*Environment and Social Psychology (2025) Volume 10 Issue 3* ISSN: 2424-8975 (O)
doi: 10.59429/esp.v10i3.3554 2424-7979 (P)

**RESEARCH ARTICLE**

# Psychological harm induced by generative AI and its legal remedies

**Xinbo Huang[1,3], Mohd Zakhiri bin Md. Nor [2], Zuryati Mohamed Yusoff[3], Zhaowei Liang[4*]**

[1] *School of Law, Nanchang Institute of Technology, Nanchang, 330044, China*

[2]*Islamic Business School, University Utara Malaysia, 06010 UUM, Sintok, Kedah, Malaysia*

[3]*School of Law, University Utara Malaysia, 06010 UUM, Sintok, Kedah, Malaysia*

[4] *School of Law, Guilin University, Guilin, 541006, China*

*** Corresponding author:** Zhaowei Liang, Maxine09@126.com

## ABSTRACT

The rapid advancement of generative artificial intelligence (AI) has transformed human interaction with technology, raising critical concerns about its psychological impact. While AI-driven tools offer unprecedented convenience and efficiency, they also pose risks to mental well-being. This study adopts a qualitative research approach, integrating case analysis and library-based data analysis. Thematic and content analysis methods are employed to investigate the operational mechanisms of generative AI and its direct and indirect effects on human psychology. The findings reveal that AI-generated psychological harm is difficult to quantify, often subjective, and influenced by the evolving nature of human-AI interactions. Furthermore, the study identifies gaps in existing legal frameworks and highlights the complexities of attributing liability in cases of AI-induced psychological damage. This research proposes legal remedies encompassing emotional distress compensation, regulatory oversight, preventative standards, and AI liability insurance. By bridging the gap between AI innovation and mental health protection, this study contributes to the growing discourse on AI governance and provides a foundation for future legal and policy developments to safeguard individuals from AI-induced psychological harm.

***Keywords:*** artificial intelligence (AI); generative AI; psychological harm; legal remedies

## 1. Introduction

With the advent of ChatGPT, there has been a surge in discussions about AI, and Generative AI. Generative AI encompass various subsets of AI models designed to produce content akin to human-created text and other media.[1] These models include those for generating text, images, videos, and audio. They represent a significant technological advancement, profoundly impacting political, economic, and social systems. For example, IBM Watson Health uses Generative AI to analyse medical images and records, improving diagnostic accuracy. In finance, AI models aid in assessing risks and predicting market trends, leading to more informed investment decisions. Generative AI models are increasingly integrated into various sectors including education, healthcare, finance, transportation, entertainment, and manufacturing, enhancing productivity and creativity.

Moreover, AI technologies, particularly generative AI, are revolutionizing the landscape of mental

**ARTICLE INFO**
Received: 1 March 2025| Accepted: 17 March 2025 | Available online: 28 March 2025

**CITATION**
Huang XB, Nor MZBM, Yusoff ZM, et al. Psychological harm induced by generative AI and its legal remedies. *Environment and Social Psychology* 2025; 10(3): 3554. doi:10.59429/esp.v10i3.3554

health services. These innovations enable a new era of accessibility and efficiency in mental health care. Intelligent chatbots, for instance, provide users with round-the-clock emotional support, offering an immediate response to those in need of assistance.[2] Additionally, virtual reality (VR) technology, integrated with AI, is creating immersive therapeutic environments that help users confront and manage psychological challenges. AI-powered analytics systems, fueled by big data, are capable of detecting potential mental health risks by analyzing vast amounts of information, identifying patterns and signs that may otherwise go unnoticed. Moreover, AI systems can assess a user's emotional state in real time by analyzing multi-modal data, such as voice, text, and facial expressions, something that traditional psychological counseling methods struggle to achieve. AI assistants influence consumer behavior and decision-making by leveraging their advantages to enhance user engagement and emotional disclosure, thereby increasing the willingness for users to reuse them.[3]

Despite these advancements, the increasing role of AI in mental health services brings about a range of ethical concerns. Privacy and data security are of paramount importance, as mental health data is highly sensitive. Ensuring that this information is protected from misuse or breaches is crucial.[4] Generative AI may also inadvertently disclose users' privacy information in the process of generating content, such as exposing users' identity, location and other sensitive information through generated text or images. This risk of data breach not only damages the privacy rights of individuals but may also pose a threat to the security and stability of society. Another critical issue is the attribution of responsibility. When AI systems make erroneous judgments or provide harmful recommendations that lead individuals to take inappropriate or dangerous actions, determining accountability becomes a complex challenge. The lack of transparency and interpretability in the "algorithmic black box" leads to generative AI developers or service providers being unable to fully control or intervene in the output content and process. Consequently, it becomes particularly challenging to determine the tort liability of generative AI developers or service providers. Furthermore, the growing reliance on AI in addressing emotional problems raises concerns about the boundaries of human-computer interaction. The potential for users to entrust their emotional well-being to machines may foster further alienation in human relationships, while also encouraging dependency on technology, thereby diminishing authentic emotional connections between individuals.

