CaseFlow Automation Ltd

AI Safety & Security Policy

How we govern AI behaviour, prevent hallucination, and protect your data

Classification: Internal - Compliance & Client Assurance | Last Updated: 25 June 2026

CaseFlow Automation Ltd uses AI to help credit hire professionals analyse insurer correspondence, draft responses, and receive strategic case guidance. This document explains the controls we have in place to ensure AI outputs are safe, accurate, and secure.

1. AI Governance Model

Approved Use Cases

AI is used exclusively for **`decision-support`** - it drafts, analyses, and suggests. It never makes legal decisions, sends correspondence on your behalf, or takes autonomous action.

Feature	AI Role	Human Role
Correspondence Analysis	Identify insurer arguments & cited cases	Review, verify, and decide response strategy
Reply Generation	Draft a response cross-referenced against our curated case law database	Edit, approve, and send
Live Case Advice	Provide conditional strategic guidance	Apply professional judgement to specific facts
Argument Letter Generation	Draft structured legal arguments	Review citations, adapt to case specifics

Every AI output is presented as a **`draft requiring human review`**, never as a final document.

2. Anti-Hallucination Controls

Legal AI carries a specific risk: fabricated case names, invented citations, or misattributed principles. We address this with multiple layers of control:

Control	How It Works
Closed Knowledge Base	The AI can only cite cases and authorities from our curated, pre-loaded database. It is explicitly instructed not to cite anything outside this set.
Explicit System Instructions	Every AI prompt includes directives such as *"Do NOT invent case names"*, *"Only cite cases from the provided knowledge base"*, and *"If no authority exists, say so."*
Deterministic Model Settings	All AI calls use the platform's standard low-creativity configuration, which favours deterministic, factual responses over creative completion.
Knowledge Isolation	GTA claims receive only GTA protocol and rate tables - no case law. Non-GTA claims receive case law only - no GTA data. This prevents cross-contamination of authority sources.
Mandatory Limitation Language	When the knowledge base contains no relevant authority, the AI is required to state this explicitly rather than fill the gap with speculation.

⚠ **No AI system can guarantee zero hallucination.** These controls are designed to significantly reduce the risk, but users should always independently verify case law citations before relying on them in legal proceedings.

3. Data Privacy & PII Protection

Our three-layer privacy architecture is designed to minimise the personal data that reaches the AI model and to ensure that any data persisted in our database is masked at rest. For full technical detail, see How We Protect Your Data. For our formal data processing commitments, see our Privacy Policy.

Layer 1 - Local PDF Processing

Text-based PDFs are processed entirely in the user's browser using Mozilla's PDF.js. The original file does not leave the device; only the extracted text is submitted for analysis, and only when the user explicitly chooses to do so. Scanned (image-only) PDFs that the browser cannot read are uploaded to a private temporary storage bucket for OCR, then deleted immediately after processing. A scheduled daily job removes any object older than 24 hours as a safety net.

Layer 2 - PII Masking Gateway (in transit to the AI)

Before any text reaches the AI model, it passes through a mandatory server-side masking gateway that detects and replaces common UK identifiers (person names with or without a title, emails, phone numbers, VRMs, NI numbers, bank details, policy and claim references, street addresses, etc.) with neutral placeholders. The AI works on masked text, not raw personal data. **`UK postcodes are deliberately preserved`** because they are material to basic-hire-rate and locality arguments. A separate output-side scrubber removes any postcode the AI introduces that was not present in the original input.

Layer 3 - At-Rest Masking (in the database)

Saved records (correspondence, AI replies, case advice and pre-emptive letters) are masked at the database level via PostgreSQL triggers calling `public.mask_pii()` before persistence. Even users with direct row access see placeholders rather than raw identifiers in saved records.

Bare Name Detection

Titled names (Mr/Mrs/Ms/Miss/Dr/Prof + surname) are always masked. Bare/untitled names are masked only in high-precision contexts: field labels (e.g. *`Claimant:`*, *`Driver:`*, *`Insured:`*, *`Renter Name:`*, *`Full Name:`*), salutations (e.g. *`Dear John Smith`*), sign-offs, and "Re <name>" in email subjects. A non-person token list and a case-citation guard (any string containing " v " or " vs ") protect legal party names and company names so case authorities remain intact in AI outputs. A bare first name and surname in free prose with none of those cues may not be caught.

What We Don't Mask (and Why)

- **`Case party names in citations`** - e.g. "Lagden v O'Connor" is intentionally preserved so the AI can apply binding authority correctly.

- **`Contextual data`** - information that is only personally identifiable in combination (e.g. "blue Ford Focus") requires semantic understanding beyond pattern matching.

`Post-Upload Scrubbing`

If residual personal data is identified in a saved record after processing, the underlying row can be re-masked or deleted on request. AI provider prompts are processed under zero-retention enterprise terms and are not retained for training or storage by the model provider, so there is no persistent prompt history to scrub at the model layer.

4. Data Handling & Retention

CreditHire Assist is designed **`not to retain personal case data`**. Uploaded text and the outputs the tool generates are deleted immediately when the user is finished, on session close. A continuous scheduled purge runs as a backstop, so nothing is ever retained beyond 24 hours. Users save anything they need to keep to their own systems (a copy button is provided in-app) and re-run the task if required.

