Paper Proposal:

The Algorithmic Panopticon: How AI Amplifies Security Exploits in Online Communities

1. Abstract

Modern online platforms, despite their security measures, harbor inherent vulnerabilities rooted in third-party modifications and features that provide a false sense of privacy, such as ephemeral messaging. While these risks are significant on their own, this paper argues that the integration of Artificial Intelligence (AI) acts as a "force multiplier," transforming niche exploits into tools for mass surveillance and manipulation. We will investigate how AI can automate and scale the harvesting of supposedly private or deleted data and weaponize it for sophisticated social engineering attacks. By synthesizing a technical analysis of platforms like Discord with a robust ethical framework, this research exposes the escalating tension between user autonomy, platform responsibility, and the emergent threat of AI-driven exploitation.

2. Introduction & Problem Statement

The document highlights two critical areas of digital vulnerability: the unsanctioned power of third-party client modifications (e.g., Vencord for Discord) and the "myth of ephemerality" in secure messaging. Plugins that log deleted messages or bypass phishing warnings create significant security holes. Simultaneously, users who rely on "disappearing messages" are often unaware that their data remains recoverable.

Our research extends this foundation by introducing AI as a catalyst that dramatically elevates the threat level. The central problem is no longer just that these vulnerabilities exist, but that AI can exploit them systematically, autonomously, and at a previously unimaginable scale. This creates a new paradigm of security risks with profound ethical implications for privacy, consent, and digital safety. This paper will address the core question: **How does the application of AI to existing platform vulnerabilities fundamentally alter the landscape of digital ethics and security?**

3. Proposed Methodology & Division of Labor

Our team will employ a mixed-method approach, blending technical investigation, socio-technical analysis, and ethical inquiry.

- **Phase 1: Technical & Systems Analysis (two members)**
  - **Task:** Conduct a technical audit of the Vencord plugin ecosystem, specifically identifying functionalities that bypass platform-native security and privacy controls (e.g., MessageLogger, AlwaysTrust).
  - **Task:** Analyze data persistence in messaging apps, documenting how "deleted" or "ephemeral" content can be forensically recovered from local caches or network data.
  - **Task:** Develop conceptual models for two AI systems:
    1. An AI that ingests data from tools like MessageLogger to build psychological profiles of users and identify prime targets for social engineering.
    2. An AI-powered forensic tool designed to automatically reconstruct and analyze supposedly ephemeral conversations at scale.
- **Phase 2: Ethical Framework & Discourse Analysis (two members)**
  - **Task:** Develop a comprehensive ethical framework based on principles of digital consent, the right to be forgotten, and distributive justice. This framework will be used to evaluate the harm potential of the conceptual AI systems.
  - **Task:** Analyze public discourse on platforms like Reddit and GitHub regarding modding tools and privacy features. The goal is to understand user motivations and their mental models of security, contrasting them with the technical reality amplified by AI.
- **Phase 3: Case Studies (two members)**
  - **Case Study 1 : AI as the Ultimate Phishing Lure.** This study will explore how an AI, armed with data from privacy-violating plugins, could craft and execute hyper-personalized phishing

attacks. It will analyze the ethical implications of using intimate, undeleted conversation data to manipulate users, effectively automating betrayal of trust.

- **Case Study 2 : AI, Ephemerality, and Vulnerable Populations.** This study will investigate how AI-driven data recovery shatters the protective illusion of ephemeral messaging. We will focus on the disproportionate harm to marginalized groups, activists, or individuals in abusive situations who rely on these features for their safety. The analysis will cover the ethical responsibility of platforms that market a misleading sense of security and how AI-powered forensics could be used by oppressive actors to silence dissent or control individuals.

4. Expected Contributions & Originality

This research project will make three primary contributions to the field of AI ethics and digital security:

1. **A Novel "Force Multiplier" Framework:** We will systematically frame AI's role not as the source of new vulnerabilities, but as an amplifier of existing, human-centric security flaws in online platforms.
2. **A Socio-Technical Analysis of Emergent AI Threats:** By bridging the technical capabilities of client-side mods with the scaling power of AI, we provide a forward-looking analysis of how user consent, privacy, and safety are redefined in this new paradigm.
3. **An Actionable Ethical Framework:** The paper will produce a clear set of ethical principles and recommendations for platforms, policymakers, and users to mitigate the risks of AI-driven exploitation, urging a shift from reactive security to proactive digital safety design.

5. Ethical Considerations for Research

Our team is committed to the highest ethical standards. All research will be conducted using publicly accessible data.

- The AI models will remain theoretical; no functional code or malicious tools will be developed.
- For the discourse analysis, all personally identifiable information (usernames, etc.) from platforms like Reddit and GitHub will be anonymized. All quotes will be paraphrased to prevent re-identification via search engine linkage.
- This proposal will be submitted to the university's Institutional Review Board (IRB) for full review and approval before any data collection begins.

6. Preliminary Timeline

- **Weeks 1-2:** Literature Review & Finalizing Methodology. Team finalizes scope and roles. IRB proposal submission. Data Collection & Technical Analysis (Phases 1 & 2). Two Members conduct technical audits. Two members can begin discourse analysis.
- **Weeks 3-4:** Case Study Development & Analysis (Phase 3). The remaining two members develop their case studies based on findings from the previous phase. Team-wide integration of technical and ethical frameworks.
- **Weeks 5-7:** First Draft Writing. Each member/pair writes their assigned section.
- **Week 8:** Peer Review & Revisions. The team collaboratively reviews and edits the full draft, Finalization & Submission. Final proofreading and submission of the paper.

7. Conclusion

The weaponization of AI to exploit latent security vulnerabilities represents a significant and under-examined ethical challenge. By dissecting how AI can turn user-centric platform modifications and flawed privacy features into tools for mass manipulation, this paper aims to illuminate the urgent need for a more robust ethical and technical approach to platform governance. Our findings will serve as a critical warning and a guide for building a more secure and trustworthy digital future.