



ROMENTER AI

# MODULE 01 // EXECUTIVE BRIEFING

## The “Garbage In” Protocol

Prepared By :

**Ben Jerome**

Founder of Romenter.ai

[www.romenter.ai](http://www.romenter.ai)

If you feed the world's most advanced AI a diet of low-resolution scans, mislabeled files, and corrupted text, it will not give you insights. It will give you confident, hallucinated lies.



This is The Garbage In Protocol. It is not a suggestion. It is the minimum viable standard to prevent your AI investment from becoming a litigation generator. Until your data meets these standards, your "AI Strategy" is effectively a strategy for automated negligence.

**The Threat:** OCR Errors are Legal Landmines You likely have thousands of PDFs: scanned contracts, handwritten site diaries, and stamped drawings. To read these, AI uses OCR (Optical Character Recognition).

**OCR is not magic;** it is a guess. When you scan a crumpled variation order at low quality, the AI does not see what you see.

### **The Liability Scenario:**

**The Document:** A signed Variation Order states: "Contractor is not entitled to an extension of time."

**The Garbage Data:** A coffee stain or a fold line obscures the word "not. " The low-quality scan makes the text faint.

**The OCR Read:** The AI reads: "Contractor is now entitled to an extension of time."

**The Result:** The AI summarizes this document as an admitted liability. You approve a payout or concede a negotiation point based on this summary.

**The Verdict:** You are now legally exposed because you didn't check the source.

If your data is dirty, your AI is a saboteur.

DATA SOVEREIGNTY: Public LLMs (like ChatGPT) learn from your uploads. Romenter isolates your data. When you upload drawings to the Insight Engine, they are vectorized locally within your secure partition. Your intellectual property never trains the public model.

The 5-Step Decontamination Protocol, you must execute these five steps to scrub your data environment.

**Step 1: Enforce "Born Digital" Supremacy** Stop printing documents just to sign them, scan them, and upload them as "image-only" PDFs. This destroys data integrity.

**The Rule:** All contracts, RFIs, and schedules must remain in their native digital format (Word, Excel, native PDF).

**The Ban:** "Scan-to-email" is banned for any document originating within your organization. E-signatures (DocuSign, Adobe Sign) are now mandatory.

**Step 2: The "Pixel-Perfect" Scan Standard** For legacy paper documents that must be scanned (incoming correspondence, old drawings), the "Eco-Mode" setting on the office printer is forbidden.

**The Rule:** All scans must be at minimum 300 DPI (Dots Per Inch) and saved as OCR-ready PDFs.

**Why:** Below 300 DPI, an "8" looks like a "3", and a "0" looks like an "O". In a budget spreadsheet, that is a disaster.

**Step 3: The "Unsearchable" Purge** Run a script across your document server. Identify every PDF that does not have a selectable text layer.

**The Action:** If you cannot highlight the text with your mouse cursor, the AI cannot read it. These files are "Dark Data."

**The Fix:** You must run a batch OCR process (using Adobe Acrobat Pro or Bluebeam) to textrecognize these files before the AI ever touches them. If the OCR confidence score is low, the file must be flagged for human manual review.

**Step 4: Metadata Hygiene (The "Scan001" Ban)** AI uses filenames and metadata to understand context. A folder full of files named Scan\_290126.pdf, New\_Doc\_Final.pdf, and Meeting.pdf is useless.

**The Rule:** Enforce a strict naming convention: [YYYY-MM-DD]\_[Project Code]\_[Doc Type]\_[Description].

**Example:** 2026-01-29\_CP105\_VO\_Chiller\_Variation.pdf.

DATA SOVEREIGNTY: Public LLMs (like ChatGPT) learn from your uploads. Romenter isolates your data. When you upload drawings to the Insight Engine, they are vectorized locally within your secure partition. Your intellectual property never trains the public model.



**The Consequence:** Files without compliant metadata are rejected from the AI ingestion pipeline.

Step 5: Segmentation (Don't Feed the Beast the Wrapper) Do not dump a 500-page PDF containing the Contract, the Appendices, the Safety Manual, and the Insurance Certificate into the AI as a single file.

The Problem: The AI dilutes the importance of the "Liability Clause" (Page 5) because it is drowning in "Safety Boot Requirements" (Page 400).

The Rule: Split large PDFs by document type. The Contract is one file. The Safety Manual is another. This is called Document Segmentation.

Conclusion You would not pour concrete into a trench filled with mud and trash. Do not pour AI into a data lake filled with unreadable scans and bad filenames.

The cost of cleaning your data now is high. The cost of an AI hallucination triggering a lawsuit later is infinitely higher.

DATA SOVEREIGNTY: Public LLMs (like ChatGPT) learn from your uploads. Romenter isolates your data. When you upload drawings to the Insight Engine, they are vectorized locally within your secure partition. Your intellectual property never trains the public model.



ROMENTER AI

THEORY INTO PRACTICE  
LAUNCH INSIGHT ENGINE