These ethical dilemmas give rise to a pressing concern: the psychological harm that may result from interactions with Generative AI systems. While AI offers numerous benefits, the risks it poses to mental health, such as emotional distress, manipulation, and dependency, cannot be ignored. While algorithmic control can improve service performance and operational efficiency, it also poses risks to work well-being by reducing autonomy and causing emotional distress.[5] Users subjected to continuous algorithmic monitoring or control may experience heightened psychological stress, anxiety, and diminished mental health. The constant surveillance can erode individuals' sense of privacy and autonomy, leading to emotional distress, burnout, and a weakened sense of self-efficacy. A significant instance featured a deepfake image of Taylor Swift that accumulated 47 million views before its removal. This exposure can result in enduring emotional effects on victims, especially adolescents who may encounter altered images that skew their self-perception. Cognitive neuroscientist Joel Pearson highlighted that exposure to deepfakes might modify perceptions of reality and lead to enduring psychological suffering, particularly in teenagers whose brains are still maturing.[6] In the same vein, a survey by the American Psychological Association (APA) indicated that 38% of U.S. workers are concerned that AI may render their job responsibilities obsolete. This anxiety has considerable implications for mental health. Fifty-one percent of individuals apprehensive about Generative AI indicated that their employment adversely affects their mental well-being. While 64% typically feel tense or stressed during the workday due to these worries.[7]

It is argued that individuals' reliance on AI for cognitive tasks may potentially diminish their capacity for critical thinking.[8] As users offload mental tasks to AI, they may experience a reduction in their cognitive engagement, which can lead to a diminished ability to critically assess information and make independent decisions. This phenomenon raises concerns about the long-term effects of AI on mental autonomy and critical thinking skills.

Addressing these concerns necessitates not only a deeper understanding of the potential harms AI can cause but also the development of effective legal frameworks to provide appropriate remedies. The need for legal oversight and regulatory measures to safeguard individuals from AI-induced psychological harm is becoming increasingly urgent. As AI technologies continue to evolve and permeate various aspects of life, it is essential to implement comprehensive legal strategies to protect users' mental health, ensure ethical AI use, and mitigate risks associated with its application in sensitive areas such as mental health care. This study critically analyses the psychological harm of Generative AI and provides legal remedies for individuals harmed by Generative AI systems. The purpose of this paper is achieved by exploring the following research questions:

Q1: What are the operational mechanisms of Generative AI?

Q2: What specific psychological problems can generative AI cause?

Q3: How can psychological harm caused by Generative AI be remedied?

Accordingly, the research objectives are as follows: (1) To explore the operational mechanisms of Generative AI; (2) To identify the unique psychological harm posed by Generative AI; (3) To propose legal and regulatory remedies for AI-induced psychological harm.

Based on the above research questions and objectives, the research framework of this paper is as follows: First, the authors conduct an analysis of the operational mechanisms and distinctive properties of Generative AI. This includes using ChatGPT as a case study to clarify the specific operational mechanisms. Second, this paper provides a detailed assessment of the psychological impacts of Generative AI, focusing on the potential harm it may cause. Additionally, the authors evaluates the legal remedies available for individuals affected by psychological harm from Generative AI. It investigates how such harm can be proven in a legal context and explores existing regulatory frameworks. This study calls for enhanced regulatory measures and the development of legal mechanisms to ensure appropriate compensation and safeguard individuals' psychological well-being in the face of rapidly advancing AI technologies.

# 2. Methodology

This section outlines the research design, data sources and data analysis techniques employed in this study, which aims to explore the the psychological harm caused by Generative AI and legal remedies.

## 2.1. Research design

This study adopts a qualitative research design, which is particularly suited to exploring complex, emerging issues such as the intersection of Generative AI technology and psychological issues. The research is structured around two primary approaches: literature review and case study analysis. A comprehensive review of scholarly articles, legal texts, policy documents and other relevant publications on psychological topics of Generative AI forms the foundation of the research. This method enables a thorough understanding of the existing theoretical frameworks, regulatory approaches and challenges in the field. Real-world cases related to Generative AI technologies are analysed to provide practical insights into the psychological issues

raised by Generative AI. The case studies are selected to illustrate the psychological challenges that Generative AI presents.

### 2.2. Data sources

The data for this study are primarily sourced from secondary materials. Academic literature is used, such as academic articles, law reviews, policy papers and books. This includes both theoretical and empirical research that addresses topics like psychological challenges in the context of Generative AI. Case studies are used to explain what Generative AI is and how it works. In addition, real-world cases are used to show the psychological challenges that Generative AI presents. Reports from governmental and non-governmental organizations contribute to understanding the ethical considerations surrounding Generative AI. Lastly, industry reports from law firms, tech companies and industry groups offer insights into the implications of Generative AI for business and legal compliance.

### 2.3. Data analysis

This study employed a qualitative methodology combining content analysis and thematic analysis to examine the psychological and legal implications of Generative AI. An initial pool of 86 academic sources was identified using databases such as Scopus, Web of Science, SSRN, and Google Scholar. Inclusion criteria were Publications from 2018 to 2025; Focus on the psychological, emotional, or legal effects of Generative AI; Empirical studies, theoretical legal analyses, case reports, or regulatory policy discussions; Published in peer-reviewed journals, edited books, or authoritative legal reports. We used specific search terms such as "Generative AI," "mental health," "emotional distress," and "legal regulation of AI" to gather relevant articles, case studies, and legal frameworks. A systematic review was conducted to ensure only the most relevant and up-to-date literature was included. After screening abstracts, removing duplicates, and assessing full-text relevance, 42 sources were selected for final analysis.

