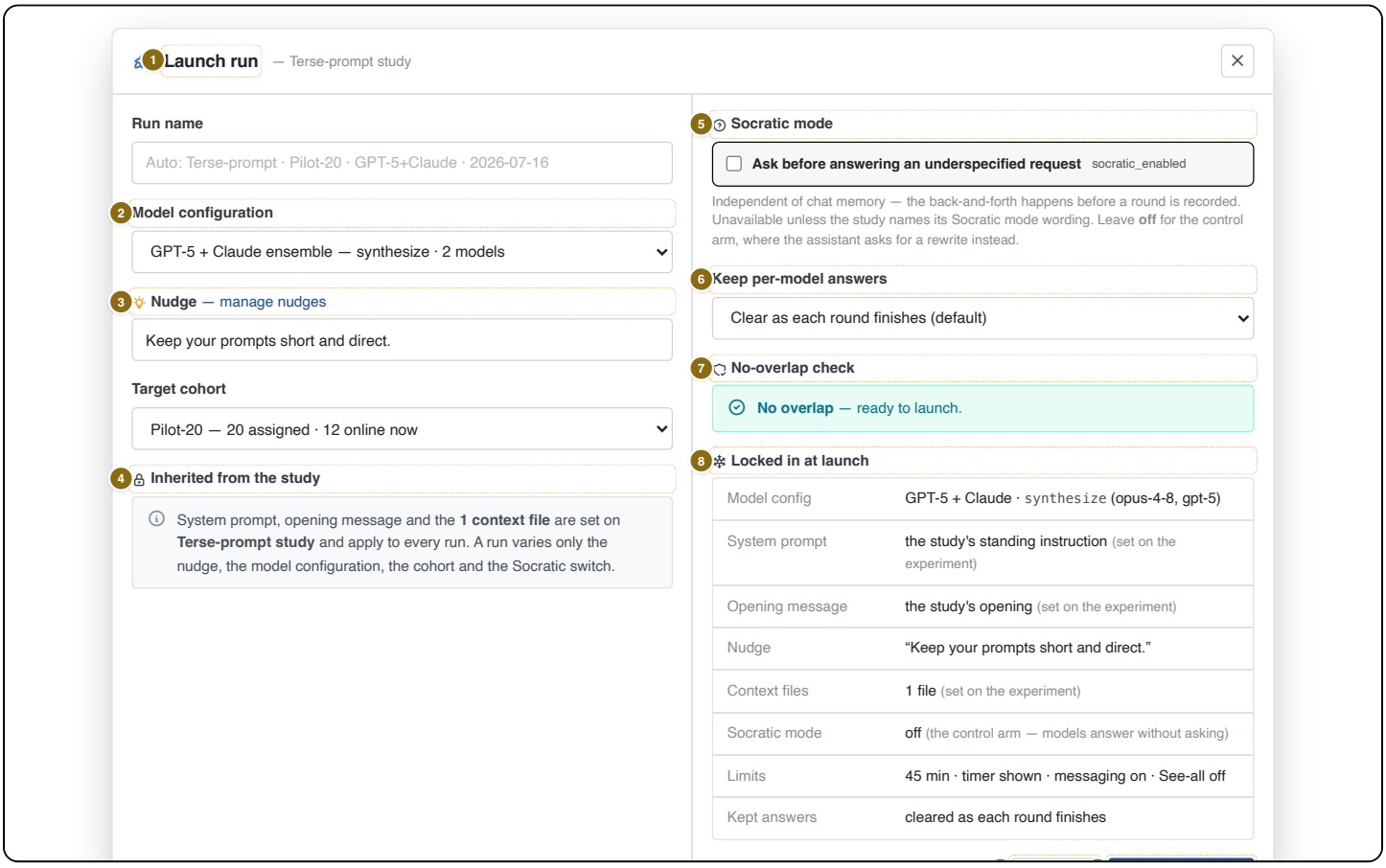
Launch run

launch-run.html · 1160×720



Annotations

1. **Defaults, then frozen** Pre-filled from the experiment's `run_defaults`. This modal is where you pick a cohort and pass the overlap gate; launching **freezes** the condition.

2. **Reusable model set** `model_config_id` — a versioned, named bundle of models + combine method, reused across runs.

3. **The one moving part** `experiment_run.nudge_id` — the single per-run steering variable. Sibling runs in the series vary **only this**. The one steering hint this run tries out; other runs in the series change only this.

4. **Set on the study, not the run** A run is one cell of a comparison grid: it varies the **nudge**, the **model configuration**, the **cohort** and the **Socratic switch**, and holds everything else constant. The standing instruction (`experiment.system_prompt`), the opening message and the context documents (`experiment_context_file`) therefore belong to the study and cannot be overridden here — there is no per-run prompt override and no run-level context file. Changing any of them means editing the study, which is locked once it has runs, or cloning it. The one prompt layering that does exist is per **model**, not per run: `model_config_model.system_prompt`, appended BEFORE the study's instruction. The standing instruction, opening message and context documents come from the study and are the same for every run of it.

5. **The fourth per-run variable** `experiment_run.socratic_enabled` (boolean, pre-filled from `experiment.run_defaults`). It decides how the **sufficiency gate** replies when it stops a request. On: the trusted judge asks for one missing piece at a time and the request is assembled through dialogue, which the models later receive as question-and-answer pairs. Off: the judge names everything missing and asks for one rewritten, self-contained request, and only that final request reaches the models. Detection is identical either way, so the arms are repaired differently rather than held to different standards. It is a **switch, not prompt text**: both phrasings live in `experiment.socratic_version` and are frozen. The models under test never ask anything themselves; the two settings are compared on `run_result.retry_stats` — how many attempts each needed before an answer existed (ALGORITHMS §30). When a request is missing something, have the assistant ask for it one piece at a time instead of asking for a rewrite. Vary this between runs to measure whether it helps.

6. **Housekeeping** `experiment_run.keep_candidates` (boolean, default `false`) — whether the per-model, per-iteration rows behind “See all” are retained after each round finalizes. False clears them as the run proceeds, which is the leanest storage and still leaves the “See all” viewer working live during a round; true keeps them for later side-by-side review. It is pre-filled from `experiment.run_defaults` and can be changed here. Candidate rows go anyway when their round or run is deleted.

7. **Overlap gate** The disjoint-cohort gate blocks launch when the chosen cohort overlaps enrollees already live in another running/paused run — e.g. “All enrollees” would overlap 3 active in **Politeness · SLM**, blocking launch until you pick a disjoint cohort. Launch is blocked if this cohort shares enrollees with another running or paused run — so no one is in two live runs at once.

8. **Frozen by immutability, not by a copy** There is **no** snapshot column. The run stores foreign keys to the three swept definitions (`nudge_id`, `model_config_id`, `cohort_id`) plus its parent experiment, and the database holds every one of those rows unchangeable while any run references them (an `ON DELETE RESTRICT` rule plus read-only substantive fields). Re-resolving through those keys therefore reproduces the exact condition at any later date. A copy would be a second source of truth that can silently disagree with the first, which is why the design does not keep one. Values that are *not* protected this way are frozen where they are read instead — model prices into `chat_round_candidate.token_cost`, the context window into `chat_round.history_window`. Editing the study or the configuration later never affects this run — the definitions it points at are held unchangeable for as long as it references them.

9. Discard changes and close.

10. Start this run for online enrollees now.