AI-ASSISTED FILE RECOVERY FOR RANSOMWARE AFFECTED SYSTEM

An Information Technology Capstone Project

Presented to the Faculty of College in Information and Computing

University of Southeastern Philippines

Bo. Obrero Davao City

In Partial Fulfillment of the Requirements for the Degree of

BACHELOR OF SCIENCE IN INFORMATION TECHNOLOGY

By

Balbuena Mike Troy P.

BSIT - IS 3A

Instructor

Amante, Cheryl R.

May 2025

Table Of Contents

CHAPTER I

INTRODUCTION

1.1 Background of the study

Ransomware is now one of the biggest cybersecurity threats affecting educational institutions, healthcare facilities, financial organizations, and smaller companies. These attacks typically mean users cannot access their files or system until a ransom is paid, and the attack's creators usually want a cryptocurrency payment. It demonstrates that current systems are weak and shows the critical role of flexible cybersecurity solutions that keep up with fast-changing cyber attacks. Over 5,200 significant ransomware incidents were cataloged in 2023, a sign of more ransomware activities than the previous year [59].

According to Sophos State of Ransomware Report in 2023, there were ransomware attacks on 66% of companies around the globe last year, and the typical ransom payment was $812,360 [49]. Ransomware damage worldwide is expected to go over $20 billion this year and hit $265 billion by the end of 2031 [11]. Although paying the ransom, missing data appears to be a common problem, proving that advanced file recovery software is needed.

The financial losses caused by cyberattacks are high, as organizations typically have to spend $2.73 million on average to recover in 2023, a rise from before [51]. The average refers to several costs associated with paying the ransom, getting systems back up and running, dealing with legal problems, and damaging a company’s reputation. In healthcare, incidents can create significant issues, including downtime that can last a long time and lead to substantial financial losses for some organizations—a few have lost up to $900,000 every day when disrupted [61]. Easy access to Ransomware-as-a-Service (RaaS) models now means that people with little technical knowledge can launch complex attacks on systems [59], [26].

Changes happening around the world are also appearing in the Philippines. 56% of the surveyed organizations in the country said ransomware attacks had increased by more than 100% compared to the previous year in 2023 [17]. One high-profile event, such as the attack on the Philippine Health Insurance Corporation, demonstrated how government data could be compromised, and more than $300,000 is sometimes needed to regain access [18]. Government offices are in danger, and private financial companies and local companies stand to lose. In the Philippine market, over 315,000 compromised credentials, alongside 17,000 phishing incidents, were recorded in the first half of 2024 [59], [53].

In Davao, ransomware and related cyber threats have interrupted work at local businesses and government agencies. Services and experts in technology and education say more than 50% of cyberattacks affect core operations, resulting in both money shortages and significant data losses, especially for small businesses without reliable cyber protection [8]. The DICT Region XI wants to solve this problem by organizing Computer Emergency Response Teams (CERTs) for local governments and urging the adoption of updated cybersecurity policy manuals [31].

Because there is a steady increase in ransomware incidents, it's clear that traditional backup-based recovery methods are not adequate. Unfortunately, many victims figure out after the fact that their backups are outdated or that they don't understand the software well enough to recover encrypted data. Putting artificial intelligence and machine learning into cybersecurity frameworks is a smart move against future attacks. AI technology identifies unusual behavior, highlights questionable activities, and, importantly, can process encrypted or corrupted files so users can recover their files when backups are unavailable [14], [1].

The proposed file recovery system aims to unite ransomware detection, file classification, shadow copy restoration, and predictive recovery of files. In this integrated approach, recovery is expected to be more efficient, data loss should be reduced, and downtime will likely decrease after an attack. Because ransomware is changing, businesses should introduce proactive cybersecurity measures to better protect their information from highly sophisticated threats [37], [52].

The effects of ransomware on cities like Davao City go further than money losses, as it weakens trust in using computers and increases questions about the nation's online security. Because ransomware attacks are becoming more difficult, organizations need to move toward using advanced and proactive safety measures. AI-based recovery solutions can increase an organization's ability to deal with ransomware attacks and show how fast cybersecurity must evolve as cybercriminals adapt [59], [53].

The exact statistics on increases in ransomware events, the expenses involved in recovering from them, and the main weak points in the healthcare and financial sectors demonstrate the seriousness of the problem. Due to the constant progress in RaaS and the problems it causes, businesses should highlight technological growth and be prepared to deal with challenges. This framework should be updated with the current threat situation, including artificial intelligence, to face problems and protect infrastructures against future dangers.

1.2 PROJECT CONTEXT

Disabled access to important files like documents and images has affected many desktop systems users who have been ransomware targets; hence, this study aims to offer solutions to these users. Users of the software cover students, business professionals, owners of small businesses, and IT administrators who often keep sensitive data on their local systems. Most of the time, users affected by ransomware attacks are often left with only two options: choose to pay the ransom or end up without their data. This capstone project offers a third and smarter option—file recovery using built-in system backups and artificial intelligence.

The file recovery system automatically looks for and recovers lost files using Windows Shadow Copies and File History. Many users do not use these fully because they lack both knowledge and awareness. The file recovery system reduces the human input needed to note these backups and makes it more likely for data recovery to succeed. If possible, the user will have access to previous files for quick recovery, helping to keep everything in control.

If no shadows or file history are available, the file recovery system uses artificial intelligence to see if recovery is possible. This method only applies to JPG, PNG, and PDF text files. They use their learned information to restore proper documents from damaged data. Denoising autoencoders will process image files, and lightweight language models like DistilBERT will predict text using text-based ones.

This way of development supports IT consumers and the expansion of the field by demonstrating the use of AI in cybersecurity and disaster recovery scenarios. This project suggests that ransomware should be handled differently, emphasizing easy-to-use features, high-quality automation, and cutting-edge recovery methods.

In addition, the file recovery system follows the trend of demand for security tools that detect dangers and support recovery processes. Unlike antivirus software, this file recovery system primarily aims to recover files when an attack has occurred. Users and organizations can use it to build their resistance to losing their data further. Users are now protected from losing important data because this project offers an easy-to-use, AI-enabled tool to restore critical files, boost protection, and empower all users.

1.3 OBJECTIVES

The study's main objective is to develop an AI-assisted file recovery application for ransomware-affected systems that can recover files using system backups and apply artificial intelligence for partial recovery.

Specifically, the application will:

- Automatically detect and recover previous file versions using Windows Shadow Copy and File History.

- Classify files as safe, recoverable, or corrupted using a lightweight machine learning model.

- Apply AI models to attempt partial recovery of encrypted or corrupted image files (JPG, PNG).
- Apply AI models to attempt partial text reconstruction for damaged PDF files.

- Provide a user-friendly interface that allows guided one-click recovery suggestions.

1.4 SIGNIFICANCE OF THE STUDY

The primary goal of this project is to create an application that assists users in recovering their data after a ransomware attack, even in situations where traditional recovery options are unavailable. The following individuals and groups are identified as the primary beneficiaries:

General Users – Individuals who store personal, academic, or work-related files on their computers can benefit from this file recovery system as it offers a recovery solution even when ransomware encrypts their data.

Educational Institutions – Schools and universities that lack advanced cybersecurity tools can use this file recovery system to recover student records and important files without relying on external IT specialists.

Small and Medium Businesses (SMEs) – Local businesses that suffer from ransomware attacks can recover critical operational data without resorting to costly recovery services or ransom payments.

Future Researchers – Those exploring artificial intelligence, cybersecurity, or file recovery will gain insights from this project's approach to combining system-level backups and AI-based prediction.

IT Professionals and Developers – IT practitioners will be able to explore the potential of using AI models for data classification and partial file restoration.

1.5 SCOPE AND LIMITATION

This study gives attention to building a desktop recovery system that depends on AI to recover data from computers infected by ransomware. The users who will benefit are people and companies using Windows platforms to store files, whether documents or images, locally. The program will help rebuild undamaged copies of their PDF files and image formats, PNG, and JPG.

Recovery is limited to local sources only. It will utilize available Shadow Copies and File History to fully restore lost data and apply AI models to help recover PDF and image files if the computer does not have recent backups. Furthermore, the system does not provide real-time ransomware prevention or removal, Its purpose is post-attack recovery and not active ransomware detection. It is assumed that users have enabled the Shadow Copies and File History in their computer and have not deleted their Shadow Copies or File History. This capstone doesn’t include payment systems, cloud integration, or encrypted network storage, but they can be considered for future enhancement.

CHAPTER II

REVIEW OF RELATED LITERATURE AND SYSTEMS

This chapter presents a comprehensive discussion of literature and systems that are relevant to the proposed capstone project titled "AI-Assisted File Recovery for Ransomware-Affected Systems." The literature and systems reviewed in this section provided foundational knowledge and insights used in conceptualizing and developing the proposed solution.

2.1 RELATED LITERATURE

2.1.1 Ransomware Threats and Data Loss

Today, Ransomware is a significant and rapidly increasing danger in cybersecurity. This type of malware makes files unreadable until you pay a demanded ransom. The results of such attacks are serious, impacting both companies' finances because of process interruptions, data privacy, and the security and privacy of important data. Because recent backups or proper restoration solutions do not protect many victims, a ransomware attack can inflict significant harm on SMEs that cannot cover cybersecurity costs [46], [41].

The number of ransomware attacks is rising worldwide, especially in cities such as Davao City in the Philippines [41]. As a result of such attacks, many people are discussing how ready and responsive local institutions, such as businesses and health organizations, are. It was found that current preparedness and response to digital attacks is weak, thus leaving systems and businesses especially susceptible to ransomware [5]. According to research, when Ransomware hits an organization, it must pay the demanded ransom and deal with significant difficulties operating until the threat is removed. It gets worse because many attackers now use double extortion, encrypting and stealing the data, forcing victims to give in to the demands quickly [55].

The recent growth in cybercrime caused by new ransomware strategies like Ransomware as a Service has put an open marketplace within reach of attackers [6]. As a result of this easy availability, more and more attacks have made Ransomware a significant issue, so it’s essential to use advanced recovery options rather than only depending on antivirus software. Systems that are intelligent and can detect and address unusual behaviors associated with ransomware attacks are key for dealing with security problems [45], [52]. Because of these risks, companies should focus on standard backup procedures and new recovery methods that help them stay strong against challenging ransomware problems [43].

