

Building Trustworthy Legal AI – Preventing Hallucinations

This document sets out how the SOF Generator, ET1 Automation, and Schedule of Loss Generator are designed to be trustworthy and to avoid the risks of AI hallucinations in legal practice.

Human-in-the-Loop

All AI outputs are clearly labelled as drafts. Lawyers must review, edit, and approve before any document is considered final. The AI never files or submits documents autonomously.

Evidence Anchoring

Every fact in the SOF or SoL is linked back to a source document such as a payslip, client transcript, or uploaded evidence. This ensures traceability and provides an audit trail for verification.

Strict Prompting & Guardrails

Models are instructed explicitly: 'Do not invent facts. Only use provided data. If uncertain, flag as missing information.' Outputs are constrained to structured formats like JSON or tables to reduce ambiguity. Narrative drafts are built only from structured facts.

Validation Layers

Automated checks validate outputs before they reach the lawyer. Dates must be real and consistent, monetary values must add up correctly, and party names must match across documents. Inconsistencies are highlighted rather than hidden.

Transparency & Explainability

Lawyers can view the structured data and assumptions behind each draft. All calculations in the Schedule of Loss are documented step-by-step, ensuring no 'black box' results.

Continuous Improvement

Feedback loops allow lawyers to flag errors or misleading outputs. The system improves prompts, templates, and extraction logic over time. Regular audits are conducted to maintain accuracy and compliance.

Benefits

Trustworthy drafts grounded in evidence. Reduced risk from hallucinated facts. Defensible outputs with traceable sources. Professional integrity maintained by keeping ultimate responsibility with the lawyer.