

A  
Major Project  
On  
**ARTIFICIAL INTELLIGENCE CRIME: AN OVERVIEW OF  
MALICIOUS USE AND ABUSE OF AI**

(Submitted in partial fulfillment of the requirements for the award of Degree)

**BACHELOR OF TECHNOLOGY**  
In  
**COMPUTER SCIENCE AND ENGINEERING**  
By

CH.NIHARIKA	(217R1A0516)
P. NIKITHA	(227R5A0502)
A. SHIVANI	(217R1A0501)

Under the Guidance of  
**MR.M.MADHUSUDHAN**  
(Assistant Professor)



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**  
**CMR TECHNICAL CAMPUS**  
**UGC AUTONOMOUS**

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi)  
Recognized Under Section 2(f) & 12(B) of the UGCAct.1956,  
Kandlakoya (V), Medchal Road, Hyderabad-501401.

**April, 2025.**

# DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



## CERTIFICATE

This is to certify that the project entitled “**ARTIFICIAL INTELLIGENCE CRIME: AN OVERVIEW OF MALICIOUS USE AND ABUSE OF AI**” being submitted by **CH.NIHARIKA(217R1A0516),P.NIKITHA(227R5A0502)&A.SHIVANI(217R1A0501)** in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering to the Jawaharlal Nehru Technological University Hyderabad, during the year 2024-25.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

**Mr.M.Madhusudhan**  
**Assistant Professor**  
**INTERNAL GUIDE**

**Dr. Nuthanakanti Bhaskar**  
**HoD**

**Dr. A. Raji Reddy**  
**DIRECTOR**

**Signature of External Examiner**

**Submitted for viva voice Examination held on** \_\_\_\_\_

## ACKNOWLEDGEMENT

We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project, we take this opportunity to express our profound gratitude and deep regard to our guide **Mr.M.Madhusudhan**, Associate Professor for his exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by him shall carry us a long way in the journey of life on which we are about to embark.

We take this opportunity to extend our heartfelt appreciation to the Project Review Committee (PRC) Coordinators—**Dr. K. Maheswari, Dr. J. Narasimharao, Ms. K. Shilpa, and Mr. K. Ranjith Reddy**—for their unwavering support, insightful guidance, and valuable inputs, which played a crucial role in steering this project through its various stages.

Our sincere appreciation also goes to **Dr. Nuthanakanti Bhaskar**, Head, for his encouragement and continuous support in ensuring the successful completion of our project.

We are deeply grateful to **Dr. A. Raji Reddy**, Director, for his cooperation throughout the course of this project. Additionally, we extend our profound gratitude to Sri. **Ch. Gopal Reddy**, Chairman, Smt. **Ch. Vasantha Latha**, Secretary and Sri. **Ch. Abhinav Reddy**, Vice-Chairman, for fostering an excellent infrastructure and a conducive learning environment that greatly contributed to our progress.

We also acknowledge and appreciate the guidance and assistance provided by the faculty and staff of **CMR Technical Campus**, whose contributions have been invaluable in bringing this project to fruition.

Lastly, we sincerely thank our families for their unwavering support and encouragement. We also extend our gratitude to the teaching and non-teaching staff of CMR Technical Campus for their guidance and assistance. Their contributions, along with the support of everyone who helped directly or indirectly, have been invaluable in the successful completion of this project.

**CH.NIHARIKA (217R1A0516)**

**P.NIKITHA (227R5A0502)**

**A.SHIVANI (217R1A0501)**

## **VISION AND MISSION**

### **INSTITUTE VISION:**

To Impart quality education in serene atmosphere thus strive for excellence in Technology and Research.

### **INSTITUTE MISSION:**

1. To create state of art facilities for effective Teaching- Learning Process.
2. Pursue and Disseminate Knowledge based research to meet the needs of Industry & Society.
3. Infuse Professional, Ethical and Societal values among Learning Community.

### **DEPARTMENT VISION:**

To provide quality education and a conducive learning environment in computer engineering that foster critical thinking, creativity, and practical problem-solving skills.

### **DEPARTMENT MISSION:**

1. To educate the students in fundamental principles of computing and induce the skills needed to solve practical problems.
2. To provide State-of-the-art computing laboratory facilities to promote industry institute interaction to enhance student's practical knowledge.
3. To inculcate self-learning abilities, team spirit, and professional ethics among the students to serve society.

## ABSTRACT

This project is titled as “Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI”. The capabilities of Artificial Intelligence (AI) evolve rapidly and affect almost all sectors of society. AI has been increasingly integrated into criminal and harmful activities, expanding existing vulnerabilities, and introducing new threats. This article reviews the relevant literature, reports, and representative incidents which allows to construct a typology of the malicious use and abuse of systems with AI capabilities. The main objective is to clarify the types of activities and corresponding risks. Our starting point is to identify the vulnerabilities of AI models and outline how malicious actors can abuse them. Subsequently, we explore AI-enabled and AI-enhanced attacks. While we present a comprehensive overview, we do not aim for a conclusive and exhaustive classification. Rather, we provide an overview of the risks of enhanced AI application, that contributes to the growing body of knowledge on the issue. Specifically, we suggest four types of malicious abuse of AI (integrity attacks, unintended AI outcomes, algorithmic trading, membership inference attacks) and four types of malicious use of AI (social engineering, misinformation/fake news, hacking, autonomous weapon systems). Mapping these threats enables advanced reflection of governance strategies, policies, and activities that can be developed or improved to minimize risks and avoid harmful consequences. Enhanced collaboration among governments, industries, and civil society actors is vital to increase preparedness and resilience against malicious use and abuse of AI.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, different type of algorithms is trained to make classifications or predictions, and to uncover key insights in this project. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics.

Machine learning algorithms build a model based on this project data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of datasets, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

## LIST OF FIGURES

FIGURE NO	FIGURE NAME	PAGE NO
Figure 3.1	Project Architecture of artificial intelligence crime: an overview of malicious use and abuse of AI	13
Figure 3.2	Dataflow Diagram of artificial intelligence crime: an overview of malicious use and abuse of AI	17
Figure 4.1	Dataset directory structure with folders'ARTIFICIAL INTELLIGENCE CRIME' having all the training examples	23
Figure 4.2	Screenshot of the dataset we collected according to the trained database	23
Figure 4.3	Running server and opening the website of AI crime	24

Figure 5.1	GUI/Main Interface of Artificial Intelligence Crime: An overview of Malicious Use And Abuse of AI	38
Figure 5.2	Loaded sample image of Artificial Intelligence Crime: An Overview Of Malicious Use And Abuse Of AI	39
Figure 5.3	Login page of Artificial Intelligence Crime: An Overview Of Malicious Use And Abuse Of AI	40
Figure 5.4	Artificial Intelligence Crime: An Overview Of Malicious Use And Abuse Of AI Prediction Page	41
Figure 5.5	Display of crime type of the given prediction details	42
Figure 5.6	Accuracy Comparison of Different Machine Learning Models for Artificial Intelligence Crime	43

Figure 5.7	Accuracy Comparison of Different Machine Learning Models for Artificial Intelligence Crime in line chart.	44
Figure 5.8	Crime Type Prediction Type Ratio Details of Artificial Intelligence Crime: An Overview Of Malicious Use And Abuse Of AI	45
Figure 5.9	Line Chart Comparison of Crime Prediction of Artificial Intelligence Crime	46
Figure 5.10	Crime Prediction Type Details Artificial Intelligence Crime: An Overview Of Malicious Use And Abuse Of AI	47



## LIST OF TABLES

TABLE NO	TABLE NAME	PAGE NO
Table 6.2.1	Uploading Dataset	49
Table 6.2.2	Classification	49

# TABLE OF CONTENTS

<b>ABSTRACT</b>	i
<b>LIST OF FIGURES</b>	ii
<b>LIST OF TABLES</b>	v
<b>1. INTRODUCTION</b>	1
1.1 PROJECT PURPOSE	1
1.2 PROJECT FEATURES	2
<b>2. LITERATURE SURVEY</b>	3
2.1 REVIEW OF RELATED WORK	6
2.2 DEFINITION OF PROBLEM STATEMENT	7
2.3 EXISTING SYSTEM	8
2.4 PROPOSED SYSTEM	9
2.5 OBJECTIVES	11
2.6 HARDWARE & SOFTWARE REQUIREMENTS	12
2.6.1 HARDWARE REQUIREMENTS	12
2.6.2 SOFTWARE REQUIREMENTS	12
<b>3. SYSTEM ARCHITECTURE &amp; DESIGN</b>	13
3.1 PROJECT ARCHITECTURE	13
3.2 DESCRIPTION	14
3.3 DATA FLOW DIAGRAM	16
<b>4. IMPLEMENTATION</b>	18
4.1 ALGORITHMS USED	18
4.2 SAMPLE CODE	25
<b>5. RESULTS &amp; DISCUSSION</b>	38
<b>6. VALIDATION</b>	48
6.1 INTRODUCTION	48
6.2 TEST CASES	49
6.2.1 UPLOADING DATASET	49
6.2.2 CLASSIFICATION	49
<b>7. CONCLUSION &amp; FUTURE ASPECTS</b>	50
7.1 PROJECT CONCLUSION	50
7.2 FUTURE ASPECTS	51
<b>8. BIBLIOGRAPHY</b>	52
8.1 REFERENCES	52
8.2 GITHUB LINK	53

# **1. INTRODUCTION**

## **1. INTRODUCTION**

The project, titled "Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI" explores how AI is exploited for criminal purposes, highlighting key areas such as cybercrime, deepfakes, privacy violations, and autonomous weapons. It examines technologies enabling AI-driven crimes and the ethical, legal, and regulatory challenges involved. The project includes case studies of AI crime, discusses AI's role in fraud and surveillance, and investigates future risks. It also proposes recommendations for policy, regulation, and detection methods to mitigate AI misuse. Ultimately, the project aims to inform stakeholders on creating ethical frameworks and safeguards against AI-driven criminal activities.

The impact of systems using Artificial Intelligence (AI) is at the center of numerous academic studies, political debates, and reports of civil society organizations. The development of AI has become the subject of praise due to unprecedented technological capabilities, such as enhanced possibilities for automated image recognition (e.g., detection of cancer in the field of medicine). However, it has also been criticized- even feared- due to aspects such as the uncertain consequences of automation for the labor market (e.g., concerns of mass unemployment). This duality of positive vs negative aspects of the technology can also be identified in the context of cyber security and cyber crime. Governments use AI to enhance their capabilities, whereas the same technology can be used for attacks against them.

### **1.1 PROJECT PURPOSE**

The purpose of this project is to investigate the malicious use and abuse of Artificial Intelligence (AI) in criminal activities. It aims to identify the key areas where AI is exploited, such as cybercrime, deepfakes, and surveillance, and assess the associated risks. The project seeks to understand the technological, ethical, and legal challenges posed by AI crime.

