iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.77985

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

The Psychology Behind Online Anonymityand Toxic Behaviour

Ishank Batra[1], Shaurya Saraswat[2], Krrish Dang[3], Yogita Thareja[4]
[1, 2, 3]*Research Scholar,* [4]*Assistant Professor, Vivekananda Institute of Professional Studies, Delhi,*

Abstract: ***The quick development of digital communication platforms has opened up new areas where people's behavior is profoundly and frequently shaped by online anonymity. This paper looks at the psychological understanding of online anonymity and how it can lead to harmful behaviors like hate speech, harassment, cyberbullying, and flaming. In order to provide a thorough understanding of why people behave differently when protected by anonymity in digital environments, the study shows findings from theoretical literature, drawing on important theoretical frameworks.***

The research paper shows three main factors of toxic online disinhibition: anonymity, invisibility and lack of eye-contact. The long-held belief that toxic behavior is solely motivated by anonymity is challenged by experimental findings showing that the single biggest cause of negative disinhibition is absence of eye-contact. The study also shows how cyberbullying appears in online forums, how anonymity is adversely correlated with aggressive attacks, and how algorithmic amplification and social media platform design amplify these effects.

The study recognizes the dual nature of anonymity in addition to the risks: although it encourages toxic behavior, it also decreases barriers to self-disclosure, assists vulnerable people seeking mental health support, and creates positive anonymous networks. The effects of toxic online behavior are examined at the individual level, such as despair, anxiety, and suicidal thoughts among victims, as well as at the communal and societal levels, such as the dissemination of false information and heightened social division.

AI-powered content filtering, digital literacy instruction, identity verification systems, and platform design reform are some of the methods to lessen toxic behavior that are covered.

Keywords: online anonymity, toxic behaviour, cyberbullying, online disinhibition effect, flaming, digital psychology, deindividuation, eye-contact, social media

I. INTRODUCTION

A. Background of the Study

Instant messaging, video chats and social networks are just few examples of online communication frameworks that have shown rapid advancement and growth over the last decade. Such frameworks have become accessible to larger groups of Internet users, who can connect with one another from their home computer, laptop or mobile smartphone.

The rise of anonymous online interactions and the increasing prevalence of digital intimacy are reshaping interpersonal dynamics in profound ways. Digital intimacy, as explored by Massimiano Bucchi, emphasizes the human aspect of technology use, highlighting how individuals adapt and integrate technological novelties into their intimate lives.

The COVID-19 pandemic has exacerbated issues of online sexual abuse, particularly among youth, as digital spaces become central to communication and identity formation, leading to increased online harassment and necessitating new digital safety initiatives. The anonymity afforded by digital platforms, such as Twitter, has been linked to increased cyber sexual harassment, as anonymity can embolden users to engage in harmful behaviors.

B. Problem Statement

One phenomenon that has been shown to be characteristic of online communication participants is the online disinhibition effect, defined as a lowering of behavioral inhibitions in the online environment (Lapidot-Lefler & Barak, 2012). Many of the human behaviors that are witnessed in cyberspace, including violence, incitement, flaming, and verbal attacks, on the one hand, and self-disclosure, philanthropy, and the dispensing of help and advice, on the other, may be attributed to the online disinhibition effect.

Cyberbullying is a growing concern in online communications. Cyberbullying has negative impacts such as distress or suicide of a victim. One common type of cyberbullying attack utilizes aggressive forum posts to insult or threaten a victim.

Everyday users on the Internet - as well as clinicians and researchers - have noted how people say and do things in cyberspace that they wouldn't ordinarily say and do in the face-to-face world. They loosen up, feel less restrained, and express themselves more openly. So pervasive is the phenomenon that a term has surfaced for it: the online disinhibition effect.

C. Research Objectives

The present research studied the impact of three typical online communication factors on inducing the toxic online disinhibition effect: anonymity, invisibility, and lack of eye-contact. Using an experimental design with 142 participants, we examined the extent to which these factors lead to flaming behaviors, the typical products of online disinhibition.

This paper explores the psychological underpinnings of anonymous chatting and its impact on intimacy, emotional self-disclosure, and trust. Drawing from social penetration theory and the online disinhibition effect, the study examines how anonymity fosters deep self-revelation while also posing risks such as deception, emotional dependency, and digital addiction.

D. Research Questions

The significance of studying anonymous intimacy in the digital era lies in understanding the complex interplay of benefits and risks associated with technology-mediated practices. This paper seeks to explore the following research questions: In what ways does anonymity influence emotional self-disclosure? To what extent can deep intimacy develop through anonymous communication? What psychological factors facilitate or impede intimacy in anonymous interactions?.

E. Significance of the Study

By integrating diverse academic perspectives, this review provides a comprehensive understanding of the paradox of anonymous intimacy, offering both freedom and vulnerability, while emphasizing the need for responsible technological advancements that balance privacy with accountability in digital interactions.

