Cyber Warfare, Cyberbullying and Psychological Warfare: Ethical and Anticipated Ethical Issues

Richard L Wilson[1] and Noah Donnelly[2]
[1]Department of Philosophy/Computer Science and Information Sciences, Towson University, Baltimore, Maryland, USA
[2]Computer Science and Information Sciences, Towson University, Baltimore, Maryland, USA

wilson@towson.edu
ndonnel1@students.towson.edu

Abstract: This analysis takes as its starting point the view that ideas that originate in the civilian arena can migrate to the area of warfare and cyber warfare. The central idea is that distinctions that apply for cyberbullying can be applied to issues in cyber warfare. The analysis addresses the escalation of individual-level psychological warfare exhibited in cyberbullying to the level of psychological warfare in cyber warfare by analyzing the intersection of social engineering and algorithmic vulnerabilities. It is in this way that cyberbullying is related to cyber warfare. This discussion employs a method that draws upon distinctions taken from computer science, conceptual ethical analysis and case studies. Utilizing case studies from the domain of cyberbullying, this analysis examines four distinct incidents involving Tyler Clementi, Amanda Todd, Jay Taylor and Elijah Heacock. Cyberbullying represents a form of psychological operation that can also operate in cyberwarfare because it weaponizes digital communications to manipulate emotions, erode morale, and destabilize individuals and groups. The goal of the analysis is to identify the evolution of threat modalities related to psychological warfare involving cyber bullying from unauthorized webcams to AI powered velocity sextortion and gamified harassment groups, to the domain of cyber warfare. The cyberbullying cases are analyzed through the lens of Anticipatory Ethics, which is ethical analysis focused on developing technologies, specifically highlighting the failure of social media platforms to uphold the "Sociotechnical Imperative" and the "Post-Deployment Mandate." To mitigate against human factor risks associated with psychological warfare and cyberbullying and cyber warfare, various defensive AI frameworks can be implemented: Behavioral Graph AI to break grooming funnels, Stylometric Analysis to prevent ban evasion and Acoustic Coercion Analysis to detect the real-time sources of psychological distress. This research, employing analysis and definitions of standard concepts from cyberbullying and Cyberwarfare, concludes that proactive, AI driven detection grounded in ethical analysis and the ACM Code of Ethics is essential for securing the safety of the human element in digital infrastructure. Future analysis will explore more directly how cyber bullying is being employed in cyber warfare.

Keywords: Cyber warfare, Psychological operations, Cyberbullying, Anticipatory ethics, Artificial intelligence, Sextortion

1. Introduction

In the modern digital landscape, the distinction between individual level cyber harassment and state sponsored information warfare has grown increasingly less clear. While traditional definitions of bullying emphasize physical intimidation, violence and verbal abuse within a confined physical space, cyberbullying represents a fundamental shift in the manner and space within which aggression takes place. It is defined as "willful and repeated harm inflicted through the use of computers, cell phones, and other electronic devices." (Kresic, 2020) However, when analyzed through the lens of cybersecurity studies cyberbullying is revealed to be a form of psychological operation (PSYOP). Like military PSYOPS cyberbullying weaponizes digital communication to manipulate emotions, degrade morale, and destabilize the psychological state of targeted individuals.

The importance of addressing the issue of cyberbullying is complimented by the escalation of cyberbullying's reach and the ubiquity of digital aggression. Statistics indicate that 1 in 6 school aged children now experience cyberbullying (WHO, 2024). Unlike schoolyard bullying which ends when the victim returns home, cyberbullying is unique in three distinct ways: anonymity, perpetrators can hide behind screens; permanence, digital footprints are hard to erase; and reach, which grants aggressors 24/7 access to their victims.

This analysis investigates the evolution of these threatening characteristics of cyber bullying through four seminal case studies: Tyler Clementi, Amanda Todd, Jay Taylor, and Elijah Heacock. These cases provide a technological trajectory that exhibits how methods from cyber bullying can be applied to cyber warfare from simple hardware misuse to complex, AI-powered attacks. The following cyber bullying cases give examples of techniques that can be adapted to cyber warfare.

- Tyler Clementi (2010) represents the era of unauthorized surveillance and privacy violation.
- Amanda Todd (2012) demonstrates the weaponization of persistence and cross-platform harassment.
- Jay Taylor (2022) marks the shift towards "gamified harassment," where organized groups like "764" treat human suffering as a scored competition.