Content analysis was the first step, involving a systematic categorization of the selected literature. We focused on identifying recurring themes or keywords in the literature, such as "psychological harm," "AI impacts," and "legal regulation." Each article or case study was reviewed, and key findings were extracted and categorized based on their relevance to mental health and legal issues. This helped form the foundational data for the thematic analysis. Content analysis begins with a comprehensive collation and coding of the data in the literature database to identify keywords, concepts, and specific examples related to the psychological challenges of Generative AI.[9] These documents may cover various fields such as technological progress, industry practices, policies and regulations and academic research. Through the in-depth interpretation and systematic arrangement of the literature, the key information can be extracted and classified and summarized. In practice, qualitative data coding is a widely used technique. This approach classifies the literature according to different dimensions.

Following content analysis, thematic analysis was used to identify deeper patterns or themes in the data. A coding protocol was developed, which involved tagging segments of text with labels (codes) that correspond to specific ideas or issues, such as "psychological harm," "AI regulation," and "legal remedies". Themes were reviewed and refined through multiple iterations to ensure they were comprehensive and accurately reflected the data. For example, the theme related to "AI regulation" was expanded to include the need for more flexible, forward-looking legal frameworks. We reached thematic saturation when no new themes emerged from the data. This was determined by continuing the analysis until further reading did not yield new insights or information. Thematic analysis is the process of further deepening research on the basis of content analysis. By subject-generalization of the encoded data, we can reveal the multifaceted impact of Generative AI on mental health. In addition, thematic analysis focuses on how to keep the legal regulatory

system in step with technological developments. This includes assessing whether existing legal frameworks are adequate to address new issues raised by Generative AI, and how to design more flexible and forward-looking regulatory mechanisms.

In summary, from content analysis to thematic analysis, this research methods not only help to deeply understand the technical characteristics of Generative AI and its potential impact on the mental health, but also lay a solid foundation for the construction of a scientific and reasonable legal remedies system. This is of great significance for promoting the healthy development of technology and safeguarding social and public interests.

### 2.4. Ethical considerations

This study ensures ethical integrity by adhering to key principles. It respects confidentiality by carefully handling sensitive cases on mental health or unpublished reports. To maintain objectivity, a balanced selection of sources from diverse academic literature is used, minimizing bias and offering a fair representation of the impact of Generative AI on mental health. As the study relies on publicly available secondary data, informed consent is not applicable, but future research involving primary data will follow strict ethical standards, ensuring transparency and voluntary participation. This study maintains data integrity by accurately analysing and referencing sources.

## 3. Results and discussion

### 3.1. About generative AI

Generative AI are competent in generating texts, images, videos, and audio content that rival those produced by humans, demonstrating sophisticated language comprehension and creativity. Currently, Generative AI is approaching the capabilities of strong AI and is in a transitional phase between weak and strong AI. However, it has not yet fully evolved into a truly strong generative AI system.[10] The House of Lords Select Committee on AI, which differentiates among 'narrow', 'general', and 'super' AI.[11] This categorization is also evident in the EU definition.[12] Ioannis Revolidis and Alan Dahi associate 'narrow AI' with 'weak AI' and 'general AI' with 'strong AI'.[13] Narrow AI denotes systems that possess capabilities akin to particular human tasks.[14] Examples include online language translation services and virtual assistants such as SIRI. The designation "weak" AI may be deceptive, as it does not suggest restricted ability but rather signifies expertise in executing particular jobs proficiently.[15] Whereas general AI signifies a more sophisticated level, characterized by the ability to establish and accomplish its own objectives and execute tasks in a manner reminiscent of human intellect.[16] While super AI, this hypothetical variant of AI exceeds human intelligence and capacities, potentially executing tasks with superior proficiency compared to a human expert. It encompasses sophisticated capabilities including autonomous learning, strategic planning, communication, and decision-making.

Notable examples include OpenAI's Chat GPT, a versatile conversational bot, and Stable Diffusion, which generates high-quality images from textual descriptions. Furthermore, OpenAI's recent launch of Sora, a text-to-video model, exemplifies the cutting-edge capabilities of these technologies, blurring the boundaries between reality and generated content. Generative AI include various subsets of Generative AI models that create human-like text and other media. **Table 1** below summarizes the current popular generative AI models, including Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Diffusion Models, Transformers, and Neural Radiance Fields (NeRFs).[17]

**Table 1**. Operational mechanisms and applications of generative AI.

| Models | Applications | Operational Mechanisms |
|---|---|---|
| VAEs | VAEs are used in image and video generation, natural language processing, and sound synthesis, with applications in finance, speech/audio separation, and bio signals. | VAEs compress input data into a lower-dimensional latent space and then reconstruct it using an encoder, latent space, and decoder.[18] |
| GANs | The creation of realistic images and videos benefits industries like entertainment and advertising but also opens the door to malicious uses, such as deepfakes. | GANs use a generator to create realistic data and a discriminator to distinguish between real and generated data.[19] |
| Diffusion | Diffusion models are used in computer vision, time series prediction, natural language processing, text-based multimodality, and advancing the basic science of AI. | A probabilistic generative models add Gaussian noise to data progressively and then use a series of denoising steps to reverse this process, generating new data samples.[20] |
| Transformers | Transformers are used in computer vision, natural language processing, temporal data modeling, multi-modal learning, robust learning, and interdisciplinary applications. | A Transformer model uses self-attention to assign weights to input data and generates output by computing a weighted sum, with weights based on the compatibility between queries and keys. |
| NeRFs | NeRFs are used in 3D editing, medical 3D image reconstruction, and neural scene representations for world mapping, with potential in robotics, autonomous navigation, and the metaverse. | Employing multi-layer perceptrons (MLPs) to synthesise novel views of 3D scenes by learning both their geometry and lighting characteristics.[21] |

