Workstream Comparison: Use of Cumulative Sums of Squares (GPT-5 vs GPT-5-nano)

Inputs

- Baseline (archive): `reports/archived/workflow-report-1308532de7a9446d813e57129826aa71.json`
- Variant (current): `reports/workflow-report-34c88b2bf0194f5b9b72793845290a52.json`

Schema

Field	Type	Purpose
`comparison_group_id`	`text`	Logical key that groups runs into one composed workstream com
`source_bucket`	‘enum(archive	current)‘
`report_path`	`text`	Path to the source workflow-report JSON used for this row.
`run_id`	`text`	Unique workflow execution identifier.
`papers_dir`	`text`	Input paper path(s) processed by the run.
`started_at`	`timestamptz`	Workflow start timestamp.
`finished_at`	`timestamptz`	Workflow end timestamp.
`duration_sec`	`numeric`	Run duration in seconds for cost/latency comparison.
`chat_model`	`text`	Primary LLM used by the run.
`embedding_model`	`text`	Embedding model used for retrieval/chunk similarity.
`report_question_set`	`enum`	Question set mode (structured
`report_question_count`	`integer`	Number of structured/agentic report question records emitted.
`agentic_status`	`text`	Top-level status for agentic stage.
`index_status`	`text`	Top-level status for index stage.
`index_reason_or_error`	`text`	Failure/skipped reason to diagnose differences.
`report_store_status`	`text`	Whether report was stored to Postgres.
`prep_corpus_hash`	`text`	Fingerprint of corpus content; validates same input set.
`config_hash`	`text`	Effective configuration hash; validates run comparability.
`confidence_mean`	`numeric`	Mean confidence score across structured questions.
`confidence_labels`	`text`	Label distribution summary for quick quality profile.
`final_answer_preview`	`text`	Truncated synthesis preview for side-by-side semantic comparison

Records

comparison_group_id	source_bucket	report_path
icss-gpt5-vs-gpt5-nano-v1	archive	reports/archived/workflow-report-1308532de7a9446d813e5712982
icss-gpt5-vs-gpt5-nano-v1	current	reports/workflow-report-34c88b2bf0194f5b9b72793845290a52.jso

Comparison Notes

- Both runs process the same paper path: `papers\Use_of_Cumulative_Sums_of_Squares_for_Re.pdf`.

- Baseline model: `gpt-5`; variant model: `gpt-5-nano`.
- Baseline duration: `778.32` sec; variant duration: `388.08` sec.
- Baseline confidence mean: `0.27249828745446014`; variant confidence mean: `0.2724441085027672`.
- Baseline report-question count: `83`; variant report-question count: `83`.
- Baseline index status: `skipped` (db_unreachable); variant index status: `indexed` (None).
- Baseline report store: `skipped`; variant report store: `stored`.