- Elijah Heacock (2025) shows the emergence of "velocity sextortion," where perpetrators use AI tools to enable professional criminal organizations to execute sextortion scripts in under 45 minutes.

This paper analyzes these incidents through the lens of Anticipatory Ethics and the application of the ACM Code of Ethics, arguing that social media platforms have failed to uphold the "Sociotechnical Imperative" – the obligation for developers to consider the social systems in which their artifacts are embedded (Miller et al., 2011). Furthermore, it proposes that the only feasible defense against this form of psychological warfare is the implementation of defensive AI frameworks. Key defensive frameworks include Stylometric Analysis and Acoustic Coercion Analysis, which aim at detecting and intercepting threats before psychological damage to intended targets becomes permanent.

2. Technical Issues

Understanding cyberbullying as a type of psychological warfare requires an understanding of the technology that facilitates it. The core technical issue lies in the misuse of consumer-grade technology. Everyday devices and platforms, designed for connecting people, are repeatedly weaponized for unauthorized surveillance while allowing for the amplification of harassment.

In the case of Tyler Clementi, the technology was a standard webcam. Webcams are intended for video communication, yet this technology was repurposed to allow for the unauthorized surveillance of a private, intimate act. This represents a failure of physical privacy controls in the digital age. In the case of Amanda Todd, social media served as the "amplification engine" for digital harassment. Social media platforms make abuse public permanent, and inescapable, violating the victim's human right to dignity. These same distinctions can apply to cyber warfare.

However, as we move towards the more recent cases of Taylor and Heacock, technology has evolved and become much more sophisticated. The threat landscape for cyber bullying is now defined by the introduction of the hierarchy of Artificial Intelligence:

1. Artificial Intelligence (AI): The broad science of building machines that mimic human intelligence.
2. Machine Learning (ML): Systems that learn from data without explicit programming.
3. Generative AI (GAI): The specific subset of AI is used to create new content.

Generative AI has introduced a new avenue of "velocity" to psychological warfare.

In the Heacock case, predators utilized AI tools to generate sexually explicit fake images of the victim. This lowers the barrier of entry for sextortionists, they no longer require real images, they can simply generate fake ones. Fake images can also be created for use in the context of cyber warfare The rapid advance of technology requires a shift in ethical analysis from reactive judgement to Anticipatory Ethics, which focuses on the study of the development of technological artifacts and how they will develop and work in the future.

3. Ethical Issues

To understand these ethical issues, we introduce distinctions developed by Furrow. As Furrow states, ethics is related to evaluating actions, and actions are performed by those capable of being moral agents. As Furrow says, "When we evaluate an action, we can focus on various dimensions of the action. We can evaluate the person who is acting, the intention or motive of the person acting, the nature of the act itself, or the consequences." (Furrow, 2005) A moral agent can justify their intentions, actions and outcomes produced by actions by referring to ethical principles. The themes discussed below revolve around how the widespread use of techniques of cyberbullying can easily be extended into the domain of psychological warfare and cyberwarfare. This ethical analysis applies standard ethical frameworks to the three primary stakeholder groups: the victims, the perpetrators, and the platforms.

The ethical defense of the victims is grounded in Kantian Deontology and Rights Ethics. In the case of Tyler Clementi, under Kantian Deontology categorical imperative 1, morality is derived from universal duties. Clementi had the right to be treated with dignity and privacy. This holds for victims of cyber bullying. This duty was also violated along with ACM Principle 1.6: *Respect Privacy.* The case of Amanda Todd can be viewed through Rights Ethics, which focuses on fundamental human rights. Todd possessed the rights to security of person and freedom from degrading treatment (Alabama Policy Institute, 2020). The ruthless digital harassment directed at her violated these rights, violating ACM Principle 1.1: *Contribute to society and human well-being*. The dual cases of Elija Heacock and Jay Taylor invoke the 'Vulnerability Principle' which dictates that special ethical care is owed to vulnerable populations, such as minors. Due to his age Heacock was unable to discern that he was interacting

with professional criminals and a professional script. Furthermore, the Sanctity of Life principle argues that life itself has inherent value. The "764" group, an international online network involved in child exploitation, coercion, and extremist activities, violated this principle by treating Jay Taylor's life as a "toy" to be broken for amusement. This is a violation of ACM Principle 1.2: *Avoid Harm*.