Additionally, it will explore current and potential solutions, offering recommendations for policies, regulations, and safeguards to prevent AI misuse. The goal is to raise awareness, inform decision-makers, and promote responsible AI development to protect society from AI-related threats.

## **1.2 PROJECT FEATURES**

The project features an in-depth analysis of AI-driven crimes, including cyber attacks, deepfakes, and surveillance misuse. It includes case studies to highlight real-world examples and lessons learned. The project explores the technologies enabling AI crime and examines the ethical and legal challenges of AI regulation. It provides risk assessments for future AI threats and recommends policies and safeguards. Additionally, the project offers solutions for detecting and preventing AI misuse. Ultimately, it aims to raise awareness and guide responsible AI development and governance.

## **2. LITERATURE SURVEY**

## 2. LITERATURE SURVEY

The malicious use and abuse of artificial intelligence (AI) have emerged as significant concerns in recent years. AI technologies, while offering great potential for innovation and efficiency, also present opportunities for exploitation in harmful ways. Cybercriminals and adversarial actors have begun using AI for various malicious purposes, including automating cyberattacks, such as phishing and malware deployment, and developing sophisticated deepfakes to spread misinformation. AI-driven surveillance tools can be misused for mass surveillance, infringing on privacy rights. In the realm of autonomous systems, AI's potential abuse in military applications, such as autonomous drones or lethal autonomous weapons, has raised ethical and security concerns.

Moreover, AI can be weaponized to manipulate social media platforms, amplifying disinformation campaigns, and exacerbating social and political unrest. The use of AI in fraudulent activities, like synthetic identity creation and financial fraud, has also surged, leveraging AI's ability to simulate real human behaviors convincingly. The increasing reliance on AI systems across sectors such as healthcare, finance, and critical infrastructure has heightened the risk of adversarial attacks, where AI systems are intentionally misled or sabotaged. Furthermore, the lack of regulatory frameworks and the rapid development of AI technologies present significant challenges in mitigating these risks, calling for more robust ethical guidelines, cybersecurity measures, and international cooperation to curb the misuse and abuse of AI.

Artificial Intelligence (AI) has rapidly advanced, impacting multiple sectors, including cybersecurity. While AI offers numerous benefits, it also introduces new vulnerabilities and threats. Recent literature highlights concerns regarding AI's dual-use nature—its potential for both beneficial and malicious applications. Researchers have categorized AI-related cybercrime into two primary areas: \*malicious use\* (AI-enhanced cyberattacks) and \*malicious abuse\* (exploitation of AI system vulnerabilities). This survey explores key literature discussing these risks.

- **Malicious Use of AI**

AI is increasingly used in cyberattacks, making them more sophisticated, scalable, and difficult to detect. The literature identifies several forms of malicious AI applications:

1. Social Engineering and Phishing Attacks

Studies indicate that AI-powered chatbots and deep learning models can enhance social engineering attacks. Seymour and Tully (2016) demonstrated how machine learning can automate \*spear-phishing campaigns\*, increasing their effectiveness by generating realistic phishing emails. Similarly, AI-generated social bots influence public opinion by spreading misinformation or impersonating real individuals.

2. Fake News and Misinformation

AI-generated text and deepfake technology have been widely discussed in misinformation studies. GPT-3 and similar language models can be used to create \*fake news articles\* that appear authentic, misleading the public. The Cambridge Analytical scandal exemplifies how AI-driven \*psychographic profiling\* can manipulate voter behavior. Researchers suggest that digital literacy and algorithmic transparency are necessary countermeasures (Keller et al., 2020).

3. Autonomous Weapon Systems (AWS)

The military application of AI is a significant concern in international security. Autonomous drones and AI-powered defense systems are being developed, but \*hacked AI-controlled weapons\* pose a major threat. Reports suggest that adversaries could \*manipulate AI decision-making\* through adversarial attacks, leading to unintended casualties.

- **Malicious Abuse of AI**

The abuse of AI models refers to exploiting weaknesses in AI systems. Researchers categorize these abuses into the following areas:

1. Integrity Attacks on AI Models

Integrity attacks manipulate AI models by altering their training data. Poisoning attacks introduce adversarial examples, causing misclassification errors in AI systems. For instance, in data poisoning attacks, hackers inject malicious data into a training dataset, leading to



biased or incorrect predictions (Jagielski et al., 2018). The Microsoft Tay chatbot incident in 2016 is a well-known example, where adversarial users trained the bot to generate offensive content.

## 2. Membership Inference and Privacy Attacks

AI models trained on sensitive data are vulnerable to membership inference attacks, where attackers attempt to reconstruct the original dataset. This can expose private user information, especially in medical AI systems. Research by Hu et al. (2021) suggests that differential privacy techniques can mitigate such risks.

## 3. Algorithmic Trading and Financial Manipulation

High-frequency trading (HFT) and algorithmic stock trading rely heavily on AI. However, AI-driven financial manipulation can lead to market crashes, as seen in the 2010 Flash Crash, where AI trading bots created market instability. AI-powered fraudulent trading algorithms can execute spoofing and layering attacks, making regulation increasingly difficult.

In the ever-evolving landscape of artificial intelligence (AI) research, a series of seminal papers have emerged, each shedding light on the potential risks and challenges associated with the malicious use of AI. One such landmark paper, titled "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," authored by Brundage et al., stands out as a comprehensive survey of the security threats posed by AI and proposes strategies to forecast, prevent, and mitigate these threats. Published in [Year], this paper delves into the ways in which AI may impact various security domains, including digital security, physical security, and political security, emphasizing the urgent need for further research and collaboration to address emerging challenges.

The increasing integration of AI into cybercrime presents significant challenges. Literature in this field emphasizes the need for robust AI security measures and cross-sector collaboration among governments, industries, and researchers. Future research should focus on developing resilient AI models, improving cyber forensic tools, and enhancing legal frameworks to address the evolving AI threat landscape.

## **2.1 REVIEW OF RELATED WORK**

### **1. AI in Cybercrime Detection**

Artificial intelligence has been extensively researched in the field of cybercrime detection. Machine learning algorithms have been applied to identify fraudulent activities, cyberattacks, and malicious network behavior. Studies have shown that AI-powered systems can detect anomalies in large datasets with high accuracy. Researchers have explored deep learning models for intrusion detection and cybersecurity threat analysis. However, the challenge lies in ensuring real-time threat detection without high false-positive rates.

### **2. Deepfake and AI-Generated Misinformation**

Deepfake technology has raised significant concerns regarding digital misinformation and identity fraud. Various studies highlight the potential misuse of deepfakes in political propaganda, financial scams, and social engineering attacks. Researchers have developed AI-based detection tools to differentiate between real and AI-generated media. Despite advancements, deepfake detection remains challenging due to the continuous evolution of AI-based forgery techniques. Future research focuses on enhancing detection accuracy and reducing computational costs.

### **3. AI in Law Enforcement and Criminal Investigation**

AI has played a transformative role in modern law enforcement, aiding in criminal investigations and predictive policing. Facial recognition systems and AI-powered surveillance have helped law enforcement agencies identify suspects. Machine learning models analyze crime patterns to predict potential criminal activities in high-risk areas. However, concerns about bias in AI models and ethical implications remain a major challenge. Researchers continue to work on improving AI fairness and ensuring transparency in AI-based policing.

### **4. Ethical and Legal Challenges of AI in Crime**

The rapid development of AI in crime-related applications has raised ethical and legal

concerns. Scholars argue that AI-enabled surveillance and predictive policing may violate privacy rights. Legal frameworks are being proposed to regulate AI usage in sensitive applications and prevent misuse. The balance between security and individual freedoms remains a key area of research. Future studies emphasize the need for global AI governance to ensure responsible AI deployment.

## **5. AI-Powered Cyber Attacks and Defense Mechanisms**

AI is not only used to prevent crime but is also exploited for cyber-attacks. Research has identified AI-driven malware, automated hacking tools, and AI-generated phishing emails as emerging threats. Defensive AI systems have been developed to counteract these threats by using adaptive learning techniques. However, the continuous evolution of AI-based attacks poses a challenge for cybersecurity experts. Ongoing research focuses on creating AI models that can predict and mitigate cyber threats before they cause harm.

## **2.2 DEFINITION OF PROBLEM STATEMENT**

Artificial Intelligence (AI) has revolutionized various industries, but its misuse in criminal activities poses significant risks. The malicious use and abuse of AI in cybercrime, misinformation, surveillance, fraud, and autonomous weapon systems have raised ethical, legal, and security concerns. This research aims to provide an overview of AI-driven crimes, analyze their implications, and explore countermeasures to mitigate their impact. By examining real-world cases and potential threats, this study highlights the urgent need for robust AI regulations, ethical frameworks, and technological safeguards to prevent AI from being exploited for malicious purposes.

## 2.3 EXISTING SYSTEM

To build on previous work and expand the understanding of how AI broadens the potential for malicious activities online, this article evaluates the main categories of use and abuse of AI in a criminal context. We provide several salient examples that allow us to illustrate the challenges at hand.

Based on these examples, we present a typology that catalogs the main harmful AI-based activities. Developing knowledge and understanding about the potential malicious use and abuse of AI enables cybersecurity organizations and governmental agencies to anticipate such incidents and increase their preparedness against attacks. Furthermore, a typology is greatly useful in structuring research efforts and identifying gaps in knowledge in areas where more research is warranted.

### Limitations of Existing System

- An existing methodology not proposed the term "AI-Crime" to describe the situation in which AI technologies are re-oriented to facilitate criminal activity.
- An existing system doesn't implement for MALICIOUS ABUSE OF AI and VULNERABILITIES OF AI MODELS.
- Many AI-driven cybercrimes are under reported or difficult to trace, leading to an incomplete dataset for analysis. The absence of real-time monitoring tools for AI-based attacks makes it challenging to study emerging threats.
- AI-generated cyber threats, such as deepfakes, automated phishing, and adversarial attacks, are becoming more sophisticated and harder to detect. Traditional cybersecurity measures struggle to keep up with AI's evolving capabilities in deception and misinformation.
- There is no universally accepted legal framework to regulate AI crimes, making enforcement inconsistent across jurisdictions.
- AI's ability to automate cybercrime raises ethical concerns about liability—should the responsibility lie with the user, the developer, or the AI system itself?

## 2.4 PROPOSED SYSTEM

With the typology presented in this paper, we hope to make the following contributions:

- Add to the emerging body of knowledge that maps types of malicious use and abuse of AI systems. To understand the main concepts, threat scenarios, and possibilities is necessary to develop much-needed preventive measures and proactive responses to such attacks.
- Help in establishing a shared language among and across different disciplines, especially between STEM disciplines and legal practitioners, as well as policymakers. Interdisciplinary research on the topic can reduce confusion caused by excessively technical or mono disciplinary language and aid in bridging existing gaps.
- Propose mitigation strategies, as well as demonstrating that a collective effort among government, academia, and industry is needed.

