NORTHWIND COMMERCE · AI GOVERNANCE

AI Policy — Version 1.1

Effective 2026-04-22 | Owner: Aria · AI Governance Lead | ISO/IEC 42001:2023 aligned

---

Northwind Commerce is a mid-size direct-to-consumer retailer (apparel, home goods, 140 staff, 9 markets in EU + UK). This policy tells our people **exactly which AI tools to use, which data to put in them, and what to do when something goes wrong**. It is opinionated on purpose. A policy that says "use AI responsibly" is a policy that says nothing.

If you are reading this at 11pm wondering whether to paste a customer email into a chatbot, the answer is in Section 3. Stop reading at Section 3 if you need to.

1. Scope & Intent

Applies to every Northwind employee, contractor, and agency partner who uses, builds, or buys AI. Covers generative AI (text, code, image, voice, video), predictive models (fraud, forecasting, recommendations), and autonomous agents.

> **Our stance.** AI is a force multiplier for a 140-person company competing with 10,000-person incumbents. We adopt it deliberately, we pick specific vendors, and we never let it become the reason a customer loses trust in us.

2. Principles (the non-negotiables)

1. **Human accountability.** A named human is accountable for every AI-influenced decision that affects a customer, employee, or financials. AI never signs contracts, fires people, sets final prices, or approves refunds above threshold.
2. **Transparency.** If a customer is talking to AI, we tell them. AI-generated content is labelled where context matters.
3. **Data minimisation.** Anonymise, aggregate, or mask before paste. Always.

4. **Proportional control.** Rules scale with impact.
5. **Continual improvement.** Reviewed quarterly. Incidents feed the next version.

3. Data Handling — the Three-Light Rule

Light	Data type	Where it may go
🟢 GO	Product catalogue, public marketing copy, anonymised/aggregated analytics, generic how-to questions, our own public docs	Any approved tool in Section 4
🟡 PAUSE	Internal non-customer data (financial drafts, supplier terms, internal wiki, unreleased roadmap)	**Claude Enterprise** or **ChatGPT Team/Enterprise** only. Never free/personal tiers.
🔴 STOP	Customer PII (name, email, address, phone, order history with identifiers), payment data, passwords, API keys, health data, data about children, employee performance data, legal/HR records	**Never.** Not in any LLM tool. Not even "just to test". Not even masked if re-identification is plausible.

Screenshots count. Code that contains real data counts. CSV exports count. If you can mentally re-identify a person from it, it's STOP-tier.

4. Approved LLM Providers & Tools

We are opinionated about which models touch our data. The decision is about **where the request goes, who controls the server, and whose laws apply to it.**

4.1 Approved providers

Provider	Status	What we use it for
Anthropic (Claude) Zero-retention, no-training on Enterprise. US, EU option available.	✓ PRIMARY	Default for drafting, analysis, code, customer support, internal agents
OpenAI (ChatGPT Team/Enterprise, API) Enterprise tier zero-retention + no-training. EU region on Enterprise.	✓ APPROVED	Image generation (GPT Image), specific workflows, Claude fallback
GitHub Copilot (Business/Enterprise) Microsoft/OpenAI pipeline, no-training on Business+.	✓ APPROVED	Code suggestions only

Perplexity Enterprise Cited research; Enterprise tier does not train on queries.	✓ APPROVED	Named research tasks. Never with customer data.
Google Gemini, Mistral	⏸ ASK ARIA	Not banned, not approved. Reassessed each quarterly review.

4.2 Providers we do NOT use

> **No Chinese-hosted models.** DeepSeek, Qwen (Alibaba), Kimi (Moonshot), Doubao, GLM, ERNIE, or any model hosted inside the PRC. Data may be legally compelled by Chinese state authorities under the National Intelligence Law (Art. 7), Data Security Law, and PIPL. No reliable DPA we can enforce in EU/UK jurisdiction. Not worth the category of risk, however capable the models are.

