Automated Penetration Testing for Android

By

Adam Toms-Sheridan

Submitted to
The University of Roehampton

In partial fulfilment of the requirements
for the degree of

BACHELOR OF ENGINEERING IN COMPUTING

Declaration

I hereby certify that this report constitutes my own work, that where the language of others is used, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions, or writings of others.

I declare that this report describes the original work that has not been previously presented for the award of any other degree of any other institution.

Adam Toms-Sheridan
Date: 02/12/2024

Acknowledgements

Introduction

Penetration testing is a process done by cyber security specialists to attempt to gain access to a device or network by attempting to breach some or all of the systems security by using the same tools and methodologies that a potential attacker may use. An automated penetration testing device aims to automate the process of penetration testing and produce results that are useful to security experts in order to highlight the vulnerabilities within security.[1]

Penetration testing is a very expensive and can cost between $10,000 and $35,000. An automated penetration tester would vastly reduce the cost of penetration tests for a company. If a company does not have a good level of security, not only does it affect the company, but it also affects the customers and other stakeholders of the company too. Should a data breach occur, all sensitive information stored within the company's network is vulnerable to different types of abuse. This can lead to huge fines for the company due to breaches in GDPR, as well as loss of investors and customers as some customers will have lost faith in the company's ability to keep their data secure. Every 39 seconds, a cyber-attack occurs.[3] This can affect anyone and anywhere. It can happen to anyone that has an electronic device at any time of the day. A hack is unpredictable and needs to be defended against before it can occur. It is important for this problem to be solved as roughly half of companies will only do 1 pen test per year.

The aim of this project is to automate the process of at least one manual penetration testing tool on android devices in an effective manner, produce a report based on the output of the penetration testing tool, and ensure that the process requires minimal human interaction.

To achieve these aims:

- Architecture designed for this project will be strictly focused on modularity of the software to ensure that the tool can be expanded at a later date.
- The tool used will be researched and tested manually before selection. This is to ensure that the tool is suitable for automation.
- Software will then be developed to automate the process of penetration testing tool in the terminal.
- A UI will then be created to run the tool without providing an output.
- A report generator will then be made, designed to be modular, as it will read the outputs provided and then turn the data into a report.
- The report generated will highlight areas of weakness that has been identified by the pen testing tool.
- During testing, the tool will produce its raw output as well as the report for the software. These will be compared to ensure that the report is accurate and contains all of the significant and important information from the test.

This software has the possibility of breaking UK laws. Section 1 of the Computer Misuse Act 1990 explains that the intent to secure data from any computer, whether or not the individual owns the computer, in an unauthorised manner is a crime. This does not need to be targeted at a specific program, but rather the attempt to gain access to data. This software, although not the primary function of it, also has the potential to be modified in order to commit further crimes. This would fall under section 2 and potentially section 3 of the Computer Misuse Act.[5] In order to reduce the likelihood of the misuse of this software, it will have security

implementations in the program to ensure that it will not work should any alterations occur to the software.

This project aims to have a positive impact on the social aspect its stakeholders. As the intended audience of this software is companies that require penetration tests, this will have a beneficial effect on the customers of these companies. This aims to ensure that their data is safe to an acceptable standard and will give peace of mind to the customers of the companies that are protecting their data.

This project has a lot of ethical implications. As this project involves hacking, should it end up in the wrong hands, it could be weaponised and give attackers the vulnerabilities detected within their target's software. The users of this software should be trusted and have a valid reason to use it, therefore whoever wishes to gain access to this software would have to go through a multi-stage background check including verifying the company and the representative who reached out for the purchase of the software. There are other ethical implications which include the storage of data from the penetration tests and how it is handled by us. All data will not be stored other than for the compilation of the report and never stored after the program is run. The user of the software will have the responsibility of maintaining the confidentiality of the report and how it is handled.

This tool is specifically designed for a professional workplace and is designed for a member of the IT team to be able to identify any issues present in the software which they can then take appropriate action by sending the issues to the software development team or sending the data to an external firm to fix the issues present. The tool should not be available for personal use as it would be significantly easier for a malicious user to gain access to it which could lead to many ethical issues arising.

