

AI Output QA Checklist

HOLD Check what the AI in your stack produces.

The common failure with AI in a stack is not bad output. It is output nobody checked, going straight to a customer or into a record that drives a decision. One control matters most: a human looks before it ships or takes effect. You cannot hand-check everything, so match the scrutiny to the stakes.

Every AI output.

- ☐ Accurate. No invented specifics, numbers, or names.
- ☐ On brand and appropriate.
- ☐ Safe. No leaked data, no compliance problem, nothing biased.
- ☐ Reviewed by a person before it went out or took effect.
- ☐ Traceable. You can tell it was AI-generated and audit it later.

By output type.

- ☐ Content (emails, copy): claims accurate, brand voice holds, personalization resolves, a person reviews before send.
- ☐ Lead scores: spot-checked against leads you know are good or bad, not silently misrouting, owned by someone.
- ☐ Data enrichment: writes correct values, does not overwrite good data, source noted, writes sample-checked.
- ☐ Chat and agent replies: accurate and on brand, escalates to a person when it should, does not promise what it should not.
- ☐ Summaries feeding decisions: accurate, does not overstate confidence, the people deciding know it came from AI.

Match review to stakes.

Output	Review level
Customer-facing or decision-driving	A person reviews every time
Low-stakes internal	Spot-check, not every item

When in doubt, review it. A bad output in front of a customer costs more than a check.