INDIAN GOVERNANCE AND POLICY PROJECT

AI AND CRIMES

Understanding the Taxonomy of AI-Related Crimes

Authors
Aparna Bhatnagar
Dedipyaman Shukla

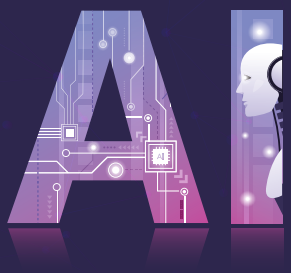
Edited by
Soumya AK

The taxonomy of AI-related crimes is classified into two key categories: **'AI as a Tool' for crimes** and **'AI as a Target' of crimes**.

This classification helps understand the dual roles AI plays in cybersecurity—both as an enabler of crime and a vulnerable entity itself.


Computer as Target
Computer as Tool
AI AS A TOOL FOR CRIME
ADVANCED CYBERCRIME
PHYSICAL CRIME
AI CRIME
TARGET OF CRIME
TARGET OF CRIME
AI AS TARGET OF CRIME
Training System Attack
Training System Theft
Inference System Cracking
Inference System Abuse

Reference

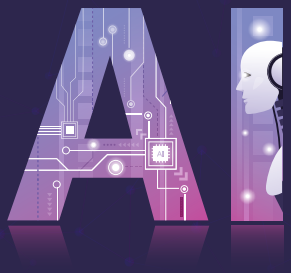


AS A TOOL FOR CRIME

AI is increasingly being used as a tool to commit or facilitate criminal activities. In the digital realm, it enables advanced cybercrimes, such as automated phishing attacks, adaptive malware that evades detection, and deepfake technology used for fraudulent purposes. However, AI's misuse extends beyond cybercrime into physical crimes. For instance, drones equipped with facial recognition can be used for stalking, and vulnerabilities in autonomous vehicles can be exploited for malicious purposes.

AI also plays a direct role in automating illegal activities, such as generating deepfakes or facilitating identity theft. Additionally, AI algorithms are being leveraged to exploit cybersecurity weaknesses, including optimizing Distributed Denial-of-Service (DDoS) attacks.

A cybercrime in which the attacker floods a server with internet traffic to prevent users from accessing connected online services and sites.



AS THE TARGET OF CRIME

In contrast, AI is increasingly becoming a target of criminal activities. Instead of using AI as a tool, criminals focus on compromising AI systems themselves. This can begin during the training phase, where malicious data is introduced to corrupt the system's learning process.

Once deployed, AI systems can be hacked to expose their inner workings or manipulate them to produce harmful outcomes. For instance, content moderation algorithms may be exploited to suppress legitimate content.

UNPACKING TYPES OF AI-ENABLED CRIMES

1 Cybercrime

2 Identity-Based Crimes and Fraud

3 Information Manipulation and Exploitation

4 Financial Crimes

1 Cybercrime

AI refines and intensifies traditional cyberattacks, making them more sophisticated and harder to counter.

Examples

Advanced malware adapts in real-time to evade detection (e.g., AI-driven ransomware targeting high-value victims).

Automated hacking uses AI to identify system vulnerabilities faster than human hackers.

AI-enhanced phishing generates highly convincing fake messages using natural language capabilities, increasing the likelihood of success.

2 Identity-Based Crimes and Fraud

These crimes leverage AI to manipulate or fabricate identities, posing significant risks for individuals and organizations.

Examples

Deepfakes create hyper-realistic media to misrepresent individuals, often for blackmail or misinformation campaigns.

Biometric spoofing replicates fingerprints or facial features to bypass security systems.

AI-driven identity theft analyzes online data to impersonate individuals for malicious purposes.

3 Information Manipulation and Exploitation

AI is weaponized to influence public opinion and destabilize societies.

Examples

Disinformation campaigns use AI to generate and amplify fake news on social media platforms.

AI-enhanced social engineering schemes target individuals with highly personalized attacks.