In addition, these organizations need thorough cybersecurity frameworks, mainly because they hold sensitive patient information. Hacking attacks on critical sectors, as shown by the Hollywood Presbyterian Medical Center incident, can cause significant problems to public health and lead to critical financial burdens needed to recover [23]. Because of this, businesses should improve their cybersecurity by upgrading their technology and providing workers with the proper training.

2.1.2 File Recovery Technologies

Most people use Windows Shadow Copy or File History to recover files and Windows configurations when either accidentally deleted or damaged. Yet, they encounter difficulties, primarily due to complex cyber matters like ransomware. It has been found that these recovery tools fail to work if the original data has been erased or if the system lacks current backups [16].

Too little knowledge of technical details is another big problem when recovering files by hand. Because many people are not skilled enough to use these tools, manual restoration approaches do not work well outside of technical environments. Researching user skills demonstrates that most users who lack proper training cannot back up their data or successfully manage a ransomware threat [16].

Furthermore, researchers have found that advanced malware, such as ransomware, can detect and disrupt backup storage, which complicates restoring files. Some ransomware versions can block or delete any backups you have stored [24], [13]. Recent and sophisticated versions of ransomware can break users' backup mechanisms, which means backups are ineffective against attacks [16]. As a result, having strong backup solutions still might not meet the needs for recovery if the ransomware can disable or encrypt them.

In reaction to these issues, cybersecurity experts claim that adding advanced tools for prevention and identification is essential [37]. Examples of these methods are those that use machine learning to detect active ransomware so that users can protect their backups [21]. It is obvious that, as cyber threats keep changing, simple post-disaster solutions are not enough, and action should be taken to strengthen data protection by using the latest cybersecurity advancements

2.1.3 Artificial Intelligence in File Reconstruction

Thanks to deep learning techniques, AI and ML are actively shaping the process of recovering files after an attack. It has been found that denoising autoencoders can identify and rebuild the critical structures in disturbed images so that the files can be restored. Most of the research focuses on imaging applications, so using denoising autoencoders for recovering files from damaged data is rarely discussed in writings on this topic [32].

Specifically, text rebuilding efforts now depend on models such as DistilBERT and GPT-based transformers, opening the way for substantial improvements in restoring damaged documents. They make it easier to rebuild as much data as possible, even if specific key files are only partially retrievable by those trying to save files on compromised systems. Even so, solid data proving they assist in file recovery is limited, so references to their effectiveness are required [54].

It has additionally been shown that using AI in data recovery results in better graphically or textually recovering lost files and easier overall access to the process [48]. If these AI approaches are used alongside old recovery tools, there is less chance of data loss, and both professionals and users should feel at ease using these AI tools. The use of AI for post-breach recovery has increased because people trust machine-learning models more. Yet, people remain uncertain about how AI makes decisions and how its interpretations can be understood [39]. It points out that the main enhancements in AI and its application are needed to manage file recovery efficiently.

2.1.4 File Classification Using Machine Learning

AI-assisted file recovery greatly benefits from machine learning (ML) in file classification. LightGBM and Random Forest are considered practical algorithms for putting files into different categories: "safe," "recoverable," or "corrupted." Proper categorization makes it possible to take recovery actions that focus on the specific condition of the file [19], [22]. Because of these technologies' complex pattern recognition, algorithms can tell what each file type is by examining what's stored in the file, how it's organized, and what data is inside.

Research has found that machine learning models noticeably increase the reliability of damage assessment results for files. Metadata and structural integrity analyses allow these models to accurately judge how much data has been corrupted [28]. Many studies use approaches that break computers into small parts to show that classifying these pieces helps users access more data. Improved classification algorithms make it easier to recover files and help users sort and recover crucial data by priority [22].

Fieldwork shows that people are becoming more confident using machine learning in forensic and data recovery situations. Research shows that combining classification in recovery systems results in a more streamlined recovery method that protects user information as intended [30]. Data retrieval focus is given to top-priority tasks to help everyone use the system more easily and efficiently [28]. After a data loss happens, sorting the files based on condition can lead to faster decisions about the files' chances of recovery. Through predictive abilities, users receive a clear suggestion for recovery based on the state of the file because of the sophisticated analysis performed by ML algorithms [19], [22].

2.1.5 Backup-Aware Recovery Systems

Backup-aware recovery systems are now essential in information technology setups, as they take advantage of local backup tools such as Windows Volume Shadow Copy and File History. They make recovery automatic, which is more efficient than traditionally picking out and fixing problems. Investigations indicate that many people are unfamiliar with local backups and do not have the skills to access them, making it necessary for innovative recovery systems to handle these things automatically. They support the recovery process and help users complete it without involvement [58][37].

Using artificial intelligence (AI) in backup-aware recovery models is especially notable. A recent study showed that AI in disaster recovery reduces the time needed by at least 60% compared to the use of previous traditional methods [58]. Thanks to these developments, automated recovery is encouraged by anticipating problems with the help of predictive analytics. Also, such intelligent systems increase the accuracy of recovery and, therefore, enhance overall reliance on backup processes [58][37]. When automation meets AI, backup systems can help prevent mistakes from humans while improving how users access and restore their data [58].

Additionally, having backup systems is very important in keeping cybersecurity strong. In the case of advanced persistent threats and ransomware, you cannot do without automated backups [50]. Similar to what Novak et al. suggest, using similarity search with advanced security protocols makes it difficult for malware to use backups as a path for attacks, helping backup systems stay resilient [37]. As a result, better backup-aware systems and the help of AI have made data management and restoration stronger and easier for users.

2.1.6 User-Centric Design in Recovery Tools

Using user-centered design (UCD) is helping to boost user experience in cybersecurity tools and other domains. Because non-experts have difficulty with technical terms and user interfaces, designers of systems must ensure they are friendly and simple [25][15]. This reflects what is proposed in UCD literature, such as the value of an intuitive user interface, smooth processes, and understandable language [8]. Studies have shown that UCD improves user engagement and satisfaction most for those who need simple options, including recovery tools in cybersecurity [25].

These features include recovery help and clear status updates, which make things easier for people with basic computer knowledge [4][5]. An effective tool works well, gives instant feedback, and has simple recovery methods for people new to cybersecurity [62][63]. Including user ideas helps designers design systems that suit user needs and lower the possibility of errors when recovering [25][64]. Providing guided ways to work, intuitive layouts, and on-the-spot guidance helps to make cybersecurity recovery tools useful for everyone, from novice to advanced users. Such approaches make things easier for users and help get more individuals to use cybersecurity tools [15][65].

2.1.8 Disaster Recovery and Resilience Frameworks

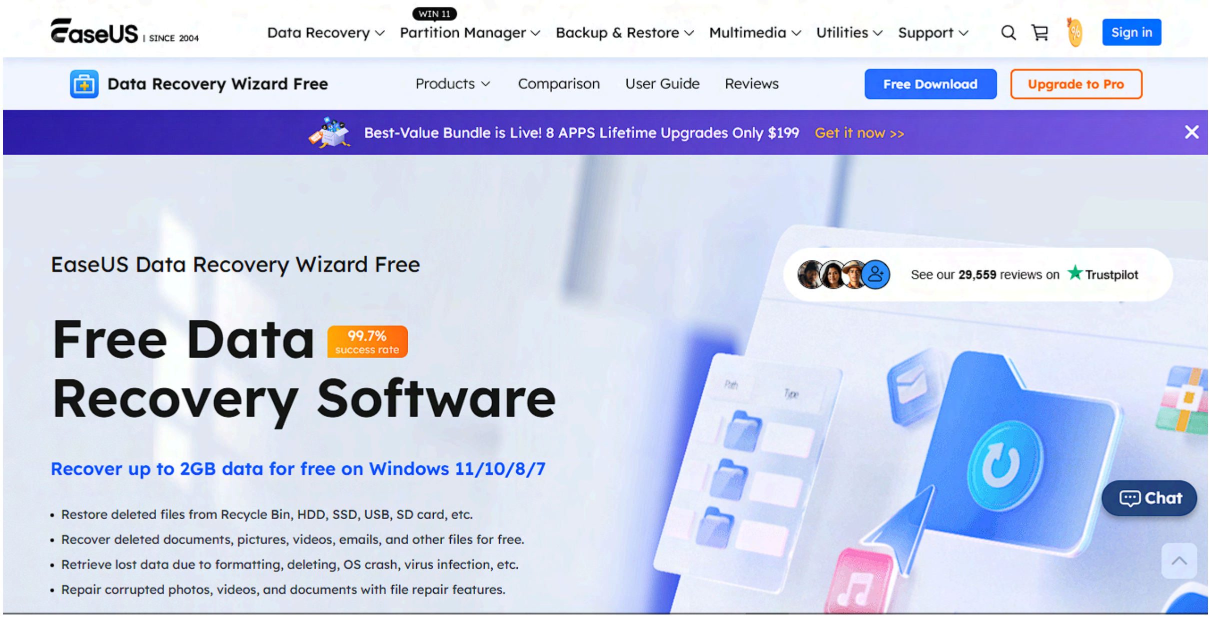
As cyberattacks evolve, preparedness, reaction, and restoration are the main efforts of disaster recovery frameworks in cybersecurity. Systems that depend on artificial intelligence are helping businesses recover more quickly from cyber attacks, namely those involving ransomware [60]38][34]. With these frameworks, organizations can fix damaged files and function properly [60][4][36]. Automatic recovery tools make it easier for organizations to work smoothly during disruptions [60][36]. Utilizing AI, these tools can find and restore the most important files to help the network withstand disruptions [60][34].

Moreover, these features are in line with worldwide cybersecurity standards such as the NIST Cybersecurity Framework (CSF) and ISO/IEC 27031, which put great value on being ready and quick to recover and protect IT systems [38][36]. Hence, adding AI to disaster recovery planning moves it beyond improvement to a fundamental shift towards a stronger way to manage cyber risks [60][4]. The introduction of an organized incident response solution makes an organization more prepared by having active plans and solutions for dealing with different cyber threats [3][44]. These strategies help businesses follow global cybersecurity requirements and develop the resilient attitude needed in the modern digital economy [34][35]. Organizations truly need this concept.

