Workstream Comparison: Alcott Food Deserts (GPT-5 vs GPT-5-nano)

Sources

- Baseline (archive): `reports/archived/workflow-report-d1617aa16048423daa22ca7e29c3526c.json`
- Variant (current): `reports/workflow-report-e53e1c45434547afaedf7a0d079d5615.json`

Workstream Table Schema

Column	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

Composed Workstream Rows

comparison_group_id	source_bucket	report_path
alcott-food-deserts-gpt5-vs-gpt5-nano-v1	archive	reports/archived/workflow-report-d1617aa160484
alcott-food-deserts-gpt5-vs-gpt5-nano-v1	current	reports/workflow-report-e53e1c45434547afaedf7a0

Focused Comparison (Model)

Metric	GPT-5 (baseline)	GPT-5-nano (variant)
run_id	`d1617aa16048423daa22ca7e29c3526c`	`e53e1c45434547afaedf7a0d079d5615`
chat_model	`gpt-5`	`gpt-5-nano`
duration_sec	`845.17`	`406.22`
report_question_count	`84`	`84`
confidence_mean	`0.3041450240118245`	`0.3042301084348596`
confidence_labels	`high=0, medium=18, low=65`	`high=0, medium=18, low=65`
index_status	`skipped`	`failed`
report_store_status	`skipped`	`stored`

Notes

- Config hash differs between runs (expected when model defaults changed).
- GPT-5-nano run completed full agentic+structured flow (`report_question_set=both`).
- New nano audit report: `reports/audit-workflow-report-e53e1c45434547afaedf7a0d079d5615.md`.