## **✅ Feasibility of Each Step**

| **Step** | **Function** | **Can it be implemented without ML?** | **Tools/Tech** |
| --- | --- | --- | --- |
| **1. Heuristic Authenticator** | Parse magic numbers, detect spoofing | ✅ Yes | C/C++, Rust, or Go + magic number DB |
| **2. Steganography Detector** | Entropy & LSB analysis | ✅ Yes | Manual entropy calc, stego signature DB |
| **3. Behavior Logger** | Run in sandbox, monitor activity | ✅ Yes | Use **Cuckoo Sandbox**, strace, procmon, WinAPI hooks |
| **4. Byte Pattern Analyzer** | Compare byte distribution to known formats | ✅ Yes | Cosine similarity, n-gram match |
| **5. Structure Fingerprinter** | Check for a valid archive format | ✅ Yes | Manual format decoders for ZIP, EXE, RAR |

## **🧩 Integration Design: All in One Tool**

### **🖼️ Frontend (Drag & Drop GUI)**

* **Frameworks**:  
  + Electron.js or Tauri (Rust) for a desktop app
  + HTML5 + JavaScript for web GUI
* **UI Features**:  
  + Drag & drop/upload file
  + Display: File metadata, type confidence, spoof detection, entropy graphs, sandbox behavior summary
  + One-click export of full report

### **⚙️ Backend (Core Analysis Engine)**

* **Languages**: Rust (safe), Go (fast), or C++ (powerful), depending on your preference
* Modular engine with:  
  + module\_magic(): magic number & extension consistency
  + module\_entropy(): calculate entropy and LSB analysis
  + module\_behavior(): interface with Cuckoo or custom sandbox
  + module\_pattern\_matcher(): manual byte pattern similarity
  + module\_structure(): check headers, footers, section alignment

### **🧪 Sandbox/Behavior Module**

* You can:  
  + Integrate **Cuckoo Sandbox** via API (for Windows behavior)
  + Or build a lightweight runner for .exe or .py in an isolated VM or Docker container
* Capture:  
  + Network calls
  + File system changes
  + Process tree

## **🧾 Output: Unified Report**

* **Type Guess**: e.g., “Disguised Executable (EXE)”
* **Spoof Detection**: e.g., “Fake PNG header; structure mismatch”
* **Stego Warning**: “Unusual entropy: possible embedded data”
* **Behavior Summary**: “Initiated network connection to 45.11.X.X”
* **Byte Similarity**: “Matches EXE with 89.3% confidence”
* **Archive Test**: “Invalid ZIP structure”

## **📌 Is This Already Made?**

No, not in **one unified, drag-and-drop tool**, especially not open source. Some components exist in isolation:

* Cuckoo Sandbox (behavior)
* TrID (file type via magic number)
* binwalk (embedded file detection)
* stegdetect (basic stego)  
   But nothing combines **all 5 forensic techniques into a single UX-friendly, modular tool**.

## **🚀 What to Call This?**

Ideas:

* **FileScope** – "See beyond the extension"
* **FIRE** – File Inspection & Reverse Engineering
* **DeepInspect** – Hybrid static + dynamic inspection

## **🧱 Tool Name Suggestion**

**FileScope** – *"See Beyond the Extension"*

## **🧩 Tool Overview**

| **Feature** | **Description** |
| --- | --- |
| **Input** | Drag-and-drop or upload any file |
| **Output** | Real-time display of detection results + exportable PDF report |
| **Detection Modules** | Magic number check, header spoof detection, entropy & LSB scan, sandbox behavior check, byte pattern analyzer, structure validation |
| **Report** | Generates a clean PDF log with timestamp, summary, risk rating, and technical details |

## **⚙️ Architecture Overview**

mathematica

CopyEdit

[Frontend GUI (Tauri/Electron)] ←→ [Core Engine Modules (Rust/C++)]

