

AI agent job description & first-month scorecard

Describe the agent's job on one page before anything is built. Then, for its first month, have every output arrive as a draft a person approves, with nothing acting on its own, and score each week on the pages that follow.

01 The job description

Agent name / the process it owns

The goal, in one sentence

Inputs it watches (inbox, folder, CRM view, schedule)

Tools it may use (each one is a permission decision)

Rules and red lines (tone, never-do items, escalation triggers)

Where a human approves (suggested: everything that leaves the firm)

Owner (named person)

02 The first month, scored weekly

All month, every output is a draft a person approves; nothing sends, saves or acts on its own. Fill in one block at the end of each week.

Week 1

Items handled

Drafts approved unchanged / edited / rejected (three numbers)

Errors a reviewer caught, and their type

Hours returned to the team (estimate honestly)

Notes

Week 2

Items handled

Drafts approved unchanged / edited / rejected (three numbers)

Errors a reviewer caught, and their type

Hours returned to the team (estimate honestly)

Notes

Week 3

Items handled

Drafts approved unchanged / edited / rejected (three numbers)

Errors a reviewer caught, and their type

Hours returned to the team (estimate honestly)

Notes

Week 4

Items handled

Drafts approved unchanged / edited / rejected (three numbers)

Errors a reviewer caught, and their type

Hours returned to the team (estimate honestly)

Notes

03 Month-end decision

The numbers above decide; impressions do not.

Totals for the month (handled / unchanged / edited / rejected / errors / hours returned)

Decision (let it act on routine items with consequences still approved / keep approving everything / stop, and why)

Next process to give it, if any

04 Go/no-go checklist

Every action the agent takes is logged and inspectable

Spending and volume caps enforced outside the agent

The stop switch exists, is known to two people, and has been tested

The owner reviews the log on a stated cadence