The objective in this paper is to progress towards developing tools to automatically detect potential cyberbullying. An automated tool to detect cyberbullying may help reduce the occurrence of cyberbullying or negative impacts associated with cyberbullying.

II. LITERATURE REVIEW

A. Concept of Online Anonymity

1) Definition of Anonymity in Digital Environments

Anonymity, or the condition of being unknown (nameless) to others, is considered a major determinant of disinhibitive behavior. In an early experiment, unrelated to the virtual environment, Zimbardo (1969) examined behavioral disinhibition by manipulating the anonymity variable by asking participants to administer electric shocks to fellow participants. Some of those who were asked to administer the shocks had their faces covered by hoods with eyeholes and were dressed in loose-fitting smocks, while others had no hoods and were dressed in their regular attire and bore name tags. The anonymous group of shock administers behaved more aggressively, delivering longer shocks than did their non-anonymous counterparts.

As people move around the Internet, others they encounter can't easily determine who they are. Usernames and e-mail addresses may be visible, but this information may not reveal much about a person, especially if the username is contrived and the e-mail address derives from a large Internet service provider. Technologically savvy, motivated users may be able to detect a computer's IP address, but for the most part others only know what a person tells them. If so desired, people can hide some or all of their identity. They also can alter their identities. As the word 'anonymous' indicates, people can have no name or at least not their real name.

2) Types of Anonymity: Full Anonymity, Pseudonymity, Partial Anonymity

In an attempt to refine and elucidate the concept of anonymity we chose to refer to the unidentifiability aspect of anonymity rather than namelessness. That is, other than being nameless we argue that a major - actually crucial - aspect of anonymity should focus on the unidentifiability dimension, which is quite broader and more personally significant than just being nameless. The unidentifiability factor refers to the realistic condition of being unknown to online partners in terms of identifying personal details, such as gender, weight, age, occupation, ethnic origin, residential location, and so on. Online unidentifiability thus makes it possible to use real names while remaining relatively anonymous, as long as other identifying details remain concealed.

Forum posts can be labeled with a feature of anonymity to describe whether a victim is unable to identify the individual posting a forum post. A cyberbully may have psychosocial motivation for being anonymous.

Anonymity may free the cyberbully 'from normative and social constraints on their behavior' (Patchin & Hinduja, 2006) or allow the cyberbully to avoid social punishment through school or criminal law.

Anonymity in online environments, such as social networking sites (SNSs), can lower barriers to self-disclosure by reducing perceived risks and enhancing privacy, thus facilitating deeper relational development. The dynamic nature of online communication platforms has significantly altered the application of Social Penetration Theory, as users navigate the balance between privacy and openness in digital interactions.

B. Toxic Behaviour in Online Spaces

1) Definition of Toxic Behaviour

Suler (2004) coined the term toxic disinhibition to describe phenomena of online flaming and acting-out behaviors that often involve damaging the other's or even one's own self-image, without any beneficial personal growth. This can be seen in Pro-Ana and Pro-Mia blogs, social loafing in online communities, online gaming sites, hate sites, violent pornographic and pedophilic sites, cyberbullying, comments on YouTube, online interpersonal relationships, organizational conflicts, and more.

The disinhibition is not always salutary. We witness rude language, harsh criticisms, anger, hatred, even threats. Or people visit the dark underworld of the Internet - places of pornography, crime, and violence - territory they would never explore in the real world. We may call this toxic disinhibition.

2) Forms of Toxic Behaviour: Trolling, Cyberbullying, Hate Speech, Online Harassment

Flaming behavior - considered a typical manifestation of toxic online disinhibition - is defined as the use of hostile expressions toward others in online communication. It typically includes the use of a variety of textual elements, such as aggressive and hostile language, swearing, derogatory names, negative comments, threats, and sexually inappropriate comments. Flaming behavior can also be observed in the use of capital letters, in the increased use of question marks and exclamation points, and in the mixture of letters, numbers, and dingbats conceptualizing shouting or calling a derogatory name.

Cyberbullying is a new form of bullying which can occur through email messages, forum posting, web site publishing, cell phone text messages, chat rooms, computer hacking, and any other means of electronic communication. For example, cyberbullies use email, cell phone text messages, or forum posts to send insults or death threats to victims. Another example is a cyberbully creating web pages to insult a victim or perform polling to vote negatively about a victim.

Cyberbullying attacks can happen anywhere and anytime - an individual can receive messages at home at night - and attacks are impersonal compared to face-to-face bullying. Attacks using forums or websites have some degree of permanency and thus it is possible for a single cyberbully incident to achieve an effect similar to repetitive attacks.

Addressing issues like grooming, cyberbullying, and misinformation in the digital age requires a multifaceted approach that combines legal, technological, and educational strategies. Online grooming, a significant threat to minors, involves deceptive relationships developed through digital communication to exploit vulnerabilities.