The intentions and ethical failures of the perpetrators of cyber bullying reveal the psychological warfare mindset. In the Tyler Clementi case, Dhuran Ravi's actions can be analyzed as a flawed application of Act Utilitarianism. He performed a distorted analysis related to the greater good, where the amusement of his Twitter followers was weighed more heavily than the harm to Clementi. Like the professional/organized cases of the "764" Network and Sextortion Rings: these actors operate under Ethical Egoism and Nihilistic Hedonism. They reject all moral principles in favor of experiencing immediate pleasure from causing pain to victims. Leaders of the "764" group, such as "Felix" or "Rabid," engaged in Predatory Pragmatism, a criminalized form of efficiency where the victim is dehumanized by treating victims according to a "conversion rate." In this situation, "conversion rate" has become a sanitized corporate metric that has been manipulated for measuring the percentage of targeted participants under the legal age that have been successfully coerced into producing explicit content or committing acts of harm against themselves; thus dehumanizing a person's life into only two values: successful or unsuccessful. This represents a complete rejection of ACM Principle 1.3: *Be honest and trustworthy.*

In these cases, social media platforms (Twitter/X, Meta, Discord) bear significant ethical responsibility for actions performed on their platforms. When analyzing Twitter (now X) which failed to have checks in place to prevent the broadcasting of non-consensual intimate footage in the Clementi case, they violated Virtue Ethics. This is also a failure of ACM Principle 3.7: *Recognize and take care of systems that become integrated into society.* Facebook (now Meta) failed to apply the ethical principle entitled Duty of Care. In the Todd case they failed to connect 22 different aliases to a single user (Coban) quickly enough to stop the cyber stalking that occurred in the case. There was a lack of recognition of product Liability among all the platforms, algorithms such as "People You May Know" often facilitate the initial connection between a predator and a minor. This failure to evaluate the risk of the algorithm violates ACM Principle 2.5: *Give comprehensive evaluations of computer systems and their risks.*

4. AI in Cyber Warfare Context

Before conducting an analysis of specific cases, it is important to contextualize the role of AI in cyber bullying to AI applied to cyber warfare. This paper analyzes AI in the context of Human-Targeting Operations. The convergence of AI and cyberbullying have created what security experts refer to as Automated Social Engineering Environment (Seymour & Tully, 2016). Traditional bullying required human efforts such as typing messages, finding photos, engaging in verbal abuse. AI automates these activities using three tactics: Deep Learning Surveillance, AI algorithms now curate the "targets" for bullies by identifying vulnerable users through sentiment analysis of their posts (Ferrara et al., 2016). Generative Adversarial Networks (GANs), as seen in the Heacock case, GANs can generate realistic images that never actually existed, which creates "evidence" for blackmail where none previously existed. LLM-Driven Psyops, here Large Language Models can now conduct conversations with thousands of victims simultaneously. These models can follow a specialized "sextortion script" optimized for maximized psychological expression. These tactics move cyberbullying from being a social issue to the level of being an example of a technological weapons system.

5. Case 1: Tyler Clementi (2010)

Tyler Clementi, a freshman at Rutgers University became the target of a localized surveillance operation by his roommate Dharun Ravi. On September 19, 2010, Ravi used a remote desktop application to activate the webcam from his friend's room. Ravi used the webcam in his shared dorm with Clementi, to spy on him engaging in an intimate act with another man. Ravi broadcasted this stream and subsequently announced the violation on Twitter, inviting others to watch a planned second attempt.

Surveillance and Exposure as defined in the lexicon of cyber warfare applies to this case, because the incident represents an Insider threat utilizing Unauthorized Surveillance. The webcam, a trusted node in the room's digital environment, was repurposed to act as a sensor. The psychological impact on Clementi was the destruction of his "Sanctuary," an important concept in psychological resilience. By proving that Clementi's most private moments were subject to public observation, Ravi successfully executed an Information Dominance operation. The humiliation related to the surveillance of a private act was related to the act itself and to the

broadcasting of the act. Through the lens of cyber warfare this could be perceived to be the dissemination of 'intelligence' to a hostile audience.

The technical failure is the lack of Hardware-Level Privacy Controls. The webcam software did not provide a distinct and undeniable indicator (hard-wired LED or shutter) that could indicate that the webcam was active and recording. From the perspective of ethics, this was a failure of the Sociotechnical Imperative -- as defined by Miller et al. (2011). This imperative requires creators of technology to anticipate the social context of the use of the technology. The designers of the webcam and the remote software did not anticipate that they would be used for voyeurism in a shared living space. Tyler Clementi died by suicide on September 22, 2010, a direct result of Ravi's information operation.