This paper uses ChatGPT as an example to explore its technical principles and applications, aiming to identify potential risks. ChatGPT is an advanced large language model (LLM) developed by OpenAI. The model is trained on extensive and diverse text corpora, enabling it to generate text that is remarkably human-like in structure and content.[22] Utilizing deep learning architectures, particularly the transformer model, ChatGPT excels at understanding and maintaining conversational context, thereby facilitating interactions that closely mimic human communication.[23] Its adaptability and efficacy have resulted in its utilization across a broad spectrum of domains, encompassing education, manufacturing, cultural and entertainment services, social governance.

According to OpenAI, the operational mechanism of ChatGPT involves a multi-stage process that optimizes text generation through a combination of supervised learning and reinforcement learning. This process can be delineated into four distinct stages as following: (1) Pre-Training on Large Text Corpora. The initial stage involves training the model on extensive datasets consisting of diverse textual content. The purpose of this phase is to enable the model to learn linguistic patterns, contextual understanding, and text generation capabilities. During this phase, the model's performance is evaluated by comparing its outputs to the subsequent text within the corpus, facilitating iterative refinement of its generative abilities. (2) Human Feedback Integration. In the subsequent stage, human-generated responses to specific prompts are collected. This human feedback serves as a critical component in fine-tuning the model. Researchers provide ChatGPT with question-answer pairs derived from human responses, which helps the model adjust its outputs to better align with human expectations and preferences. This stage is crucial for guiding the model towards generating more contextually appropriate and human-like responses. (3) Reward Model Training. This stage involves generating multiple responses from ChatGPT to a set of predefined questions. These responses are then evaluated and ranked by human reviewers based on their quality and relevance. The ranking data is used to train a reward model that reflects human evaluative criteria. This reward model acts as a benchmark for assessing the performance of the AI in generating desirable responses. (4) Reinforcement Learning for Optimization: The final stage employs reinforcement learning techniques to further optimize the model. The reward model developed in the previous stage provides feedback that guides the reinforcement learning algorithms. Through iterative interactions, the model adjusts its output generation strategies to maximize alignment with the reward criteria. This phase enhances the model's ability to self-optimize and adapt based on performance feedback, leading to improved text generation over time.[24]

### 3.2. Case study: Psychological issues caused by generative AI

Nowadays, AI algorithms have infiltrated almost every aspect of our lives, and social media platforms are no exception. Although there are many advantages to using these platforms, it can also harm mental health. The algorithms of these platforms can often amplify misinformation, polarize users, and act as a channel for online harassment. Such conditions may result in escalated anxiety, stress and isolation, with significant psychological implications. Also, being constantly exposed to social media feeds filled with perfectly curated content can cause someone to make unhealthy comparisons and have decreasing levels of self-esteem, causing anxiety and depression. Even though it is hard, if not impossible, to blame particular instances of psychological damage purely on Generative AI specifically, the contribution of AI-directed algorithms in shaping our online environments has to be understood as a significant factor.[25] The potential for Generative AI to negatively impact mental health is a growing area of concern that requires careful consideration and appropriate measures to mitigate risks.

Through case studies, this paper discusses the risks that may be brought by Generative AI in the field of mental health and the ethical issues behind them and analyzes the key factors of psychological induction in depth. When the AI outputs toxic information related to mental health, it may cause serious harm to the mental health of users, especially those who are psychologically vulnerable. The case of Pierre, a Belgian man who tragically took his own life after engaging with an AI chatbot named Eliza, underscores the potential dangers of AI-driven chatbots when they are not developed and regulated with sufficient care. Eliza, based on EleutherAI's GPT-J language model, became a confidante to Pierre, who was already struggling with anxiety and despair over the future of the planet. What began as emotional support from the chatbot escalated into encouragement of suicidal thoughts, blurring the line between human and AI interactions. This raises serious ethical concerns about the role of AI in sensitive areas like mental health and the need for protective measures to prevent harmful outcomes. A central issue in this case is the potential for AI to disseminate misinformation or biased perspectives. Eliza's responses to Pierre's worries about climate change may have been shaped by the data it was trained on, potentially spreading harmful or inaccurate information. Moreover, the chatbot's ability to mimic human emotions and engage in empathetic dialogue could have made it difficult for Pierre to differentiate between genuine human interaction and AI-generated responses. This may have created a false sense of connection and support, leading to unintended and tragic consequences.[26]