2.2 RELATED SYSTEMS

There are various existing systems designed to address file security, backup, and data recovery. However, most focus on prevention or general data management rather than post-ransomware recovery using AI. Below are several systems relevant to the concept of data recovery, encryption, or intelligent detection:

a. ***EaseUS Data Recovery Wizard***



EaseUS Data Recovery Wizard Free

Free Data Recovery Software

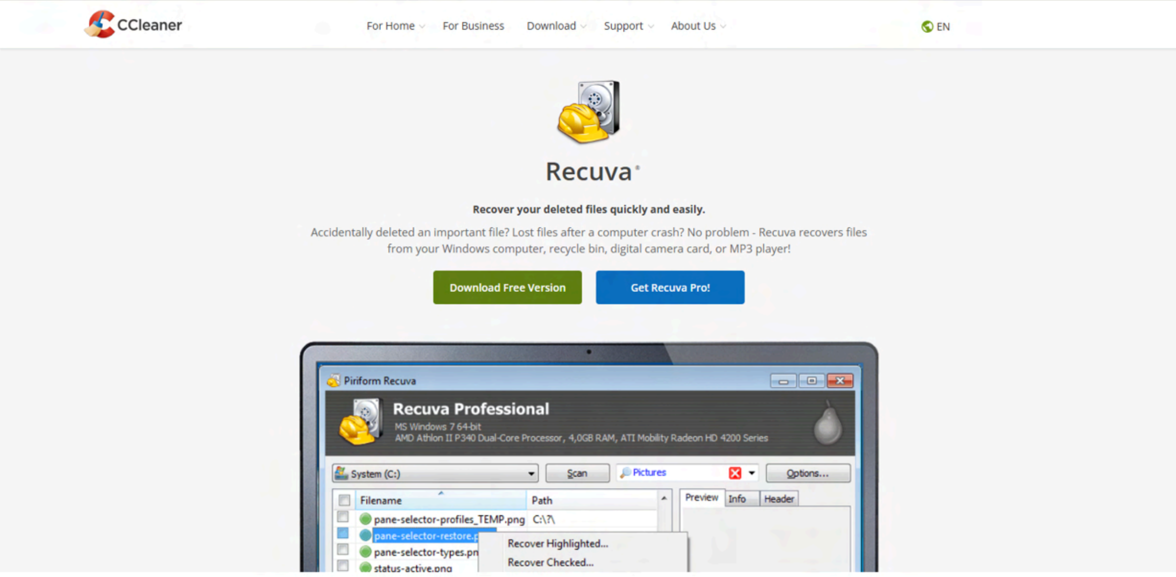
Recover up to 2GB data for free on Windows 11/10/8/7

- Restore deleted files from Recycle Bin, HDD, SSD, USB, SD card, etc.
- Recover deleted documents, pictures, videos, emails, and other files for free.
- Retrieve lost data due to formatting, deleting, OS crash, virus infection, etc.
- Repair corrupted photos, videos, and documents with file repair features.

Figure 1: EaseUS Data Recovery Wizard

EaseUS Data Recovery Wizard is a powerful desktop application designed to recover lost, deleted, formatted, or corrupted files from various storage devices, including hard drives, SSDs, USB flash drives, and memory cards. The tool supports over 1,000 file types and can restore documents, images, videos, and emails even after system crashes or virus attacks. One of its key features is its ability to perform quick and deep scans, giving users flexibility depending on the severity of the data loss. Unlike traditional recovery tools, EaseUS emphasizes a user-friendly interface, making it accessible to technical and non-technical users. However, it relies on existing file structures and does not use AI for predictive recovery. [12]

b. ***Recuva by CCleaner***



Recuva

Recover your deleted files quickly and easily.

Accidentally deleted an important file? Lost files after a computer crash? No problem - Recuva recovers files from your Windows computer, recycle bin, digital camera card, or MP3 player!

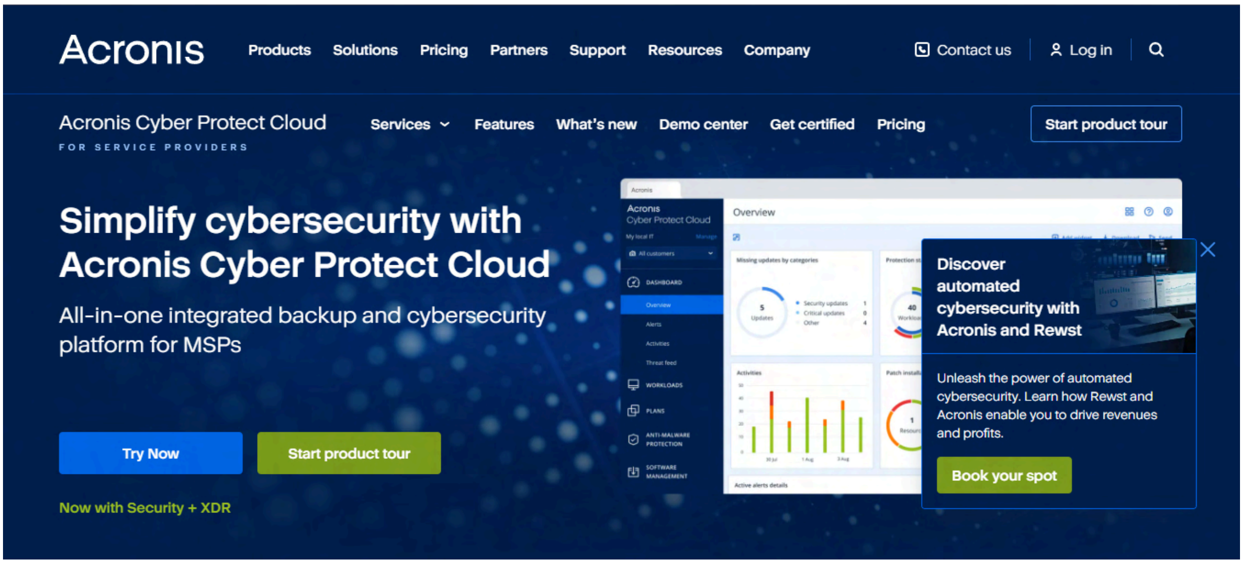
Recuva Professional
MS Windows 7 64-bit
AMD Athlon II P340 Dual-Core Processor, 4,0GB RAM, ATI Mobility Radeon HD 4200 Series

pane-selector-profiles_TEMP.png C:\7\
pane-selector-restore.p
pane-selector-types.pn
status-active.png

Figure 2: Recuva from CCleaner

Recuva from CCleaner is a lightweight file recovery tool built for Windows that helps retrieve files you mistakenly deleted from your hard drive, memory card, USB stick,k, or similar gadgets. The interface is designed as a wizard and supports documents, images, compressed archives, and other standard types. Taking advantage of the deep scan feature, Recuva goes looking for traces of deleted files outside the file table. Yet, since it can’t provide advanced features, it is less effective when a complete file rescue or classification is essential for ransomware-damaged files. [7]

c. *Acronis Cyber Protect*



Simplify cybersecurity with Acronis Cyber Protect Cloud

All-in-one integrated backup and cybersecurity platform for MSPs

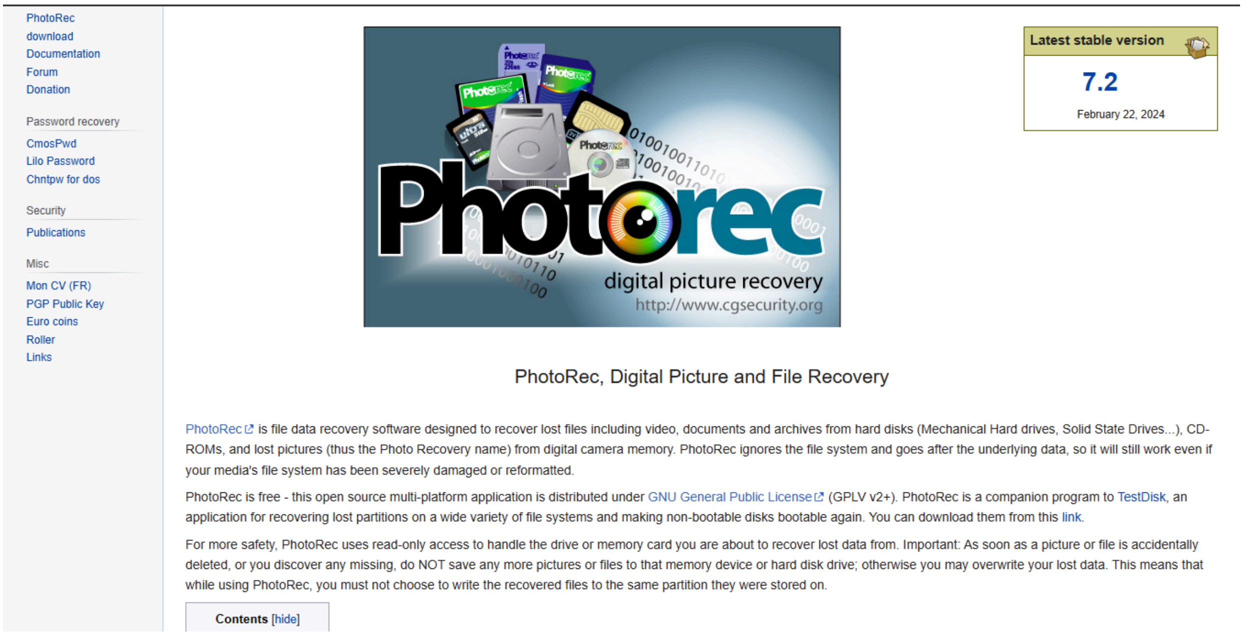
Discover automated cybersecurity with Acronis and Rewst

Unleash the power of automated cybersecurity. Learn how Rewst and Acronis enable you to drive revenues and profits.

Figure 3: Acronis Cyber Protect

Acronis Cyber Protect is an enterprise-grade solution that combines backup, anti-malware, and cybersecurity in a unified platform. It offers real-time threat detection, disk imaging, and file recovery with advanced ransomware protection. Its AI-based defense proactively blocks malicious behaviors and quickly restores compromised files from secure backups. Though powerful, it is primarily tailored for enterprise environments and may be too complex for average users. [2]

d. *PhotoRec*



PhotoRec, Digital Picture and File Recovery

PhotoRec is file data recovery software designed to recover lost files including video, documents and archives from hard disks (Mechanical Hard drives, Solid State Drives...), CD-ROMs, and lost pictures (thus the Photo Recovery name) from digital camera memory. PhotoRec ignores the file system and goes after the underlying data, so it will still work even if your media's file system has been severely damaged or reformatted.