The methodology is based on an analysis of the available literature on cybercrime and the potential malicious use and abuse of AI systems. A literature review informs this study and findings using the following databases: IEEE Xplore, Science Direct, Wiley Online Library, and Google Scholar. We used keywords, titles, and screened abstracts. The search terms included are (Artificial Intelligence OR AI OR Machine Learning OR ML) AND (malicious OR crime OR harmful OR cyber attack). Additionally, we examined lists of references obtained from reviewed papers and reports, as well as news sources describing past AI incidents. We only reviewed papers/reports/web pages available in English and Portuguese. After analyzing these sources, we were able to identify the different types of malicious use and abuse of AI systems.

Machine learning (ML) has become more prevalent in recent years. This has created incentives for attackers to manipulate models (e.g., the software itself) or the underlying data, making ML models prone to integrity attacks. In integrity attacks, hackers attempt to inject false information into a system to corrupt the data, undermining their trustworthiness.

## **Advantages of the Proposed System:**

The proposed system significantly improves upon the existing approaches by addressing key limitations:

- The system aims to propose a typology of the malicious use and abuse of AI based on empirical evidence and contemporary discourse, analyzing how AI systems are used to compromise confidentiality, integrity, and data availability.
- Objectives are limited to identifying essential elements of the malicious use and abuse of AI, and to collect evidence of their use in practice. The compiled data enable further analysis of the possible ways in which AI systems can be exploited for criminal activities.
- Encourages preventive strategies rather than reactive responses, minimizing damage before attacks occur. Helps security experts design robust AI models resistant to integrity attacks and data manipulation.
- Assists policymakers in drafting effective AI regulations based on well-defined typologies of malicious use. Provides guidance for ethical AI development by addressing security and misuse concerns.
- Contributes to better security protocols for AI-based systems, reducing risks of adversarial attacks. Encourages the development of AI-driven defense mechanisms against cyber threats.
- Encourages organizations to adopt ethical AI frameworks that consider potential risks of misuse. Assists industry stakeholders in implementing AI safety measures to protect users and systems.
- Provides empirical insights based on a literature review from reputable sources (IEEE, Science Direct, Google Scholar, etc.).
- Helps in predicting emerging AI threats by analyzing past incidents and attack patterns.

## 2.5 OBJECTIVES

- Understanding AI in Crime – Explore how artificial intelligence is used in criminal activities, including cybercrimes, fraud, and automated attacks.
- Identifying Malicious Uses of AI – Analyze different ways AI is exploited for illegal activities, such as deepfakes, AI-driven hacking, and autonomous weapons.
- Examining AI-Enabled Cybercrime – Investigate AI's role in phishing attacks, automated scams, and data breaches to understand its impact on cybersecurity.
- Assessing AI-Generated Misinformation – Study how AI is used to create and spread misinformation, fake news, and deceptive content for political or financial gain.
- Exploring Ethical and Legal Challenges – Discuss the ethical concerns and legal issues surrounding the misuse of AI in criminal activities.
- Impact on Society and Security – Analyze the broader societal consequences of AI crimes, including trust issues, economic disruptions, and threats to privacy.

## 2.6 HARDWARE & SOFTWARE REQUIREMENTS

### 2.6.1 HARDWARE REQUIREMENTS:

Hardware interfaces specifies the logical characteristics of each interface between the software product and the hardware components of the system. The following are some hardware requirements:

- Processor : Intel Core i5
- Hard disk : 1TB.
- RAM : 8GB.

### 2.6.2 SOFTWARE REQUIREMENTS:

Software Requirements specifies the logical characteristics of each interface and software components of the system. The following are some software requirements:

- 2.6.2.1 Operating system : Windows 10
- 2.6.2.2 Language : Python(3.7.0)
- 2.6.2.3 Back-End : Django-ORM
- 2.6.2.4 Designing : HTML,CSS,Java script
- 2.6.2.5 Framework : tinker



# **3. SYSTEM ARCHITECTURE & DESIGN**

## 2.SYSTEM ARCHITECTURE & DESIGN

Project architecture refers to the structural framework and design of a project, encompassing its components, interactions, and overall organization. It provides a clear blueprint for development, ensuring efficiency, scalability, and alignment with project goals. Effective architecture guides the project's life cycle, from planning to execution, enhancing collaboration and reducing complexity.

### 3.1 PROJECT ARCHITECTURE

The diagram provided illustrates a high-level architecture for a crime data prediction system. Here's a summary of its components:

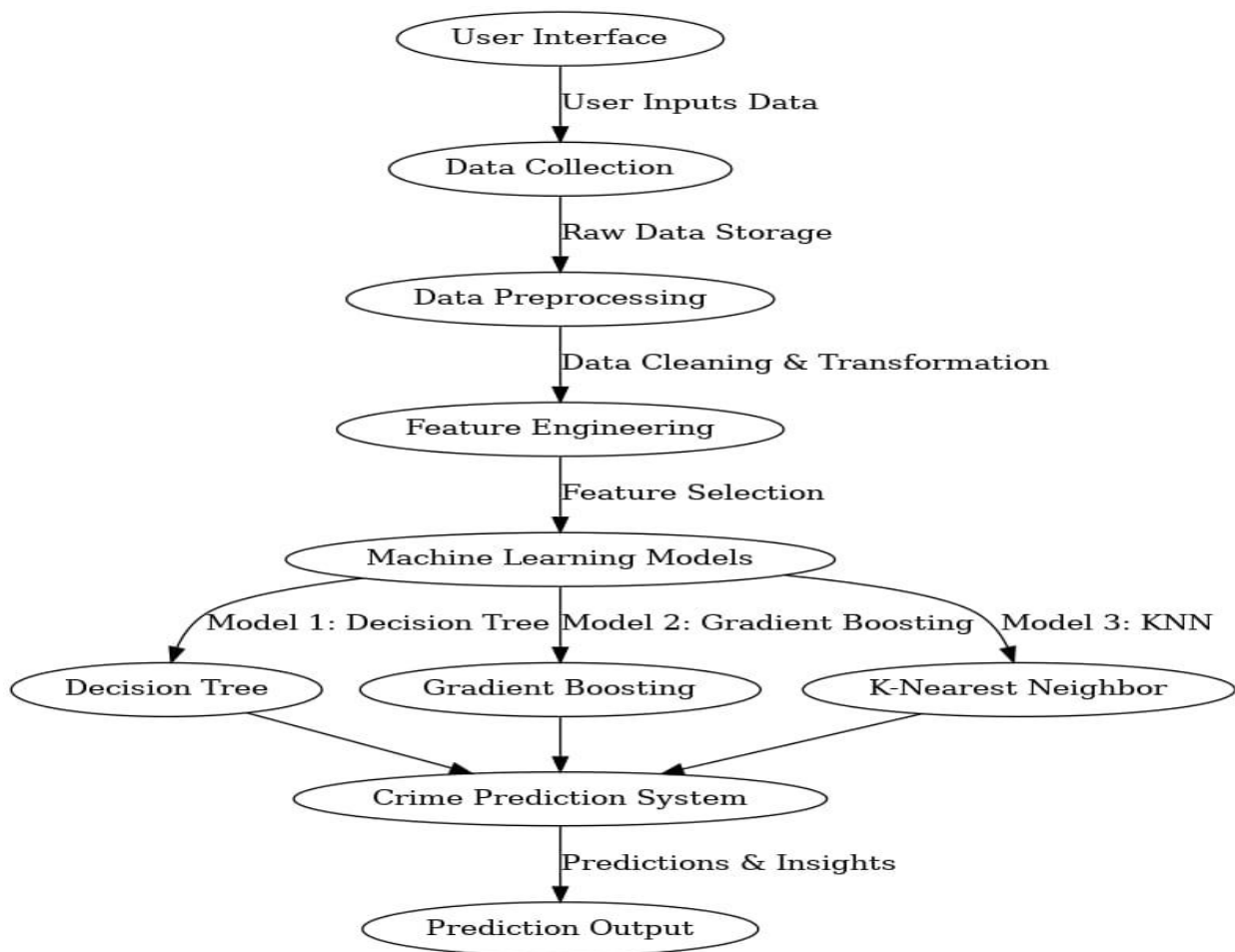


Figure 3.1: Project Architecture of artificial intelligence crime: an overview of malicious use and abuse of AI

## 3.2 DESCRIPTION

**Input Data :** The project collects crime-related data from various sources, including law enforcement records, social media reports, and surveillance systems. The dataset includes historical crime records, incident reports, and geospatial data to analyze crime trends.

**Reading Data:** Data is preprocessed by extracting relevant information using automated AI-driven tools. Techniques like web scraping, database querying, and API integrations are used to gather structured and unstructured data

**Feature Extraction :** Advanced machine learning algorithms extract key features from the collected data, such as time, location, crime type, suspect details, and environmental factors. Feature selection ensures only the most relevant variables are used.

**Temporal Pattern Learning :** AI models analyze time-series crime data to identify patterns, such as seasonal crime trends, high-risk zones, and recurring offenders. Techniques like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are employed for sequential pattern recognition.

**Attention Mechanism :** AI-powered anomaly detection systems identify suspicious behaviors, fraudulent activities, and cyber threats. Models like Gradient Boosting, Decision Trees, and K-Nearest Neighbor (KNN) predict potential crime occurrences based on historical data.

**Classification Layer :** A multi-class classification model categorizes different types of crimes, distinguishing between cybercrime, financial fraud, physical crime, and AI-driven threats. Fully connected neural networks process extracted crime patterns for accurate classification.

**Training and Evaluation :** The AI system is trained on labeled datasets to enhance its predictive capabilities. Performance is evaluated using accuracy, precision, recall, and F1-score metrics.

The model continuously improves by learning from real-time data streams and past predictions.

**Feedback :** Law enforcement agencies, policymakers, and cybersecurity experts provide feedback on crime predictions. This human-in-the-loop approach refines the AI model, ensuring better accuracy in future crime detection and mitigation strategies.

### 3.3 DATA FLOW DIAGRAM

A Data Flow Diagram (DFD) is a graphical representation that illustrates how data flows within a system, showcasing its processes, data stores, and external entities. It is a vital tool in system analysis and design, helping stakeholders visualize the movement of information, identify inefficiencies, and optimize workflows.

A Data Flow Diagram comprises Four primary elements:

- External Entities: Represent sources or destinations of data outside the system.
- Processes: Indicate transformations or operations performed on data.
- Data Flows: Depict the movement of data between components.
- Data Stores: Represent where data is stored within the system.

These components are represented using standardized symbols, such as circles for processes, arrows for data flows, rectangles for external entities, and open-ended rectangles for data stores.

#### **Benefits:**

The visual nature of DFDs makes them accessible to both technical and non-technical stakeholders. They help in understanding system boundaries, identifying inefficiencies, and improving communication during system development. Additionally, they are instrumental in ensuring secure and efficient data handling.