Data Type	Storage	Retention
Account / profile data	Encrypted at rest in your company's isolated database partition	Life of account + 30 days
Uploaded text, AI replies, case advice, pre-emptive letters and other case outputs	Encrypted at rest, RLS-isolated to your company	Deleted immediately on session close; continuous scheduled purge ensures nothing is retained beyond 24 hours
Activity / audit logs	System logs, restricted to admins	12 months
Temporary scanned-PDF files (OCR only)	Private storage bucket	Deleted on completion; max 24 hours via scheduled purge

Database backups (disaster recovery)	Managed cloud backup	Uploaded files are never included in any backup. Standard disaster-recovery backups exist for the database, but because case content is deleted within 24 hours those backups hold essentially no personal case data, and once a record is deleted there is nothing to restore it from.
AI prompts (after masking)	Transient — sent to AI provider under Zero Data Retention enterprise terms	Not retained
PII masking logs	Count and category only (e.g. "3 items: EMAIL, VRM") - no original values	12 months

We do not use your data to train AI models. Your correspondence and case details are used solely to generate the specific output you requested.

Special category data: The Service is not intended for special-category data and users are asked not to submit it. We recognise that in credit hire work, health, injury or vulnerability information can incidentally appear in free-text correspondence. Where it does, the same safeguards apply (masking, data minimisation, encryption and immediate deletion), and as data controller the customer remains responsible for the Article 9 lawful basis.

5. Access Control & Data Isolation

Company-Level Isolation

Every company on the platform operates in its own data silo. Row-Level Security (RLS) policies enforce that users can only access their own company's correspondence, case advice, templates, and history. There is no cross-company data access.

Role-Based Access

Role	Access Level
Handler / User	Own company's data - analyse, draft, and view history
Manager / Senior	Company-wide visibility - see team usage and activity
Platform Admin	User management and platform configuration only - no access to correspondence content

Authentication & Session Security

- Email-verified login with secure password hashing
- Concurrent session detection - only one active session per user
- Automatic session expiry for inactive users
- Invite-only registration - users must be invited by their company or a platform administrator

6. Prompt Security

All AI interactions are mediated through server-side functions. Users never interact with the AI model directly.

Control	Description
Server-Side Prompts	System prompts and knowledge base content are injected server-side. Users cannot modify, override, or view the underlying instructions.
Input Validation	All user inputs are validated and sanitised before being included in AI prompts.
No Direct Model Access	There is no API endpoint that allows users to send arbitrary prompts to the AI model.
PII Masking Before Transmission	The masking gateway processes all text before it reaches the model, reducing data exposure even if prompt content were intercepted.

7. Model Selection & Third-Party AI

Model Governance

- The platform uses selected models via a managed AI gateway - models are chosen for their suitability for legal text analysis.
- Model selection is controlled centrally and is not configurable by end users.
- We do not fine-tune or train models on client data.

Data Processing Agreements

AI model providers process data under their enterprise data processing terms, which prohibit the use of input/output data for model training. Combined with our PII masking, this creates a layered protection model.

8. User Safeguards & Disclaimers

The platform employs a three-tier disclaimer framework to ensure users understand the nature and limitations of AI-generated content:

Layer	When Shown	Purpose
One-Time Acceptance Modal	First use of AI features	Requires explicit acknowledgement that AI outputs are not legal advice and must be independently verified
Persistent Banners	Dashboard and Case Advice pages	Continuous reminder that outputs are decision-support drafts
Inline Notices	Every AI generation dialog and result card	Context-specific reminder at the point of consumption

 Users must **`explicitly accept`** the disclaimer before using any AI feature.

 Every AI output is labelled as a **`draft requiring review`**.

 Language throughout the platform uses terms like *`"cross-referenced"`* and *`"greater confidence"`* rather than *`"verified"`* or absolute assurances.

9. Audit & Accountability

- **Activity Logging:** All AI feature usage (analysis, reply generation, case advice, letter generation) is logged with timestamps, user IDs, and company IDs.
- **PII Masking Audit:** Each masking invocation logs the count and categories of items redacted - never the original values.
- **Usage Dashboards:** Company managers can view team usage statistics. Platform administrators can view aggregate usage across the platform.
- **No Shadow Processing:** AI is only invoked when a user explicitly requests it. There is no background AI processing of stored data.

10. Regulatory Alignment

Principle	Implementation
Data Minimisation (GDPR Art. 5(1)(c))	Three-layer privacy architecture: local processing, in-transit masking gateway, and at-rest masking
Privacy by Design (GDPR Art. 25)	Masking gateway is mandatory in the processing pipeline; no bypass path
Transparency (GDPR Art. 13/14)	Disclaimers, honest language about AI limitations, this policy document
Accountability (GDPR Art. 5(2))	Auditable logs, role-based access, documented controls
Lawful Basis	Contractual necessity for providing the Service, together with legitimate interest in providing efficient legal support tools; data minimised before external processing
Human Oversight (EU AI Act alignment)	AI is decision-support only; all outputs require human review and approval before use

11. Incident Response

In the event of a suspected AI safety issue (e.g. fabricated case law, data leakage, or unexpected model behaviour):

1. The affected AI feature can be disabled immediately at the platform level.
2. Audit logs allow identification of affected outputs and users.
3. Affected users and companies are notified with details of the issue and recommended actions.
4. Root cause analysis is conducted and controls are updated before re-enabling the feature.

This policy reflects our commitment to **responsible AI use** in a legal context. We design our systems to be transparent, auditable, and honest about their limitations - because trust is earned, not assumed.

CaseFlow Automation Ltd - Legal Processes. Streamlined.
This document describes the platform's AI safety and security architecture as implemented. It does not constitute legal advice. Organisations should obtain independent legal counsel for compliance assessments.
See also: How We Protect Your Data | Privacy Policy | Policy Change Log | Data & AI Summary