6. Case 2: Amanda Todd (2012)

Amanda Todd's case demonstrates the weaponization of Persistence. At age 15, Todd was coerced into exposing herself on a webcam by a predator, Aydin Coban. Unlike the Clementi case, which was a specific event, this was a sustained campaign of attrition. Coban utilized 22 different aliases to harass Todd across Facebook, YouTube, and other platforms. When Todd blocked one account, Coban immediately engaged with another attack vector on another platform. When she changed schools, Coban utilized Open-Source Intelligence (OSINT) tactics to locate her new peer group and distributed the explicit images to them.

Coban used the tactics of Attrition and Swarming. These tactics mirror Swarming in military doctrine, an attack from multiple directions designed to overwhelm the target's defenses. By using multiple accounts, Coban created the illusion of a mob despite the attacks being the product of a single actor. The goal was Psychological Attrition: wearing down the victim's will to resist over time. The "safe zones" (new schools and profiles) were systematically breached, creating a state of "learned helplessness" within the victim.

The technical and ethical failures here lay in Identity Management and the Post-Deployment Mandate. Platforms like Facebook had deployed "blocking" tools, but they were ineffective against a determined attacker using "Sybil attacks" (creating many fake identities). The platforms failed to monitor the artifact's effects post-deployment. These shortcomings materialized in the failure to recognize that a ban on an email address alone is insufficient in a cyberbullying/cyberwarfare context. Amanda Todd committed suicide in 2012, succumbing to attrition.

7. Case 3: Jay Taylor (2022)

The Jay Taylor case introduces the concept of Gamified Warfare into the domain of cyberbullying. Taylor, aged 13, was targeted by a recruiter for a gaming server who befriended him. He was then coerced into performing degrading acts on livestream. The perpetrators were members of the "764" network, an online group that "gamified' abuse. Members of this network earned points and status ("clout") based on how far they could push their targeted victims.

In a Cyber Warfare context, the predators used Dehumanization and C2 Structures. The "764" network operates like a decentralized paramilitary group. They utilize Encrypted Command and Control (C2) channels (specifically private channels in Discord servers) to coordinate attacks. The use of gamification in this context serves two purposes: it incentivizes attackers and desensitizes them to the suffering of their victims. This technique in psychology is known as Operant Conditioning. This is indistinguishable from the dehumanization training used to prepare soldiers for kinetic warfare. In the Jay Taylor case, the victim of the attack was no longer viewed as being human; predators have an objective of capturing and destroying a victim for a high score.

This case highlights the danger of Encrypted Enclaves. The "764" network operated in "invite only servers" where platform moderation bots couldn't scan for content. This violated ACM Principle 2.9: *Design and implement systems that are robustly and usably secure.* The platforms failed to detect the Behavioral Signature of the grooming funnel, which involved the movement of a minor from a public gaming lobby (Roblox) to a high risk encrypted private server. Jay Taylor's death was broadcast live, a final act of Propaganda of the deed by the aggressors.

8. Case 4: Elija Heacock (2025)

The death of Elijah Heacock represents the arrival of AI-Automated Kinetic Effects. In 2025, 16-year-old Elijah was contacted by a predator on social media. Unlike previous eras where compromising photos had to be stolen or coerced, the predator used Generative AI tools to fabricate fake explicit images of Elijah using just a photo of

his face. The predator then threatened to release the deep-faked images unless paid. The entire interaction lasted less than 45 minutes from the first "Hello" to Elijah's suicide (WHAS11 Staff, 2023).

In the context of Cyber Warfare this parallels what is referred to as Hyperwar (Velocity) in future warfare studies. Hyperwar refers to conflicts conducted at machine speed, faster than the reaction time of human cognition. The Heacock case is a civilian instance of Hyperwar. The sextortion ring used automated scripts and AI generation to execute the "OODA Loop" (Observe, Orient, Decide, Act) faster than Elijah could process the threat. The Velocity of the attack was the primary weapon employed in the attack. The victim was cajoled into a state of panic (a "cognitive siege") so rapidly that he could not seek help.

This highlights the failure of Foreseeability of Effect (Rule 1) from Miller et al. (2011), according to which the developers are morally responsible. The developers of the Generative AI tools failed to anticipate that their software could be used to manufacture non-consensual pornography involving minors (Tiku, 2022). Technically, the social platforms failed to detect the "Velocity signature"—the rapid escalation of a conversation from a new contact to high pressure demands within minutes.

