AI Ethics & Responsible Use Policy

1. Purpose

This policy establishes Subcontext's principles and commitments for the ethical and responsible development and deployment of AI systems.

Subcontext provides AI agents that hold conversations with our customers' own customers on web chat, SMS, WhatsApp, email and voice. Most of our customers are UK insurers, brokers and other regulated firms, so the people our agents speak to are often making decisions about financial protection, sometimes at difficult moments in their lives. This policy explains how we design the platform so that those conversations are safe, honest and fair, and what we expect of the customers who deploy our agents.

2. Scope

This policy applies to:

- every AI capability in the Subcontext platform, including customer-facing agents on all channels, the dashboard assistant, and AI features used inside the CRM, quoting, applications, policies and claims
- everyone at Subcontext who designs, builds, configures, tests or supports those capabilities
- the AI model providers we use, through our Supplier Risk Management Policy

It also sets out the responsibilities of customers who deploy agents, which are reflected in the Terms of Service and Acceptable Use Policy.

3. Definitions

Term	Meaning
Agent	An AI system configured by a customer on the Subcontext platform to converse with that customer's customers or staff
End user	The person an agent converses with, usually a customer's own customer or prospective customer
Guardrail	An automated control that screens messages and agent behaviour, and blocks or changes them when a risk is detected
Conduct rule	A rule that constrains what an agent may say or do, for example to reflect FCA conduct requirements or Consumer Duty
Human escalation	Transferring a conversation, or a decision within it, from an agent to a person
Model provider	A third party whose AI model processes prompts on our behalf, as listed in the Sub-Processor List

4. Principles

4.1 Transparency

- Customers are informed that the platform uses third-party AI models
- The Terms of Service clearly describe the autonomous capabilities and limitations of AI Agents
- We encourage customers to disclose to their end users that they are interacting with an AI system
- The platform offers an optional AI disclosure rule, so a customer can require its agents to say that they are an AI
- The models and providers we use, and where they process data, are published in the Sub-Processor List

4.2 Human Oversight

- The platform provides configurable Guardrails and Policies for controlling Agent behaviour
- Escalation mechanisms allow conversations to be transferred to human operators
- Human review is available, so a person can check an agent's work before it takes effect
- Tool approvals let a customer require a person to approve specified agent actions before they run. In the dashboard assistant, every action that changes data needs approval
- Customers are required to maintain appropriate human oversight of Agent operations
- High-risk use cases require additional safeguards as set out in the Terms of Service

4.3 Safety

- Content screening tools are available to detect and prevent harmful outputs
- Baseline safety controls are applied to all Agents and cannot be disabled. Baseline guardrails screen every inbound message for prompt injection, jailbreak attempts and attempts to extract an agent's system prompt, and cannot be detached from an agent
- Untrusted content, such as text from a website, a document or an integration, is wrapped and kept separate from an agent's instructions, so it is treated as information rather than as commands
- A compliance policy gate checks agent actions before they happen
- Conduct rules allow customers to define specific behavioural constraints. The platform provides rules covering FCA conduct, Consumer Duty, medical disclosure, vulnerable customers, complaints handling, financial promotions and data protection
- The platform supports testing of Agents before production deployment, and customers test their agents before they go live

4.4 Vulnerable Customers

Our customers have obligations to identify and support customers in vulnerable circumstances, and our agents are designed to help them meet those obligations:

- Vulnerability detection flags a person who may be vulnerable, holds back sales and outreach activity to them, and notifies the customer's team so a person can follow up
- Conduct rules for vulnerable customers shape how an agent responds once vulnerability is detected
- Health information is only collected through a dedicated consent step, so the end user knows what is being asked and why

- Conversations can be escalated to a person at any point

4.5 Data Minimisation and Protection

- AI model inputs are limited to what is necessary for the task
- Text conversations run on models hosted in AWS Bedrock, which does not store or log prompts and responses and does not use them to train models
- Text AI is pinned to UK and EU regions: Anthropic models (Claude) are called in AWS eu-west-1 (Ireland) through EU cross-region inference profiles, which may serve a request from another AWS EU region, and DeepSeek runs in eu-west-2 (London). Requests to models that would route through the United States are refused
- Voice calls use OpenAI, which does not use API data to train its models and may retain it for up to 30 days for abuse monitoring before deletion. Speech synthesis also uses ElevenLabs. Both process data in the United States, as recorded in the Sub-Processor List and the Transfer Risk Assessment
- Customer data used for AI inference is not used to train third-party models
- AI-generated memories are scoped to the customer's organisation
- Sensitive parameters passed to AI tools are stored in redacted form, and sensitive values are filtered from logs
- AI call logs are kept for 13 months and then deleted automatically, in line with the Data Retention & Deletion Policy