C. Psychological Theories Related to Online Behaviour

1) Online Disinhibition Effect

This disinhibition can work in two seemingly opposing directions. Sometimes people share very personal things about themselves. They reveal secret emotions, fears, wishes. They show unusual acts of kindness and generosity, sometimes going out of their way to help others. We may call this benign disinhibition. However, the disinhibition is not always so salutary. We witness rude language, harsh criticisms, anger, hatred, even threats.

The Online Disinhibition Effect, as conceptualized by Suler (2004), manifests in both benign and toxic forms, influencing online behavior in diverse ways. Benign disinhibition encourages honesty and openness, as evidenced by Kahlow's (2024) study, which found that anonymity on platforms like Reddit facilitates deeper self-disclosure among individuals discussing ADHD, compared to more public platforms like Facebook. Conversely, toxic disinhibition can lead to harmful behaviors, such as cyber incivility and deception.

The psychological restraints that often serve to block or conceal emotions and undisclosed needs are found to be lowered in cyberspace in various online interpersonal behaviors. The negative online disinhibition effect is the concept used to refer to the negative results of this loss of inhibitions, usually manifested in aggressive behaviors that apparently would not be exhibited in a similar scenario in the 'real world'.

2) Deindividuation Theory

Invisibility renders irrelevant stereotypes and prejudices related to gender, age, skin color, physical attributes such as weight, height, and general appearance, stigmatizing behaviors such as stuttering, and physical and sensory impairments. Diminished social presence may lead to a process of communication-induced de-individuation, which in turn may produce instances of disinhibition.

Participants placed in a dark, fully padded room, in conditions of utter invisibility, were found to exhibit greater behavioral disinhibition than did an experimental group that was left in an illuminated room under conditions of full visibility. Based on this observation, the investigators concluded that visual anonymity induced a sense of freedom from social norms and restrictions, allowing people to manifest their feelings, whether aggressive or affectionate, in more meaningful ways.

3) Social Identity Theory

The traditional Internet philosophy holds that everyone is an equal, that the purpose of the net is to share ideas and resources among peers. The net itself is designed with no centralized control, and as it grows, with seemingly no end to its potential for creating new environments, many of its inhabitants see themselves as innovative, independent-minded explorers and pioneers. This atmosphere and this philosophy contribute to the minimizing of authority.

Anonymity can influence social identity formation, as informal and anonymous interactions may reinforce the 'friend/foe' dichotomy, shaping group identities and perceptions of in-group versus out-group members.

4) Diffusion of Responsibility

As seen with the shock experiment, anonymity may cause Internet users to feel unaccountable for their negative actions, as they cannot be identified as the perpetrators of certain actions or behaviors. This loss of accountability may result in an increased level of toxic disinhibition, consequently promoting impulsive, aggressive, and abusive behaviors.

This anonymity is one of the principle factors that creates the disinhibition effect. When people have the opportunity to separate their actions online from their in-person lifestyle and identity, they feel less vulnerable about self-disclosing and acting out. Whatever they say or do can't be directly linked to the rest of their lives. In a process of dissociation, they don't have to own their behavior by acknowledging it within the full context of an integrated online/offline identity.

III. PSYCHOLOGICAL FACTORS BEHIND TOXIC BEHAVIOUR

A. Reduced Accountability

In the case of expressed hostilities or other deviant actions, the person can avert responsibility for those behaviors, almost as if superego restrictions and moral cognitive processes have been temporarily suspended from the online psyche. In fact, people might even convince themselves that those online behaviors 'aren't me at all'.

Recent research shows that cyberbullying has several differences from face-to-face bullying. In cyberbullying, communications can be anonymous - communication is associated with an unknown identifier - and attacks can come from a broader audience relative to face-to-face bullying. Cyberspace lacks supervision, attacks can be technical in nature, and cyberbullies can use technology to coordinate attacks.

Anonymity correlated with attacks positively and defends negatively. This finding indicates that attackers used anonymity more relative to non-aggressive communications, while defenders were less anonymous relative to non-aggressive communications.

B. Emotional Detachment

In many online environments, especially those that are text-driven, people cannot see each other. When people visit web sites, message boards, and even some chat rooms, other people may not even know they are present at all. This invisibility gives people the courage to go places and do things that they otherwise wouldn't. Even with everyone's identity known, the opportunity to be physically invisible amplifies the disinhibition effect. People don't have to worry about how they look or sound when they type a message.

The lack of physical presence can limit the depth of intimacy experienced, as physical touch and shared physical experiences are integral to complete intimacy. For sexual minority men, emotional intimacy is cultivated through vulnerability, empathy, and trust; these can be enhanced through digital communication but still lack the physical dimension that in-person interactions provide.

C. Power and Control

While online a person's status in the face-to-face world may not be known to others and may not have as much impact. Authority figures express their status and power in their dress, body language, and in the trappings of their environmental settings. The absence of those cues in the text environments of cyberspace reduces the impact of their authority. In many environments on the Internet, everyone has an equal opportunity to voice him or herself. Everyone - regardless of status, wealth, race, or gender - starts off on a level playing field.