9. Anticipatory Ethics

Anticipatory Ethics is a future oriented ethical framework that studies potential ethical issues that arise from new and emerging technologies before they are fully integrated into society. As defined by Phillip Brey, Anticipatory Technology Ethics (ATE) requires that we move beyond reactive judgements about the effects of technologies after an event has occurred to a proactive forecasting of the dangers that these technologies pose. This section analyzes the failure of social media platforms to apply basic Anticipatory Ethical analysis to their technologies. These failures lead to the Collingridge Dilemma: (Collingridge, 1980) the problem where impacts cannot be easily predicted until the technology is extensively developed, at which point control becomes exceedingly difficult.

The Prediction Problem and Uncertainty: The central challenge in Anticipatory Ethics is the "Problem of Uncertainty." In the development of the webcam (Clementi case) or Generative AI (Heacock case), developers typically argue they cannot foresee malicious use of their technology. However, Brey argues that ethical forecasting must include "Dark Scenarios." Dark Scenarios specifically model how a bad actor would abuse the system. In the case of generative AI, the "Dark Scenario" of non-consensual deepfakes was not a bug, it was a foreseeable capability involving the misuse of the technology distributed with the built-in ability to manipulate reality. By failing to anticipate this, developers allowed the technology to reach the "entrenched" phase of the Collingridge Dilemma. The entrenched phase is where technology is widely available and impossible to fully recall.

We apply the rules of "Moral Responsibility for Computing Artifacts" (Miller et al.) to bridge the gap between technical design and ethical obligations. According to Rule 4: The Sociotechnical Imperative: This rule states that "People who design, develop, or deploy a computing artifact can do so responsibly only when they make a reasonable effort to take into account the sociotechnical systems in which the artifact is embedded." (Miller et al., 2011). In the case of Clementi, webcam designers failed to apply the Sociotechnical Imperative. The developers designed the camera for a sterile "communication" context, ignoring the "dorm room" sociotechnical context where privacy is not always a given. A responsible design would have included a hardware-level shutter that physically blocks the lens, making remote activation impossible. Rule 2: The Post-Deployment Mandate: This rule asserts that "A person's responsibility includes being answerable for the behaviors of the artifact and for the artifact's effects after deployment." (Miller et al., 2011). In the Amanda Todd case, platforms such as Facebook failed this mandate. They deployed primitive "blocking" features that were functional but insufficient, therefore useless against a "swarming" attacker with multiple accounts. The mandate requires the platform to monitor the effect of the tool (did the harassment stop?), not just the function of the tool (did the block button work?). Rule 1: The Foreseeability of Effect: "The people who design, develop, or deploy a computing artifact are morally responsible for that artifact, and for the foreseeable effects of that artifact." (Miller et al., 2011). In the Heacock case, the creators of text to image generators bear some degree of moral responsibility. It is reasonably foreseeable that a tool can generate "any image" and could potentially be used to generate specific "harmful images." The lack of "Watermarking" or "Provenance" technologies at launch represents a failure of the developers to foresee harm and take steps to mitigate it. Rule 5: Honesty in Promotion declares that "People who design, develop, or deploy a computing artifact must not deceive the users about the artifact." (Miller et al., 2011). Gaming platforms market themselves as "community hubs" or "safe spaces for play." However, the architecture of these platforms, specifically the integration of unmoderated, encrypted channels (such as private

Discord servers) creates a zone of impunity for groups like 764. By promoting the platform as safe for minors while knowing that their encryption architecture prevents the detection of grooming rings, the developers violated Rule 5. They sold a “community” but delivered a “dark room” where any meaningful oversight was impossible.

10. Proposed Solutions: The Cognitive Security Architecture

To counter the speed, persistence and anonymity of modern psychological warfare, we recommend a Defensive AI Architecture that mirrors the OODA Loop (Observe, Orient, Decide, Act) of an AI-driven attacker which is too fast for human moderators to intercept. Therefore, we propose a shift to Automated Cognitive Security; a multilayered AI defense architecture designed to intercept and block threats at the network and application layers, preventing psychological damage to potential victims. This architecture integrates specific recommendations based on the case study analysis at 3 levels: Stylometric Analysis, Behavioral Graph Analysis, and Acoustic Coercion Analysis.