Propaganda creation involves generating fake content to promote ideologies or disrupt political stability.

4 Financial Crimes

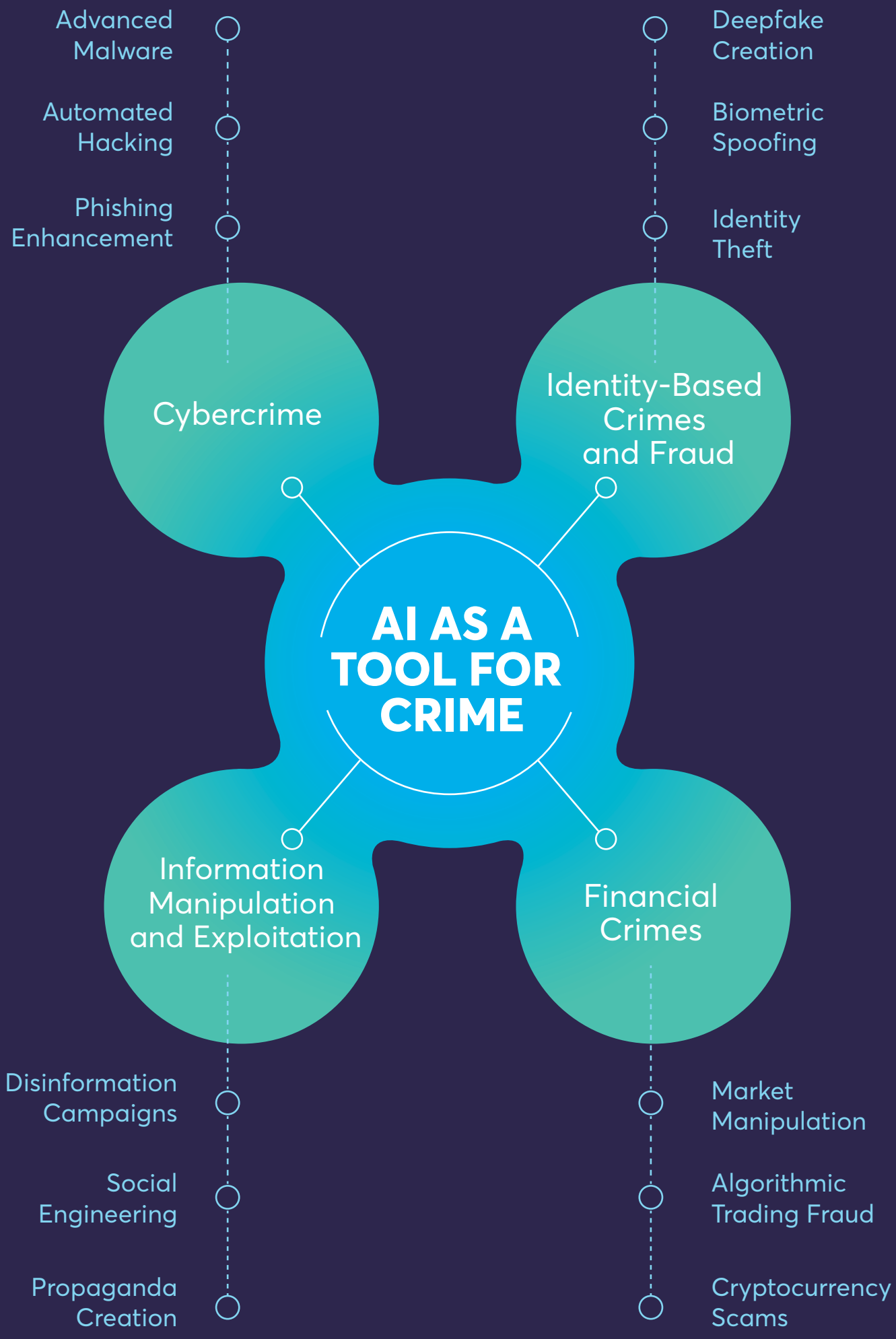
AI is exploited to manipulate financial systems and commit fraud.

