

How We Protect Your Data

How our PII masking protects your data

Last Updated: 25 June 2026

CaseFlow Automation Ltd is designed to minimise the personal data that ever reaches the AI model, while still giving you high-quality legal drafts.

Three-Layer Privacy Design

Layer 1 - Local PDF Processing

When you upload a text-based PDF, it is processed directly in your browser. The original document is not uploaded to our servers; only extracted text is sent securely for analysis. The only exception is a **scanned (image-only) PDF**, where the browser cannot read the text. In that case the file is sent to a private temporary store so it can be OCR'd, then deleted immediately afterwards (with a daily safety-net purge of anything older than 24 hours).

Layer 2 - PII Masking Gateway (in transit)

Before any text is sent to the AI model, it passes through our server-side PII masking gateway. This gateway attempts to detect and replace common identifiers with neutral placeholders, so the model sees the structure of the case but not the underlying personal details.

Layer 3 - At-Rest Masking (in storage)

The same masking is applied again at the database level via triggers, so the saved history of correspondence, replies, advice and letters contains placeholders rather than raw identifiers. If our database were ever accessed without authorisation, the personal details would already be redacted.

What We Try to Mask

Our gateway uses pattern-based rules tuned for UK claims data. It attempts to detect and replace items such as:

- Person names - titled names (Mr, Mrs, Ms, Miss, Dr, Prof + surname) are always masked. Bare/untitled names are masked in high-precision contexts only: field labels (e.g. "Claimant:", "Driver:", "Insured:"), salutations ("Dear ..."), sign-offs, and "Re <name>" in email subjects, with safeguards that protect case citations and company names. A bare first name and surname in free prose with none of those cues may not be caught.
- Email addresses and phone numbers
- National Insurance and driving licence numbers
- Vehicle registration marks (VRMs)
- Dates of birth and bank account details (where indicated by context)
- Policy and claim references - markers include POL, CLM, REF, CLAIM, POLICY, FILE, BHR, CHR, CHO, TPI, plus structured formats such as 24-ABC-123456 and ABC/2025/12345
- Street addresses where they match common "house number + street" patterns

UK postcodes are deliberately preserved. Postcodes are material to basic-hire-rate and locality arguments, so they reach the model by design. A separate output-side scrubber then removes any postcode the AI introduces that was not present in the original input.

Each masked item is replaced with a generic placeholder (for example, [EMAIL_REDACTED] or [VRM_REDACTED]), so you can still understand the draft, but the AI model does not see the raw identifier.

Always-On, Logged, and Auditable

The masking gateway is a mandatory step for all AI-related features (such as analysing correspondence, generating replies, and case advice). There is no bypass path.

Every call records how many items were masked and which categories (e.g. "3 items redacted: email, VRM, reference"), but we never log the original values.

This supports GDPR principles of data minimisation, privacy by design, and accountability, and gives you an audit trail for your own DPIA and compliance checks.

How Can You Be Sure?

We understand that "trust us, it's local" isn't enough - especially when dealing with client data. Here's a plain-English explanation of why this is verifiable, and what to tell a

sceptical client or IT team.



Think of it like a photocopier with a shredder built in.

When you load a letter into a photocopier, the machine reads it and produces a copy - but the original letter never leaves the room. CreditHire Assist works the same way. Your PDF is "read" inside your own browser. The system extracts the words, then the original document stays put. Only the extracted words - with personal details stripped out - are ever sent anywhere.

The Technical Reason

PDF processing is performed using a standard browser technology called **PDF.js**, which is the same open-source library used by Firefox to display PDFs natively. It runs entirely within the browser tab - no server connection is made during this step. This is not a policy choice; it is how the technology is architected.

What an IT Team Can Verify

Any IT professional can confirm this using their browser's built-in traffic monitor. If they watch the network traffic while a PDF is uploaded, they will see:

- **No outbound file upload request** during the PDF loading step
- The first outbound request only occurs *after* clicking Analyse
- That request contains plain text - not a file, and not binary PDF data

🔍 **In plain terms:** Your PDF document never travels across the internet. The words inside it do - but only after personal details have been replaced with placeholders like [NAME_REDACTED] . The document itself stays on your computer at all times.

What This Means for You

- The AI works on **masked text**, not raw personal identifiers.
- Original PDFs **stay on your device**; the only exception is scanned (image-only) PDFs, which are temporarily uploaded for OCR and deleted straight after.

- Saved history is **masked at rest** in the database, not just in transit to the AI.
- We **do not use your data to train AI models**, and we design our architecture to send as little personal data as possible to the external model.

Our PII masking gateway acts as an **automated safety net** - users are not required to manually redact documents before upload. The system is designed to handle data hygiene on their behalf as a mandatory, always-on processing step.

CaseFlow Automation Ltd - Legal Processes. Streamlined.

This page describes the platform's privacy architecture as implemented. It does not constitute legal advice.

See also: [AI Safety & Security Policy](#) | [Privacy Policy](#) | [Data & AI Summary](#) | [Security & Compliance Overview](#)