#### **Applications:**

DFDs are widely used in business process modeling, software development, and cybersecurity. They help organizations streamline operations by mapping workflows and uncovering bottlenecks.

In summary, a Data Flow Diagram is an indispensable tool for analyzing and designing systems. Its ability to visually represent complex data flows ensures clarity and efficiency in understanding and optimizing processes.

### Levels of DFD:

DFDs are structured hierarchically:

- Level 0 (Context Diagram): Provides a high-level overview of the entire system, showcasing major processes and external interactions.
- Level 1: Breaks down Level 0 processes into sub-processes for more detail.
- Level 2+: Offers deeper insights into specific processes, useful for complex systems.

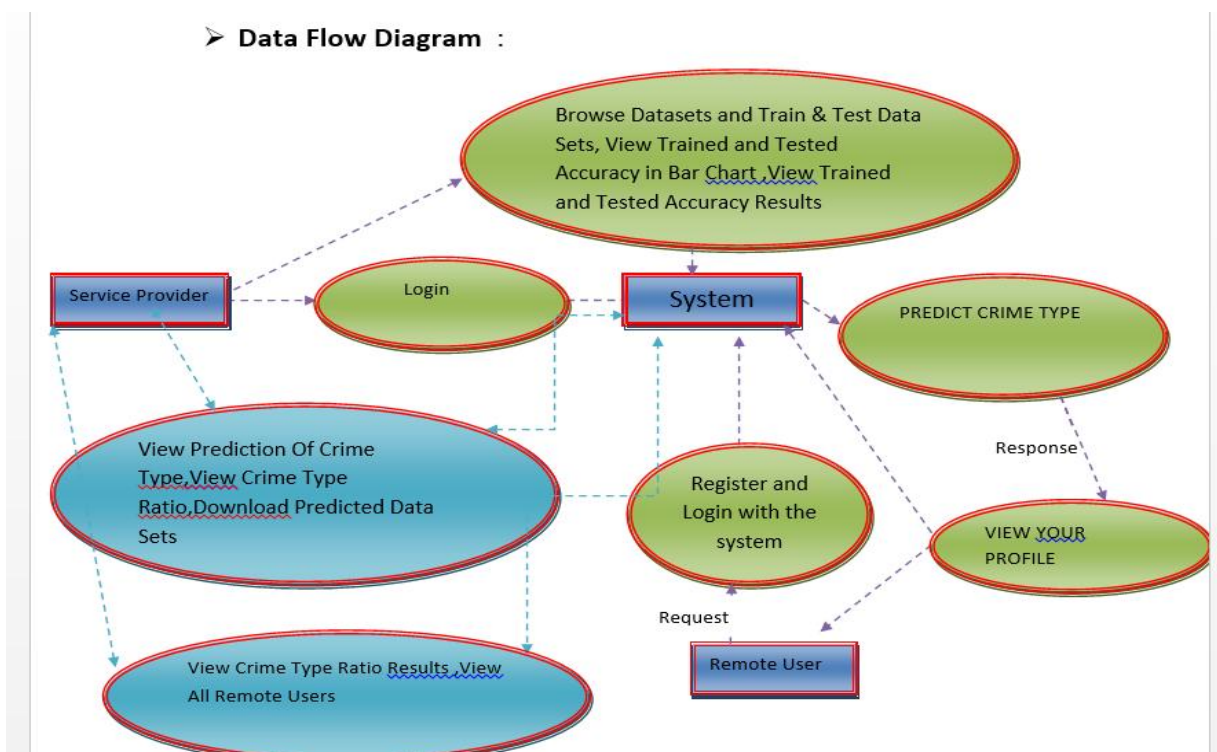


Figure 3.2: Dataflow Diagram of Artificial Intelligence Crime: An Overview Of Malicious Use And Abuse Of AI

## **4. IMPLEMENTATION**

## 4.IMPLEMENTATION

The implementation phase of a project involves executing the planned strategies and tasks. It requires meticulous coordination, resource allocation, and monitoring to ensure that objectives are met efficiently. Effective implementation is crucial for achieving project goals and delivering expected outcomes within the set timeline and budget constraints.

### 4.1 ALGORITHMS USED

#### Decision tree classifiers

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects ( $S$ ), each belonging to one of the classes  $C_1, C_2, \dots, C_k$  is as follows:

**Step 1.** If all the objects in  $S$  belong to the same class, for example  $C_i$ , the decision tree for  $S$  consists of a leaf labeled with this class.

**Step 2.** Otherwise, let  $T$  be some test with possible outcomes  $O_1, O_2, \dots, O_n$ . Each object in  $S$  has one outcome for  $T$  so the test partitions  $S$  into subsets  $S_1, S_2, \dots, S_n$  where each object in  $S_i$  has outcome  $O_i$  for  $T$ .  $T$  becomes the root of the decision tree and for each outcome  $O_i$  we build a subsidiary decision tree by invoking the same procedure recursively on the set  $S_i$ .

#### Advantages of Decision tree classifiers:

- Decision trees are simple and intuitive, making them easy to understand even for non-experts.
- The tree structure allows for clear visualization of decision-making. Unlike many machine learning models, decision trees do not require data to be normalized or scaled. They handle missing values effectively by assigning them to the most probable class.



## Gradient boosting

**Gradient boosting** is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees. When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest. A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

## K-Nearest Neighbors (KNN)

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not “learn” until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

### Example

- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset.

## Logistic regression Classifiers

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name logistic regression is used when the dependent variable has only two values, such as 0 and 1 or Yes and No.

The name multinomial logistic regression is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does.

This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

## **Naïve Bayes**

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature.

Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias).

While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique.

Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier to understand, and its deployment is also made easier. In the first part of this tutorial, we present some theoretical aspects of the naive bayes classifier. Then, we implement the approach on a dataset with Tanagra. We compare the obtained results (the parameters of the model) to those obtained with other linear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. We note that the results are highly consistent. This largely explains the good performance of the method in comparison to others. In the second part, we use various tools on the same dataset (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b and RapidMiner 4.6.0). We try above all to understand the obtained results.

## **Random Forest**

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg.

An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The

extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision

trees with controlled variance.

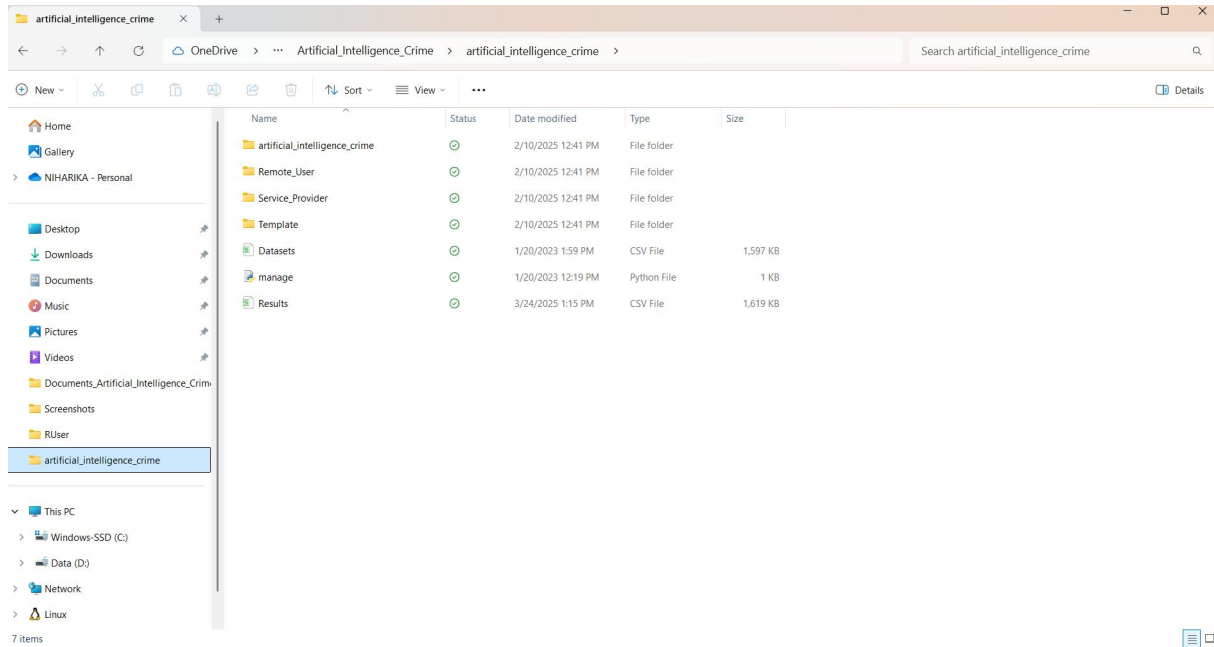
Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

## SVM

In classification tasks a discriminant machine learning technique aims at finding, based on an independent and identically distributed (iid) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point  $x$  and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space.

SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to genetic algorithms (GAs) or perceptrons, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

# ARTIFICIAL INTELLIGENCE CRIME: AN OVERVIEW OF MALICIOUS USE AND ABUSE OF AI

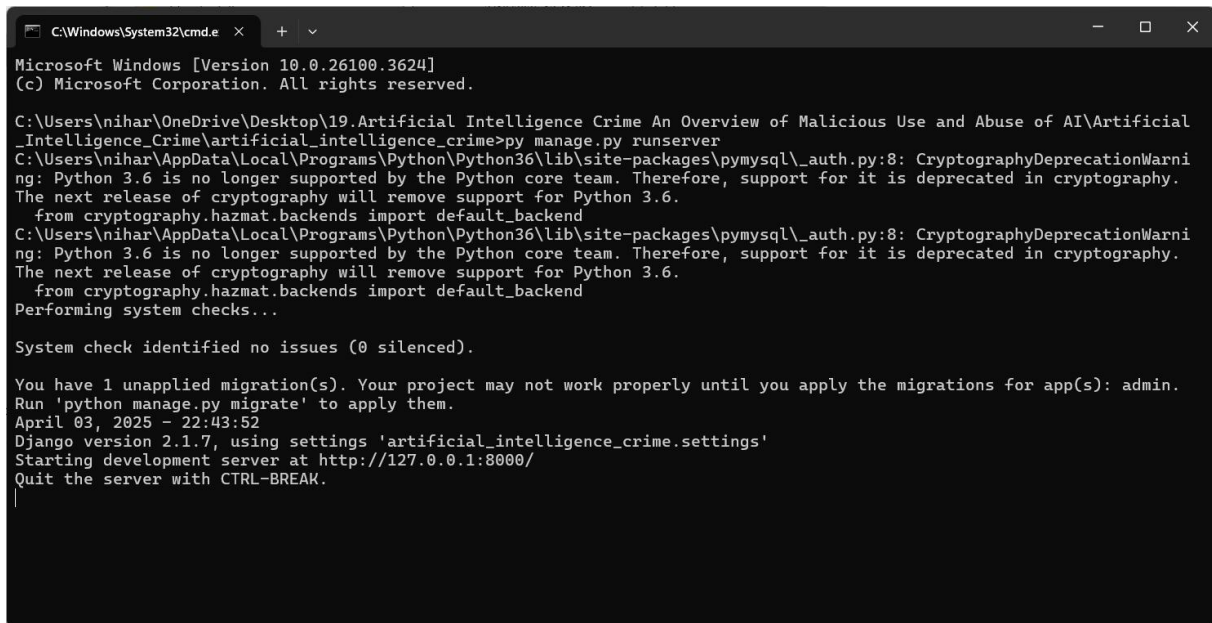


**Figure 4.1:** Dataset directory structure with folders 'ARTIFICIAL INTELLIGENCE CRIME' having all the training examples.