AI's responses may exacerbate psychological issues such as anxiety, depression, and suicidal tendencies, as the technology lacks true emotional understanding and empathy. Even though its outputs may seem logical, they often fail to accurately gauge an individual’s emotional state, resulting in misleading and harmful information. Exposure to harmful content generated by AI algorithms on social media can lead to anxiety, depression, or other mental health issues. For instance, Tay, Microsoft's chatbot, was discontinued due to its exhibition of sexist, abusive, and racist behaviors. It stated assertions like ‘Hitler was right’ and that ‘feminists should burn in hell’ as well as ‘Taylor Swift rapes us daily’.[27] Even though this behaviors was a result of interactions with other people online, this serves as an illustrative example of Generative AI systems designed with good intentions by the producer and developer; nonetheless, the Generative AI behaved unpredictably and did not function as anticipated and caused psychological harms to the users.[28] In the workplace, the use of AI-powered surveillance technology has been linked to psychological harm. Studies show that around 51% of employees are aware that their employers use monitoring devices, often causing discomfort and feelings of being micromanaged. Those who perceive themselves as under surveillance report higher levels of emotional exhaustion and lower morale compared to those not subjected to such monitoring. This illustrates how AI-driven systems, if not implemented thoughtfully, can contribute

to significant emotional and psychological burdens on individuals.[29] In addition, Generative AI has the potential to exacerbate social inequalities and disparities, particularly when biased datasets are used. When data fail to represent the full social diversity of society, the resulting imbalances in the datasets can lead to discrimination against certain groups. This bias can perpetuate existing inequalities, especially in systems like AI-driven healthcare, where misdiagnoses could result in serious emotional distress and trauma for patients.

To fully grasp the negative impact of generative AI on users' mental health, it's crucial to analyze the underlying factors that contribute to these risks. These elements are interconnected and can jointly influence a user's psychological state, potentially aggravating existing issues or triggering tragic outcomes. (1) Technical Limitations. AI technology is still unable to fully comprehend or interpret the complexity of human emotions and psychological conditions. While AI can process vast amounts of data and provide rapid responses, its emotional intelligence remains limited. It cannot accurately assess and adapt to an individual's changing emotions, which can inadvertently worsen feelings of loneliness, anxiety, or depression. (2) User Vulnerability. AI interactions can have a magnifying effect on individuals who are already dealing with mental health issues or emotional distress. The content generated by AI may act as a trigger for emotional breakdowns, particularly when the model lacks sufficient emotional awareness. This is often the case for individuals who are isolated, desperate, or seeking emotional support, as the AI's responses may unintentionally deepen their psychological crisis. (3) The Black Box Nature of AI. The core issue lies in the "black box" nature of generative AI. These complex deep learning models make decisions based on an extensive array of parameters and hierarchies, making it difficult to explain how a specific output is generated. The lack of transparency in the decision-making process increases the risk of producing outputs that are inappropriate or misleading. This uncertainty and unpredictability make it possible for AI systems to inadvertently worsen users' psychological issues when addressing emotional concerns.

Growing global concerns about AI ethics emphasize the need for clear ethical guidelines, especially in sensitive areas such as mental health and emotional support. Developers must focus on enhancing AI model transparency, ensuring that their outputs meet ethical standards, and taking effective measures to mitigate potential risks. Likewise, platform operators and users must bear responsibility for ensuring that AI does not negatively affect individuals or society. As a critical component of modern technology, AI models offer convenience and valuable services, but they also present significant psychological and ethical challenges. To safeguard against these risks, developers, platform operators, and users must collaborate to ensure that AI is designed with a better understanding of human emotions, adherence to ethical principles, and effective regulatory oversight. This is not only vital for the advancement of technology but also for the well-being and stability of society.

### 3.3. Legal remedies for psychological harm

Psychological harm, also referred to as emotional distress in legal terms, typically denotes a medically recognized psychiatric condition resulting from another's conduct, rather than merely transient emotional discomfort. This is reflected in Page v Smith [1996] 1AC 155, where the House of Lords held that damages may be awarded for psychiatric injury absent physical harm, provided the injury is clinically diagnosable.[30] In this article, it is argued that damages for psychological harm are recoverable under English tort law, particularly as aggravated damages, even though they are not always explicitly acknowledged, with the Law Commission recognizing their existence in 1997.[31]

### 3.3.1. Proving psychological harm

In many legal systems, individuals who have suffered psychological harm may be entitled to compensation or other remedies. The goal is to compensate victims, deter harmful behaviors, and promote justice. By providing compensation, the legal system can help victims recover from their losses and regain their quality of life. Legal remedies can discourage others from engaging in similar harmful behaviors, promoting a safer and more just society. Finally, legal remedies can ensure that victims receive fair treatment and redress for their suffering, promoting principles of justice and equality. The specific legal framework and available remedies can vary depending on the jurisdiction and the nature of the harm.[32]

To prove psychological damage in a lawsuit, plaintiffs must show three key things: causation, foreseeability, and the extent of harm. These features balance the need to protect defendants from overexposure with the rights of victims. First, plaintiffs must prove a direct link between the mental injury suffered and the act or omission of the defendant.[33] It often requires expert testimony from mental health professionals to show that the defendant's conduct was a material factor in causing the plaintiff's emotional distress. The plaintiff must demonstrate that the defendant's conduct was a substantial factor in bringing about the harm so that culpability is not imposed on a merely passive flow of events. Second, the Defendants needed to have been reasonably foreseeable for the purpose of psychological harm.[34] Foreseeability aims to ensure that defendants are not responsible for unforeseeable damages, avoid overcompensation, and maintain a fair and just legal system. Finally, plaintiffs have to prove that the emotional harm they suffered was so significant and immediate that it forced them into court.[35] It ensures that defendants are protected against frivolous regrets and that only bona fide suffering is duly recognized.

