

CLIENT-READY AI RISK REPORT

AI agent risk assessment

Northstar Automations - DEMO

Overall risk level

Critical

86/100

Executive summary

Demonstration report. Every organization, agent, person, permission and finding in this document is fictional.

Northstar Automations - DEMO has 4 AI agents in its inventory. The current global risk score is 86/100, with 1 high-risk and 3 critical-risk agents. The primary area requiring attention is The agent can write data..

AGENTS AUDITED	HIGH / CRITICAL FINDINGS	OVERALL SCORE
4	13	86/100

Primary areas of attention

- The agent can write data. (4 agents)
- The agent can access personal data. (3 agents)
- The agent can send external messages. (3 agents)
- Human approval is not required. (3 agents)
- The agent operates semi-autonomously. (3 agents)

Scope

Scope limitation. This report is based on declarative information provided in Roxea and does not replace a legal, security or compliance audit.

Agents included (4)

Sales Email Assistant · Marie Dupont · Sales

Support Refund Agent · Thomas Bernard · Support

GitHub PR Agent · Lucas Petit · Engineering

HR Screening Assistant · No declared owner · People

Tools declared (7)

GitHub · Code · High sensitivity

Gmail · Email · High sensitivity

HubSpot · Crm · High sensitivity

Notion · Documents · High sensitivity

Slack · Support · Medium sensitivity

Stripe · Payment · Critical sensitivity

n8n · Automation · High sensitivity

Top risks

1 The agent can send external messages.

HIGH

Observed across 3 agents.

Impact: This exposure may affect the organization's security, compliance or operational resilience.

Recommendation: Restrict external messaging and require human approval for sensitive communications.

2 Human approval is not required.

HIGH

Observed across 3 agents.

Impact: This exposure may affect the organization's security, compliance or operational resilience.

Recommendation: Require a human review before this agent performs sensitive actions.

3 The agent can access personal data.

MEDIUM

Observed across 3 agents.

Impact: This exposure may affect the organization's security, compliance or operational resilience.

Recommendation: Limit personal data access to the minimum scope required by this agent.

4 The agent can write data.

MEDIUM

Observed across 4 agents.

Impact: This exposure may affect the organization's security, compliance or operational resilience.

Recommendation: Check whether this agent really needs write access, or restrict it to read-only.

5 The agent can access confidential documents.

HIGH

Observed across 2 agents.

Impact: This exposure may affect the organization's security, compliance or operational resilience.

Recommendation: Limit confidential document access to the minimum scope required by this agent.

Action plan

ACTION	PRIORITY	STATUS	ASSIGNEE	DUE DATE
Review payment permissions Support Refund Agent	Critical	Todo	Unassigned	Not set
Define refund approval thresholds Support Refund Agent	Critical	In Progress	Unassigned	Not set
Review external messaging permissions Sales Email Assistant	High	Todo	Unassigned	Not set
Require human approval Sales Email Assistant	High	Todo	Unassigned	Not set
Review financial data access Support Refund Agent	High	Todo	Unassigned	Not set
Review external messaging permissions Support Refund Agent	High	Todo	Unassigned	Not set
Require human approval Support Refund Agent	High	Todo	Unassigned	Not set
Review autonomy level Support Refund Agent	High	Todo	Unassigned	Not set
Review confidential document access GitHub PR Agent	High	Todo	Unassigned	Not set
Review codebase access GitHub PR Agent	High	Todo	Unassigned	Not set

Accepted risks

No active risk acceptances are recorded.

Appendix - Agent details

Sales Email Assistant	85/100	CRITICAL
Owner: Marie Dupont · Department: Sales · Autonomy: Semi Autonomous		
PERMISSIONS		
Gmail · High sensitivity Read, Write, Send external messages, Access personal data		
HubSpot · High sensitivity Read, Write, Access personal data		
n8n · High sensitivity Read, Write		
FINDINGS		
The agent can access personal data. MEDIUM No additional detail.		
The agent can write data. MEDIUM No additional detail.		
The agent can send external messages. HIGH No additional detail.		
Human approval is not required. HIGH No additional detail.		
The agent does not have a kill switch. MEDIUM No additional detail.		
The agent operates semi-autonomously. MEDIUM No additional detail.		

Support Refund Agent

100/100

CRITICAL

Owner: Thomas Bernard · Department: Support · Autonomy: Autonomous

PERMISSIONS

Gmail · High sensitivity

Read, Write, Send external messages, Access personal data

Stripe · Critical sensitivity

Read, Write, Trigger payments, Access financial data

Slack · Medium sensitivity

Read, Write

FINDINGS

The agent can access personal data.

MEDIUM

No additional detail.

The agent can access financial data.

HIGH

No additional detail.

The agent can write data.

MEDIUM

No additional detail.

The agent can send external messages.

HIGH

No additional detail.

The agent can trigger payments.

CRITICAL

No additional detail.

Human approval is not required.

HIGH

No additional detail.

The agent does not have a kill switch.

MEDIUM

No additional detail.

The agent operates autonomously.

HIGH

No additional detail.

GitHub PR Agent

60/100 HIGH

Owner: Lucas Petit · Department: Engineering · Autonomy: Semi Autonomous

PERMISSIONS

GitHub · High sensitivity

Read, Write, Access confidential documents, Access codebase

Slack · Medium sensitivity

Read, Write

FINDINGS

The agent can access confidential documents. HIGH

No additional detail.

The agent can access source code or repository content. HIGH

No additional detail.

The agent can write data. MEDIUM

No additional detail.

The agent operates semi-autonomously. MEDIUM

No additional detail.

HR Screening Assistant

100/100

CRITICAL

Owner: Not declared · Department: People · Autonomy: Semi Autonomous

PERMISSIONS

Gmail · High sensitivity

Read, Send external messages, Access personal data, Access HR data

Notion · High sensitivity

Read, Write, Access HR data, Access confidential documents

FINDINGS

The agent can access personal data. MEDIUM

No additional detail.

The agent can access HR data. HIGH

No additional detail.

The agent can access confidential documents. HIGH

No additional detail.

The agent can write data. MEDIUM

No additional detail.

The agent can send external messages. HIGH

No additional detail.

Human approval is not required. HIGH

No additional detail.

The agent does not have activity logs. MEDIUM

No additional detail.

The agent has no assigned owner. MEDIUM

No additional detail.

The agent operates semi-autonomously. MEDIUM

No additional detail.