People are reluctant to say what they really think as they stand before an authority figure. A fear of disapproval and punishment from on high dampens the spirit. But online, in what feels more like a peer relationship - with the appearances of authority minimized - people are much more willing to speak out and misbehave.

D. Frustration and Aggression

The negative online disinhibition effect is the concept used to refer to the negative results of this loss of inhibitions, usually manifested in aggressive behaviors that apparently would not be exhibited in a similar scenario in the 'real world.' Cyberbullying has resulted in significant psychological and social impact for both victims and bullies. There are several publicized suicides which involved bullies attacking through cyberbullying. Cyberbullying has been identified as a contributing factor for depression and suicide of victims as well as feelings of anger by victims.

Emotional dependence and digital addiction are intricately linked, particularly when anonymous chatting becomes a coping mechanism. Emotional dependency, particularly in adolescents, correlates with internet and mobile abuse, leading to psychological consequences such as anxiety, depression, and low self-esteem.

E. Desire for Attention and Validation

If we combine the opportunity to easily escape or dissociate from what happens online with the psychological process of creating imaginary characters, we get a somewhat different force that magnifies disinhibition. Consciously or unconsciously, people may feel that the imaginary characters they 'created' exist in a different space, that one's online persona along with the online others live in a make-believe dimension, separate and apart from the demands and responsibilities of the real world.

Absent face-to-face cues combined with text communication can alter self-boundaries. People may feel that their mind has merged with the mind of the online companion. Reading another person's message might be experienced as a voice within one's head, as if that person's psychological presence and influence have been assimilated or introjected into one's psyche.

IV. ROLE OF SOCIAL MEDIA PLATFORMS

A. Algorithmic Amplification

Anonymous chat apps and platforms, such as Reddit, Soul, Omegle, and AI-driven chatbots like FASSLING, play a significant role in facilitating private and secure communication. Reddit, a popular platform, allows users to engage in discussions and share content anonymously, fostering open dialogue without revealing personal identities. Soul App is a social networking platform designed to foster anonymous yet emotionally engaging conversations and sharing moments of life. Omegle connects users randomly for anonymous text or video chats, emphasizing spontaneity and anonymity in interactions.

Forum posts are important since they appear in a variety of attacks and have features which may increase the damage of an attack such as anonymity, permanency - one post can be viewed many times - and public visibility, where public viewing allows other individuals to join an attack.

B. Platform Design and User Behaviour

In e-mail and message boards, communication is asynchronous. People don't interact with each other in real time. Others may take minutes, hours, days, or even months to reply. Not having to cope with someone's immediate reaction disinhibits people. In a continuous feedback loop that reinforces some behaviors and extinguishes others, moment-by-moment responses from others powerfully shapes the ongoing flow of self-disclosure and behavioral expression, usually in the direction of conforming to social norms. In e-mail and message boards, where there are delays in that feedback, people's train of thought may progress more steadily and quickly towards deeper expressions of benign and toxic disinhibition that avert social norms.

Cyberspace allows users to control their level of social presence and, thus, their degree of intimacy; that is, to control the amount and duration of visibility or invisibility during social interactions through the use or non-use of pictures, webcam, and videos.

Thus, the absence of visibility and the unique form of social presence inherent in online communication accelerate the processes of online disinhibition because of reduced visibility-originated interpersonal communication cues.

C. Online Communities and Echo Chambers

Cyberbullying has resulted in significant psychological and social impact for both victims and bullies. Victims of cyberbullying may feel distress as the result of online interactions and cell phone text messages which threaten or embarrass the victim. Victims can become very distressed in certain situations such as youth being cyberbullied, adults cyberbullying children, requests for pictures, or bullies attacking online coupled with offline contact. In addition to victims, cyberbullies are also at risk for psychosocial problems such as poor parent/child relationships, substance use, delinquency and depression.

This tension is further explored in online communities like r/NoSurf, where individuals discuss the paradox of digital disconnection, highlighting the ambivalent relationships people have with digital media and the complex ways in which disconnection can support digital well-being. In the realm of digital dating, particularly for gay and bisexual men, technology facilitates social and sexual intimacy but also contributes to a sense of disconnected connectedness, where the abundance of virtual interactions fails to satisfy the need for genuine, live communication.

V. CONSEQUENCES OF TOXIC ONLINE BEHAVIOUR

A. Impact on Individuals

Cyberbullying has been identified as a contributing factor for depression and suicide of victims as well as feelings of anger by victims. Victims of cyberbullying may feel distress as the result of online interactions and cell phone text messages which threaten or embarrass the victim.

Anonymity can negatively impact psychological well-being, particularly in the context of social media. It has been found to moderate the relationship between the Fear of Missing Out (FoMO) and psychological well-being, where individuals with anonymous accounts experience a stronger negative impact of FoMO on their well-being. Furthermore, anonymity can lead to a reduction in the influence of social norms, resulting in more unfair or socially undesirable behaviors, as demonstrated in experiments where anonymous environments led to less compliance with social norms.