Literature-Technology Review

Introduction- Literature Review

Penetration testing is an essential aspect of cyber security. Penetration (Pen) testing is a process which mimics the attacks from an attacker in a real-world scenario. In December 2023, Insomniac games was hacked by a group called rhysida.[6] This hack only took 20-25 minutes before the group were able to gain access to 1.67 TB worth of files including sensitive data about employees. The shockingly low amount of time it took for rhysida to gain access to Insomniac games shone light on the dire situation of cyber security within the industry today.

The cyber security industry is spread thin. In Europe, 61% of teams say that they feel that they are understaffed and 52% of teams feel that they are underfunded.[7] The cyber security field is under a lot of pressure to focus on many areas at once and require the support of their companies. An automated penetration device would help assist the industry for cyber security professionals to focus on other areas of security.

This project will focus on android applications due to high market share that android has in the industry. Currently there is roughly a 72% share hold in the market for android devices[8], with roughly 7.2 billion smartphones being used around the world.[9]., meaning there are roughly 5.2 billion android phones currently being used today, however android isn't just used on phones as many devices also run android, such as tablets or VR headsets, to smart fridges or self-service checkouts. Android is everywhere, and if a company does not ensure that their cyber security is maintained, it could be disastrous to all parties involved.

How feasible are automated penetration testing tools?

Penetration testing is a very manual process and have a much more cost-effective long-term solution to security. However, parts of the process currently are automated by experienced security experts in order to speed up the manual process.[10] Currently, automated pen testing tools are only suitable for a small variety of attacks that they are designed for.[11] The conclusions from these studies do suggest that automated penetration tests are a feasible tool, but have a lot of limitations that need to be considered before it can be implemented fully.

How much data is stored on device?

Mobile devices store a lot of data about the user. 53% of people in the U.K. use mobile banking.[12] and 25% of people write a note on their phone to store their passwords. Other information that is stored on a user's phone is also their photos, emails, texts, contacts and search history which a potential attacker can use to blackmail a victim based on various data that the attacker could access on the victim's phone.

A phone is the place where a lot of people's personal data and lives are stored on them. If an attacker was able to exploit weaknesses from applications on a device such as an android, it could be devastating to the victim as the attacker would gain access to nearly the whole life of the victim.

Technology Review

Penetration testing tools

Penetration testing tools are tools designed to find a way into a system by bypassing system security/protocols; they are used by hackers, ethical or malicious, to attempt to gain access to said system. This is a vital part of the project as any penetration testing tools used within the project will be automated. Many tools will be tested manually before settling on a tool that will be ideal for testing the security strength of particular android APKs.

Kali Linux

This is a distribution (distro) of the Linux operating system and is favoured by security professional as it is designed specifically for testing the cyber security of different systems. It is readily available and has many versions, including one designed for android devices. Kali Linux has around 600 penetration tools included at launch which will be useful for penetration testing.[14]

Unsecure APKs

An unsecured APK is a file with known vulnerabilities that have either been designed in a way that it purposely has these vulnerabilities for testing purposes or is an old version of a product that has had its vulnerabilities discovered after deployment. The reason to have an unsecure APK is to ensure that the tool that has been developed is acting as programmed. This acts as a test case, as the known outcomes can then be compared to the outcomes of the project in order to ensure correctness.

GitHub

GitHub is a version control website that utilises the Git version control software. It has a much friendlier interface for a user to interact with, as well as other features that Git does not have. GitHub is a site that allows developers to track progress with timelines, Kanban boards, and commit history. This provides an accurate picture of the development cycle of the project that will be developed. GitHub is also a widely use service and is a globally used service for the development of software.

Visual Studio Code

This is an IDE (Integrated Development Environment) used to develop the software that will automate the process of penetration testing. It has many libraries and extensions that can be found in the extensions tab in the application. Visual Studio Code (VSCode) has a wide variety of coding languages that can be used without the need to install different packages from different sites and is designed to be a versatile IDE. VSCode is a practical IDE for the development of multiple different languages thus allowing for easier integration between coding languages as it can support multiple languages at once.

Android Device

An android device is required for later stage testing of this project. It will be used as an unseen testcase for this project. This will simulate a practical environment for use in the field. As the results of the penetration test would be unknown, a manual penetration test would need to be performed on the same applications in order to determine whether the results

from the project is successful. The benefits of using an android device are that multiple applications can be installed on the device, thus allowing for multiple tests to be ran in quick succession as most software on an android will be applications.