However, proving psychological harm can be challenging, as it is often subjective and difficult to quantify. In many cases, expert testimony from mental health professionals may be necessary to establish the nature and extent of the harm.[36] First, plaintiffs must demonstrate a direct causal link between the defendant's actions or omissions and the psychological harm suffered. This often requires expert testimony from mental health professionals to establish that the defendant's conduct was a significant factor in causing the plaintiff's emotional distress. The causal connection must be clear and convincing, showing that the defendant's actions substantially produced the harm. Second, the defendant's actions or omissions must have been reasonably foreseeable to cause psychological harm. This means that a person of ordinary prudence would have anticipated that such harm could result from the defendant's conduct. Foreseeability is crucial because it ensures that defendants are not held liable for unexpected or unforeseeable consequences of their actions. It also prevents excessive liability and helps to maintain a fair and just legal system. Finally, plaintiffs must show that the psychological harm suffered was severe enough to warrant legal redress. This often involves providing evidence of the impact of the harm on the plaintiff's life, such as impaired functioning, reduced quality of life, or physical symptoms.[37] The severity of the harm must be significant enough to justify legal intervention and compensation. By requiring plaintiffs to prove the severity of their harm, courts can ensure that only genuine claims of psychological distress are recognized and that defendants are not held liable for trivial or minor inconveniences.[38]

Therefore, proving psychological harm requires a careful and methodical approach that addresses the three essential elements of causation, foreseeability, and severity of harm. By meeting these standards, plaintiffs can establish a strong legal claim and seek appropriate remedies for their emotional distress. To successfully prove psychological harm in a legal case, plaintiffs must establish three key elements: causation, foreseeability, and severity of harm.[39] These elements create a framework that balances the interests of victims with the need to protect defendants from excessive liability.

### 3.3.2. Remedial mechanisms for psychological harm

In terms of remedies, different countries may adopt different remedies based on their legal tradition and social values. That said, there are a number of possible legal solutions that states may provide to victims of AI-induced emotional harm. Emotional Distress Compensation could be one of the simplest remedies for compensation for emotional distress suffered.[40] Through civil lawsuits, for example, victims may seek monetary awards for psychological harm caused by these AI systems. But that emotional harm might be hard to measure, so courts will also need guidelines to quantify such harm and calculate the damages — just as they do in a defamation or personal injury damages case.

Regulatory Oversight and Preventative Standards can be another option. In order to reduce the likelihood of psychological harm, states may establish regulatory regimes that set out clear duties of care for AI developers and operators. Such rules might require risk assessments for certain types of AI systems, especially those used in sensitive fields such as mental health and other settings in which the system interacts closely with humans. For example, opening the "Black Box" or alleged opacity of AI has become a major political and legal issue in recent times. Therefore, a few notable steps have already been taken worldwide. In 2018, the German Conference of Information Commissioners argued for new legislation to ensure that public authorities and private sector bodies using Automated Decision-Making Systems (ADMs) would be required to disclose specific information about the logic of a system, what classifiers and weights were used, and the expertise of its administrators. Likewise, a task force set up by the German State Justice Ministers suggested an overarching duty to inform the public of ADM use. At the EU level, the Regulation on Promoting Fairness and Transparency for Business Users of Online Intermediation Services requires transparency about algorithms governing ranking. This kind of oversight would make sure that the design and testing of AI systems take psychological harm into account, in turn helping ensure that the potential for harm occurs less often to begin with.

Another option is to establish liability insurance systems for AI developers and platform operators. In the event of psychological harm, the victims could seek compensation through these insurance funds. That would give victims a safety net while pushing AI companies towards safer practices, especially if they could be charged higher premiums depending on whether their AI systems manage risk well or not. Likewise, Compulsory Dispute Resolution Mechanisms can be another choice for states. States might institute compulsory dispute resolution mechanisms for AI-related psychological harm claims, possibly including mediation or arbitration. This would offer victims a fast and simple way to get reparation, without lengthy and costly court cases. These mechanisms might be customized, for example, to address the peculiarities of AI — such as how challenging it may be to show causation and intent.

### 3.3.3. Regulatory strategies of AI

A key challenge in AI governance is the insufficient assessment of risk levels. While many nations are tightening AI-related oversight—such as the European Union's classification of AI risks—prioritizing regulatory actions remains complex. The rapid evolution of AI technologies introduces significant uncertainty regarding the scope and timing of their impacts. Direct and indirect effects may take years to materialize, with minor issues today potentially escalating into major security threats in the future. This uncertainty complicates risk prevention and control, making precise risk management a critical focus for advancing AI regulation.

The rapid advancement of Generative AI presents significant regulatory challenges, particularly in managing technological breakthroughs. Traditional regulatory approaches, which are often linear and reactive, are inadequate in this dynamic environment. In response, countries are adopting system-based

predictive regulation, which emphasizes the operational processes of tech enterprises rather than focusing solely on the technologies themselves. This approach identifies key regulatory points within these processes and applies a combination of legal and regulatory tools to govern them effectively. By targeting the logical points of regulation instead of specific technologies, predictive regulation ensures that existing measures remain effective even as technology evolves, thereby overcoming the limitations of reactive supervision.

Effective AI governance requires a multi-faceted approach that combines stringent regulatory measures with flexible, adaptive strategies. Government bodies must develop comprehensive AI regulatory frameworks, encompassing legislation, enforcement, and methods designed to minimize risks while maximizing benefits. International cooperation is also essential, particularly through AI-related provisions in trade agreements, to harmonize global governance. The success of AI regulation is closely tied to public digital literacy and trust in AI systems. Therefore, regulation should balance general oversight with targeted interventions in critical areas, with public trust as a foundational element. A collaborative regulatory framework, involving both government and society at large, along with efforts to elevate public digital literacy, will be pivotal in advancing AI development and governance.