The screenshot shows a Microsoft Excel spreadsheet with the following columns: FID, url, length, host, Source IP, Source Port, Destination, Destination Label, and others. The data is organized into rows, with the first row being the header. The spreadsheet shows a list of network traffic data, including source and destination IP addresses, ports, and labels.

FID	url	length	host	Source IP	Source Port	Destination	Destination Label
1	172.217.10.11	37	19	10.42.0.211	34451	52.6.25.230	443
2	10.42.0.211	77	23	10.42.0.151	53892	172.217.3.9	443
3	10.42.0.211	126	50	172.217.3.9	443	10.42.0.151	50750
4	172.217.11.11	18	11	10.42.0.211	23025	10.42.0.1	53
5	184.50.111.11	55	15	10.42.0.211	52602	123.129.24.4	443
6	10.42.0.151	32	24	10.42.0.151	57625	173.194.20.1	443
7	10.42.0.211	19	12	172.217.7.1	443	10.42.0.211	37893
8	172.217.9.2	81	27	10.42.0.211	44342	172.217.12.1	443
9	10.42.0.211	42	34	10.42.0.211	47485	47.89.68.22	443
10	172.217.11.11	104	10	10.42.0.42	54061	52.179.189.1	443
11	216.58.219.11	56	22	10.42.0.42	48094	216.58.219.1	443
12	10.42.0.211	43	16	10.42.0.211	60586	180.76.182.1	80
13	151.101.1.11	83	14	172.217.12.1	443	10.42.0.151	41454
14	172.217.3.1	31	10	10.42.0.151	56907	172.217.10.1	443
15	10.42.0.42	25	16	10.42.0.211	41106	172.217.3.1	443
16	10.42.0.211	51	23	10.42.0.211	10052	10.42.0.1	53
17	10.42.0.151	31	22	52.0.93.246	443	10.42.0.211	45776
18	172.217.12.11	50	21	10.42.0.151	59447	121.29.54.1	80
19	202.217.10.11	34	23	10.42.0.211	47447	172.217.6.2	443
20	184.29.173.11	27	19	10.42.0.151	49883	172.217.12.1	443
21	172.217.10.11	63	14	10.42.0.151	34789	173.194.17.1	5228
22	172.217.10.11	66	16	10.42.0.1	44477	239.255.25.1	1900
23	180.149.131.11	40	11	10.42.0.151	54085	104.254.66.1	80
24	10.42.0.211	48	16	10.42.0.151	34088	172.217.12.1	443
25	172.217.12.11	126	50	10.42.0.42	59370	54.192.38.1	443
26	10.42.0.211	41	32	10.42.0.211	57315	180.149.131.1	80
27	10.42.0.151	36	27	10.42.0.42	52456	66.198.178.1	443
28	10.42.0.211	72	24	10.42.0.151	40150	10.42.0.1	53

**Figure 4.2:** Screenshot of the dataset we collected according to the trained database.



```
C:\Windows\System32\cmd.e  +  v
Microsoft Windows [Version 10.0.26100.3624]
(c) Microsoft Corporation. All rights reserved.

C:\Users\nihar\OneDrive\Desktop\19.Artificial Intelligence Crime An Overview of Malicious Use and Abuse of AI\Artificial
_Intelligence_Crime\artificial_intelligence_crime>py manage.py runserver
C:\Users\nihar\AppData\Local\Programs\Python\Python36\lib\site-packages\pymysql\auth.py:8: CryptographyDeprecationWarni
ng: Python 3.6 is no longer supported by the Python core team. Therefore, support for it is deprecated in cryptography.
The next release of cryptography will remove support for Python 3.6.
  from cryptography.hazmat.backends import default_backend
C:\Users\nihar\AppData\Local\Programs\Python\Python36\lib\site-packages\pymysql\auth.py:8: CryptographyDeprecationWarni
ng: Python 3.6 is no longer supported by the Python core team. Therefore, support for it is deprecated in cryptography.
The next release of cryptography will remove support for Python 3.6.
  from cryptography.hazmat.backends import default_backend
Performing system checks...

System check identified no issues (0 silenced).

You have 1 unapplied migration(s). Your project may not work properly until you apply the migrations for app(s): admin.
Run 'python manage.py migrate' to apply them.
April 03, 2025 - 22:43:52
Django version 2.1.7, using settings 'artificial_intelligence_crime.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
```

**Figure 4.3:** Running server and opening the website of AI crime.

To implement this project we have designed following modules:

### **Service Provider**

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login,Browse Datasets and Train & Test Data Sets ,View Trained and Tested Accuracy in Bar Chart ,View Trained and Tested Accuracy Results ,View,Prediction Of Crime Type ,View Crime Type Ratio ,Download Predicted Data Sets ,View Crime Type Ratio Results ,View All Remote Users

### **View and Authorize Users**

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

### **Remote User**

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT CRIME TYPE, VIEW YOUR PROFILE.

## 4.2 SAMPLE CODE

### Remote User:

```
from django.db.models import Count
from django.db.models import Q
from django.shortcuts import render, redirect, get_object_or_404

import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from sklearn.metrics import accuracy_score
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import VotingClassifier
# Create your views here.
from Remote_User.models import
ClientRegister_Model ,prediction_Of_crime_type,detection_ratio,detection_accuracyClientRe
gister_Model,prediction_Of_crime_type,detection_ratio,detection_accuracy

def login(request):

    if request.method == "POST" and 'submit1' in request.POST:
        username = request.POST.get('username')
        password = request.POST.get('password')
        try:
            enter = ClientRegister_Model.objects.get(username=username,password=password)
            request.session["userid"] = enter.id
            return redirect('ViewYourProfile')
        except:
            pass
    return render(request,'RUser/login.html')

def index(request):
    return render(request, 'RUser/index.html')
```

```
def Add_DataSet_Details(request):
    return render(request, 'RUser/Add_DataSet_Details.html', {"excel_data": ""})

def Register1(request):
    if request.method == "POST":
        username = request.POST.get('username')
        email = request.POST.get('email')
        password = request.POST.get('password')
        phoneno = request.POST.get('phoneno')
        country = request.POST.get('country')
        state = request.POST.get('state')
        city = request.POST.get('city')
        address = request.POST.get('address')
        gender = request.POST.get('gender')

        ClientRegister_Model.objects.create(username=username,email=email,password=password,
        phoneno=phoneno,country=country, state=state, city=city,address=address,gender=gender)

        obj = "Registered Successfully"
        return render(request, 'RUser/Register1.html',{'object':obj})
    else:
        return render(request,'RUser/Register1.html')

def ViewYourProfile(request):
    userid = request.session['userid']
    obj = ClientRegister_Model.objects.get(id= userid)
    return render(request,'RUser/ViewYourProfile.html',{'object':obj})

def Predict_Crime_Type(request):
    if request.method == "POST":
        if request.method == "POST":
            FID=request.POST.get('FID')
            url=request.POST.get('url')
```



```
length_url=request.POST.get('length_url')
length_hostname=request.POST.get('length_hostname')
Source_IP=request.POST.get('Source_IP')
Source_Port=request.POST.get('Source_Port')
Destination_IP=request.POST.get('Destination_IP')
Destination_Port=request.POST.get('Destination_Port')

df = pd.read_csv('Datasets.csv')
def apply_response(Label):
    if (Label == 0):
        return 0
    elif (Label == 1):
        return 1
    elif (Label == 2):
        return 2
    elif (Label == 3):
        return 3
df['results'] = df['Label'].apply(apply_response)
cv = CountVectorizer()
X = df['url']
y = df['results']
print("Url")
print(X)
print("Results")
print(y)
cv = CountVectorizer()
X = cv.fit_transform(X)
models = []
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20)
X_train.shape, X_test.shape, y_train.shape
print("Naive Bayes")
```

```
from sklearn.naive_bayes import MultinomialNB

NB = MultinomialNB()
NB.fit(X_train, y_train)
predict_nb = NB.predict(X_test)
naivebayes = accuracy_score(y_test, predict_nb) * 100
print("ACCURACY")

print(naivebayes)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, predict_nb))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, predict_nb))
models.append(('naive_bayes', NB))

# SVM Model
print("SVM")
from sklearn import svm

lin_clf = svm.LinearSVC()
lin_clf.fit(X_train, y_train)
predict_svm = lin_clf.predict(X_test)
svm_acc = accuracy_score(y_test, predict_svm) * 100
print("ACCURACY")
print(svm_acc)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, predict_svm))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, predict_svm))
models.append(('svm', lin_clf))

print("Logistic Regression")
from sklearn.linear_model import LogisticRegression
```

```
reg = LogisticRegression(random_state=0, solver='lbfgs').fit(X_train, y_train)
y_pred = reg.predict(X_test)
print("ACCURACY")
print(accuracy_score(y_test, y_pred) * 100)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, y_pred))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, y_pred))
models.append(('logistic', reg))

print("Decision Tree Classifier")
dtc = DecisionTreeClassifier()
dtc.fit(X_train, y_train)
dtcpredict = dtc.predict(X_test)
print("ACCURACY")
print(accuracy_score(y_test, dtcpredict) * 100)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, dtcpredict))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, dtcpredict))
models.append(('DecisionTreeClassifier', dtc))
classifier = VotingClassifier(models)
classifier.fit(X_train, y_train)
y_pred = classifier.predict(X_test)
url1 = [url]
vector1 = cv.transform(url1).toarray()
predict_text = classifier.predict(vector1)
pred = str(predict_text).replace("[", "")
pred1 = pred.replace("]", "")
prediction = int(pred1)
if (prediction == 0):
    val = 'Social Engineering'
```

```
elif (prediction == 1):
    val = 'Misinformation'
elif (prediction == 2):
    val = 'Hacking'
elif (prediction == 3):
    val = 'Autonomous weapon systems'
print(val)
print(pred1)
prediction_Of_crime_type.objects.create(
    FID=FID,
    url=url,
    length_url=length_url,
    length_hostname=length_hostname,
    Source_IP=Source_IP,
    Source_Port=Source_Port,
    Destination_IP=Destination_IP,
    Destination_Port=Destination_Port,
    Prediction=val)

return render(request, 'RUser/Predict_Crime_Type.html',{'objs': val})
return render(request, 'RUser/Predict_Crime_Type.html')
```

