# IDP: Automated Discovery of Fraud Detection Algorithms for Images and Documents

**Project Overview**

We are building a system for automated detection of manipulations and inconsistencies in documents and images, with a focus on fraud in insurance contexts. Neural networks can be trained to recognize AI-generated images, but they are typically not explainable out of the box and need to be retrained for every newly released model. In contrast, classical image processing is interpretable and has the potential to generalize to new generators as long as the underlying architectural building blocks (for current image generative models, mainly convolutions and attention layers) remain similar.

Designing such algorithms by hand is time-consuming and requires extensive experimentation. In this IDP, we propose to explore automated discovery of fraud detection algorithms using algorithmic search methods such as ShinkaEvolve [1] and recent agentic research automation frameworks [2–7].

**Project Objectives**

**1. Research**

- Review automated scientific discovery systems [1–7].
- Summarize how these methods search for image-processing pipelines without hand-engineering.
- Revisit classical image forensics: residuals, frequency cues, simple statistics.

**2. Reproduction**

- Reproduce one selected method (e.g., ShinkaEvolve or a close derivative).
- Keep the reproduction minimal: run, verify, and ensure it produces nontrivial algorithmic proposals.

**3. Application to Fraud Detection**

- Provided a small curated dataset, define a simple reward based on detection performance.
- Run the chosen method on the fraud-detection images.
- Analyze the resulting algorithms: Are they any good for fraud detection, and where does the method fail?

**Desired Profile**

- Strong Python skills.
- Creative mindset and enjoyment of analytical, experimental work.

- Some understanding of image processing (e.g., Fourier transform, denoising, basic filters) and/or image forensics.
- Ability to call APIs, write a Dockerfile, and run automated quantitative evaluations.
- Experience with LLM-based coding assistants (e.g., Codex, Claude Code) is desired.
- Independent and structured working style and ability to document experiments clearly.

**Learning Outcomes**

- Gain hands-on experience with automated algorithm discovery methods (e.g., evolutionary or program search) applied to classical image processing for fraud detection.
- Acquire a deep understanding of the challenges and solutions associated with detecting AI-generated and manipulated images in realistic insurance-like datasets.
- Strengthen skills in literature review, algorithm and pipeline design, implementation, testing, and validation in image forensics.
- Contribute to research with high practical relevance in digital forensics and fraud detection.

**Contact**

If you are interested in joining, we would be glad to hear about your motivation and how your skills and background make you a strong fit (in 2–3 sentences or a short 1-minute video).

Please reach out to Julian at **Julian@trisure.de** 😊

---

**References**

1. ShinkaEvolve: Evolving New Algorithms with LLMs, Orders of Magnitude More Efficiently.
2. MLR-Bench: Evaluating AI Agents on Open-Ended Machine Learning Research (2025).
3. The AI Scientist-v2: Workshop-Level Automated Scientific Discovery via Agentic Tree Search (2025).
4. A Self-Improving Coding Agent (2025).
5. EXP-Bench: Can AI Conduct AI Research Experiments? (2025).
6. Alita: Generalist Agent Enabling Scalable Agentic Reasoning with Minimal Predefinition and Maximal Self-Evolution (2025).
7. R&D-Agent: Automating Data-Driven AI Solution Building Through LLM-Powered Automated Research, Development, and Evolution (2025).