↓ ↓

Drag-and-drop UI ┌────────────────────────────┐

│ Detection Modules │

├────────────────────────────┤

│ 1. Magic Number Checker │

│ 2. Entropy & Stego Scanner │

│ 3. Header Spoof Detector │

│ 4. Byte Pattern Analyzer │

│ 5. Structure Validator │

│ 6. Sandbox Behavior API │

└────────────────────────────┘

↓

[Report Generator (PDF)]

## **📁 Module Design Details**

### **1. 🔍 Magic Number + Extension Checker**

* Match the magic number against the known database
* Compared to extension (if any)
* Flag mismatch

Example: Header says PNG but content = MZ (Windows EXE)

### **2. 📊 Entropy + Steganography Detector**

* Calculate the **Shannon entropy** of the file and chunks
* Highlight high-entropy zones
* Check for common **LSB** steganography signatures

Alert if entropy > 7.8 in image files

### **3. 🎭 Spoofed Header Validator**

* Look for:  
  + Mismatched file structure (e.g., wrong offsets in PE)
  + Invalid or impossible values
  + Conflicts between the header & real content
* Optionally use binwalk logic (reimplemented)

### **4. 📈 Byte Pattern Analyzer**

* Compare n-gram frequency to known file types
* Use cosine similarity (manual math, no ML lib)
* Detects the true nature of obfuscated files

### **5. 🧱 Archive/File Structure Validator**

* ZIP, RAR, 7Z, DOCX, EXE:  
  + Parse internal structure
  + Validate magic numbers + internal table of contents

### **6. 🛡️ Behavior Sandbox (Optional, Extensible)**

* External interface to:  
  + **Cuckoo Sandbox** (via REST API)
  + **Custom Virtual Machine script** that logs process, file, and network activity
* Detects mismatched behavior:  
  + .jpg making system calls = suspicious

## **🖥️ GUI Design (Tauri or Electron)**

| **Component** | **Description** |
| --- | --- |
| Drag-and-Drop Zone | Accept file upload |
| Live Output Panel | Show type, structure, entropy, and spoof alerts |
| Entropy Graph | Display byte-wise entropy |
| File Behavior | (Optional) Show sandbox behavior log |
| PDF Export Button | Generate and save a detailed PDF report |

## **🧾 PDF Report Contents**

text

CopyEdit

==============================

FileScope Forensic Report

==============================

File: suspicious.bin

Time: 2025-06-18 13:24 IST

----------------------------------

[SUMMARY]

Detected Type: Executable (EXE)

Declared Type: PNG

Status: SPOOFED

Risk Level: HIGH

[DETAILS]

- Magic Number: MZ (Executable)

- Extension: None

- Structure Validity: Invalid PNG format

- Entropy Analysis: 8.12 (Possible Encryption/Stego)

- LSB Check: Hidden bits found in offset 4000–5000

- Behavior: Attempted network access on port 443

- Byte Pattern Similarity: 91% match to known EXE

[RECOMMENDATION]

Quarantine the file immediately. DO NOT run on production systems.

-------------------------------

Generated by FileScope v1.0

Use a **PDF generation library** like:

* **Rust**: printpdf, genpdf
* **C++**: libharu, wkhtmltopdf wrapper
* **Go**: gofpdf

## **💻 Technologies to Use**

| **Part** | **Tech** |
| --- | --- |
| Frontend GUI | Tauri (Rust-based) or Electron.js |
| Core Engine | Rust or C++ |
| PDF Report | printpdf (Rust), libharu (C++) |
| Entropy/Pattern | Custom algorithms |
| Behavior | Optional: API to Cuckoo Sandbox or Docker VM |
| File Structure | Manual format parsers or signatures (open-source format specs) |

## **🚀 Roadmap**

### **🔹 Phase 1: MVP (Core Static Analysis)**

* Magic number + extension check
* Entropy + LSB scanner
* Header validator
* Byte pattern analyzer
* PDF report

### **🔹 Phase 2: Full Prototype**

* GUI frontend
* File structure validator
* Drag-and-drop + output display
* PDF export button

### **🔹 Phase 3: Sandbox Extension**

* Cuckoo/VM behavior logger
* Network/process/file change hooks