Virtual Machine

A virtual machine (VM) is a piece of software that runs a second, isolated, operating system within a personal computer. Virtual Machines are designed to enable users to do a variety of different activities, such as run software in a more secure environment ensuring the protection of a personal machine, using different operating systems such as Linux, Windows, or MacOS. This project will be using a virtual machine to ensure that any malicious software that may be installed is quarantined away from the main machine.

Summary of Outcomes of Literature and Technology Review

Literature:

	Benefits	Limitations
Automated Penetration Tester	This will save the user of this tool more money than manual penetration testing. Automated penetration testers are much faster than manual testers as a computer can perform actions much faster than a human.	Aspects of manual penetration testing have not been automated as it is difficult to automate for a large variety of tests.
Automating Penetration Testing on Androids	There are 7.2 billion android users around the world, and most personal data is stored on a mobile phone.	There is no guarantee that system has been completely covered in depth.

Technology:

	Benefits	Limitations
Penetration Testing tools	Different tools cover a large variety of different potential attacks which is beneficial to this project.	The large number of potential tools have different complexities. This means that it will take a lot of time to find the most suitable tool for automation.
Kali Linux	This is an operating system specifically designed for penetration testing, with many penetration testers that are downloaded with the operating system from the start.	Kali Linux is resource intensive and can lead systems to lag if they lack the necessary resources to run.[15]
Unsecure APKs	This allows the ability to test the tool, while knowing the outcome of the tests.	These files can come with malicious software which could be damaging to the machine.

GitHub	This is version control and allows the user to track their progress as a project progresses.	GitHub has heavy reliance on third party services to ensure that its servers are running, which can lead to downtime, should a third party be facing issues.[16]
Visual Studio Code	This has a vast array of languages supported as well as extensions that can be used in software.	Some libraries, such as python, require external downloads as VSCode does not support certain libraries in its extensions tab.
Android Device	Android has weaker security than iPhone and thus a better option for penetration testing. It is also used by a much larger percentage of the market.	There are many different types of android device, which can make development of automation difficult if there are different requirements on each device.
Virtual Machine	This ensures security of a personal machine, if malicious software is executed on a virtual machine, the user's personal machine is likely safer due to the closed nature of a virtual machine.	Virtual machines are also very slow and resource intensive as it acts as a second operating system on a first system. If the user's machine does not have sufficient resources, the virtual machine will be slow and an unpleasant experience.

Methodology

GUIDANCE: Up to 1000 words

Design

The artifact produced within this project will be designed modularly and will follow SOLID principles. SOLID principles are defined as:

Single responsibility principle- each component should have only one reason to change.

Open-closed principle- components changed by adding new code, not modifying existing code.

Liskov substitution principle- subclass behaviour should satisfy superclass specification, likewise for implementations of interfaces.

Interface segregation principle- components should not depend on things that they don't use.

Dependency inversion principle- high-level components should not depend on low-level components, instead low-level components should depend on high-level components.

Following SOLID principles allows futureproofing of the project by enabling other developers to read over code and understand it faster than software that doesn't follow SOLID principles. The design of this artifact is very important for the scalability, future development, and readability of the software. SOLID principles aim to achieve all of this and is the highest priority of the project design.

The design of the artifact follows a 2-tier architecture having a user layer and a logic layer. There is no data layer due to security; should the device be compromised, there will be no data regarding vulnerabilities saved on it. All data will be saved directly into the user's computer to a designated file location as decided by the user.

Testing and Evaluation

When testing the artifact in this project, there will be multiple stages of testing and evaluating. There will be static tests within the software, which will ensure that different functions are being performed correctly by the machine. If the tests pass each stage during testing the software is functional and would pass this stage of testing.

During the penetration testing stage of testing, the artefact will be tested on software with known vulnerabilities to ensure that the automation is correct. As the vulnerabilities are known within the testing stage, the output results from the artifacts will be directly compared to the known vulnerabilities in the software to test whether the results are accurate.

The final testing stage is with unknown software. This software will be tested with no known vulnerabilities and will be tested both manually and automatically. Both the results of the tests, as well as the timing of both tests will be compared to measure accuracy of the test, and time it took for both tests to happen. This enables the artifact to be measured for efficiency in a real-world scenario to measure how much time it could save when running real world penetration tests.