Examples

Market manipulation through AI-generated fake trading signals or unethical practices in stock markets.

Algorithmic trading fraud involves activities like front-running trades or price manipulation in high-frequency trading.

Cryptocurrency scams use AI to create fake platforms, impersonate legitimate entities, or steal digital assets.


AI AS A TOOL FOR CRIME
Cybercrime
Advanced Malware
Automated Hacking
Phishing Enhancement
Identity-Based Crimes and Fraud
Deepfake Creation
Biometric Spoofing
Identity Theft
Information Manipulation and Exploitation
Disinformation Campaigns
Social Engineering
Propaganda Creation
Financial Crimes
Market Manipulation
Algorithmic Trading Fraud
Cryptocurrency Scams

The current Indian regulatory framework governing AI-related crimes relies on a combination of traditional laws and sector-specific regulations. Key legislations like the Indian Penal Code, 1860 (IPC) and the Bharatiya Nyaya Sanhita, 2023 (BNS) address conventional criminal activities, while the Information Technology Act, 2000 (IT Act) addresses legal issues arising from the use of information technology, particularly focusing on cybercrimes, data protection, and e-commerce. Additionally, laws like the Representation of the People Act, 1951 (RPA) and the Protection of Children from Sexual Offences Act, 2012 (POCSO) tackle AI misuse in electoral disinformation and online exploitation of children, respectively. The table below outlines specific AI-related criminal incidents reported in 2024, along with the relevant legal provisions.

Criminal Incident

AI voice cloning software

AI voice cloning software mimicked a child's voice in distress to scam a senior citizen into transferring INR 50,000 via Paytm.

Impacted Provisions

Extortion: S.383 (IPC) / S.308 (BNS) – Fear of injury, dishonestly inducing property transfer.

Cheating by Personation: S.416 (IPC) / S.319 (BNS) – Pretending to be another person for dishonest gains.

Criminal Intimidation: S.507 (IPC) / S.351(4) (BNS) – Anonymous threats to induce an act.

IT Act: S.66C (Identity Theft) & S.66D (Cheating via computer resources).

Applicability of Law / Legal Gaps

The existing provisions of the IPC, BNS, and IT Act can be effectively applied to AI-related activities, as their core elements are still met.

The nature of the relevant offences under both the IPC and BNS remains unchanged, ensuring applicability.

While identifying anonymous perpetrators may be more challenging, this issue is not new and has existed in the context of online offences even before AI came into play.

Criminal Incident

Deepfake video

Deepfake video using the voice and likeness of a famous actor critiquing government policies was circulated online to influence public opinion during elections.

Impacted Provisions

Defamation: S.499 (IPC) / S.356 (BNS) – Publishing imputations harming reputation.

Forgery and Use of Forged Records: S.463, S.469, S.471 (IPC) / S.336, S.336(4), S.340 (BNS) – Creating or using false electronic records to damage reputation or deceive.

Identity Theft: S.66C (IT Act) – Dishonest use of unique identity features.

Election-Related Offences: S.123(4) (Representation of the People Act, 1951) – False publication relating to the conduct of a candidate with intent to influence elections.

Applicability of Law / Legal Gaps

Under the IPC, BNS, and IT Act, the definition of 'electronic record' is consistent, encompassing data, images, and sounds. This broad definition allows provisions related to forgery to be applied to election manipulation through deepfakes.

Additionally, the Representation of the People Act, 1951 may apply if such deepfakes are used to influence election outcomes, depending on the specific circumstances of the case.

Criminal Incident

AI generated deepfake obscene video

An **AI generated deepfake obscene video** of a famous actor circulated online.

Impacted Provisions

Defamation: S.499 (IPC) – Publishing imputations harming reputation.

Forgery: S.463 (IPC) /S. 336 (BNS) – Making false electronic records to damage reputation.