Layer 1: Stylometric Identity Resolution. The target threat of Layer 1 is Anonymity and persistence inspired by the Amanda Todd case. The technical concept is: Current platform defenses rely on “Ban by Attribution” which blocks a specific IP address, device ID, or email address. The Amanda Todd case demonstrates that a motivated attacker can easily avoid defensive strategies by using VPN’s, burner phones, and fake accounts. To counter these tactics, we propose Stylometric Identity Resolution. This relies on the linguistic theory of “Idiolect.” Idiolect states that everyone possesses a unique, subconscious cognitive fingerprint in how they construct language. The implementation of this engine would operate using a Support Vector Machine (SVM) or a Long Short-Term Memory (LTSM) neural network trained on three feature sets (Narayanan et al., 2012): Lexical Features, the statistical distribution of vocabulary richness, unique word usage, and grammatical mistakes. Syntactic Features relate to the frequency of function words (e.g., "the," "and," "but"), punctuation patterns, and sentence length variability. Idiosyncratic Features involve specific uncommon grammar or slang unique to a user. The operational workflow would function as follows: When a user is banned for severe harassment (a “High Confidence Threat Actor”), their writeprint is hashed and stored in a “Recidivism Database.” When a new account interacts with the victim the proposed engine will analyze the first N messages. If the new accounts writeprint correlates with the banned writeprint with a confidence interval of over 95% the account will subsequently be flagged as a “Sockpuppet.” Following this identification the system executes a Shadowban, which hides the users’ profile and activity from the new accounts, effectively neutralizing the attacker. This fulfills the Post-Deployment Mandate by ensuring the “block” functionality is resilient against emerging threats (Narayanan et al., 2012).

Layer 2: Behavioral Graph Anomaly Detection. Predatory groups like the members of the “764” network rely on a specific social structure. They use a “Grooming Funnel” to move victims from high-visibility public spaces (e.g., Roblox lobbies) to low-visibility encrypted spaces (e.g., private Discord servers). Traditional content scanning fails to detect the predator’s presence because the invitation is typically inconspicuous. But what lays underneath is the metadata which reveals the threat. An implementation of this concept would look like: using Graph Theory to detect Topological Anomalies (Akoglu et al., 2015). The social network is modeled as a graph G = (V, E), where nodes (V) are users and edges (E) are interactions. This engine would calculate the Centrality Metrics to spot “Grooming Hubs.” A grooming ring will normally display a “Star Topology” where the central hub (predator) forms rapid connections with multiple unrelated leaf nodes (minors). For The “Swarm Signature” to be identified, the system looks for a specific sequence of graph updates: Cluster Formation, a group of High-Risk nodes enter a Low-Risk environment. Rapid Edge Creation, these nodes initiate friend requests with the target nodes (victims). Funneling, the target node receives a URL to an external, encrypted domain within T minutes of connection. If this “Swarm Signature” is detected, the AI triggers an off switch. The invite link is killed, and the minor’s account is placed in “isolation mode” which prevents new friend requests and messages from new accounts. This breaks the Command and Control (C2) structure of the gamified harassment group (Levine, 2025).

Layer 3: Acoustic Coercion Analysis. The threat to a target at this layer is real-time psychological distress which was demonstrated in the Jay Taylor case. In the Jay Taylor case, the abuse happened in a voice chat. Text filters cannot detect audio coercion, and recording all calls for transcriptions violates privacy laws per the GDPR/CCPA. To solve this, we recommend Acoustic Coercion Analysis using Edge AI. This approach analyses how something is said and not what is said, which preserves privacy by never transcribing words. The Acoustic Coercion Analysis model is a lightweight Convolutional Neural Network deployed on the user’s device (Edge Computing). It extracts Paralinguistic Features from the audio stream (Guyer et al., 2021): Jitter (Frequency Perturbation), micro-fluctuations in pitch. High jitter is biologically correlated with the fight or flight response which causes a lack of laryngeal control caused by fear. Shimmer (Amplitude Perturbation), which is caused by small fluctuations in

loudness. And Fundamental Frequency(F0): Sudden spikes in pitch indicating panic or screaming. The model establishes a "Baseline Calm" profile for the user. If the audio stream deviates from the baseline into the "Distress Cluster" (high jitter + high energy) for a prolonged duration (over 30 seconds), the system infers Psychological Duress. Once triggered, the defense system takes over. The system can automatically mute incoming audio from the aggressor, lower volume, or display a "Safety Check" pop-up ("Are you okay? Click here to exit call and report"). This acts as a sensor for detecting physiological signs of the psychological attack (Guyer et al., 2021).