Project Management

The project will follow the agile methodology, specifically utilising a kanban board and sprints. Each sprint will follow a 4-week deadline and should have a measurable outcome after each sprint. A timeline will be used to set sprint deadlines which can be found within GitHub.

The sprints that are divided into 4 stages: planning, implementation, testing, and review. The planning stage of agile occurs at the start of each sprint and is the part where the software, and structure of the artifact will be designed and conceptualised. After the planning has been completed, the artifact moves onto the next stage of the product life cycle.

The implementation stage follows the plans laid out in the previous part of the life cycle and is the main part of each sprint. The implementation stage focuses on building each part of the plan and ensuring that the software is at a suitable point after each sprint including designated milestones of progress for the end of each sprint.

The next stage of the product life cycle is the testing stage. This stage attempts to remove any bugs and glitches that may occur during real world use of the software. The testing phase strives to ensure that the artifact is in a usable and shippable state for the customer, or for the next sprint of the product life cycle as developers do not want to revisit software unnecessarily.

The final stage is the review stage, which focuses on what the customer wants. The customer may want to change the way the user interface (UI) looks or may wish to add additional functionality within the program. The customer may want to only see how the product is progressing and provide some feedback on progression. The review sets up how the next sprint will be completed.

Technologies and Processes

The technologies used for the artifact will be a virtual machine, Visual Studio Code, Kali Linux, a penetration testing tool, GitHub, an android software with known vulnerabilities, android apps with no known vulnerabilities.

These technologies will be used as they are designed specifically for the role that the artifact is attempting to automate. These tools are used for penetration testing, except GitHub and Visual Studio Code, which is imperative for the artifact as these technologies are what will be automated during this project.

Implementation

GUIDANCE: Up to 3000 words

The starting point of this project is to design the architecture, and modelling it for real world scenarios.