Use of Forged Documents: S.471 (IPC) /S. 340 (BNS) – Dishonestly using forged electronic records.

Obscenity: S.509B (IPC) – Sexual harassment via electronic mode by transmitting obscene content.

IT Act: S.67 – Publishing/transmitting lascivious or depraving material, S.67A (IT Act) –Publishing/transmitting sexually explicit material.

Indecent Representation of Women – S.4 (IRWPA, 1986): Prohibiting indecent representation of women in any form.

Identity Theft: S.66C (IT Act) – Dishonest use of unique identity features of any person.

Applicability of Law / Legal Gaps

The IPC, BNS, and IT Act, define 'electronic record' to cover data, records, images, and sounds. This allows forgery provisions to apply to election manipulation deepfakes.

Obscenity under S. 509B of the IPC would also be potentially applicable, however, there exists no similar provision under BNS.

S. 67A of the IT Act may also be applicable if the content meets the standard of sexually explicit act or conduct.

Although IRWPA applicability to the online domain is not explicit, various orders and offence bookings reflect its applicability.

Criminal Incident

AI-generated deepfake video

An **AI-generated deepfake video** falsely portraying a school principal making racist, antisemitic, and offensive remarks went viral, inciting death threats and causing unrest in a suburban community in USA.

Impacted Provisions

Promoting Enmity: S.153A (IPC)/ S.196 (BNS) – Promoting enmity between different groups (religion, race, caste, community) through words, visuals or electronic means.

Outraging Religious Feelings: S.295A (IPC)/ S.299 (BNS) – Deliberate acts to outrage religious feelings (through words, visuals, etc.) via electronic communication.

Identity Theft: S.66C (IT Act) – Dishonestly using someone's identity or unique features.

Applicability of Law / Legal Gaps

BNS makes electronic communication an explicit term in relation to the promotion of enmity. Relevant provisions under the IPC and IT Act also cover this class of offenses.

Criminal Incident

Use of Generative AI Chatbot to launch phishing attacks

Use of Gen AI like ChatGPT to launch phishing attacks that mimic emails from reputable travel brands (e.g., Booking.com, Airbnb), tricking consumers to share financial information.

Impacted Provisions

Cheating & Personation: S.416 (IPC)/ S.420 (IPC) – Cheating by impersonation (inducing delivery of property through dishonest means).

Forgery: S.463 (IPC)/ S.336 & S.336(4) (BNS)– Making a false electronic record to cause harm.

Using forged documents or records: S.471 (IPC)/ S.340 (BNS) – Dishonestly using forged electronic records.

Applicability of Law / Legal Gaps

While provisions of the IPC and BNS would be applicable to such misuse of GPT/GenAI, it is unclear what provisions of the IT Act may be applicable, as personation of a non-corporate entity may not be necessary to commit an offence.

Personation under the IT Act is not defined. However, S.66D may potentially remain applicable.

WHAT WE KNOW, WHAT IS EVOLVING, AND WHAT NEEDS TO BE DONE

1. At this point, our observations suggest that technology merely serves to improve the scale, efficiency or skill associated with existing criminal activity taking place either online or offline.

2. Deepfakes are the most applied technology in criminal activities due to their ability to impact social behaviour and interactions. However, AI hasn't fundamentally changed the nature of the crimes—these activities could still be conducted in the absence of the technology.

3. The real challenge lies in prosecution—how do we identify, apprehend and ensure accountability for criminals who operate under the veil of anonymity? Issues pertaining to identifying first originator of deepfake content or identifying original IP addresses/other electronic records increase the responsibility of intermediaries and service providers in the online space.

4. Cross-border challenges complicate the prosecution of AI-enabled cybercrimes, as perpetrators may be located outside India. In these cases, reliance on treaties like the Mutual Legal Assistance Treaty (MLAT) and extradition agreements will be key.

Stay tuned as we dive deeper into potential regulatory solutions in our next edition!

INDIAN
GOVERNANCE
AND POLICY
PROJECT