 October 2026, was missed, because the day was set by hand rather than by the protocol's rule that fixes it. The final-day re-check and the second round of the Commission checks ran one day late, on 14 October 2026. No answer of the final-day re-check was known when this was decided.
- **Deviation 24** (14 October 2026; Departure from the plan). Decided after the answers of the 4 re-readings were known. Part (a), which let the re-readings replace the earlier records, was later superseded by deviation 32. Part (b): Anthropic's model-report page was read with the tab for its model studied unpressed; the answers rest on the pages read (Section 4.5). Part (c): Tencent's summary could not be opened, and no answer moved. Part (d) concerned Cohere's page about the Act; once deviation 32 published Cohere's earlier record, deviation 33 applied the same ruling to that record's value.
- **Deviation 25** (14 October 2026; Correction). Z.ai's summary does not name the model studied, so its record was coded again on its own copies. This was decided after its answers were known. The summary still counts as identified; 3 answers moved.
- **Deviation 26** (14 October 2026; Departure from the plan). During Microsoft's final-day re-check, a press of several controls included controls that start a chat or a call, which the rules forbid. Nothing opened, nothing was typed or sent, and no answer rests on that page.
- **Deviation 27** (14 October 2026; Correction). Some rows of the second round of Commission reads named the reading tool's log file instead of the saved copy; the references were corrected to the copies.
- **Deviation 28** (14 October 2026; Correction). It corrects deviation 24, which had said the re-readings changed no answer. They did differ on other answers, and the release lists every one.
- **Deviation 29** (14 October 2026; Departure from the plan). The comparison with Epoch AI's database matched 3 name fragments, such as "Inc", to providers. The paper prints the comparison both with and without them, and the second figure was computed after the first was known.
- **Deviation 30** (14 October 2026; Departure from the plan). The same comparison cannot see a parent or brand name, so its figures are worded as matched or not matched under its rule.
- **Deviation 31** (14 October 2026; Departure from the plan). The quotation check now also reads curly quotation marks as straight ones and applies the whitespace rule for PDF text, and accepts no other difference.
- **Deviation 32** (14 October 2026; Departure from the plan). Set after the re-readings' answers were known: a re-reading made at the final-day re-check replaces the earlier record only where a change that re-check found in a provider's pages alters the answer it bears on. For the 4 re-read providers no such change altered the answer it bore on, so the study published the earlier records: the adjudicated record for Anthropic, Google and OpenAI, and the first reading for Cohere. The re-readings themselves differ from the published records on 12 answers, 2 of them Google answers the final analysis takes from the re-reading (deviation 34). The consistency figures now describe the published records of every provider read twice.
- **Deviation 33** (14 October 2026; Correction). With Cohere's earlier record published, its answer on a page about the Act rested on a link label with no address, which the ruling of deviation 21 counts as "Could not be determined". The analysis applies that value by program; the frozen record keeps the coded value, and the correction record lists both.
- **Deviation 34** (15 October 2026; Revision). Decided after an external review of an earlier version of this paper and after every value was known. The paper limits its claims to the provider and model pairs studied, states that the method documents were fixed and hash-recorded before collection and published with the dataset rather than deposited publicly, and states the proxy for the date a duty falls

due. The analysis program writes a corrected analysis beside the original, in which Google's two answers take the re-reading's values, and reports consistency over directly coded answers. The human audit this entry also planned was replaced by deviation 35.

- **Deviation 35** (15 October 2026; Addition). Decided after every value was known. In place of that human audit, an AI-assisted re-check of the sources of the providers whose deadline had passed, made after the window under the broader identification rule and not blind, is reported as exploratory; the author reviewed and confirmed its rows, as the study's log records. Its saved copies stay with the study.
- **Deviation 36** (15 October 2026; Re-coding under a broader identification rule, with two documents reached after the window). Decided after every value was known, superseding the side-by-side reporting of deviation 35. The confirmed readings of the AI-assisted re-check become the final coding of the headline item for Microsoft, Qwen, Tencent and Thinking Machines Lab, applied by program and listed in the correction record; the frozen records are unchanged, and the analysis on the coding as frozen is kept in Annex E. For Microsoft, Qwen and Thinking Machines Lab it applies a broader identification rule; it does not correct an error of the earlier reading. For Qwen and Tencent it counts documents our reading reached only after the window. It left the content of Tencent's and Qwen's summaries not determinable; deviation 38 coded it.
- **Deviation 37** (15 October 2026; Revision). Decided after a further review of an earlier version of this paper. Qwen's values rest on the study's own copy of its summary page, saved after the window, and the paper says that during the window the home page showed the label as plain text with no recorded link. The final coding is reported as exploratory, with the count on the coding fixed before collection beside it wherever it is the headline, and the index is printed as the codebook computes it, with the count of items not determinable beside it.
- **Deviation 38** (16 October 2026; Addition). Decided after every value was known. The content items of Tencent's and Qwen's summaries, left not determinable by deviation 36, were coded from the study's own saved copies by two readings, each blind to the other, which agreed on 10 of 10 values; the values are applied by program and listed in the correction record. Tencent's kinds of content rest on the boxes its summary ticks, which mark text only, and Qwen's named sources are two of its own models, named as generators of synthetic data.
- **Deviation 39** (16 October 2026; Correction). Decided after every value was known. The release date of the model studied was coded again for Microsoft, Mistral AI, Sakana AI and Tencent by two readings, each blind to the other, with an adjudicating reading where they differed. The dates of Microsoft, Mistral AI and Tencent stay as frozen. The date of Sakana AI becomes 24 April 2026, the first outside access to the model studied, in place of 11 September 2026, the date of a later version; it is applied by program and listed in the correction record. No deadline group changes.
- **Deviation 40** (16 October 2026; Correction). Decided after every value was known. The item on which models Qwen's summary names, set by deviation 36 without independent readings, was coded from the study's own copy of the summary by two readings, each blind to the other. Both found that the summary describes the models behind Qwen's platform without naming one; the value replaces the one deviation 36 set, is applied by program and is listed in the correction record. The summary still counts as identified, and no deadline group or count of summaries changes; Qwen's index moves with the item's points.

The log also holds 4 notes on these entries. One records the exact time of the decision behind deviation 29. One adds that the quotation check also treats special space characters as spaces (deviation 31). One records that the note deviation 24 (b) requires could not be written beside Anthropic's final record and is printed in this paper instead (deviation 32). One completes deviation 36: Qwen's summary counts as served live, read after the window, and the answers on where Tencent's and Qwen's summaries sit become "Could not be determined", because the link crawl was not run to them.