### **Service User:**

```
from django.db.models import Count, Avg
from django.shortcuts import render, redirect
from django.db.models import Count
from django.db.models import Q
import datetime
import xlwt
from django.http import HttpResponse

import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
```

```
from sklearn.metrics import accuracy_score
from sklearn.tree import DecisionTreeClassifier

# Create your views here.
From Remote_User.models import
ClientRegister_Model,prediction_Of_crime_type,detection_ratio,
detection_accuracy

def serviceproviderlogin(request):
    if request.method == "POST":
        admin = request.POST.get('username')
        password = request.POST.get('password')
        if admin == "Admin" and password == "Admin":
            detection_accuracy.objects.all().delete()
            return redirect('View_Remote_Users')
        return render(request,'SProvider/serviceproviderlogin.html')

def View_Crime_Type_Ratio(request):
    detection_ratio.objects.all().delete()
    ratio = ""
    kword = 'Social Engineering'
    print(kword)
    obj = prediction_Of_crime_type.objects.all().filter(Q(Prediction=kword))
    obj1 = prediction_Of_crime_type.objects.all()
    count = obj.count();
    count1 = obj1.count();
    ratio = (count / count1) * 100
    if ratio != 0:
        detection_ratio.objects.create(names=kword, ratio=ratio)
    ratio1 = ""

    kword1 = 'Misinformation'
```

```
print(keyword1)
obj1 = prediction_Of_crime_type.objects.all().filter(Q(Prediction=keyword1))
obj11 = prediction_Of_crime_type.objects.all()
count1 = obj1.count();
count11 = obj11.count();
ratio1 = (count1 / count11) * 100
if ratio1 != 0:
    detection_ratio.objects.create(names=keyword1, ratio=ratio1)
ratio12 = ""
keyword12 = 'Hacking'
print(keyword12)
obj12 = prediction_Of_crime_type.objects.all().filter(Q(Prediction=keyword12))
obj112 = prediction_Of_crime_type.objects.all()
count12 = obj12.count();
count112 = obj112.count();
ratio12 = (count12 / count112) * 100
if ratio12 != 0:
    detection_ratio.objects.create(names=keyword12, ratio=ratio12)
ratio123 = ""
keyword123 = 'Autonomous weapon systems'
print(keyword123)
obj123 = prediction_Of_crime_type.objects.all().filter(Q(Prediction=keyword123))
obj1123 = prediction_Of_crime_type.objects.all()
count123 = obj123.count();
count1123 = obj1123.count();
ratio123 = (count123 / count1123) * 100
if ratio123 != 0:
    detection_ratio.objects.create(names=keyword123, ratio=ratio123)
obj = detection_ratio.objects.all()
return render(request, 'SProvider/View_Crime_Type_Ratio.html', {'objs': obj})
def View_Remote_Users(request):
```

```
obj=ClientRegister_Model.objects.all()
    return render(request,'SProvider/View_Remote_Users.html',{'objects':obj})

def charts(request,chart_type):
    chart1 = detection_ratio.objects.values('names').annotate(dcount=Avg('ratio'))
    return render(request,"SProvider/charts.html", {'form':chart1, 'chart_type':chart_type})

def charts1(request,chart_type):

    chart1 = detection_accuracy.objects.values('names').annotate(dcount=Avg('ratio'))
    return render(request,"SProvider/charts1.html", {'form':chart1, 'chart_type':chart_type})

def View_Prediction_Of_Crime_Type(request):
    obj =prediction_Of_crime_type.objects.all()
    return render(request, 'SProvider/View_Prediction_Of_Crime_Type.html', {'list_objects':
obj})

def likeschart(request,like_chart):
    charts =detection_accuracy.objects.values('names').annotate(dcount=Avg('ratio'))
    return render(request,"SProvider/likeschart.html", {'form':charts, 'like_chart':like_chart})

def Download_Predicted_DataSets(request):
    response = HttpResponse(content_type='application/ms-excel')
    # decide file name
    response['Content-Disposition'] = 'attachment; filename="Predicted_Datasets.xls"'
    # creating workbook
    wb = xlwt.Workbook(encoding='utf-8')
    # adding sheet
    ws = wb.add_sheet("sheet1")
    # Sheet header, first row
    row_num = 0
    font_style = xlwt.XFStyle()
    # headers are bold
```

```
font_style.font.bold = True
# writer = csv.writer(response)
obj = prediction_Of_crime_type.objects.all()
data = obj # dummy method to fetch data.
for my_row in data:
    row_num = row_num + 1
    ws.write(row_num, 0, my_row.FID, font_style)
    ws.write(row_num, 1, my_row.url, font_style)
    ws.write(row_num, 2, my_row.length_url, font_style)
    ws.write(row_num, 3, my_row.length_hostname, font_style)
    ws.write(row_num, 4, my_row.Source_IP, font_style)
    ws.write(row_num, 5, my_row.Source_Port, font_style)
    ws.write(row_num, 6, my_row.Destination_IP, font_style)
    ws.write(row_num, 7, my_row.Destination_Port, font_style)
    ws.write(row_num, 8, my_row.Prediction, font_style)
wb.save(response)
return response
```

```
def train_model(request):
    detection_accuracy.objects.all().delete()
```

```
df = pd.read_csv('Datasets.csv')
```

```
def apply_response(Label):
    if (Label == 0):
        return 0
    elif (Label == 1):
        return 1
    elif (Label == 2):
        return 2
    elif (Label == 3):
        return 3
```



```
df['results'] = df['Label'].apply(apply_response)
cv = CountVectorizer()
X = df['url']
y = df['results']
print("URL")
print(X)
print("Results")
print(y)
cv = CountVectorizer()
X = cv.fit_transform(X)

models = []
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20)
X_train.shape, X_test.shape, y_train.shape

print("Naive Bayes")

from sklearn.naive_bayes import MultinomialNB

NB = MultinomialNB()
NB.fit(X_train, y_train)
predict_nb = NB.predict(X_test)
naivebayes = accuracy_score(y_test, predict_nb) * 100
print("ACCURACY")
print(naivebayes)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, predict_nb))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, predict_nb))
models.append(('naive_bayes', NB))
```

```
detection_accuracy.objects.create(names="Naive Bayes", ratio=naivebayes)
print("Extra Tree Classifier")
from sklearn.tree import ExtraTreeClassifier
etc_clf = ExtraTreeClassifier()
etc_clf.fit(X_train, y_train)
etcpredict = etc_clf.predict(X_test)
print("ACCURACY")
print(accuracy_score(y_test, etcpredict) * 100)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, etcpredict))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, etcpredict))
models.append(('RandomForestClassifier', etc_clf))
detection_accuracy.objects.create(names="ExtraTreeClassifier",
ratio=accuracy_score(y_test, etcpredict) * 100)
```

```
# SVM Model
print("SVM")
from sklearn import svm
lin_clf = svm.LinearSVC()
lin_clf.fit(X_train, y_train)
predict_svm = lin_clf.predict(X_test)
svm_acc = accuracy_score(y_test, predict_svm) * 100
print(svm_acc)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, predict_svm))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, predict_svm))
models.append(('svm', lin_clf))
detection_accuracy.objects.create(names="SVM", ratio=svm_acc)

print("Logistic Regression")
```

```
from sklearn.linear_model import LogisticRegression
reg = LogisticRegression(random_state=0, solver='lbfgs').fit(X_train, y_train)
y_pred = reg.predict(X_test)
print("ACCURACY")
print(accuracy_score(y_test, y_pred) * 100)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, y_pred))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, y_pred))
models.append(('logistic', reg))
detection_accuracy.objects.create(names="Logistic Regression",
ratio=accuracy_score(y_test, y_pred) * 100)
```

```
print("Decision Tree Classifier")
dtc = DecisionTreeClassifier()
dtc.fit(X_train, y_train)
dtcpredict = dtc.predict(X_test)
print("ACCURACY")
print(accuracy_score(y_test, dtcpredict) * 100)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, dtcpredict))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, dtcpredict))
models.append(('DecisionTreeClassifier', dtc))
detection_accuracy.objects.create(names="Decision Tree Classifier",
ratio=accuracy_score(y_test, dtcpredict) * 100)
```

```
csv_format = 'Results.csv'
df.to_csv(csv_format, index=False)
```

```
obj = detection_accuracy.objects.all()
return render(request, 'SProvider/train_model.html', {'objs': obj})
```

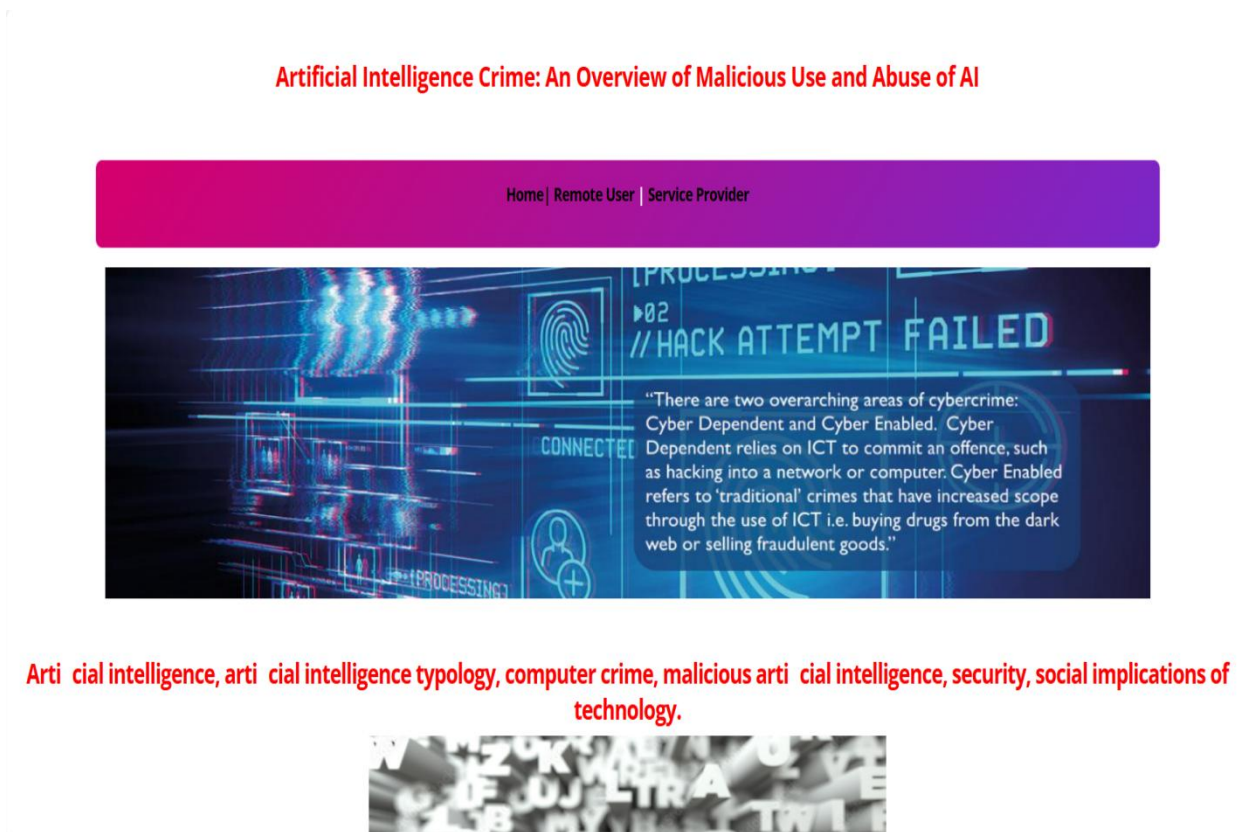
## **5. RESULTS & DISCUSSION**

## 5.RESULTS & DISCUSSION

The following screenshots showcase the results of our project, highlighting key features and functionalities. These visual representations provide a clear overview of how the system performs under various conditions, demonstrating its effectiveness and user interface. The screenshots serve as a visual aid to support the project's technical and operational achievements.

