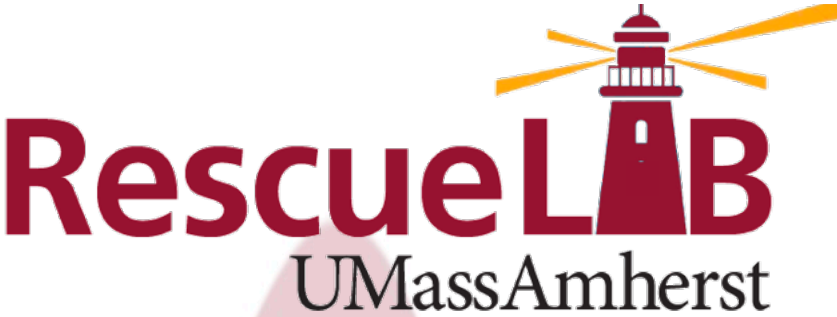


A Preview of RescueBox

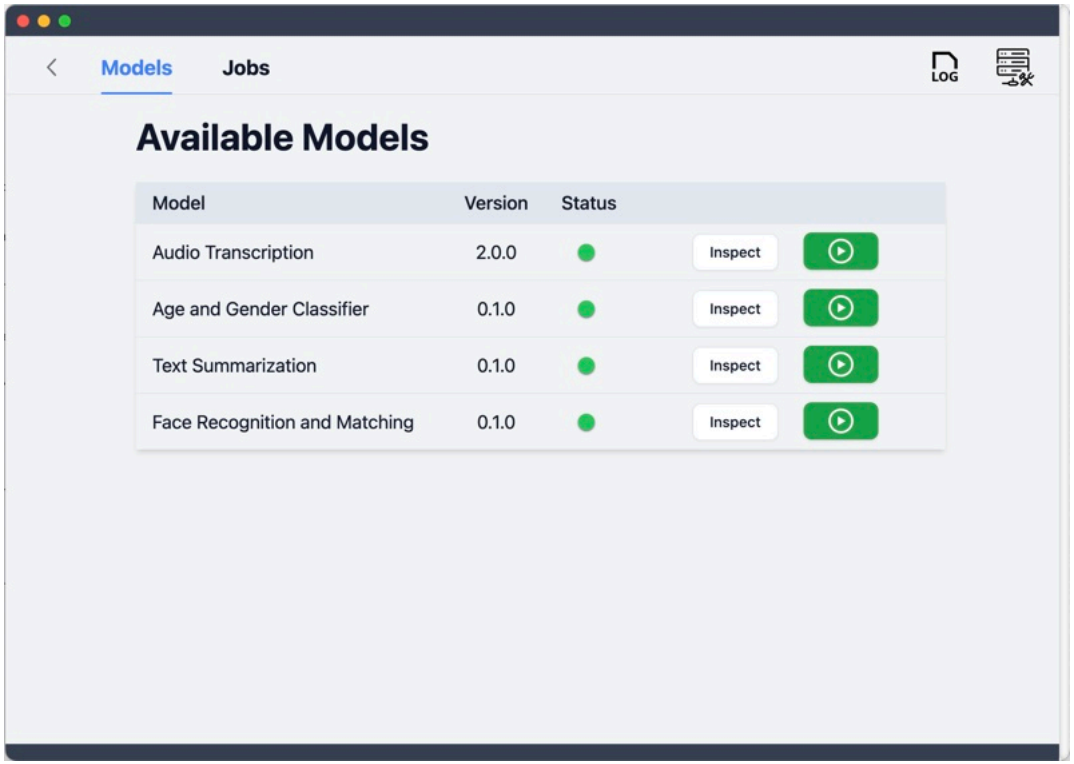
Open-Source Software for Investigations of Child Exploitation

Sahil Sharma
Research Fellow

Manning College of Information & Computer Sciences
University of Massachusetts Amherst

What is RescueBox

- RescueBox is a free, open-source desktop tool for forensic investigators.
- Built to run completely offline, ensuring data never leaves the investigator's machine.
- Designed to be extensible and user-friendly for non-technical users.
- Signups for Beta version