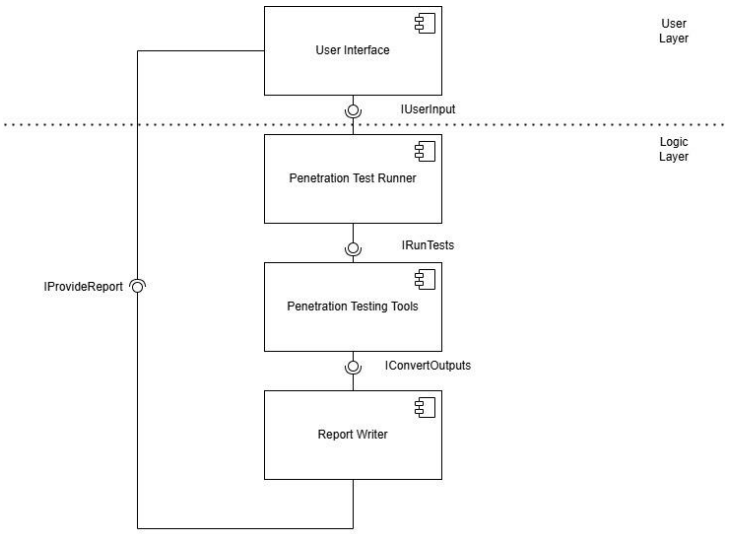

figure 1: Overall Architecture

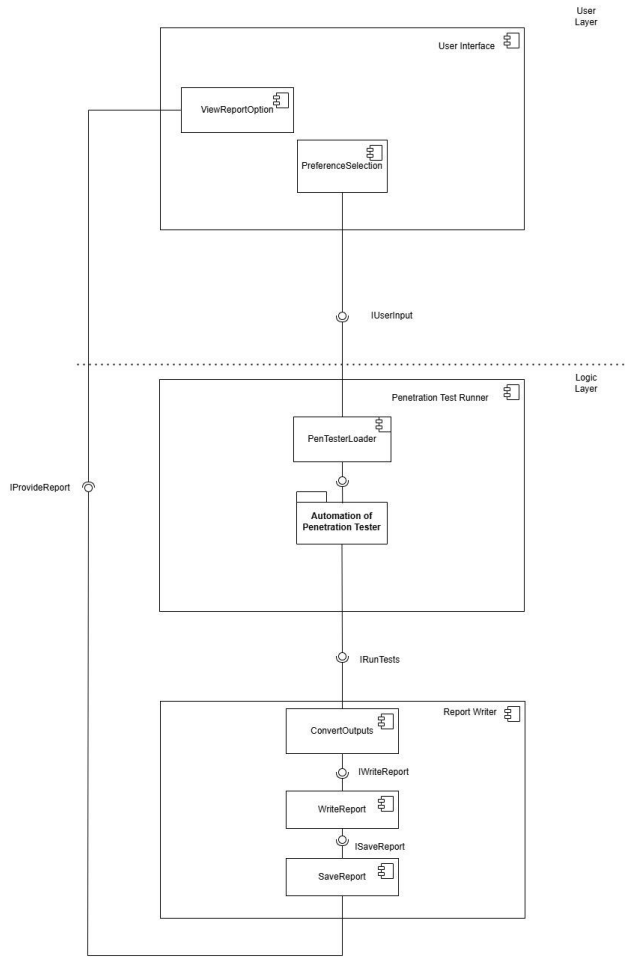

figure 2: Component Diagram

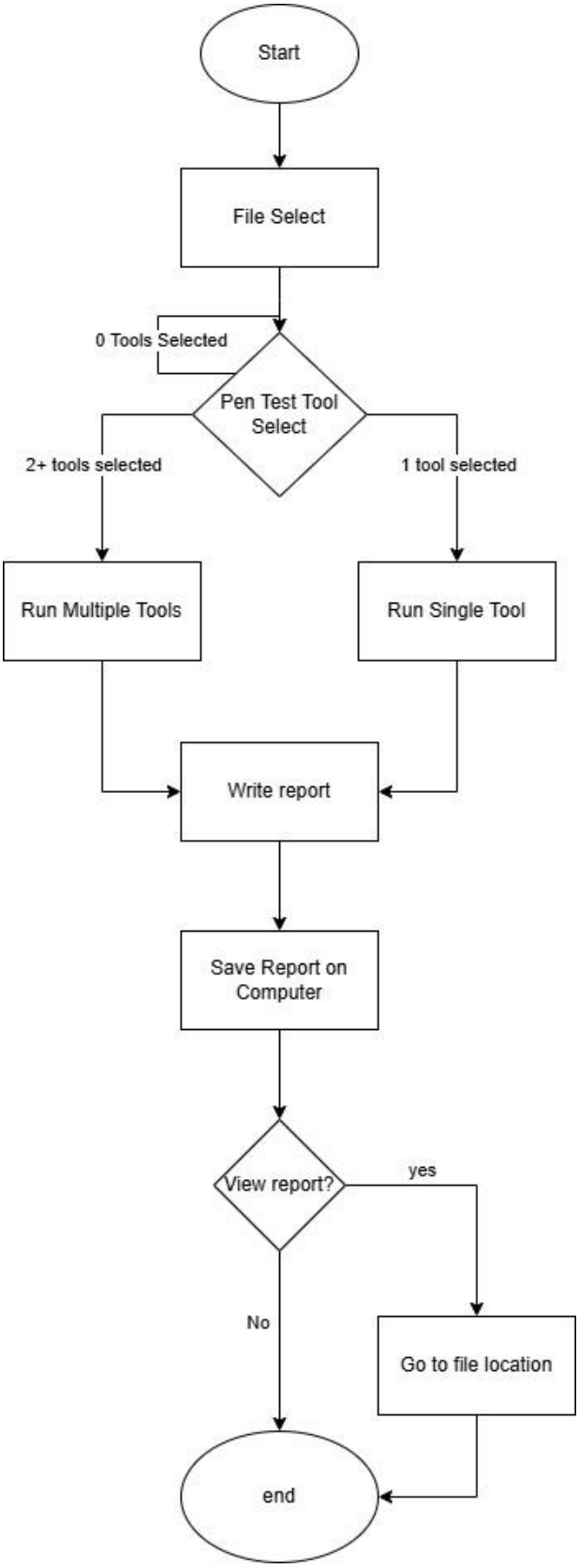

figure 3: Flow Diagram

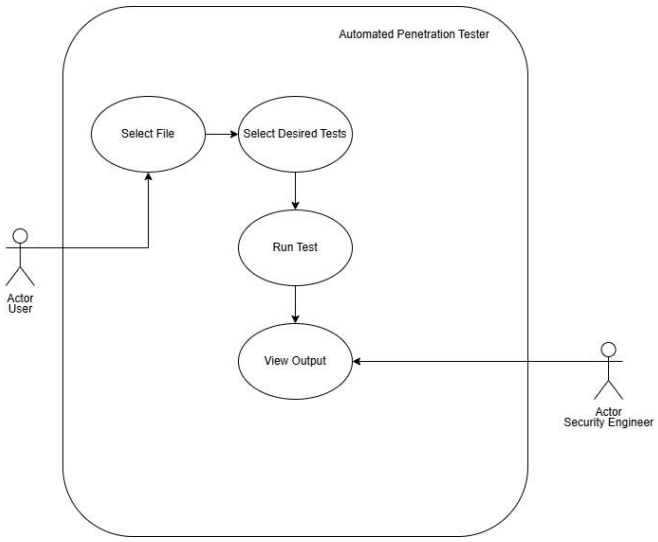

figure 4: Use case diagram

Evaluation and Results

GUIDANCE: Up to 2000 words

Conclusion

GUIDANCE: Up to 1500 words

References

[1]National Cyber Security Centre (2017, August 8) Penetration Testing https://www.ncsc.gov.uk/guidance/penetration-testing

[2] Ewelina Baran (2023, May 15) Pricing Insights – How Much Does Penetration Testing Cost? https://www.blazeinfosec.com/post/how-much-does-penetration-testing-cost/

[3] Elsie Boskamp (2023, Jun 15) 30 Crucial Cybersecurity Statistics [2023]: Data, Trends and more https://www.zippia.com/advice/cybersecurity-statistics/

[4] worldometer (2024, Oct 27) U.K Population (LIVE) https://www.worldometers.info/world-population/uk-population/

[5]CPS (2023, August 03) Computer Misuse Act https://www.cps.gov.uk/legal-guidance/computer-misuse-act#:~:text=There%20are%20two%20elements%3A,data%20held%20in%20a%20computer.

[6] Rachel Davies (2023, December 21) Insomniac Games hack: what happened? https://www.standard.co.uk/news/tech/insomniac-games-hack-what-happened-rhysida-malware-b1128434.html

[7] ISACA (2024, October 1) Cybersecurity teams can't keep up with growing levels of cyberattacks, new research reveals https://www.isaca.org/about-us/newsroom/press-releases/2024/cybersecurity-teams-cant-keep-up-with-growing-levels-of-cyberattacks-new-research-reveals#:~:text=These%20figures%20have%20dropped%20only,most%20amongst%20today's%20cybersecurity%20professionals.