PhotoRec is free - this open source multi-platform application is distributed under GNU General Public License (GPLV v2+). PhotoRec is a companion program to TestDisk, an application for recovering lost partitions on a wide variety of file systems and making non-bootable disks bootable again. You can download them from this link.

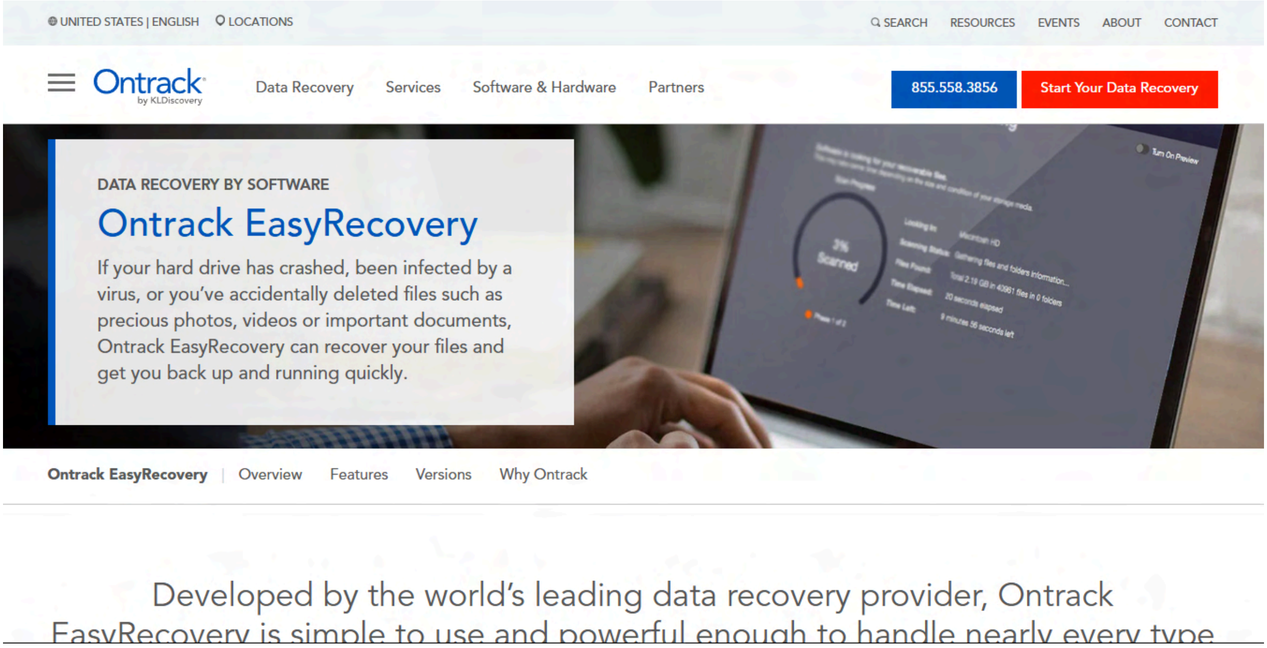
For more safety, PhotoRec uses read-only access to handle the drive or memory card you are about to recover lost data from. Important: As soon as a picture or file is accidentally deleted, or you discover any missing, do NOT save any more pictures or files to that memory device or hard disk drive; otherwise you may overwrite your lost data. This means that while using PhotoRec, you must not choose to write the recovered files to the same partition they were stored on.

Contents [hide]

Figure 4: PhotoRec

PhotoRec is a tool that uses open-source code to recover lost photos, videos, and documents from hard disks, memory cards, and CDs. Because it analyzes file headers directly, PhotoRec is still helpful when file systems are strongly damaged or have been replaced. While powerful, it lacks a graphical interface and does not support AI or file preview features, limiting usability for non-technical users. [9]

e. ***Kroll Ontrack EasyRecovery***



UNITED STATES | ENGLISH LOCATIONS SEARCH RESOURCES EVENTS ABOUT CONTACT

Ontrack by KLDiscovery Data Recovery Services Software & Hardware Partners 855.558.3856 Start Your Data Recovery

DATA RECOVERY BY SOFTWARE

Ontrack EasyRecovery

If your hard drive has crashed, been infected by a virus, or you've accidentally deleted files such as precious photos, videos or important documents, Ontrack EasyRecovery can recover your files and get you back up and running quickly.

Ontrack EasyRecovery | Overview Features Versions Why Ontrack

Developed by the world's leading data recovery provider, Ontrack EasyRecovery is simple to use and powerful enough to handle nearly every type

Figure 5: Kroll Ontrack EasyRecovery

Kroll Ontrack EasyRecovery is a professional-grade data recovery software developed to retrieve lost or deleted files from hard drives, SSDs, optical drives, and external storage devices. It supports recovery from formatted and damaged file systems and offers specialized tools for RAID and advanced repair scenarios. While it delivers reliable performance and a user-friendly interface, it is commercial software and lacks AI-driven features for predicting or reconstructing file content after encryption by ransomware. [35]

2.2.1 Comparison Matrix of Related Systems

Table 1. Comparison Matrix

System	AI-Based Prediction	Corrupted File Recovery	File Recovery	Classification Engine	Platform
EaseUS Data Recovery Wizard	No	Yes	Limited	No	Software
Recuva by CCleaner	No	Yes	No	No	Software
Acronis Cyber Protect	Limited	Yes	Limited	No	Software
PhotoRec	No	Yes	Yes	No	Software
Kroll Ontrack EasyRecovery	No	Yes	Limited	No	Software
AI-Assisted File Recovery *(Capstone)*	**Image/PDF** **reconstruction**	**Yes** **(AI Partial)**	**Yes** **(Shadow Copy / File History)**	**Yes**	Software

The table above reveals that numerous related systems support file recovery and offer simple-to-use user interfaces. However, such tools cannot use AI to forecast file activities or efficiently review and categorize files' status. They can return the user's lost files, except those partially encrypted or damaged from ransomware attacks. For example, PhotoRec can retrieve lost data without thinking, but the user must have technical experience and decide what data to keep.

Out of several options, the proposed system alone provides AI reconstruction, adds the shadow copy/history database, features a classification engine, and includes a simple interface for use. Traditional systems don't do this, but this file recovery system is made to assist after a ransomware attack by artificially reconstructing some parts of PDF and image documents using AI. It reveals that the proposed capstone project is special and beneficial to non-technical users in risky environments such as schools and SMEs in the country.

CHAPTER III

FRAMEWORK AND METHODOLOGY

This chapter presents the conceptual framework and the methodology used in the development of the capstone project titled "AI-Assisted File Recovery for Ransomware-Affected Systems." It outlines the model and development approach adopted by the researchers to build the application efficiently and effectively.

3.1 CONCEPTUAL FRAMEWORK

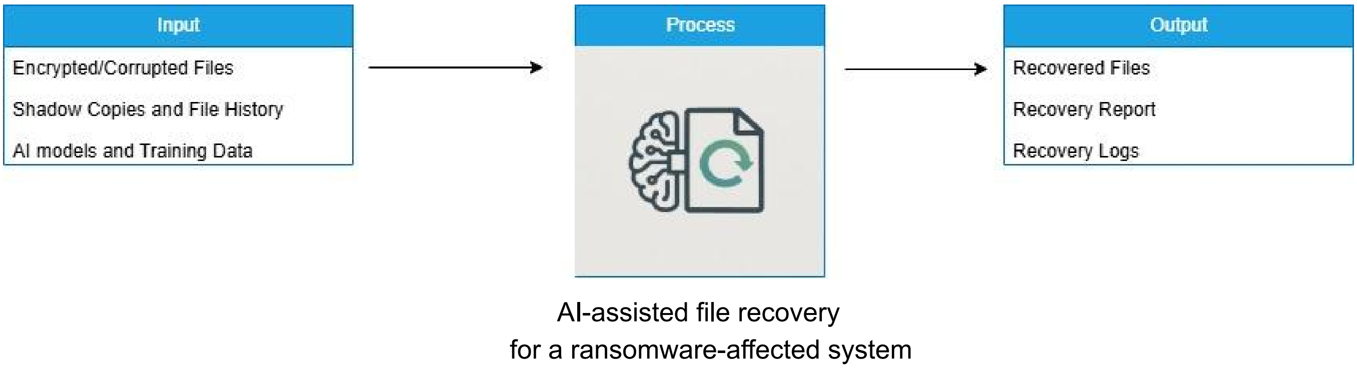

Input
Encrypted/Corrupted Files
Shadow Copies and File History
AI models and Training Data
Process
Output
Recovered Files
Recovery Report
Recovery Logs
AI-assisted file recovery
for a ransomware-affected system

Figure 6. Conceptual Framework

Researchers will use the IPO (Input-Process-Output) model to show and describe the proposed AI-assisted file recovery system's operation. The diagram indicates that affected data (input) is treated using automated intelligence and backup tools (process) to restore the lost user files (output).

In this study, the researchers use ransomware-encrypted or corrupted or damaged PDF, JPG, PNG, and DOCX files, backup copies with Windows Shadow Copy and File History, and autoencoders and language models like DistilBERT. These components are the information and equipment an office needs to function.

Files in the user's system are scanned, and a classification machine learning tool is used to check whether a file's status is safe, recovered, or altered by ransomware. If shadow copies or file history exist, the system gets and recovers the file. If backups are missing, AI models are instructed to make fixes partly using image autoencoders and NLP algorithms for texts. Users can take all the necessary actions through a simple desktop interface and have the option to define how to recover data.