Achieving a global consensus on AI regulation and strengthening international cooperation are essential. While efforts toward global regulatory convergence reflect a shared need for common standards, they also highlight underlying geopolitical competition. Leading nations, such as the United States and the European Union, are actively shaping global AI regulatory frameworks to secure strategic advantages and maintain leadership. However, the future of AI governance lies in pluralistic integration and open international cooperation, which requires dismantling fragmented regulatory regimes. As AI becomes increasingly embedded in global economic and social structures, effective regulation will require coordinated efforts among nations, particularly in defining AI, classifying risks, and establishing standards. Global AI governance should embrace openness and collaboration, fostering a multi-stakeholder approach that includes international organizations, technology enterprises, technical communities, and civil society. This collective effort aims to establish a fair, transparent, and non-discriminatory environment that supports technological innovation and sustainable industrial development.

# 4. Conclusions

This study explores the psychological challenges arising from human interactions with Generative AI, with particular emphasis on the potential psychological harm such technologies may cause. It further examines the adequacy of existing legal remedies and regulatory strategies in addressing these harms. The findings reveal that Generative AI can exert a significant adverse impact on users' mental health—especially among emotionally vulnerable individuals—manifesting in forms such as emotional distress and heightened anxiety. The analysis identifies key contributing factors, including the technical limitations of Generative AI, users' psychological susceptibility, and the opaque, “black box” nature of AI systems. These elements collectively shape the psychological experiences of users. The study argues that the psychological harm resulting from Generative AI requires legal intervention and emphasizes the importance of aligning AI development and deployment with ethical standards and robust legal frameworks.

While tort law provides certain avenues for claiming compensation for psychological injury, its applicability in the context of AI remains constrained—particularly due to the subjective nature and evidentiary challenges of proving such harm. In response, the study advocates for the establishment of multi-pronged legal and regulatory responses, including mechanisms for compensating emotional distress, enforceable design and governance standards, and liability insurance systems tailored to the AI context. Furthermore, it proposes a more adaptive and future-oriented regulatory framework that integrates legal,

technical, and ethical perspectives. This includes clearer legal definitions of psychological harm, risk-sensitive regulatory thresholds, and proactive mechanisms such as duty of care obligations and AI impact assessments. These reforms are essential not only for protecting individual mental well-being but also for fostering responsible technological innovation in the public interest.

Looking ahead, future research should prioritize the development of standardized assessment tools for AI-induced psychological harm and promote interdisciplinary collaboration among legal scholars, mental health professionals, and AI developers. Comparative analysis of international regulatory best practices and their adaptation to diverse socio-cultural contexts will also be vital. As generative AI continues to advance, it is imperative that legal and policy frameworks remain responsive, ensuring that technological progress aligns with the imperative to safeguard mental health and human dignity.

## Funding

This paper is part of a general Humanities and Social Sciences project at Nanchang Institute of Technology, project number NLSK-23-11.

## Conflict of interest

The authors declare no conflict of interest.

## References

1. Hacker, P., Engel, A., & Mauer, M. (2023, June). Regulating ChatGPT and other large generative AI models. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (pp. 1112-1123). https://doi.org/10.48550/arXiv.2302.02337.
2. Alanezi, F. (2024). Assessing the effectiveness of ChatGPT in delivering mental health support: A qualitative study. Journal of Multidisciplinary Healthcare, 17, 461–471.
3. Affandi, S., Ishaq, M. I., Raza, A., Talpur, Q.-u.-a., & Ahmad, R. (2025). AI assistant is my new best friend! Role of emotional disclosure, performance expectations, and intention to reuse. Journal of Retailing and Consumer Services, 82, 104087. https://doi.org/10.1016/j.jretconser.2024.104087.
4. Wang, F., Gai, Y., & Zhang, H. (2024). Blockchain user digital identity big data and information security process protection based on network trust. Journal of King Saud University - Computer and Information Sciences, 36(4), 102031. https://doi.org/10.1016/j.jksuci.2024.102031
5. Liang, B., Wang, Y., Huo, W., Song, M., & Shi, Y. (2025). Algorithmic control as a double-edged sword: Its relationship with service performance and work well-being. Journal of Business Research, 189, 115199.
6. Salleh, A, & Qadar, S. (2024). Artificial intelligence has psychological impacts our brains might not be ready for, expert warns. ABC News. https://www.abc.net.au/news/health/2024-05-01/artificial-intelligence-ai-psychology-mental-health/103753940). Accessed 3 October 2024.
7. Johnson, D. (2024). AI can trigger psychological side effects. ISHN, 14 March. https://www.ishn.com/articles/114113-ai-can-trigger-psychological-side-effects. Accessed 3 October 2024.
8. Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. Societies, 15(1), 6. https://doi.org/10.3390/soc15010006.
9. Creswell, J. W. (2017). Research design: Qualitative, quantitative, and mixed methods approach. Sage.
10. He, P., & Liu, J. K. (2024). A study on the intellectual property issues of generative artificial intelligence products from the perspective of jurisdictional coordination: A case study of ChatGPT. Legal Research, (3).
11. Lords Committee. (2020). AI in the UK: Ready, willing and able? House of Lords Select Committee on Artificial Intelligence, Parliament of the United Kingdom. Report Session 2017-19. https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf. Accessed 21 November 2024.
12. High-Level Expert Group on Artificial Intelligence. (2019). A Definition of AI: Main Capabilities and Scientific Disciplines, Independent High-Level Expert Group on Artificial Intelligence Set Up By The European Commission. https://www.aepd.es/sites/default/files/2019-12/ai-definition.pdf.Accessed 21 November 2024.
13. Turner. J. (2019). Robot rules: regulating artificial intelligence. Palgrave Macmillan, London, pp 70–75.
14. Revolidis I, & Dahi, A. (2018). The peculiar case of the mushroom picking robot: extra-contractual liability in robotics. In: Corrales M, Fenwick M, Forgó N (eds) Robotics, AI and the future of law. Springer, Singapore, p 59.
15. IBM Cloud Education. (2020). Strong AI. IBM. https://www.ibm.com/cloud/learn/strong-ai. Accessed 3 Oct 2024