4.6 Fairness and Bias

- We recognise that AI models may produce biased or unfair outputs
- Customers are responsible for monitoring Agent interactions for bias and taking corrective action
- We do not use customer data to fine-tune or train AI models
- We provide tools for customers to review and adjust Agent behaviour
- We consider the risk of unfair outcomes when we design new AI features, in line with our Equality, Diversity & Inclusion Policy

4.7 Accountability

- Every agent action and tool call is attributable to the agent, conversation and organisation it came from
- Compliance decisions made on the platform are written to a hash-chained decision trail, kept for 6 years by default and configurable from 3 to 25 years, so a customer can show what happened and why
- Changes made by people, including Subcontext administrators acting on a customer's behalf, are recorded in the audit log

5. Prohibited Uses

As set out in the Terms of Service and Acceptable Use Policy, the platform must not be used for:

- Impersonating real individuals without consent
- Generating deliberately deceptive content
- Social engineering or manipulation
- Automated decisions with significant legal effects without human oversight

- Any use prohibited by applicable law

Where we become aware of a prohibited use, we will act in line with the Terms of Service, which may include suspending the agent or account concerned.

6. Roles and Responsibilities

Who	Responsibilities
Directors	Own this policy, approve new AI use cases with a higher risk profile, and decide on model provider changes
Everyone building AI features	Apply this policy when designing features, keep baseline safeguards in place, and raise concerns about safety or fairness
Security Officer	Assesses the security of model providers and AI features, including prompt injection risks
Customers	Configure agents for their own use case, test them before going live, maintain human oversight, monitor for bias, and meet their own regulatory obligations to their customers

7. Introducing or Changing AI Features

Before a new AI capability, or a significant change to an existing one, is released:

1. The intended use, the people it will affect and the data it will process are described
2. Risks are assessed, including harm to vulnerable people, unfair outcomes, prompt injection, data leakage and over-reliance on automated output
3. The safeguards in section 4 are confirmed to apply, and any additional controls are designed in
4. The feature is tested, including with adversarial inputs, before release
5. Where the change affects how personal data is processed, the Data Protection Impact Assessment is reviewed
6. A director approves the release of features with a higher risk profile

Changes follow our Change Management Policy and Secure Development Policy.

Worked example

A customer wants an agent to collect health information during a protection quote. The agent can only ask health questions after the end user has given consent through the health-data consent step. If the person mentions a recent bereavement, vulnerability detection flags them, outreach is held back and the customer's team is notified. The medical disclosure and vulnerable customer conduct rules govern what the agent says, and the person can ask to speak to a human at any point.

8. Model Selection

- We select AI model providers based on capability, safety, data handling practices, and regulatory compliance
- We maintain the ability to substitute models and do not create lock-in to any single provider
- Model provider changes are communicated to customers, and new providers are added to the Sub-Processor List

- Model providers are assessed under the Supplier Risk Management Policy before use
- Where a smaller model is good enough for a task, we use it

9. Regulatory Landscape

We monitor developments in AI regulation, including the EU AI Act and the UK's approach to AI regulation, which relies on existing regulators such as the FCA and the Information Commissioner's Office applying principles of safety, transparency, fairness, accountability and contestability within their own remits.

- We map our safeguards against these frameworks and against relevant regulator guidance as it develops
- We help customers meet their own obligations, for example by supporting AI disclosure, human oversight and record keeping
- We do not claim any formal certification or conformity status under the EU AI Act. Where its obligations apply to us or our customers, we will take the steps needed to meet them

10. Raising Concerns

Anyone, inside or outside Subcontext, can raise a concern about the behaviour of an agent or the use of our AI by contacting hello@subcontext.com. Staff can also raise concerns under the Whistleblowing Policy. Concerns about AI safety are treated as a priority and, where they involve a security weakness, handled under the Incident Response Plan.

11. Continuous Improvement

- We monitor developments in AI regulation (including the EU AI Act and UK AI governance frameworks)
- We review the effectiveness of our guardrails and conduct rules in the light of incidents, customer feedback and new attack techniques
- This policy is reviewed at least annually and updated as standards and regulations evolve

12. Related Documents

- Information Security Policy
- Data Protection Impact Assessment
- Data Retention & Deletion Policy
- Sub-Processor List
- Transfer Risk Assessment
- Supplier Risk Management Policy
- Secure Development Policy
- Change Management Policy
- Equality, Diversity & Inclusion Policy
- Complaints Handling Policy
- Whistleblowing Policy

13. Document Control

Item	Detail
Owner	A director of Subcontext Ltd
Approved by	The directors
Review	At least annually, and after any significant change to our AI features, providers or applicable regulation
Effective	October 2026

Subcontext Ltd, registered in England and Wales, Company No. 17019645. Registered office: 124 City Road, London, EC1V 2NX. The current version of this document is published at subcontext.com.