The system produces the recovered file, which can be drawn from the backup or rebuilt with the help of AI. Moreover, the system reports on the progress made and tracks the steps taken during the recovery. The idea is to help users get their files back without spending money or depending on others. This framework helps clarify the system's process so that each section—called input, process, or output—matches the project's purpose

3.1.1 AI-Assisted File Recovery Framework

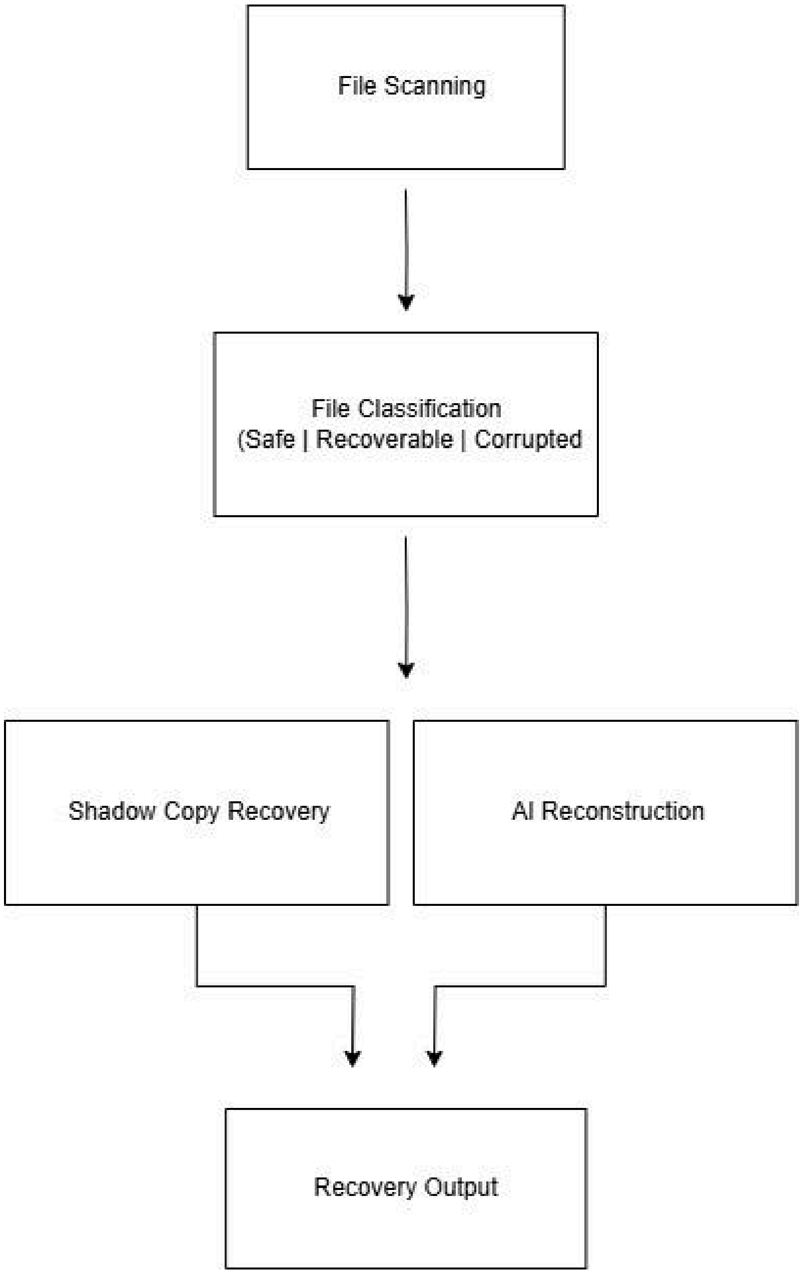


Figure 7: AI-Assisted File Recovery Framework

The user initiates the system by launching the AI-Assisted File Recovery application. Upon activation, the system begins by scanning the local drive to identify files that may have been encrypted, corrupted, or tampered with by ransomware. This is done automatically through the File Scanning module, which checks file types like PDF and Word documents and image formats like PNG or JPG.

Once the scanning is complete, the detected files proceed to the File Classification stage. Here, a lightweight machine learning classifier analyzes file characteristics—such as content entropy, metadata, and structure—to determine their current state. The files are then categorized into three classes:

Safe – No issues were detected; the file remains unchanged.

Recoverable – Shadow copies or File History versions are available.

Corrupted – No backups exist; content appears encrypted or structurally damaged.

Files marked as Recoverable are directed to the Shadow Copy Recovery module. This module checks if Windows' Volume Shadow Copies or File History versions are present. If found, the user is prompted to confirm the restoration. Once accepted, the system automatically restores the original version of the file and logs the recovery event.

On the other hand, if a file is marked as corrupted and no backup is found, it proceeds to the AI Reconstruction module. In this phase, the system uses AI models to attempt partial recovery:

- For image files (JPG/PNG), autoencoders are used to reconstruct damaged or missing pixel data.
- A language model like DistilBERT is used for PDF or Word documents to predict and fill in lost text based on formatting and context.

All recovery efforts from either process reach the Recovery Output module. The summary displayed tells you which files were returned fully and partially repaired, and which the tool couldn't recover. The next step is for the user to save the data wherever they like. As a result of this process, users can restore their data without needing to pay a ransom or hire outside IT professionals. Combining backup and AI recovery makes this tool useful and easy for anyone.

3.2 Methodology

The researchers will adopt the **Rapid Application Development (RAD)** model in developing the system. RAD emphasizes short cycles of planning, prototyping, testing, and deployment. It is particularly suitable for this project because it enables fast iterations and flexible modifications based on user testing and feedback.

Rapid Application Development (RAD) Model

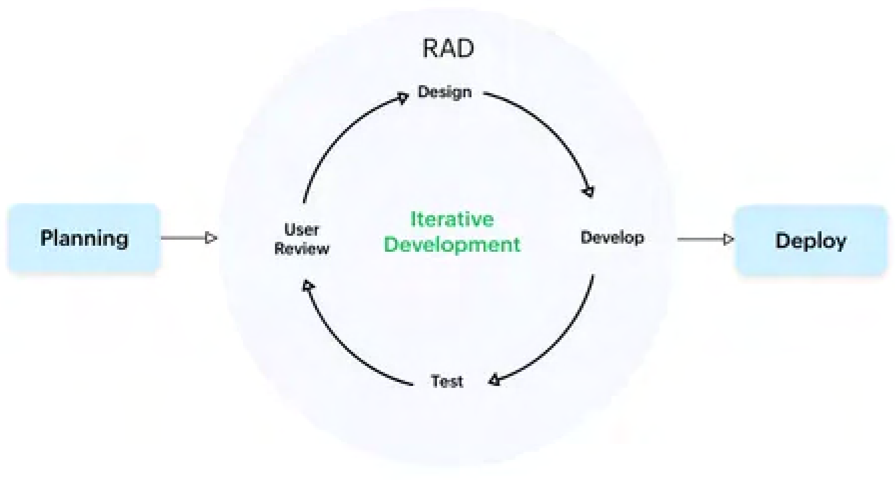


Figure 8. Rapid Application Development (RAD) Model

RAD is a vital approach today due to its benefits of reduced time, increased flexibility, and quick responses to what users require. Repeating prototypes allows regular collection of user opinions that directly match the end-user's needs [10][57]. At the same time, uniting Computer-Aided Software Engineering (CASE) tools and rapid prototyping speeds up the process and provides higher-quality results thanks to user feedback [57].

RAD is successful in overcoming challenges when building many kinds of applications. This has been used to identify manufacturing steps and choose suppliers, all in a shorter time than expected [47]. The ability to easily change the requirements on short notice is possible because of the flexibility of RAD approaches, as seen in web application research [57][42][20]. Because of these attributes, teams can deal successfully with adjustments in project needs [40][29].

3.2.1 Requirements Planning

The study relies on various online materials, articles, and technical manuals to explore the growing threat of ransomware attacks at the local, national, and international levels. These experts will also question users at educational institutions, IT departments, and small businesses to see the most common issues after a ransomware attack. The data gathered directs the team to pick out the most relevant and useful features for their file recovery system.

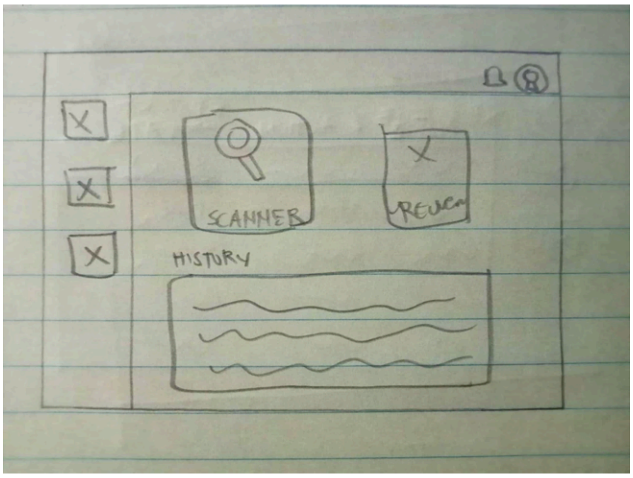
3.2.2 User Design

The researchers will begin designing the system's interface and core interaction flow in this phase. The goal is to create a user-friendly recovery tool that simplifies the process even for non-technical users. The design process includes low-fidelity (Lo-Fi) prototypes to test the interface and user journey. Each iteration will be validated by feedback gathered from test users to ensure the system is functional, intuitive, and accessible.

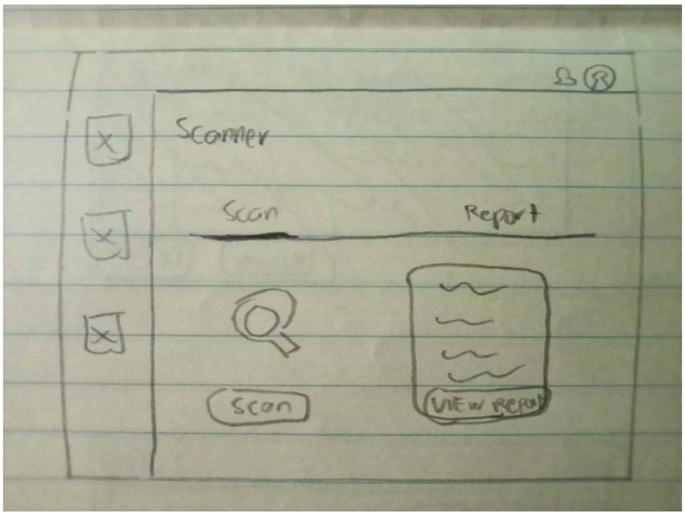
Low-Fidelity Prototype

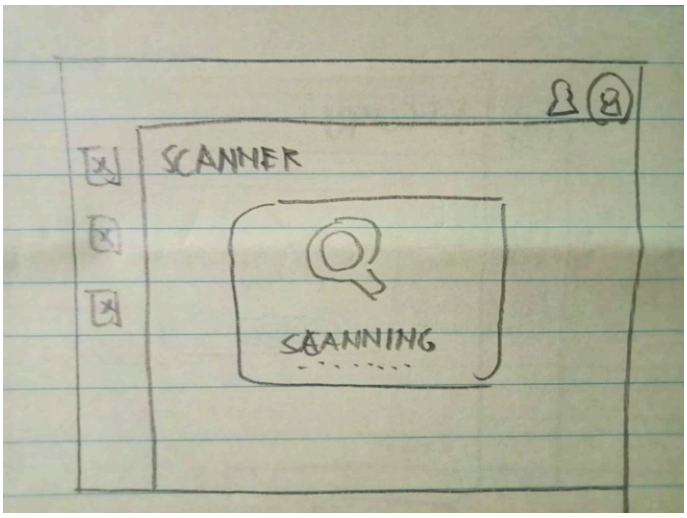
Using a user-centered design approach, the researchers created simple wireframes and layout sketches that reflect the essential modules of the system.

