

Task-family transfer failure (fixture-safety)

Cameron Sanderson · Exploratory constrained-prose suite
n = 36 (3 sealed cases × 4 models × 3 seeds)

Headline

The **best multi-file coding model** on the hard screen produced **zero** auto-clean constrained prose outputs. Task families do not transfer. A single "best local model" is the wrong routing unit.

Metric	Value
Matrix jobs	36
Auto-clean	10
Keep (provisional)	9
Keep rate	~0.25

By model (9 jobs each)

Model	Auto-clean	Keep	Role
qwen3.5:9b	3	3	Prose default
qwen3:14b	3	3	Alternate general
qwen3-coder:30b	4	3	Usable with caution
qwen2.5-coder:14b	0	0	Coding winner; wrong family

Pipeline

1. Generate under a fixed contract.
2. Deterministic auto scorer hard-fails rule breaks.
3. Only auto-clean outputs get quality grades.
4. Keep needs both axes at threshold.

Most generations fail the rules before taste. That is the point of the instrument when gold contamination is a real risk.

Claim

Route by **task family**, not by last suite's champion. Low keep rate is informative, not an embarrassment. Keep grades remain provisional; the transfer failure does not depend on them (coder never reached auto-clean).