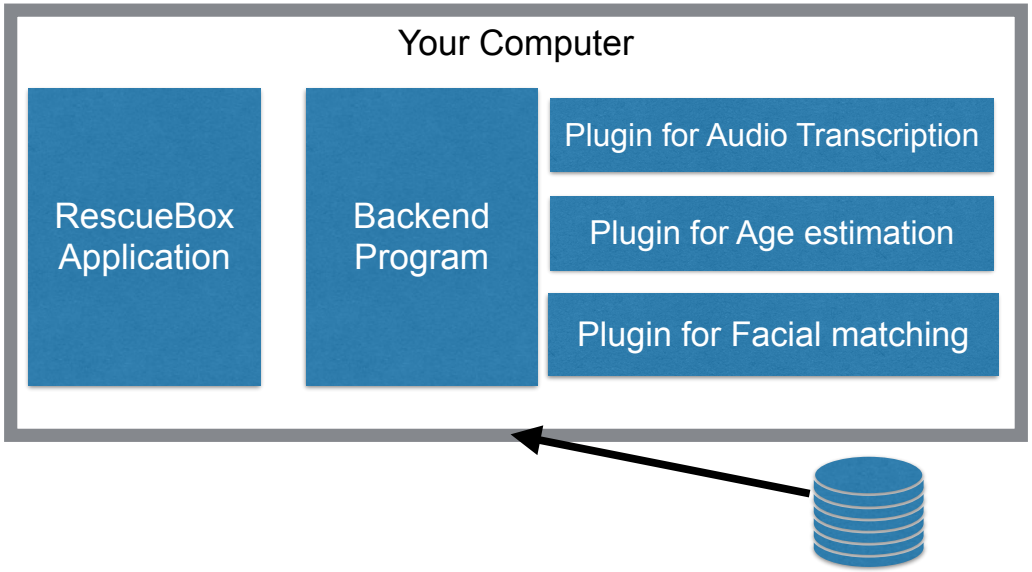


Motivation Summary

- Machine learning tools are useful, and available for free
  - but evidence cannot be uploaded to ChatGPT and the cloud!
- Computer scientists know that free ML models are perform well:
  - but hard to make part of a wholistic forensics system.
- Forensic examiners require a efficient, reliable, and repeatable **workflow** to go from raw data to useful evidence.
- To solve this: you need programmers who understand ML and forensics.
  - That's RescueLab!
  - Our goal is to provide machine learning services for all for free, running locally.

High-level Architecture

- RescueBox consists of 3 core components:
  - User-facing desktop app.
  - Backend program that does the work
  - A set of machine learning plugins
- Goal is to continually add plugins
  - And update ML models
- We package models so that they run on variety of GPUs (via *onnx*):
  - NVIDIA and Apple GPUs, and others.



- Access to cutting edge open “large language models” for text processing: *Deepseek*, *Gemma*, *Llama* (via *ollama*)

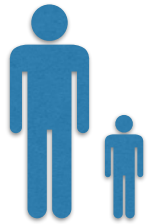
Use Case

- Our goal is to handle **bulk processing** of a lot of data.
- For example, you've seized 80 TB of data from a search warrant.
- Point RescueBox at the root-level directory of forensic image.
  - It will find all images with faces so that you can set aside 75% of the data.
  - Or summarize the contents of the videos, so that you prioritize watching relevant content.

RescueBox's Machine Learning Strategy

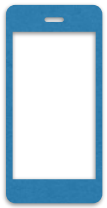
- We only use free, open models.
- We run models locally.
- RescueBox performance will improve as models are released.
- We don't oversell what machine learning models can do for you.
  - We will provide a report of the performance of each model.
  - For example, the error rate of our face detection model.

Key Features in the First release

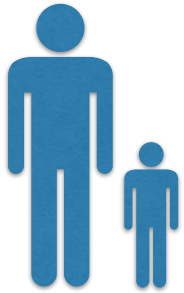
- Face Match 
- Deepfake Detection 
- Audio Transcription 
- Text Summarization 
- Age Estimation 

What's coming next

- Mounting of UFDR files as drives
- Video summarization: Extracting concise summaries of long videos.
- Speaker separation in audio transcription.
- Text message analysis for detecting elements of a crime.



- Better models for age estimation from images.



Resource Requirements

- All machine learning runs on GPUs
- RescueBox runs on GPUs!
- NVIDIA
- Apple GPUs
- We haven’t tried others, but we expect it to work.
- We expect this to run on a single air-gapped computer
  - Not hard to support a server environment in future releases.

How to Contribute

- **Try RescueBox and let us know what features you need.**
- We welcome contributions!
  - To contribute directly to the project, please follow instructions at https://github.com/UMass-Rescue/RescueBox/blob/main/CONTRIBUTING.md
  - Wiki has detailed steps on how you can add a new plugin to the project.
- We greatly appreciate your feedback. If you want to reach out, please send an email to sahilsharma@umass.edu.

Contributors

- Sahil Sharma
- Prasanna Lakkur Subramanyam
- Jai Kumar
- Brian Levine
- UMass Amherst CMPSCI 596E students

RescueBox was funded in part by the National Science Foundation (DGE-2043084).

Sign up

- Preview version available by next couple weeks if you sign up.
- Please give us your early feedback.
- We expect the first Beta version available in June.
- Signup URL: https://www.rescue-lab.org/blog-2/rescuebox-preview

Summary

- Provides cutting-edge, open-source ML tools for free.
- Designed for investigators — grounded in field research and user feedback.
- Encourages a community-driven development model — feedback will directly shape its future.
- assimilate desired features through available popular open-source projects and provide updates, ease of use.
- We can't be as good as a tool that costs $250k!
  - We aim to provide most of the benefit for all who need these tools.

https://www.rescue-lab.org/blog-2/rescuebox-preview

RescueLAB UMassAmherst