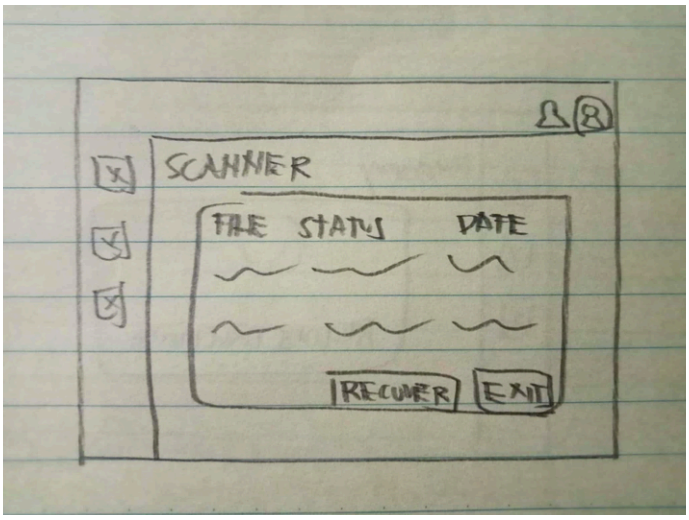
Dashboard

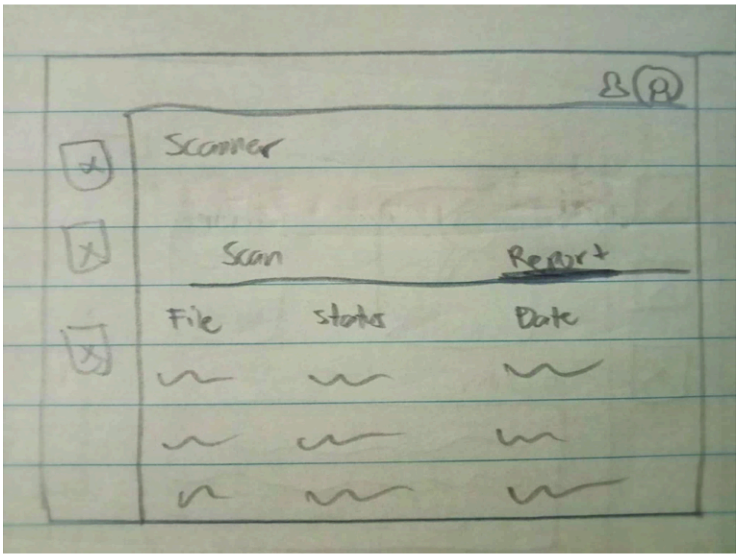


Scanner Panel

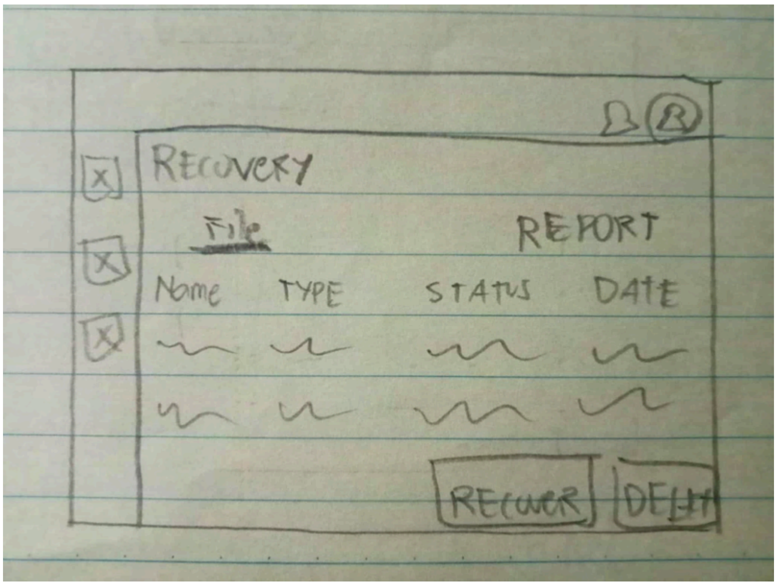


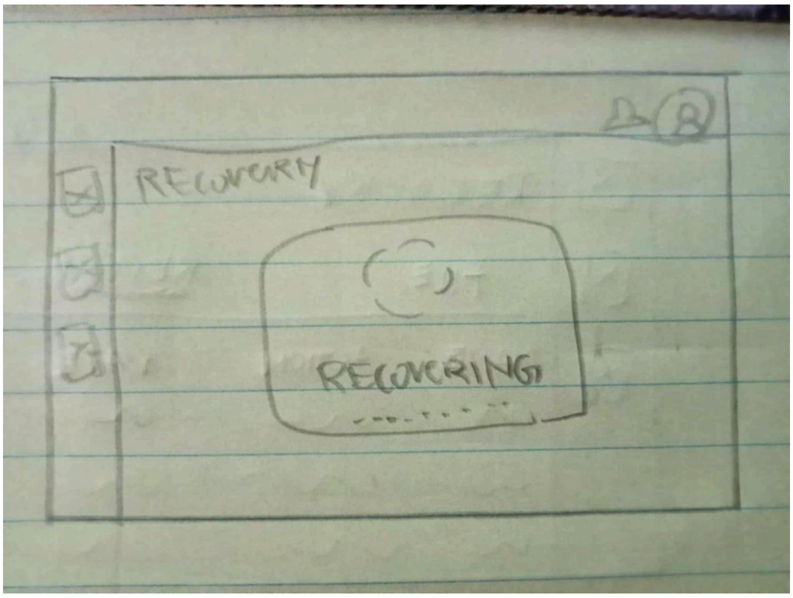

SCANNER
SCANNING

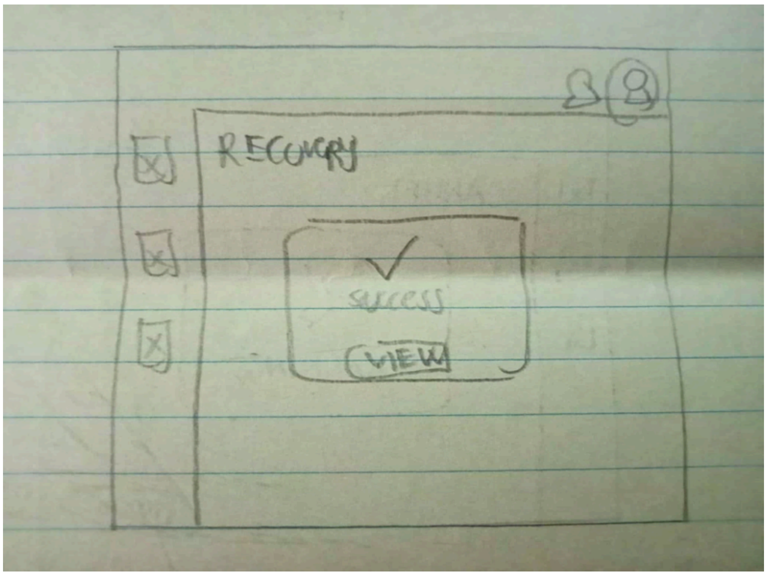

SCANNER
FILE
STATUS
DATE
RECOVER
EXIT


Scanner
Scan
Report
File
Status
Date

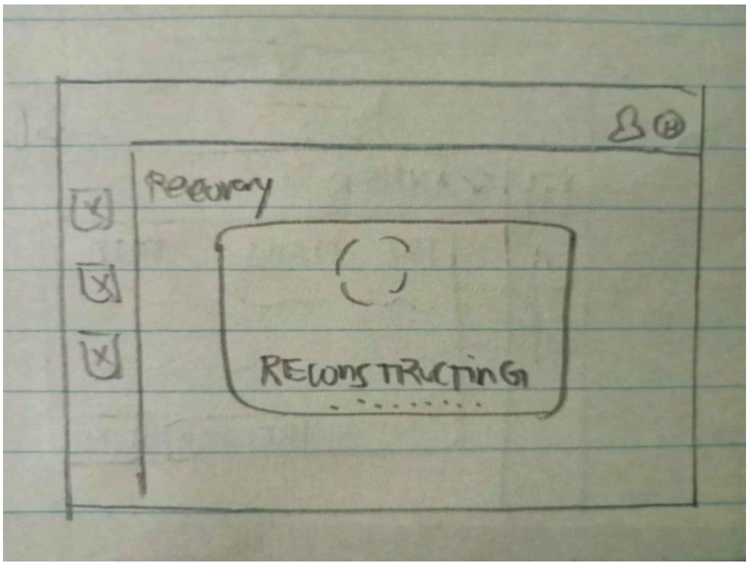
File Recovery Panel

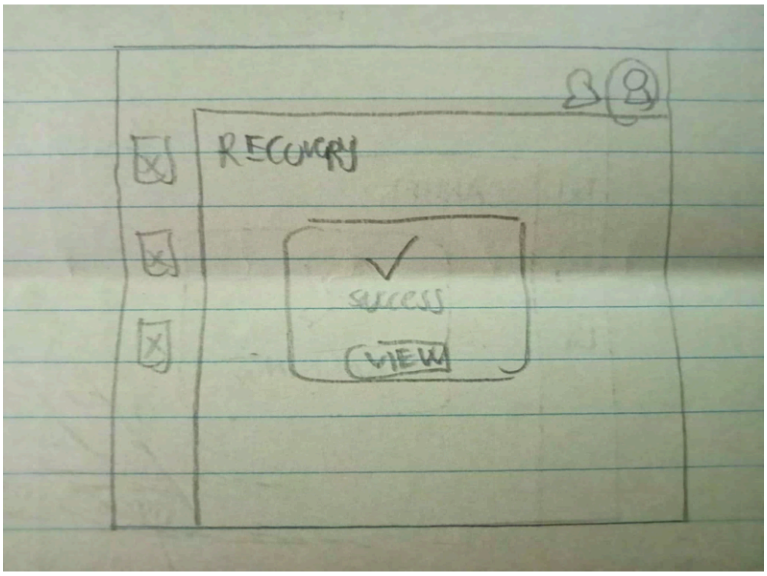




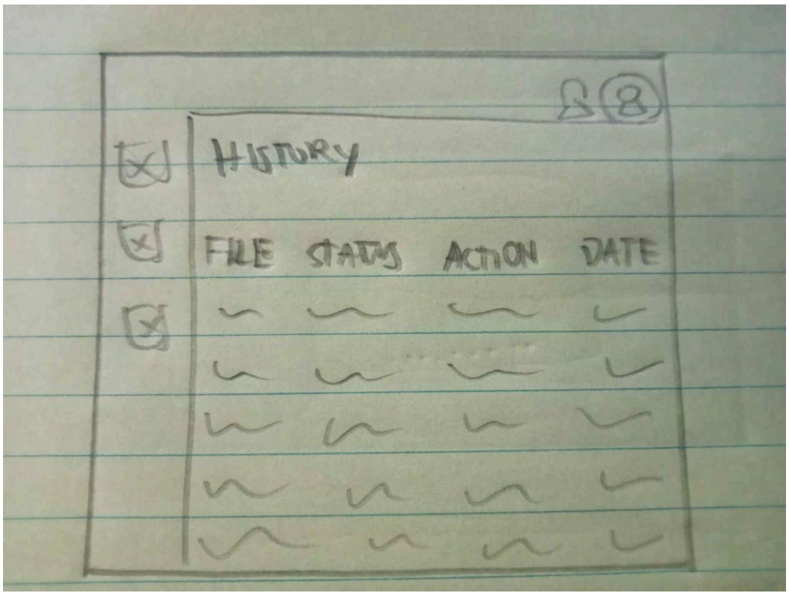


For AI Reconstructing (Images/PDF)





History Panel



These first-stage models allow the team to confirm their ideas before moving to expensive development phases. The information from usability testing helps make the system easier to use and more dependable. The quick prototypes have helped the researchers check that their design meets users' needs before proceeding with extensive design work. The early user comments helped the researchers pinpoint what improvements were needed in the design and how it works. With this input, the team will next work on creating and implementing a high-quality prototype of the AI-assisted file recovery system, guaranteed to match the standards of those using it.

3.2.3 Construction

The researchers will assemble the system by combining modules such as the file scanner, file classifier, backup retriever, and AI engine. AI and file management libraries in Python are why they will be the most used programming language. As the system works to get back to normal, any available Shadow Copies and File History versions will be found and accessed on their own. AI models will try to restore part of the files using training data and their predictive abilities if they aren’t backed up.

The researchers will apply autoencoders for images and DistilBERT for text in PDF or Word files to improve the accuracy of recovering data from damaged files. With this technology, the system can advise users using artificial intelligence and help heal damaged data, even if the user does not have a full backup. It will also involve security features to check all restored files before storing them to protect against future infection or erasing existing data.

3.2.4 Cutover

It is the point when the system is installed and equipped with the necessary tools for operation. The transition to the new system will include changing test data, carrying out the last tests, explaining the software to users, and getting all needed documentation ready. At this point, the researchers carefully watch the system to detect and solve bugs and issues people may have using it.

Once the application is polished and complete, it will be distributed as a desktop recovery app. Any changes, enhancements, or improvements will be made according to what people report in the test environment. Because RAD uses iterative development, the team hopes to make file recovery for everyone affected by ransomware dependable and straightforward.

REFERENCES

[1] A. Alraizza and A. Algarni, "Ransomware detection using machine learning: a survey", Big Data and Cognitive Computing, vol. 7, no. 3, p. 143, 2023. https://doi.org/10.3390/bdcc7030143

[2] Acronis, “Acronis Cyber Protect: Cybersecurity and backup integration,” Acronis, 2024. [Online]. Available: https://www.acronis.com/en-us/products/cyber-protect/

[3] A. C., "Strengthening india’s cybersecurity and data privacy landscape: a comprehensive overview", Indian Journal of Public Administration, vol. 70, no. 3, p. 466-478, 2024. https://doi.org/10.1177/00195561241271616

[4] A. García-Pérez, J. Navarro, M. Sallos, E. Martínez, & A. Chinnaswamy, "Resilience in healthcare systems: cyber security and digital transformation", Technovation, vol. 121, p. 102583, 2023. https://doi.org/10.1016/j.technovation.2022.102583

[5] A. Kesarwani and S. Gochhayat, "Ransomware attacks in the healthcare industry", Journal of Student Research, vol. 12, no. 4, 2023. https://doi.org/10.47611/jsrhs.v12i4.5799

[6] A. Yadav, A. Kumar, A. Totuka, D. Sharma, & M. Jain, "Cyber extortion revealed: an analysis of ransomware's development, strategies, obstacles, and prospects for the future",, 2024. https://doi.org/10.52783/eel.v14i2.1417

[7] CCleaner, “Recuva - Restore deleted files quickly and easily,” CCleaner, 2024. [Online]. Available: https://www.ccleaner.com/recuva