Not allowed	Reason
DeepSeek, Qwen, Kimi, Doubao, GLM, ERNIE (any Chinese-hosted LLM)	PRC data-compulsion regime. No enforceable EU/UK DPA.
Free/personal consumer tiers (ChatGPT free, Gemini free, Claude free web, Copilot Free)	Their terms allow training on your prompts. What you paste can end up in someone else's answer.
Random "ChatGPT wrapper" SaaS	Usually a one-person shop proxying OpenAI with no DPA, no SOC 2. If it's not on the approved list, ask Aria first.
Open-source models on personal laptops with company data	Not the model's fault — the problem is the laptop, the backups, and the person who leaves.

4.3 Do we need on-prem / in-house inference?

> **Short answer: no, and you probably don't either.** We are a 140-person retailer. A cluster of H100s depreciates faster than a phone, needs a dedicated ML-ops engineer, and still gives you a smaller model than Claude Sonnet. Running your own inference only makes sense for HIPAA-regulated healthcare handling raw PHI, classified government/defence work, or >$1M/year API

> spend with a measured ROI case. Revisit annually or when our API spend crosses $500K/year, whichever comes first.

5. Role-Based Rules

5.1 Customer Service

- **Tool**: Claude Enterprise via our Internal AI Desk. Ticket → AI draft → human edits → send.
- A human reviews and edits **every** reply that leaves Northwind. Non-negotiable.
- Refunds above €200, account deletions, anything involving a minor: human-only decision.
- When a customer asks "am I talking to a human?", the honest answer is given.
- **Never paste** customer name + email + order details into any tool outside the Internal AI Desk — the Desk already has full context.

5.2 Marketing & Content

- **Tools**: Claude (drafts), Claude Projects (brand voice), ChatGPT with GPT Image / Seedream (visuals), Midjourney (moodboards only).
- A human editor is accountable for every piece published — hallucinations that reach customers are on the editor, not the model.
- AI-generated images never depict identifiable real people without consent, and never imitate another brand's visual identity.
- Label AI-generated product photography. Stylised illustration does not need labelling if clearly illustrative.

5.3 Engineering

- **Tools**: Claude Code (primary), GitHub Copilot Business (inline), Claude in Cursor / VS Code.
- **Never in prompts**: production API keys, database credentials, real customer records, `.env` contents. If you see a secret being pasted to a chat, stop your colleague.
- AI-generated code is reviewed by a human before merge. Same bar as human-written code.

- Infra changes (Terraform, migrations, IAM) require a human to read every line AI generated before apply. "Claude said it was fine" is not a defence.

5.4 Merchandising & Pricing

- Forecasting, inventory, and recommendation models run autonomously inside documented bounds.
- **Price changes above ±8% of baseline require human approval.**
- Dynamic pricing does not use postcode, device, inferred income, or any feature correlating with a protected characteristic.

5.5 HR & Hiring

- **No AI makes a hire/no-hire decision. Period.**
- AI may help screen CVs against explicitly stated role skills; the ranked shortlist is reviewed and justified in writing by a human recruiter.
- Performance reviews, compensation, and terminations are not drafted by AI.

5.6 Finance & Fraud

- Fraud models may auto-block. Customers always get a human-reviewable appeal path within 24h.
- Accounting automation welcomed. Monthly reconciliation is a human sign-off.
- Never paste payroll files, bank statements, or vendor invoices into any LLM tool.

6. Practical Tips — How We Actually Use Claude

Claude is our primary model. Staff get faster results if they use it the way it is designed.

6.1 claude.ai (web chat)

- **Use Projects.** A Project bundles system prompt + reusable files + chat history for a recurring task. Cuts setup to near zero.
- **Attach files** rather than pasting long text — better context handling, cleaner chats.
- **Use Artifacts** for deliverables you iterate in-place (docs, HTML, React components).

- **Data posture**: on Claude Enterprise, nothing is used for training. PAUSE-tier data is fine. STOP-tier is still never fine anywhere.

