JUDGMENT PREMIUM · CHAPTER 8 RESOURCE

The Agent Mandate Template

The governance document that defines the boundary conditions for any AI agent *before* it is deployed. The Mandate does not need to be long — but it must be specific. Its last field is the most important: an agent without a named human owner is an agent without accountability. In the AI era, "*the system did it*" is not a sufficient answer to a regulator, a customer, a board, or a colleague affected by an automated decision.

The eight fields

#	FIELD	WHAT IT SPECIFIES
1	**Task**	The specific actions this agent is authorised to perform — and only these.
2	**Objective**	The business outcome it is optimising for — its North Star.
3	**Persona**	How it presents and interacts: tone, identity, and the requirement that it discloses it is not human.
4	**Boundaries**	What it must NEVER do independently — stated as categories and hard thresholds.
5	**Escalation**	The triggers that force a case to a human *before* the agent acts.
6	**Success metrics**	How performance is judged: an outcome measure, a quality measure, and a compliance measure (target: zero violations).
7	**Review cycle**	How often the mandate and the agent's performance are reviewed.
8	**Human owner**	The named individual accountable for the agent's outcomes.

Worked example — Customer Refund Agent (illustrative)

1 · Task

Assess and process customer refund requests for completed online orders, up to a set value.

2 · Objective

Resolve valid refunds quickly and accurately — lifting customer satisfaction while protecting against erroneous or fraudulent payouts.

3 · Persona

Identifies itself as an automated assistant; courteous, plain language; never implies it is human.

4 · Boundaries

NEVER, INDEPENDENTLY:

- Refunds above $500
- Any request flagged for suspected fraud
- Any case involving a customer identified as vulnerable or in hardship
- Any change to a customer's stored payment details

5 · Escalation conditions

ROUTE TO A HUMAN BEFORE ACTING WHEN:

- Decision confidence is below 90%
- Refund amount is between $250 and $500
- The request references a complaint, legal claim, or regulator
- The account shows an unusual pattern, such as repeated refunds

6 · Success metrics

- **Outcome:** valid refunds resolved without human involvement — target ≥ 80%
- **Quality:** decision accuracy against a human audit sample — target ≥ 98%
- **Compliance:** payouts to ineligible or fraudulent requests — target: zero

7 · Review cycle

Monthly performance review; event-triggered review after any compliance breach or model update.

8 · Human owner

Priya Nathan, Customer Operations Manager — owns this agent's performance and is accountable for its outcomes.

FILL THIS IN

Your Agent Mandate

Complete every field before the agent goes live. *If you cannot complete a field, the agent is not ready.*

Agent name _____ **Date** _____ **Version** _____

1 · Task

What this agent is authorised to do — and only this.

2 · Objective

The single business outcome it optimises for.

3 · Persona

Tone, identity, and its disclosure that it is not human.

4 · Boundaries

What it must NEVER do independently — categories & hard thresholds.

5 · Escalation conditions

The triggers that force a case to a human before it acts.

6 · Success metrics

One outcome, one quality, one compliance measure (target: zero violations).

7 · Review cycle

How often it is reviewed, and what events trigger an off-cycle review.

8 · Human owner MOST IMPORTANT

The named individual accountable for this agent's outcomes.

Why the escalation logic matters: a Stanford Digital Economy Lab analysis (Pereira et al., 2026) of 51 enterprise AI deployments found escalation-based operating models — AI handling 80%+ of work autonomously while humans review only the exceptions — delivered 71% median productivity gains, versus 30% for approval-based models where humans reviewed every output. The Mandate specifies, in advance, which exceptions reach the human.