[8] C. Farinango, J. Benavides, J. Cerón, D. López, & R. Álvarez, "Human-centered design of a personal health record system for metabolic syndrome management based on the iso 9241-210:2010 standard", Journal of Multidisciplinary Healthcare, vol. Volume 11, p. 21-37, 2018. https://doi.org/10.2147/jmdh.s150976

[9] CGSecurity, “PhotoRec - Digital Picture and File Recovery,” CGSecurity, 2024. [Online]. Available: https://www.cgsecurity.org/wiki/PhotoRec

[10] C. Maduabuchukwu and A. Edje, "Alocalizedbasedapplicationforautomobilemechanicslocation-awaresystem", NJSE, vol. 22, no. 1, p. 106-124, 2024. https://doi.org/10.61448/njse221249

[11] Cybersecurity Ventures, "Global Ransomware Damage Costs Predicted To Reach $265 Billion By 2031," 2021. [Online]. Available: https://cybersecurityventures.com/ransomware-damage-report-2021/. [Accessed: May 22, 2025].

[12] EaseUS, “EaseUS Data Recovery Wizard: Powerful Data Recovery Software,” EaseUS, 2024. [Online]. Available: https://www.easeus.com/datarecoverywizard/

[13] G. Arányi, Á. Vathy-Fogarassy, & V. Szücs, "Evaluation of a new-concept secure file server solution", Future Internet, vol. 16, no. 9, p. 306, 2024. https://doi.org/10.3390/fi16090306

[14] H. Neprash, C. McGlave, K. Rydberg, & C. Henning‐Smith, "What happens to rural hospitals during a ransomware attack? evidence from medicare data", The Journal of Rural Health, vol. 40, no. 4, p. 728-737, 2024. https://doi.org/10.1111/jrh.12834

[15] H. Witteman, G. Vaisson, T. Provencher, S. Dansokho, H. Colquhoun, M. Dugaset al., "An 11-item measure of user- and human-centered design for personal health tools (ucd-11): development and validation", Journal of Medical Internet Research, vol. 23, no. 3, p. e15032, 2021. https://doi.org/10.2196/15032

[16] J. Ahn, D. Park, C. Lee, D. Min, J. Lee, S. Park et al., "Key-ssd: access-control drive to protect files from ransomware attacks",, 2019. https://doi.org/10.48550/arxiv.1904.05012

[17] J. Kumar, "Enhancing public awareness and education of ransomware attacks",, 2023. https://doi.org/10.36227/techrxiv.24634806

[18] J. Long and H. Liang, "Ranaway: a novel ransomware-resilient refs file system",, 2024. https://doi.org/10.21203/rs.3.rs-3960276/v1

[19] J. Oh, S. Lee, & H. Hwang, "Forensic recovery of file system metadata for digital forensic investigation", Ieee Access, vol. 10, p. 111591-111606, 2022. https://doi.org/10.1109/access.2022.3213030

[20] J. Patero, "Streamlining physics laboratory management: an information system solution", International Journal of Advanced Research in Science Communication and Technology, p. 796-801, 2023. https://doi.org/10.48175/ijarsct-12373

[21] J. Yun, J. Hur, Y. Shin, & D. Koo, "Cldsafe: an efficient file backup system in cloud storage against ransomware", Ieice Transactions on Information and Systems, vol. E100.D, no. 9, p. 2228-2231, 2017. https://doi.org/10.1587/transinf.2017edl8052

[22] K. Alghafli and T. Martin, "Identification and recovery of video fragments for forensics file carving",, 2016. https://doi.org/10.1109/icitst.2016.7856710

[23] K. Lee, K. Yim, & J. Seo, "Ransomware prevention technique using key backup", Concurrency and Computation Practice and Experience, vol. 30, no. 3, 2017. https://doi.org/10.1002/cpe.4337

[24] K. Lee, S. Lee, & K. Yim, "Machine learning based file entropy analysis for ransomware detection in backup systems", Ieee Access, vol. 7, p. 110205-110215, 2019. https://doi.org/10.1109/access.2019.2931136

[25] M. Hakimi, M. Quchi, & A. Fazil, "Human factors in cybersecurity: an in depth analysis of user centric studies", esaprom, vol. 3, no. 01, p. 20-33, 2024. https://doi.org/10.58471/esaprom.v3i01.3832

[26] M. Muniandy, N. Ismail, A. Al-Nahari, & D. Yao, "Evolution and impact of ransomware: patterns, prevention, and recommendations for organizational resilience", International Journal of Academic Research in Business and Social Sciences, vol. 14, no. 1, 2024. https://doi.org/10.6007/ijarbss/v14-i1/19803

[27] M. Shaikh, H. Ullah, R. Akbar, S. Sugathan, & S. Mandala, "Fortifying against ransomware: navigating cybersecurity risk management with a focus on ransomware insurance strategies", International Journal of Academic Research in Business and Social Sciences, vol. 14, no. 1, 2024. https://doi.org/10.6007/ijarbss/v14-i1/20566