Ghosting, characterized by the sudden and unexplained termination of communication, has significant psychological impacts on individuals across various contexts, including personal, professional, and legal domains. In digital communications, ghosting disrupts effective communication strategies, leading to decreased psychological safety and adversely affecting well-being, resilience, and satisfaction.

B. Impact on Online Communities

The toxic aspect of disinhibition has been shown to cause various problematic behaviors in cyberspace, such as flaming and a negative atmosphere. Online disinhibition is a basic, wide-spread effect, associated with a wide range of behaviors that have been observed among many people engaged in online activities and interactions. Accumulated research shows that disinhibition significantly affects modes of discussion, information-searching behavior, online learning, interpersonal contacts and relationships, and behavior in groups.

Despite the digital age's promise of enhanced connectivity, individuals often experience a paradoxical sense of isolation and emotional distance. This phenomenon, termed 'digital loneliness,' is particularly pronounced among vulnerable populations such as older adults and people with disabilities, who face unique challenges in navigating digital spaces.

C. Societal Impact

Catfishing, a form of online deception where individuals create fabricated identities, poses significant dangers, particularly in the realm of online relationships. The motivations behind catfishing are diverse, including entertainment, emulating an ideal self, seeking meaningful interaction, and financial gain. This deceptive practice is often employed in romantic scams, exploiting the digital space's unique tools to create alternative forms of intimacy. The psychological impact on victims is profound, with emotions such as suspicion, love, depression, anger, embarrassment, and feelings of stupidity being commonly reported.

The effect of this dissociative imagination surfaces clearly in fantasy game environments in which a user consciously creates an imaginary character, but it also can influence many dimensions of online living.

For people with a predisposed difficulty in distinguishing personal fantasy from social reality, the distinction between online fantasy environments and online social environments may be blurred. In our modern media-driven lifestyles, the power of computer and video game imagination can infiltrate reality testing.

VI. STRATEGIES TO REDUCE TOXIC BEHAVIOUR

A. Platform Moderation

An automated tool to detect cyberbullying may help reduce the occurrence of cyberbullying or negative impacts associated with cyberbullying. The automated tool should analyze, for example, a set of forum posts and produce whether cyberbullying occurs in the forum posts. In order to facilitate the detection of cyberbullying, the automated tool should provide annotated forum posts which most strongly indicated the potential for cyberbullying. One approach to identify cyberbullying is to build a classifier to identify patterns in the set of forum posts which match patterns from examples of cyberbullying.

Technology can facilitate safer and more meaningful anonymous intimacy by integrating advanced privacy-preserving measures and ethical AI development. End-to-end encryption, decentralized identity management, and zero-knowledge proofs can enhance security and ensure that users maintain control over their digital footprints while engaging in intimate anonymous exchanges. Additionally, AI-driven moderation systems can detect and prevent harmful interactions without compromising user anonymity.

B. Digital Literacy and Awareness

Cyberbullying has been associated with name-calling or verbal attacks in textual messages, referred to as 'aggressive forum posts.' One possibly relevant feature is the use of profanity in a forum post which may indicate that the forum post is aggressive. Another feature is the target of an aggressive forum post. The forum post may be an aggressive attack against a victim or an aggressive defense of the victim. Cyberbullying is also described as often being anonymous. Thus, another relevant feature is whether an attacker uses an identifier or has some degree of anonymity.

Understanding cyberbullying in forum posts may help build an understanding of cyberbully attacks and provide a step towards modeling cyberbully attacks. Analysis of forum posts is also a promising direction since there are a variety of automated techniques for identifying features of text contained in forum posts.

C. Identity Verification Systems

The present findings suggest that one can think of anonymity as an assemblage of different levels of online unidentifiability, in which non-disclosure of personal details, invisibility, and absence of eye-contact compose the most significant assemblage; these components appear to combine in different degrees, thus yielding a variety of 'anonymities.' The new concept we refer to - online sense of unidentifiability - can be understood as spanning a range, in which three major factors are considered.

A key aspect of improving anonymous intimacy is balancing anonymity with accountability. Current anonymous platforms struggle with trust-related issues due to a lack of verification mechanisms. Technology could introduce pseudonymous reputation systems, where users build credibility over time based on positive interactions while maintaining anonymity. This could encourage safer and more genuine connections without the risks of deception or catfishing.

D. AI and Automated Moderation Tools

Text-based therapy and crisis support services, particularly those utilizing AI-based chatbots, have shown promising results in addressing mental health issues such as depression and anxiety. AI chatbots provide a cost-effective and accessible means of therapy, overcoming geographical and time constraints, and have demonstrated significant short-term improvements in depressive and anxiety symptoms.