6.2 Claude Code (terminal / IDE)

- **Put a `CLAUDE.md`** at the root of every repo. 20 lines. Stack, conventions, how to run tests. Claude reads it first and stops guessing.
- **Use `/init`** in a new repo — generates a first-pass `CLAUDE.md` you then edit.
- **Small focused tasks**. Great at a 50-line diff, wobbly at 2000-line refactors.
- **Let it run the tests.** After changes, let Claude Code run `bun test` / `pytest`. Saves a review round.
- **Check in the chat log** (`/export`) when a reviewer needs to see the reasoning, not just the diff.

6.3 Claude Cowork (MCP + autonomous workflows)

- Use Cowork when Claude should run on schedule or react to external events (Slack, support ticket, calendar).
- Start small — one trigger, one action, one Slack notification with a human approval button.
- Every Cowork agent has a **named human owner** in its config (see AI register in Section 9).

6.4 Good-chat hygiene

- **One task per chat.** Context rot is real; a week-long chat confidently remembers things from five topics ago.
- **Ask Claude to cite sources** on factual claims. If it can't, treat the claim as a hypothesis to verify.
- **If the answer feels too confident**, ask "what could make this wrong?" Claude is good at auditing its own output when asked.

6.5 Prompt templates we keep around

Stored in the Aria Project on Claude Enterprise; anyone can copy-paste:

- "Summarise this support ticket thread for a manager — decisions, open questions, blockers."

- "Draft a reply to this customer. Warm tone. Do not commit to anything I haven't told you is true."
- "Read this PR diff. Flag security issues, performance regressions, missed edge cases."
- "Given these three options, argue for each, then tell me which you'd pick and why."

7. Incident Response — the 5-Minute Rule

> 1. **SPOT** — notice it.
> 2. **SCREENSHOT** — prompt, output, tool, timestamp.
> 3. **POST** in `#ai-incidents` within 5 minutes. No blame. Just facts.
> 4. **PAUSE** the tool for that task until Aria clears it.
>
> **Hiding an incident to avoid embarrassment is the only AI-related behaviour at Northwind that results in disciplinary action.**

8. Vendors & Third Parties

Any supplier who touches Northwind customer data with AI must:

- Contractually **not train** on our data.
- Disclose model family, training-data origin, update cadence, sub-processors.
- Share incident logs involving our data within 72 hours of detection.
- Follow our labelling rules, including the "no Chinese-hosted models" rule.

9. AI System Register (ISO/IEC 42001 Annex A alignment)

For every AI system Northwind builds or deploys, Aria maintains a one-page record:

- **Purpose & impact** (A.5.2, A.5.4)
- **Provider & model** — Claude Sonnet 4.6, GPT-5, etc. (A.4.4)
- **Data flows** — in, out, storage location (A.7.2–A.7.6)
- **Owner** — named human (A.3.2)
- **Metrics & monitoring** (A.6.2.6)
- **Retirement plan** (A.6.2.5)

10. Training & Awareness

- New starters: 30-minute induction in week one, including claude.ai Projects + Internal AI Desk walkthrough.
- Everyone: 20-minute refresher every six months.
- Wall posters in every office — Three-Light Rule, Approved Tools, Incident Flow, Do's & Don'ts.
- Friday office hour with Aria (15:00–16:00, `#ask-aria`).

11. Enforcement & Review

- Approved by the Northwind Executive Team. Reviewed quarterly by Aria. Next review: 2026-07-22.
- Most violations are coaching moments — we want people using AI, not afraid of it.
- Deliberate disregard for Section 3 (data), Section 5.5 (hiring), or Section 7 (incident reporting) is serious misconduct.

Questions? aria@northwind.com · `#ask-aria` on Slack · Office hours Fridays 15:00–16:00

Prepared by Aria, Northwind's AI Governance Lead, drawing on ISO/IEC 42001:2023, the EU AI Act, and UK ICO guidance. Intentionally shorter than any of them — because a policy no one reads is a policy that doesn't work.