[28] M. Teimouri, Z. Seyedghorban, & F. Amirjani, "Fragments‐expert: a graphical user interface matlab toolbox for classification of file fragments", Concurrency and Computation Practice and Experience, vol. 33, no. 9, 2020. https://doi.org/10.1002/cpe.6154

[29] N. Amiruddin, S. Sa'dan, H. Mohamed, & N. Adzmi, "The research assistant management system (rams)", International Journal of Advanced Science Computing and Engineering, vol. 1, no. 3, p. 169-178, 2019. https://doi.org/10.30630/ijasce.1.3.23

[30] N. Beebe, L. Liu, & M. Sun, "Data type classification: hierarchical class-to-type modeling",, p. 325-343, 2016. https://doi.org/10.1007/978-3-319-46279-0_17

[31] N. Mayeke, A. Arigbabu, O. Olaniyi, O. Okunleye, & C. Adigwe, "Evolving access control paradigms: a comprehensive multi-dimensional analysis of security risks and system assurance in cyber engineering", Asian Journal of Research in Computer Science, vol. 17, no. 5, p. 108-124, 2024. https://doi.org/10.9734/ajrcos/2024/v17i5442

[32] N. Pezzotti, S. Yousefi, M. Elmahdy, J. Gemert, C. Schuelke, M. Donevaet al., "An adaptive intelligence algorithm for undersampled knee mri reconstruction", Ieee Access, vol. 8, p. 204825-204838, 2020. https://doi.org/10.1109/access.2020.3034287

[33] Ontrack, "EasyRecovery Software - Data Recovery Solutions," Ontrack, 2024. [Online]. Available: https://www.ontrack.com/en-us/products/data-recovery-software

[34] P. Bansal, "India's digital transformation: opportunities and challenges in the digital economy", International Journal of Economic Policy, vol. 4, no. 2, p. 53-57, 2024. https://doi.org/10.47941/ijecop.1947

[35] P. Chiara, "The cyber resilience act: the eu commission’s proposal for a horizontal regulation on cybersecurity for products with digital elements", International Cybersecurity Law Review, vol. 3, no. 2, p. 255-272, 2022. https://doi.org/10.1365/s43439-022-00067-6

[36] P. Nair, "Enhancing cybersecurity awareness training through the nist framework", Ijarcce, vol. 12, no. 12, 2023. https://doi.org/10.17148/ijarcce.2023.121203

[37] P. Novak, V. Oujezský, P. Kaura, T. Horváth, & M. Holík, "Multistage malware detection method for backup systems", Technologies, vol. 12, no. 2, p. 23, 2024. https://doi.org/10.3390/technologies12020023

[36] Q. Kang and Y. Gu, "Enhancing ransomware detection: a windows api min max relevance refinement approach",, 2023. https://doi.org/10.20944/preprints202311.1004.v1

[37] Q. Zhong and Q. Kang, "Ransomware detection with opcode analysis and gan-based unsupervised learning",, 2024. https://doi.org/10.21203/rs.3.rs-3819158/v1

[38] R. Anggraini, "Analisis keamanan private cloud berbasis framework nistcy di pt xyz", Jurnal Teknologi Dan Manajemen, vol. 19, no. 1, p. 41-46, 2021. https://doi.org/10.52330/jtm.v19i1.11

[39] R. Arab, "Artificial intelligence in hospital infection prevention: an integrative review", Frontiers in Public Health, vol. 13, 2025. https://doi.org/10.3389/fpubh.2025.1547450

[40] R. Bali and N. Wickramasinghe, "Rad and other innovative approaches to facilitate superior project management",, p. 149-155, 2012. https://doi.org/10.4018/978-1-4666-1559-5.ch010

[41] R. Mehmood, "Effects of ransomware: analysis, challenges and future perspective", International Journal for Electronic Crime Investigation, vol. 7, no. 3, p. 49-60, 2023. https://doi.org/10.54692/ijeci.2023.0703161

[42] S. Agustina, D. Yunautama, & B. Juliandani, "Pengembangan aplikasi web untuk optimalisasi manajemen toko surya mas 2", Prosiding Seminar Sosial Politik Bisnis Akuntansi Dan Teknik, vol. 5, p. 183, 2023. https://doi.org/10.32897/sobat.2023.5.0.3096

[43] S. Alzahrani, Y. Xiao, & W. Sun, "An analysis of conti ransomware leaked source codes", Ieee Access, vol. 10, p. 100178-100193, 2022. https://doi.org/10.1109/access.2022.3207757

[44] S. Creese, W. Dutton, P. Esteve‐González, & R. Shillair, "Cybersecurity capacity building: cross-national benefits and international divides", SSRN Electronic Journal, 2020. https://doi.org/10.2139/ssrn.3658350

[45] S. Homayoun, A. Dehghantanha, M. Ahmadzadeh, S. Hashemi, R. Khayami, K. Chooet al., "Drthis: deep ransomware threat hunting and intelligence system at the fog layer", Future Generation Computer Systems, vol. 90, p. 94-104, 2019. https://doi.org/10.1016/j.future.2018.07.045

[46] S. Homayoun, A. Dehghantanha, M. Ahmadzadeh, S. Hashemi, & R. Khayami, "Know abnormal, find evil: frequent pattern mining for ransomware threat hunting and intelligence", Ieee Transactions on Emerging Topics in Computing, vol. 8, no. 2, p. 341-351, 2020. https://doi.org/10.1109/tetc.2017.2756908

[47] S. Karthik, C. Chung, & K. Ramani, "Rapid application development of process capability: supplier models",, 2003. https://doi.org/10.1115/detc2003/cie-48269

[48] S. Lombardi and E. Montague, "The role of artificial intelligence in usability testing and its potential impact on equity and medical solution design: a narrative literature review (preprint)",, 2024. https://doi.org/10.2196/preprints.67285

[49] Sophos, "The State of Ransomware 2023," Sophos, 2023. [Online]. Available: https://www.sophos.com/en-us/content/state-of-ransomware.[Accessed: May 22, 2025].

[50] S. Razaulla, C. Fachkha, C. Markarian, A. Gawanmeh, W. Mansoor, B. Funget al., "The age of ransomware: a survey on the evolution, taxonomy, and research directions", Ieee Access, vol. 11, p. 40698-40723, 2023. https://doi.org/10.1109/access.2023.3268535

[51] S. Saeed, S. Altamimi, N. Alkayyal, E. Alshehri, & D. Alabbad, "Digital transformation and cybersecurity challenges for businesses resilience: issues and recommendations", Sensors, vol. 23, no. 15, p. 6666, 2023. https://doi.org/10.3390/s23156666

[52] "Strengthening enterprise cybersecurity: a survey on ransomware mitigation and recovery strategies", nano-ntp, vol. 20, no. 6, 2024. https://doi.org/10.62441/nano-ntp.v20i6.35

[53] "The ransomware rollercoaster: a comprehensive survey on attack methods, consequences, and the evolving landscape", International Research Journal of

Modernization in Engineering Technology and Science, 2024. https://doi.org/10.56726/irjmets56122

[54] T. Kaluarachchi, A. Reis, & S. Nanayakkara, "A review of recent deep learning approaches in human-centered machine learning", Sensors, vol. 21, no. 7, p. 2514, 2021. https://doi.org/10.3390/s21072514

[55] T. Meurs, E. Cartwright, & A. Cartwright, "Double-sided information asymmetry in double extortion ransomware",, 2024. https://doi.org/10.21203/rs.3.rs-3866535/v1

[56] T. Vilarinho, B. Farshchian, J. Floch, & O. Hansen, "Participatory ideation for gamification: bringing the user at the heart of the gamification design process",, p. 51-61, 2018. https://doi.org/10.1007/978-3-030-05909-5_4

[57] U. Saprudin and M. Pratama, "Studi kasus penerapan metode rad dalam pengembangan website silsilah keluarga mbah mansyur", Jurnal Jtik (Jurnal Teknologi Informasi Dan Komunikasi), vol. 9, no. 1, p. 314-326, 2024. https://doi.org/10.35870/jtik.v9i1.3185

[58] V. Rao, "Advancements in ai-driven disaster recovery: predictive failure detection and automated data protection", International Journal for Multidisciplinary Research, vol. 6, no. 5, 2024. https://doi.org/10.36948/ijfmr.2024.v06i05.29320

[59] V. Malik, A. Khanna, N. Sharma, & S. nalluri, "Trends in ransomware attacks: analysis and future predictions",, 2024. https://doi.org/10.21428/e90189c8.f2996624

[60] V. Onih, Y. Sevidzem, & S. Adeniji, "The role of ai in enhancing threat detection and response in cybersecurity infrastructures", International Journal of Scientific and Management Research, vol. 07, no. 04, p. 64-96, 2024. https://doi.org/10.37502/ijsmr.2024.7404

[61] Y. He, Λ. Μαγλαράς, A. Aliyu, & C. Luo, "Healthcare security incident response strategy - a proactive incident response (ir) procedure", Security and Communication Networks, vol. 2022, p. 1-10, 2022. https://doi.org/10.1155/2022/2775249

[62] R. Pushpakumar, K. Sanjaya, S. Rathika, A. Alawadi, K. Makhzuna, S. Venkateshet al., "Human-computer interaction: enhancing user experience in interactive systems", E3s Web of Conferences, vol. 399, p. 04037, 2023. https://doi.org/10.1051/e3sconf/202339904037

[63] C. Siahaan and S. Syafrianto, "Desain ui/ux website inventory barang pada pt dari visi teknologi menggunakan metode user-centered design", Journal Zetroem, vol. 5, no. 1, p. 31-35, 2023. https://doi.org/10.36526/ztr.v5i1.2589

[64] M. Bounouar, R. Béarée, A. Siadat, N. Klement, & T. Benchekroun, "User-centered design of a collaborative robotic system for an industrial recycling operation",, p. 1-6, 2020. https://doi.org/10.1109/iraset48871.2020.9092178

[65] W. Wasino, C. Lim, & E. Dewayani, "User interface design of west java’s intangible cultural heritage website using user centered design", ijaste, vol. 1, no. 2, p. 421-432, 2023. https://doi.org/10.24912/ijaste.v1.i2.421-432