AI chatbots, employing techniques such as sentiment analysis and emotion recognition, provide a secure environment for users to express their emotions and receive tailored advice, thereby enhancing the therapeutic experience and directing users to appropriate resources. Overall, text-based therapy and crisis support services, particularly those enhanced by AI, offer a valuable supplement to traditional mental health care, improving accessibility and engagement for diverse populations.

The current findings suggest that previous definitions of anonymity did not take into account all the factors that characterize the online communication environment, specifically invisibility and absence of eye-contact. Thus, it seems advisable that future studies define the online social setting carefully and precisely so that the effects of anonymity on the behavior of communicants in cyberspace can be evaluated alongside the effects of other online situational variables.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

VII. DISCUSSION

The results of the study show that eye-contact has a significant main effect on negative online disinhibition, as well as on self-reported flaming incidents and threats. When eye-contact was enforced, participants experienced fewer occurrences of flaming and fewer threats than when there was no eye-contact. Anonymity, however, produced a significant main effect for one measure only (i.e., threats). Similarly, visibility exerted a significant main effect on a single measure only (i.e., negative atmosphere). Of the interactive effects, four were significant, all involving eye-contact.

Contrary to the widely accepted belief, the research found that it was possible that the eye-contact factor impacted more strongly on the online disinhibition effect, not the anonymity factor. One possible explanation for this finding may be rooted in the differences between previous definitions of online anonymity and the definition used in the current study. Anonymity implies appearing in public without an associated name or pseudonym. Derlega and colleagues (1993) suggested that anonymity was equivalent to a state of unidentifiability. Thus, the meaning of anonymity changes according to context.

The disinhibition effect can then be understood as the person shifting, while online, to an intrapsychic constellation that may be, in varying degrees, dissociated from the in-person constellation, with inhibiting guilt, anxiety, and related affects as features of the in-person self but not as part of that online self. This constellations model, which is consistent with current clinical theories regarding dissociation and information processing, helps explain the disinhibition effect as well as other online phenomena, like identity experimentation, role playing, multitasking, and other more subtle shifts in personality expression as someone moves from one online environment to another.

The self does not exist separate from the environment in which that self is expressed. If someone contains his aggression in face-to-face living, but expresses that aggression online, both behaviors reflect aspects of self: the self that acts non-aggressively under certain conditions, the self that acts aggressively under other conditions. When a person is shy in person while outgoing online, neither self-presentation is more true. They are two dimensions of that person, each revealed within a different situational context.

Anonymous intimacy offers a paradoxical mix of freedom and vulnerability. It enables uninhibited self-disclosure, emotional release, and meaningful connections untainted by prejudice. However, it also introduces risks such as deception, emotional dependency, and diminished real-world social skills. The challenge lies in designing digital spaces that preserve the liberating aspects of anonymity while mitigating its dangers. The balance between connection and isolation is especially pertinent in the age of AI-driven companionship.

VIII. RESEARCH METHODOLOGY

A. Research Design

This literature review adopts a systematic and integrative approach to exploring the psychological mechanisms, risks, and potential of anonymous online behaviour in the digital age. Given the interdisciplinary nature of the topic, which intersects psychology, communication studies, and technology ethics, the review synthesizes empirical and theoretical studies to provide a multidimensional understanding. By drawing from established research on digital intimacy, self-disclosure, trust, and online disinhibition, this study is structured around key psychological frameworks such as social penetration theory, the online disinhibition effect, and attachment theory. These frameworks help to analyze how anonymity influences interpersonal relationships and the potential ethical concerns associated with anonymous digital interactions.

B. Data Collection

The literature search was conducted systematically across multiple academic databases, including Google Scholar, PubMed, PsycINFO, ScienceDirect, and IEEE Xplore. A combination of relevant keywords and Boolean search operators ensured a comprehensive retrieval of peer-reviewed sources. Keywords such as 'anonymous chatting,' 'digital intimacy,' 'self-disclosure online,' 'online disinhibition effect,' 'trust and vulnerability in digital communication,' 'AI chatbots and virtual companionship,' 'cybersecurity and privacy in online interactions,' 'catfishing and online deception,' 'emotional dependency and digital addiction,' and 'therapeutic applications of anonymity' were strategically used to refine search results. The inclusion timeframe was set to cover publications from the past two decades (2004-2024), ensuring a balance between foundational theories and recent advancements in the field.

C. Inclusion and Exclusion Criteria

To maintain academic rigor and relevance, strict inclusion and exclusion criteria were applied. Studies considered for inclusion were limited to peer-reviewed journal articles, conference proceedings, and academic book chapters that explored digital intimacy, online anonymity, trust formation, and associated psychological risks. Research specifically examining AI-driven chatbots in anonymous intimacy was also prioritized. In contrast, non-academic sources such as blog posts, opinion pieces, and non-peer-reviewed articles were excluded. Additionally, studies solely focused on online harassment or cybersecurity without clear relevance to digital intimacy were omitted, as were those addressing anonymity in non-digital contexts unless they offered insights applicable to the digital realm.