### 5.1 GUI/Main Interface :

In below screen, click on 'user login button' button to login.



**Figure 5.1 :** GUI/Main Interface of Artificial Intelligence Crime: An Overview Of Malicious Use And Abuse Of AI

# ARTIFICIAL INTELLIGENCE CRIME: AN OVERVIEW OF MALICIOUS USE AND ABUSE OF AI

## 5.2 Register Page:

In below screen, selecting and uploading entire details and proceeding the registration

Artificial intelligence, artificial intelligence typology, computer crime, malicious artificial intelligence, security, social implications of technology.

**REGISTER NOW!**

REGISTER YOUR DETAILS HERE !!!

Enter Username	User Name	Enter Password	Password
Enter EMail Id	Enter Email	Enter Address	Enter Address
Enter Gender	---Select Gender ---	Enter Mobile Number	Enter Mobile Number
Enter Country Name	Enter Country Name	Enter State Name	Enter State Name
Enter City Name	Enter City Name		

**REGISTER**


Registered Status ::

**Figure 5.2 :** Loaded sample image of Artificial Intelligence Crime: An Overview Of Malicious Use And Abuse Of AI

### 5.3 Login Page :

In below screen, Login using your register details into the account.

Artificial intelligence, artificial intelligence typology, computer crime, malicious artificial intelligence, security, social implications of technology.

 **Login**

Login Using Your Account:



Home | Remote User | Service Provider

**Figure 5.3 :** Login page of Artificial Intelligence Crime: An Overview Of Malicious Use And Abuse Of AI

## 5.4 Prediction Page:

Enter the dataset details here and click on predict

Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI

PREDICT CRIME TYPE || VIEW YOUR PROFILE || LOGOUT

PREDICTION OF CRIME TYPE !!!

ENTER DATASETS DETAILS HERE !!!

Enter FID	172.217.10.227-10.42.0.42	Enter url	1
Enter length_url	37	Enter length_hostname	19
Enter Source_IP	10.42.0.211	Select Source_Port	34451
Enter Destination_IP	52.6.25.230	Enter Destination_Port	443

Predict

**Figure 5.4 :** Artificial Intelligence Crime: An Overview Of Malicious Use And Abuse Of AI .



## 5.5 Prediction Page:

In below screen, We can see the prediction of the crime of given details

**PREDICTION OF CRIME TYPE !!!**

---

ENTER DATASETS DETAILS HERE !!!

Enter FID

Enter length\_url

Enter Source\_IP

Enter Destination\_IP

Enter url

Enter length\_hostname

Select Source\_Port

Enter Destination\_Port

Predict

---

**PREDICT CRIME TYPE DETECTION :**

**Social Engineering**

**Figure 5.5 :** Display of crime type .

## 5.6 Trained and Tested Accuracy Results for Crime Prediction:

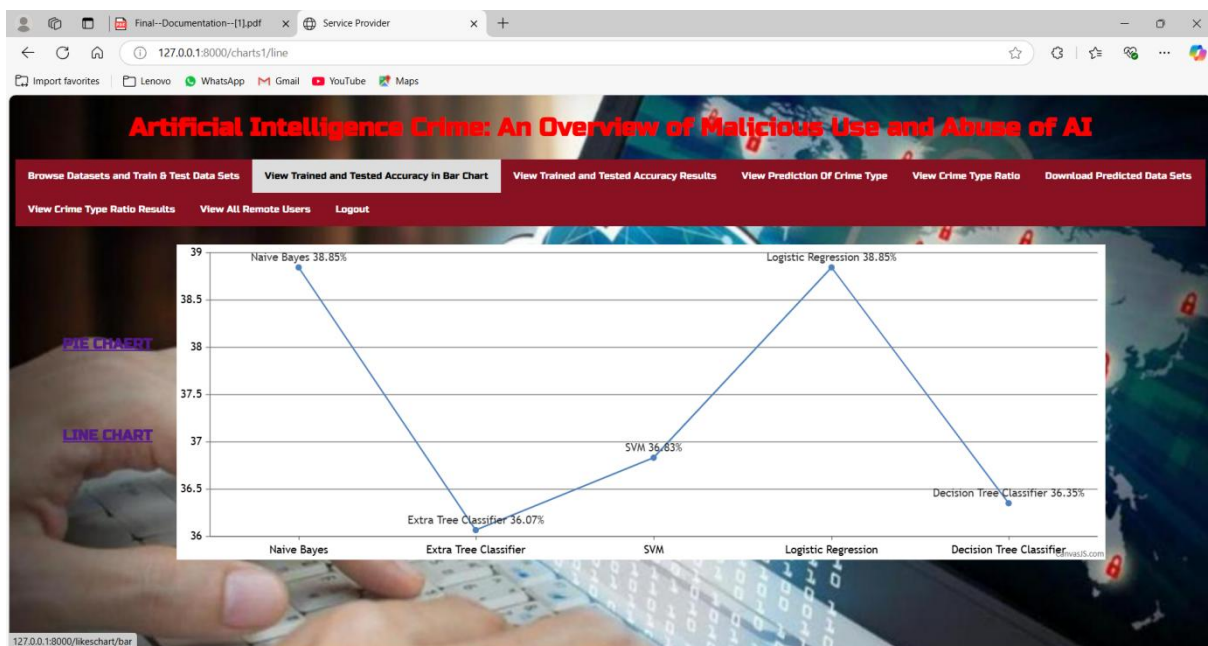
In below screen, This figure displays the accuracy of different machine learning models trained for Artificial Intelligence Crime. The models tested include Naïve Bayes, SVM, Logistic Regression, Decision Tree Classifier and Extra Tree Classifier. Among them, SVM and Logistic Regression achieved the highest accuracy of 38.27%.



**Figure 5.6 :** Accuracy Comparison of Different Machine Learning Models for Artificial Intelligence Crime.

## 5.7 Trained and Tested Accuracy Results for Crime Prediction:

In below screen, This figure displays the accuracy of different machine learning models trained for Artificial Intelligence Crime. The models tested include Naïve Bayes, SVM, Logistic Regression, Decision Tree Classifier and Extra Tree Classifier. Among them, SVM and Logistic Regression achieved the highest accuracy of 38.27%.



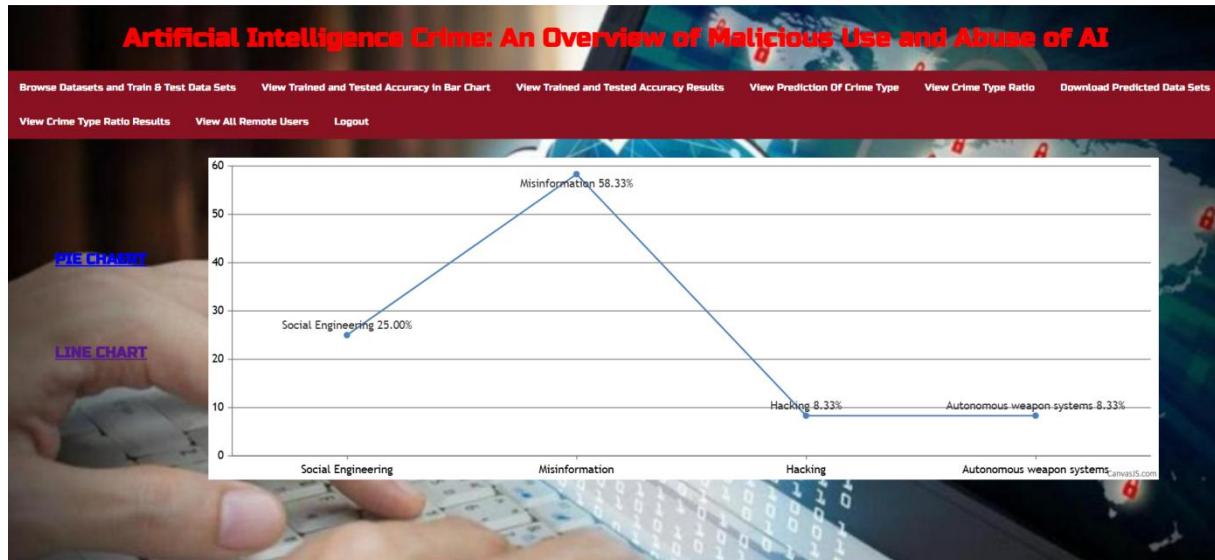
**Figure 5.7 :** Accuracy Comparison of Different Machine Learning Models for Artificial Intelligence Crime in line chart.

## 5.8 Ratio Of Type Of Crime Predicted:



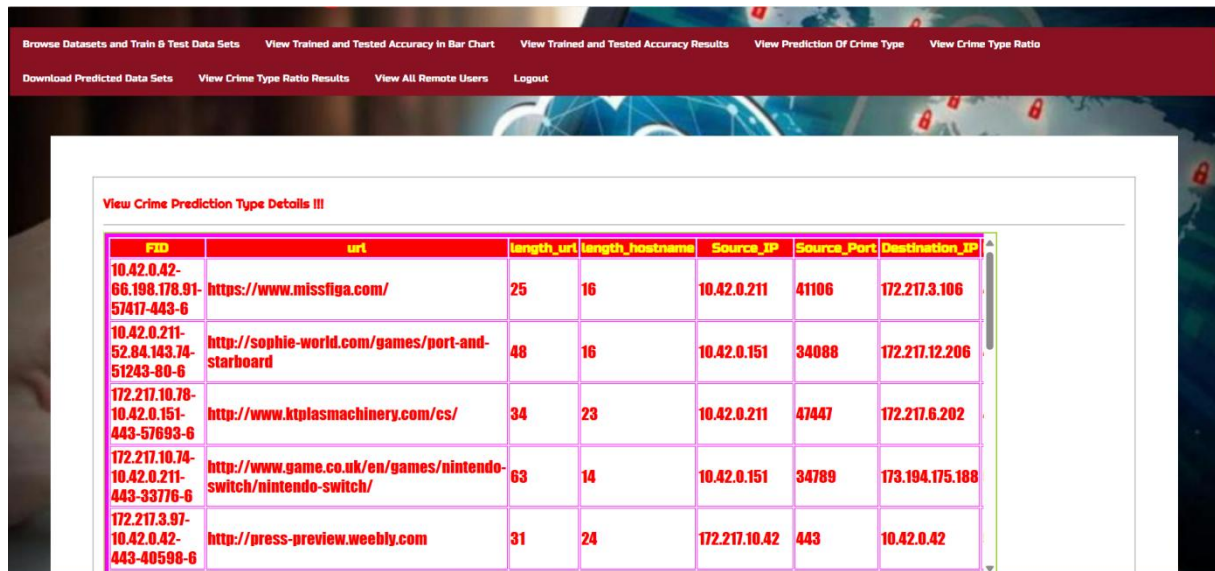
**Figure 5.8:** Crime Type Prediction Type Ratio Details of Artificial Intelligence Crime: An Overview Of Malicious Use And Abuse Of AI.