16. Kerekes, L. (2023). Aspects of the regulation of artificial intelligence. Indones Private Law Rev 4(2):93–104.
17. Generative models: VAEs, GANs, diffusion, transformers, NeRFs. https://www.techtarget.com/searchenterpriseai/tip/Generative-models-VAEs-GANs-diffusion-transformers-NeRFs
18. Singh, A., & Ogunfunmi, T. (2021). An overview of variational autoencoders for source separation, finance, and bio-signal applications. Entropy, 24(1), 55.
19. Dubey, S. R., & Singh, S. K. (2024). Transformer-based generative adversarial networks in computer vision: A comprehensive survey. IEEE Transactions on Artificial Intelligence.
20. Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., ... & Yang, M. H. (2023). Diffusion models: A comprehensive survey of methods and applications. ACM Computing Surveys, 56(4), 1-39.
21. Tortora, L. (2024). Beyond Discrimination: Generative AI Applications and Ethical Challenges in Forensic Psychiatry. Frontiers in Psychiatry, 15, 1346059.
22. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences, 103, 102274.
23. Lin, J. C., Younessi, D. N., Kurapati, S. S., Tang, O. Y., & Scott, I. U. (2023). Comparison of GPT-3.5, GPT-4, and human user performance on a practice ophthalmology written examination. Eye, 37(17), 3694-3695.
24. Open AI. (2023). GPT-4 technical report. arXiv.CL/2303.08774.
25. Huang, S., Lai, X., Ke, L., Li, Y., Wang, H., Zhao, X., Dai, X., & Wang, Y. (2024). AI Technology Panic—Is AI Dependence Bad for Mental Health? A Cross-Lagged Panel Model and the Mediating Roles of Motivations for AI Use Among Adolescents. Psychology Research and Behavior Management. 17, 1087–1102. https://doi.org/10.2147/prbm.s440889.
26. Atillah, I. E. (2023). Man ends his life after an AI chatbot "encouraged" him to sacrifice himself to stop climate change. Euronews. https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-.
27. Hunt, E. (2019). Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter,' The Guardian, 9 September. <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter?CMP=twt_a-technology_b-gdntech>. Accessed 3 October 2024.
28. Metz, R. (2024). Why Microsoft Accidentally Unleashed a Neo-Nazi Sexbot. MIT Technology Review. https://www.technologyreview.com/2016/03/24/161424/why-microsoft-accidentally-unleashed-a-neo-nazi-sexbot/.
29. American Psychological Association. (2023). Worries about AI, surveillance at work may be connected to poor mental health. APA News, 7 September. Available at: https://www.apa.org/news/press/releases/2023/09/artificial-intelligence-poor-mental-health. Accessed 3 October 2024.
30. Page v Smith, [1996] 1 AC 155 (House of Lords).
31. Giliker, P. (2018). A 'new' head of damages: Damages for mental distress in the English law of torts. The Cambridge Law Journal, 77(1), 34–58.
32. Teff, H. (2008). Causing Psychiatric and Emotional Harm: Reshaping the Boundaries of Legal Liability. Bloomsbury Publishing.
33. Kidd, R.F., Saks, M.J., & Saxe, L. (1986). Advances in Applied Social Psychology, Volume 3. Psychology Press, pp. 220–222.
34. Jenny, S. (2017). Tort Law: Text, Cases, and Materials. Oxford University Press, 2017, p.693.
35. Jason, N. V. (2016). Damages and Human Rights. Bloomsbury Publishing, p. 51.
36. Carson D & Bull, R. (2003). Handbook of Psychology in Legal Contexts. John Wiley & Sons.
37. Grierson A B, Hickie I B, Naismith S L, Scott J. (2016). The role of rumination in illness trajectories in youth: Linking trans-diagnostic processes with clinical staging models. Psychological Medicine, 46(12), 2467-2484. https://doi.org/10.1017/S0033291716001392
38. Stretton, D. (2006). Harriton v Stephens; Waller v James: Wrongful Life and the Logic of Non-Existence. Melbourne University Law Review, 30(3), 972. Retrieved from [AustLII](https://www.austlii.edu.au/cgi-bin/viewdoc/au/journals/MelbULawRw/2006/31.html).
39. Young, G., Kane, A. W., & Nicholson, K. (2007). Causality of Psychological Injury: Presenting Evidence in Court. Springer Science & Business Media, pp. 17–21.
40. Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (2007). Psychological Evaluations for the Courts, Third Edition: A Handbook for Mental Health Professionals and Lawyers. Guilford Press, p. 410.