Layer 4: LLM-Based script Interdiction. The target threat is velocity and automation that were tactics used against Elijah Heacock. The "Velocity Sextortion" attack on Elijah Heacock succeeded because it was faster than the speed of human cognitive processing. The attackers used a professional script. To counter this, we treat sextortion scripts as Malware Signatures. To solve this problem, we can utilize a Large Language Model (LLM) trained on a dataset of known sextortion transcripts. Unlike simple keyword matching, the LLM uses Vector Embeddings to understand intent (Huang et al., 2023). It scans Direct Message requests from unknown users for the "Sextortion Flow":

- 1. Phase 1: Grooming (Hyper-complimentary language).
- 2. Phase 2: Escalation (Request for image or platform migration).
- 3. Phase 3: The Turn (Threat of exposure)

If an incoming direct message sequence matches this semantic flow with high probability, the message is pre-flagged. The recipient (victim) receives a modified version of the message covered by a "Safety Blur" with a warning message: *"This message matches patterns used by financial extortion scams. Do not send images."* This intervention will happen in milliseconds, breaking the OODA Loop of the attacker before the victim can be psychologically compromised.

11. Limitations and Ethical Considerations of Cognitive Security Architecture

While Defensive AI offers a great shield against psychological warfare for both cyberbullying and cyber warfare, it introduces unique ethical and technical challenges. First, it can create a False Positive Dilemma. In a "Cyber Warfare" context, a False Negative (failing to stop an attack) can result in loss of life. Inversely a False Positive can result in censorship. With cyber bullying, Sylometric Analysis might flag a sibling using the same computer as a banned user. To mitigate against this, we propose an escalation policy where AI can quarantine an attacker, but permanent bans require human review. Additionally, privacy concerns are raised when discussing Acoustic Coercion Analysis, which requires constant monitoring of the microphone. Even if processing is done "on the edge" (system side) and data is never sent to the cloud, this still raises surveillance concerns. The implementation of Acoustic Coercion Analysis requires Transparent Consent. Gamers must opt-in to "Safety Monitoring" not dissimilar from how driver's opt-in to telematics insurance (Nissenbaum, 2009). The third and final issue addressed here is an Adversarial Arms Race. Just as anti-virus software spurred the creation of polymorphic viruses, Defensive AI will most likely spur the development of Adversarial AI. Technologically proficient aggressors or online black-market merchants may develop tools to obfuscate their writeprints or use voice changers to defeat acoustic analysis. This necessitates a continuous cycle of anticipatory ethics applied to design, where developers must constantly Red Team their own systems to find vulnerabilities to patch before attackers exploit these vulnerabilities.

12. Conclusion

The evolution of the form of threats in this study from unauthorized surveillance of Tyler Clementi (2010) to the swarming attrition of Amanda Todd (2012), to the gamified harassment of Jay Taylor (2022), and to the machine-speed sextortion of Elijah Heacock (2025) all demonstrate that cyberbullying is now more than a social nuisance. It has evolved into localized, asymmetrical Psychological Warfare. This analysis has shown that the tactical procedures of modern cyberbullies mirror military Information Operations and vice versa. They utilize surveillance to locate targets, use attrition to degrade morale, and "Hyperwar" velocities to overwhelm human cognitive defenses. The integration of Generative AI and Encrypted Command and Control (C2) channels effectively multiply the force of the perpetrators, making tradition human-centric moderation problematic. Through the lens of Anticipatory Ethics, we conclude that the failure to protect these victims was not an inevitable tragedy but a breakdown of the Sociotechnical Imperative. Developers of computer hardware, social platforms and AI models failed to anticipate the darker scenarios that could develop because of their creations. Therefore, this paper argues that the security industry must adopt a new mandate: Cognitive Security. User reports are insufficient for attacks that occur in milliseconds. We must deploy the Cognitive Security Architecture

proposed. Stylometric Identity Resolution to strip anonymity, Behavioral Graph Analysis to destroy grooming funnels, and Acoustic Coercion Analysis to detect real-time distress. These tools transfer the burden of defense from the vulnerable individual to the proposed system. Ultimately, the defense of the "Human Element" is no longer a matter of policy, it is a matter of engineering. Firewalls protect the network layer, antiviruses protect the application layer, Cognitive Defense AI must now protect the psychological layer. It is the moral obligation of the computing profession, grounded in the ACM Code of Ethics, to build this shield before the next evolution of psychological warfare claims another child's life.