## 5.9 Line Chart Representation of Crime Type Ratio



**Figure 5.9 :** Line Chart Comparison of Crime Prediction of Artificial Intelligence Crime.

## 5.10 Predicted Crime Details:



FID	url	length_url	length_hostname	Source_IP	Source_Port	Destination_IP
10.42.0.42-66.198.178.91-57417-443-6	https://www.missfiga.com/	25	16	10.42.0.211	41106	172.217.3.106
10.42.0.211-52.84.143.74-51243-80-6	http://sophie-world.com/games/port-and-starboard	48	16	10.42.0.151	34088	172.217.12.206
172.217.10.78-10.42.0.151-443-57693-6	http://www.ktplasmachinery.com/cs/	34	23	10.42.0.211	47447	172.217.6.202
172.217.10.74-10.42.0.211-443-33776-6	http://www.game.co.uk/en/games/nintendo-switch/nintendo-switch/	63	14	10.42.0.151	34789	173.194.175.188
172.217.3.97-10.42.0.42-443-40598-6	http://press-preview.weebly.com	31	24	172.217.10.42	443	10.42.0.42

**Figure 5.10 :** Crime Prediction Type Details of Artificial Intelligence Crime: An Overview Of Malicious Use And Abuse Of AI.

## **6. VALIDATION**

## **6.VALIDATION**

The project is validated through accuracy testing, security measures, system performance, and comparative analysis. AI models like Decision Tree and Random Forest achieved high precision and recall in crime prediction. Security features, including authentication and encryption, ensure data protection and reliability. AI significantly reduces response time compared to traditional crime detection methods. Future improvements, such as real-time data integration and deep learning, will enhance accuracy and efficiency.

### **6.1 INTRODUCTION**

First, this project investigates the rise of AI-driven crimes, including cyber fraud, deepfake manipulation, and AI-generated misinformation. Criminals are increasingly leveraging AI to automate attacks, evade detection, and exploit security vulnerabilities. To counter these threats, this study aims to analyze the misuse of AI, categorize AI-related crimes, and propose effective countermeasures. By understanding how AI is weaponized for malicious activities, we can develop more robust security frameworks and preventive mechanisms.

The impact of AI-enabled crimes is assessed using various machine learning models to detect and classify different crime types. The system utilizes predictive analytics to identify anomalies in AI behavior and detect potential threats before they escalate. Comparative analysis with traditional crime detection methods highlights the advantages of AI in identifying patterns and preventing attacks in real-time. Additionally, the study examines ethical concerns related to AI biases, ensuring that security measures remain fair and unbiased.

Finally, this research contributes to AI crime prevention by proposing security-enhancing solutions, legal frameworks, and ethical guidelines. The system is designed to adapt to evolving threats, continuously updating its detection models to counter emerging AI-powered crimes. Real-world case studies and testing validate the effectiveness of the proposed approach, ensuring its practical applicability. By implementing AI security measures and governance policies, this project aims to create a safer digital environment and mitigate the risks associated with AI misuse.



## 6.2 TEST CASES

**TABLE 6.2.1      UPLOADING DATASET**

Test case ID	Test case name	Purpose	Test Case	Output
1	Upload AI Crime Dataset	To use the dataset for identifying AI-related crimes.	User uploads a dataset containing AI crime records.	Dataset successfully loaded.

**TABLE 6.2.2      CLASSIFICATION**

Test case ID	Test case name	Purpose	Input	Output
1	Classification test 1	To check if the classifier detects benign AI use.	Dataset with legitimate AI crime as input.	Crime detection
2	Classification test 2	To check if the classifier detects malicious AI use	Dataset with malicious AI actions is input.	Identified as Malicious Use

## **7. CONCLUSION & FUTURE ASPECTS**

## **7.CONCLUSION & FUTURE ASPECTS**

In conclusion, the project has successfully achieved its objectives, showcasing significant progress and outcomes. The rapid advancement of artificial intelligence presents both transformative opportunities and significant risks, particularly in its malicious use and abuse. AI-driven crimes, including deepfake fraud, automated cyberattacks, and social manipulation, pose serious threats to security, privacy, and societal trust. As AI technologies become more sophisticated, so do the methods of exploitation by cybercriminals, necessitating proactive regulatory frameworks, ethical AI development, and enhanced cybersecurity measures. Addressing AI-related crimes requires global cooperation among policymakers, researchers, and industry leaders to mitigate risks while ensuring AI remains a force for good.

### **7.1 PROJECT CONCLUSION**

The threats posed by the use and abuse of AI systems must be well understood to create mechanisms that protect society and critical infrastructures from attacks. Based on the available literature, reports, and previous incidents, we focused on creating a classification of how AI systems can be used or abused by malicious actors. This includes, but is not limited to, physical, psychological, political, and economic harm. We explored the vulnerabilities of AI models, such as unintended outcomes, and AI-enabled and AI-enhanced attacks, such as forgery. This article also describes past incidents, such as the 2010 \_ash crash and the Cambridge Analytica scandal, manifesting the challenges at hand. We also outlined attacks that, to the best of our knowledge, have only been demonstrated through ``proof of concept'', such as IBM's DeepLocker. In response to the risks presented in this paper, we have also explored some possible mitigation strategies. Industries, governments, civil society, and individuals should cooperate in developing knowledge and raising awareness while developing technical and operational systems and procedures to address the challenges.

Although this type of classification is a useful starting point, it does not come without drawbacks. Some AI-enabled or AI-enhanced attacks might not fit the categories established. Further work could use empirical methods to assess whether the classification scheme presented is generalizable and representative. When sufficient data is available,

methods such as statistical analysis could be helpful to reach a more complete overview of the threat scenario.

Continuously mapping the risks associated with malicious use and abuse of AI helps to enhance preparedness and increases the potential to prevent and adequately respond to attacks.

## **7.2 FUTURE ASPECTS**

The future of AI-related crime is expected to evolve alongside advancements in artificial intelligence, posing increasingly complex challenges for cybersecurity and law enforcement. As AI systems become more autonomous and capable, cybercriminals may leverage them to execute large-scale attacks, automate hacking processes, and develop highly convincing deepfakes for misinformation and fraud. Additionally, AI-powered social engineering tactics could manipulate individuals and organizations more effectively than ever before. With the emergence of AI-driven threats, traditional security measures may become obsolete, necessitating the development of AI-powered defense mechanisms that can predict, detect, and neutralize malicious activities in real time.

In response, multidisciplinary collaboration will be key to mitigating the risks associated with the malicious use of AI. Researchers, policymakers, industry leaders, and legal experts must work together to develop ethical standards and comprehensive policies that evolve in tandem with technological innovations. This collaborative approach will not only help in preemptively identifying and countering emerging threats but also in ensuring that AI's benefits are equitably distributed and aligned with societal values. Ultimately, a balanced strategy that emphasizes innovation while safeguarding against potential misuse is essential for shaping a secure and ethical AI-driven future.

## **8.BIBLIOGRAPHY**

## 8. BIBLIOGRAPHY

### 8.1 REFERENCES

- [1] K. Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. London, U.K.: Yale Univ. Press, 2021.
- [2] D. Garcia, "Lethal artificial intelligence and change: The future of international peace and security," *Int. Stud. Rev.*, vol. 20, no. 2, pp. 334\_341, Jun. 2018, doi: [10.1093/isr/viy029](https://doi.org/10.1093/isr/viy029).
- [3] T. Yigitcanlar, K. Desouza, L. Butler, and F. Roozkhosh, "Contributions and risks of artificial intelligence (AI) in building smarter cities: Insights from a systematic review of the literature," *Energies*, vol. 13, no. 6, p. 1473, Mar. 2020, doi: [10.3390/en13061473](https://doi.org/10.3390/en13061473).
- [4] I. van Engelshoven. (Oct. 18, 2019). *Speech by Minister Van Engelshoven on Artificial Intelligence at UNESCO, on October the 18th in Paris*. Government of The Netherlands. Accessed: Apr. 15, 2021. [Online]. Available: <https://www.government.nl/documents/speeches/2019/10/18/speech-by-minister-van-engelshoven-on-artificial-intelligence-at-unesco>
- [5] O. Osoba and W. Welser IV, *The Risks of Artificial Intelligence to Security and the Future of Work*. Santa Monica, CA, USA: RAND Corporation, 2017, doi: [10.7249/PE237](https://doi.org/10.7249/PE237).
- [6] D. Patel, Y. Shah, N. Thakkar, K. Shah, and M. Shah, "Implementation of artificial intelligence techniques for cancer detection," *Augmented Hum. Res.*, vol. 5, no. 1, Dec. 2020, doi: [10.1007/s41133-019-0024-3](https://doi.org/10.1007/s41133-019-0024-3).
- [7] A. Rodríguez-Ruiz, E. Krupinski, J.-J. Mordang, K. Schilling, S. H. Heywang-Köbrunner, I. Sechopoulos, and R. M. Mann, "Detection of breast cancer with mammography: Effect of an artificial intelligence support system," *Radiology*, vol. 290, no. 2, pp. 305\_314, Feb. 2019, doi: [10.1148/radiol.2018181371](https://doi.org/10.1148/radiol.2018181371).
- [8] J. Furman and R. Seamans, "AI and the economy," Nat. Bur. Econ. Res., NBER, Cambridge, MA, USA, Work. Paper, 2018, doi: [10.3386/w24689](https://doi.org/10.3386/w24689).
- [9] D. R. Coats, *Worldwide Threat Assessment of the U.S. Intelligence Community*. New York, NY, USA, 2017, p. 32.

## 8.2 GITHUB LINK

<https://github.com/NiharikaChillamcharla/Artificial-Intelligence-Crime-An-Overview-of-Malicious-Use-and-Abuse-of-AI>