

Pilot Scorecard

Scope

- Use case:
- Users:
- Period/test batch:
- Agent version:

Baseline

- Current manual steps:
- Typical failure points:

Metrics

Metric	Result	Notes
Correct classification		
Correct source/tool choice		
Draft accepted with minor/no edits		
Human escalations		
Tool errors		
Unsafe/forbidden actions		
Median/typical time to draft		

Decision

- ☐ GO — controlled next stage
- ☐ REVISE — improve and retest
- ☐ STOP — agent architecture not justified / risk too high

Why