D. Experimental Approach

The present research studied the impact of three typical online communication factors on inducing the toxic online disinhibition effect: anonymity, invisibility, and lack of eye-contact. Using an experimental design with 142 participants, random pairs of participants were presented with a dilemma for discussion and a common solution through online chat. The effects were measured using participants' self-reports, expert judges' ratings of chat transcripts, and textual analyses of participants' conversations. A 2 x 2 x 2 factorial MANOVA was employed to analyze the findings.

Participants were recruited by means of advertisements through online and offline bulletin boards that specified the time required and the compensation offered. All participants volunteered for a 1-hour study. The study included 142 participants (71 men and 71 women), all of them students enrolled in a variety of majors and degree programs at two Israeli academic institutions. Their ages ranged from 18 to 34 years. Forum posts were also labeled in an online forum for cyberbullying-related features, and text matching techniques were applied to identify aggressive posts including both attacks and defends.

E. Thematic Analysis

Following the selection process, a thematic analysis was employed to categorize and synthesize findings. The literature was divided into key thematic areas, including psychological mechanisms of anonymous intimacy, which explored self-disclosure, vulnerability, and trust formation; risks associated with digital anonymity, which examined deception, cyberbullying, emotional dependency, and digital addiction; the role of AI and chatbots in intimacy, focusing on human-chatbot interactions and their psychological implications; and therapeutic applications of anonymity, assessing how anonymous platforms facilitate emotional support and mental health interventions. Each study was systematically evaluated based on research methodology, sample size, theoretical framework, and key findings, allowing for a comparative analysis to identify patterns and discrepancies across different sources.

F. Ethical Considerations

Ethical considerations were carefully addressed despite this being a literature-based study rather than an empirical one involving human participants. The review ensured that all included research adhered to ethical standards, particularly studies involving human subjects. Discussions on digital privacy and online ethics were critically analyzed to highlight concerns surrounding trust, anonymity, and deception in digital interactions. The ethical implications of AI-driven intimacy and chatbot companionship were also examined, particularly in relation to user emotional dependency and data privacy concerns. All participants in the experimental component freely volunteered and were ensured full confidentiality of the information collected during the research.

IX. CONCLUSION

As demonstrated, the most notable outcomes of this study are the inclusion of variables that refine our understanding of the conditions associated with negative online disinhibition. Lack of eye-contact plays a major role in triggering behaviors related to negative online disinhibition. Thus, the professional literature's overwhelming attention to the anonymity factor appears to be misplaced: the definition of anonymity reflected in past studies and theoretical conceptualizations may, perhaps, be too general. The current study suggests employing a new concept: Online Sense of Unidentifiability. This term is broader than anonymity, yet it includes specific components, namely, non-disclosure of personal data, invisibility, and lack of eye-contact.

Cyberbullying is a growing concern as the usage of the internet is growing among youth. One step towards identifying cyberbullying in online forums is the identification of features in forum posts which may be related to cyberbullying. The authors applied an automated technique to label aggressiveness and anonymity of forum posts. The authors also manually labeled the aggressiveness and role of the forum posts, i.e., attack, defend, and neutral.