13. Future Work

While the study focused on individual-level psychological warfare, the defensive frameworks proposed have broader applications. Future research should focus on applying this same technical analysis, specifically Behavioral Graph Anomaly Detection and Stylometric Identity Resolution, to state-level Cyber Warfare cases. By analyzing "writeprints" and "swarm topologies" of Advanced Persistent Threats (APTs), researchers may identify overlapping operational modes of attack between individual aggressors and state-sponsored aggression at the level of cyber warfare.

Ethics Declaration: No human participants or personally identifiable information were involved. All data sources were publicly available.

AI Tools Declaration: ChatGPT 5.1 for drafting and refinement. Human authors verified all content. Gemini 3.0 pro used for aiding in sourcing.

References

Alabama Policy Institute. (2020, November). *Understanding the Difference Between Positive and Negative Rights*. Alabama Policy Institute. https://alabamapolicy.org/wp-content/uploads/2020/11/GTI-Brief-Positive-Negative-Rights-1-1.pdf

Akoglu, L., Tong, H., & Koutra, D. (2015). Graph based anomaly detection and description: a survey. Data Mining and Knowledge Discovery, 29(3), 626-688

Association for Computing Machinery. (2018). *ACM Code of Ethics and Professional Conduct*. https://www.acm.org/code-of-ethics

Brey, P.A.E. Anticipatory Ethics for Emerging Technologies. *Nanoethics* 6, 1–13 (2012). https://doi.org/10.1007/s11569-012-0141-7

Collingridge, D. (1980). *The Social Control of Technology*. St. Martin's Press.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96-104

Guyer, J. J., et al. (2021). Paralinguistic Features Communicated through Voice can Affect Appraisals of Confidence. *Journal of Nonverbal Behavior*, 45, 479–504.

Huang, Y., et al. (2023). Are Large Language Models Good Evaluators for Intent Detection? Findings of the Association for Computational Linguistics: ACL 2023

Mike Levine (2025, November 18). '10 minutes of murder': Why one family is speaking out about the online extremist network 764. *ABC News*. https://abcnews.com/US/10-minutes-murder-family-speaking-online-extremist-network/story?id=127039503

Kresic, M. (2020). Bullying through the Internet - Cyberbullying. *Psychiatria Danubina*, 32(Suppl 2), 250–253. https://pubmed.ncbi.nlm.nih.gov/32970646/

Miller, K.W., et al. (2011). Moral Responsibility for Computing Artifacts: 'The Rules'. *IT Professional*, 13(3), 57–59. https://doi.org/10.1109/mitp.2011.46

Narayanan, A., et al. (2012). On the Feasibility of Internet-Scale Author Identification. *IEEE Symposium on Security and Privacy*.

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.

Parker, I. (2012, January 30). The Story of a Suicide. *The New Yorker*. https://www.newyorker.com/magazine/2012/02/06/the-story-of-a-suicide

Proctor, J. (2022, August 6). Dutch man Aydin Coban found guilty of extortion in Amanda Todd case. *CBC News*. https://www.cbc.ca/news/canada/british-columbia/aydin-coban-amanda-todd-verdict-1.6543853

Seymour, J., & Tully, P. (2016). Weaponizing data science for social engineering: Automated E2E spear phishing on Twitter. Black Hat USA, https://blackhat.com/docs/us-16/materials/us-16-Seymour-Tully-Weaponizing-Data-Science-For-Social-Engineering-Automated-E2E-Spear-Phishing-On-Twitter-wp.pdf

Somers, M. (2020, July 21). Deepfakes, explained. *MIT Sloan*. https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained

Tiku, N. (2022, September 28). AI can now create any image in seconds, bringing wonder and danger. *The Washington Post*. https://www.washingtonpost.com/technology/interactive/2022/artificial-intelligence-images-dall-e/

WHAS11 Staff. (2023, April 27). *Predators targeted a Kentucky teen for sextortion. He died less than an hour later* [Video]. YouTube. https://www.youtube.com/watch?v=R3BQAsAmgoc

World Health Organization. (2024, March 27). *One in Six School-Aged Children Experiences Cyberbullying, Finds New WHO/Europe Study*. https://www.who.int/europe/news/item/27-03-2024-one-in-six-school-aged-children-experiences-cyberbullying--finds-new-who-europe-study

Yang, Z., & Radke, R. J. (2025). Proceedings of the Winter Conference on Applications of Computer Vision (WACV) Workshops, pp. 1083-1092.