

# AI-Based Fake Web Shop Detection

Updated on October 4



# Introduction

- F-Secure's products are designed to protect our users as they carry out their daily life online.
- Underpinning F-Secure's protection capabilities are tools and technologies that we use to analyse websites.
- Phishing, scam and fake web shops are prevalent threats driven by increasingly automated tools that carry a high risk for financial loss.
- According to our consumer surveys, people don't always know how to spot a fake web shop
- We want to build more efficient and smart website scraping and analysis capabilities in order to better protect our users' online safety.



# Project Goals

- A tool for automatically scraping/crawling websites:
  - Mocking different devices, screens and locations.
  - Identify important features in the website (logo, colors, company name, nature of the business, texts, html elements, and such).
  - Ability to intercept, block and allow certain request types.
  - Ability to create automatic interactions with websites.
- The tool can have multiple modes:
  - User driven scraping with UX elements etc.
  - Fast scraping vs rendered scraping.
  - Auto/bulk scraping functionality.
  - Crawling and following links.
- The tool has AI based “analysis modules” to classify websites:
  - Fake webshops (third party academic paper and code available)
  - Phishing sites (F-Secure POC code available)
  - Adult sites (F-Secure POC code available)

# Stretch Goal

## Implement pre-existing AI classification model for fake-shop detection

Once a platform for efficient and smart website scraping exists, students can work on one or more different “analysis modules” for the scraped output.

F-Secure has a number of different POCs for these modules and there is also a third party academic paper that we would be interested in investigating for this use.

Reference paper: Real-Time Detection of Fake-Shops through Machine Learning.

[https://www.researchgate.net/publication/346317401\\_Real-Time\\_Detection\\_of\\_Fake-Shops\\_through\\_Machine\\_Learning](https://www.researchgate.net/publication/346317401_Real-Time_Detection_of_Fake-Shops_through_Machine_Learning)

and code: [https://github.com/mal2-project/fake-shop-detection\\_models](https://github.com/mal2-project/fake-shop-detection_models)

Students could:

- Review example paper and implement the methodology using the example code.
- Compare their results to the original paper.

# Technologies

- There is example python code that students can start with (F-Secure internal and third party academic open source)
- F-Secure can supply mentoring.
- All necessary software should be open-source (permissive license for commercial use) and can run on local machines.

# Requirements for the students

Students should have:

- Background or interest in website development
- Experience with Python (other languages can be proposed if motivated)
- Knowledge of parallelism and code optimizations
- OOP and unit tests
- Background or interest in ML theory, ML implementation (in Python) and willingness to learn. This is what makes the project particularly demanding.

# Legal Issues

## Intellectual Property Rights (IPR):

- A. All IPRs to all Results will be transferred to the Client.

## Confidentiality:

- A. The client will share some confidential information with the students.

Any other legal issues, e.g. if the default contract template does not cover something that needs to be agreed

# Client

- F-Secure is a Finnish and globally operating cybersecurity company. F-Secure designs and offers award-winning security and privacy products and services that help tens of millions of consumers to protect themselves against online threats.
- **F-Secure's representatives**
  - Markku Siikala, business owner
  - Khalid Alnajjar, technical owner (and product owner in terms of Scrum)
  - Sean Robinson, technical coach
  - Tuuli Lindroos & Marko Komssi, project owners

Contact details: [firstname.lastname@f-secure.com](mailto:firstname.lastname@f-secure.com)



# Additional Information

- Any other details, if needed.