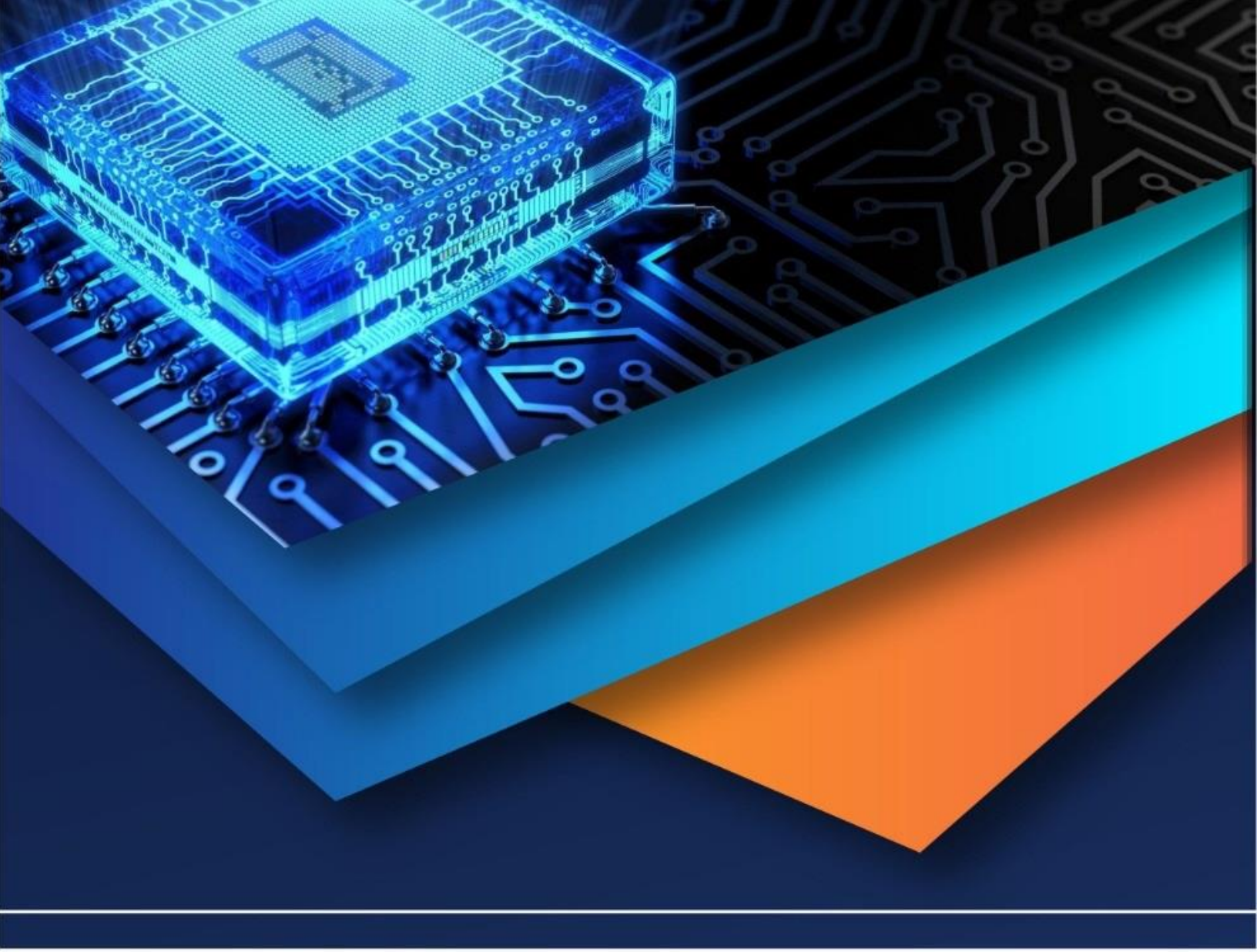
Ultimately, the future of anonymous digital interactions depends on our ability to leverage technology responsibly. By prioritizing security, ethical AI development, and psychological well-being, we can cultivate a digital landscape where anonymous interactions enrich, rather than erode, human connection. While anonymous chatbots and virtual companions can provide solace, they should complement rather than replace human relationships. Ethical considerations must guide the development of these technologies to ensure they support genuine emotional well-being rather than fostering escapism or reinforcing social disconnection.

Different modalities of online communication, such as e-mail, chat, and video, and different environments, such as social, vocational, and fantasy, may facilitate diverse expressions of self. Each setting allows us to see a different perspective on identity. Neither one is necessarily more true than another. Based on a multidimensional analysis of the various psychological features of online settings, a comprehensive theory on the psychotherapeutics of cyberspace can explore how computer-mediated environments can be designed to express, develop, and if necessary, restrain different constellations of self-structure.

The study comes to the conclusion that creating safer and more productive digital environments requires a balanced strategy that protects the legal liberties provided by anonymity while imposing accountability.

REFERENCES

[1] Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. Computers in Human Behavior, 28, 434–443. https://doi.org/10.1016/j.chb.2011.10.014
[2] Moore, M. J., Nakano, T., Enomoto, A., & Suda, T. (2012). Anonymity and roles associated with aggressive posts in an online forum. Computers in Human Behavior, 28, 861–867. https://doi.org/10.1016/j.chb.2011.12.005
[3] Suler, J. (2004). The online disinhibition effect. CyberPsychology & Behavior, 7(3), 321–326.
[4] Zhu, Y. (2025). Anonymous intimacy in the digital age: Psychological mechanisms, risks, and potential. Proceedings of the International Conference on Advanced Research in Social Sciences, 2(1), 11–29. https://doi.org/10.33422/icarss.v2i1.1056
[5] Joinson, A. N. (1998). Causes and implications of disinhibited behavior on the Internet. In J. Gackenbach (Ed.), Psychology and the internet (pp. 43–60). Academic Press.
[6] Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. American Psychologist, 39, 1123–1134.
[7] Patchin, J. W., & Hinduja, S. (2006). Bullies move beyond the schoolyard: A preliminary look at cyberbullying. Youth Violence and Juvenile Justice, 4(2), 148–169.
[8] Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. Nebraska Symposium on Motivation, 17, 237–307.
[9] Suler, J. (2004). The Online Disinhibition Effect. CyberPsychology & Behavior.
[10] Lapidot-Lefler, N. & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. Computers in Human Behavior.
[11] Moore, M.J. et al. (2012). Anonymity and roles associated with aggressive posts in an online forum. Computers in Human Behavior.
[12] Zhu, Y. (2025). Anonymous Intimacy in the Digital Age: Psychological Mechanisms, Risks, and Potential. ICARSS.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)