[8] Ahmed Sherif (2024, December 18) Market share of mobile operating systems worldwide from 2009 to 2024, by quarter. https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/

[9] Josh Howarth (2024, June 13) How Many People Own Smartphones? (2024-2029) https://explodingtopics.com/blog/smartphone-stats

[10] IEEE (2023, October 20) Automated Penetration Testing, A Systematic Review https://ieeexplore.ieee.org/abstract/document/10278377?casa_token=NVieOETjCLQAAAAA:Zz-6Esf7r6FbpkzXHrh9rRUrTy4e0wPIt7seOHCi12BmwOApMNBdxRdN-32eQQDtl-Qjdx_6Ni4

[11] IEEE (2023, October 20) Automated Penetration Testing, A Systematic Review https://ieeexplore.ieee.org/abstract/document/10278377?casa_token=NVieOETjCLQAAAAA:Zz-6Esf7r6FbpkzXHrh9rRUrTy4e0wPIt7seOHCi12BmwOApMNBdxRdN-32eQQDtl-Qjdx_6Ni4

[12] Sophie Barber (2025, January 9) Digital banking statistics 2025: How many Brits use online banking? https://www.finder.com/uk/banking/digital-banking-statistics

[13] Brett Cruz (2024, December 19) 2024 Password manager Industry Report and Statistics https://www.security.org/digital-safety/password-manager-annual-report/

[14] Kali (No date) Kali Tools https://www.kali.org/tools/

[15] javatpoint (No date) Advantages and Disadvantaged of Kali Linux https://www.javatpoint.com/advantages-and-disadvantages-of-kali-linux

[16] Meghana (2023, August 24) GitHub Pro's and Con's https://blog.aiensured.com/github-pros-and-cons/