

Privacy and Cybersecurity Risks of Sharing Personal Photos and Videos with Generative AI Platforms: A Study of User Awareness and Protective Practices.

Kaushik Musale¹, Aryan Rajput², Pooja Mehta³

¹²Student, Department of MCA, D. Y. Patil Deemed to be University, School of Humanities and Sciences, Belapur, Maharashtra, India.

³Assistant Professor, D. Y. Patil Deemed to be University, School of Humanities and Sciences, Belapur, Maharashtra, India.

Abstract - Generative Artificial Intelligence (AI) platforms have rapidly become part of everyday digital activities, allowing users to interact with AI systems by sharing text, images, videos, documents, and other forms of personal information. While these capabilities provide significant convenience and functionality, sharing personal photos and videos with AI platforms may introduce privacy and cybersecurity concerns, particularly when such media contains identifiable individuals, personal details, documents, locations, or other sensitive information. Users may also have limited awareness of how their uploaded content is processed, retained, protected, or governed by platform privacy policies. This research investigates the privacy and cybersecurity risks associated with sharing personal photos and videos with generative AI platforms, with a particular focus on user awareness and protective practices among young adults aged 18–30 (N=130). The study considers commonly used generative AI platforms, including ChatGPT, Google Gemini, and Microsoft Copilot. It examines users' awareness, sharing behaviour, perception of privacy and cybersecurity risks, understanding of platform privacy practices, and adoption of protective measures. In addition, the study examines the privacy safeguards and relevant policies provided by selected AI platforms. The research adopts an exploratory and descriptive cross-sectional survey design (N=130) combined with an analysis of relevant literature and platform privacy documentation. The study aims to identify potential privacy and cybersecurity risks, understand gaps between users' awareness and actual sharing behaviour, and examine the protective practices that can reduce exposure when interacting with generative AI systems. The findings are intended to contribute to improved user awareness, platform-level Privacy-by-Design safeguards, and more responsible handling of personal visual information in AI-enabled environments.

Keywords: Generative Artificial Intelligence, Multimodal AI, Digital Visual Privacy, Cybersecurity Risk Perception, User Awareness, Data Retention Policies, Deepfake & Synthetic Media Misuse.

1. INTRODUCTION

Generative Artificial Intelligence (AI) has rapidly become an important part of modern digital life. AI platforms such as ChatGPT [1], Google Gemini [2], and Microsoft Copilot [3] allow users to interact with AI systems not only through text but also by uploading images, videos, documents, and other forms of digital content. These capabilities have expanded the practical applications of AI, including education, communication, content creation, information analysis, and personal productivity. As these platforms become increasingly accessible, users are also becoming more comfortable sharing different types of personal information with AI systems.

Among the different forms of information shared with AI, personal photos and videos require particular attention from a privacy and cybersecurity perspective. A photograph or video may contain more information than what is immediately visible to the user. It can reveal a person's face, location, surroundings, identity documents, contact information, vehicle details, workplace or educational information, and other contextual information. Images may also contain metadata or visual characteristics that can contribute to the identification or profiling of individuals. Consequently, uploading such content to an AI platform can create privacy considerations that users may not fully recognize at the time of interaction.

The privacy implications of generative AI are also connected to the way AI systems process user-provided information. Generative AI models and their supporting services may involve processes such as data transmission, storage, processing, security controls, and other forms of data handling. The specific practices can differ between platforms and may depend on account settings, product features, applicable policies, and the type of service being used. Therefore, understanding the privacy policies and safeguards of AI platforms is important when evaluating the risks associated with sharing personal visual information.

From a cybersecurity perspective, exposure of personal images and videos can create additional concerns. Visual information containing personally identifiable or sensitive details may contribute to identity exposure, social engineering, impersonation, unauthorized profiling, or other forms of misuse if improperly handled or exposed. The increasing capabilities of generative AI also make the broader protection of visual information important because digital images and videos can potentially be analyzed, transformed, or reused in ways that users may not anticipate.

However, technological safeguards alone do not determine the level of privacy risk. User behaviour and awareness are also important factors. Users may share personal images with AI systems without carefully considering what information is visible in the image, whether the content is necessary for the task, what the platform's privacy practices are [1–3], or what protective settings are available. Understanding the relationship between users' awareness, their actual sharing behaviour, and their adoption of protective practices is therefore an important part of studying AI-related privacy risks.

This research focuses on the privacy and cybersecurity implications of sharing personal photos and videos with generative AI platforms. It particularly examines young adults aged 18–30 and considers their awareness, usage behaviour, perception of privacy and cybersecurity risks, and protective practices. The study also considers selected generative AI platforms, including ChatGPT [1], Google Gemini [2], and Microsoft Copilot [3], with attention to their relevant privacy practices and safeguards. Rather than assuming that sharing visual information with AI is inherently harmful, the study seeks to identify potential risks, understand user awareness and behaviour, and examine measures that can help users make more informed decisions when sharing personal visual information with generative AI systems.

1.1 Problem Statement

The increasing use of generative AI platforms has made it common for users to upload personal photos, videos, screenshots, and other visual information while interacting with AI systems. Although these capabilities provide useful services such as image analysis, content generation, and information extraction, users may not always understand the privacy and cybersecurity implications associated with sharing personal visual data. Personal photos and videos can contain identifiable faces, locations, documents, contact information, institutional details, and other sensitive information. When such content is submitted to an AI platform, users may have limited understanding of how the information is processed, stored, protected, or governed under the platform's privacy policies [1–3, 3]. Differences in platform policies, available privacy controls, data-handling

practices, and user settings can further complicate the assessment of potential risks.

At the same time, users' actual behaviour may not correspond with their awareness of these risks. Some users may share personal visual information without evaluating the information contained within an image or understanding the relevant privacy safeguards. This creates a need to examine not only the technical and privacy risks associated with AI-based image and video sharing, but also the level of user awareness and the protective practices adopted by users.

Therefore, the problem addressed by this research is the lack of a clear understanding of the relationship between personal image/video sharing behaviour, user awareness, platform privacy practices [1–3], and potential privacy and cybersecurity risks associated with generative AI platforms. This study seeks to investigate these aspects among young adults aged 18–30 and identify measures that can support more informed and responsible sharing of personal visual information with AI systems.

1.2 Research Gap

While extensive scholarship has examined privacy risks associated with social media sharing and text-based generative AI interactions, a significant empirical gap remains concerning how young adult users navigate the multimodal visual upload features of contemporary AI platforms. Specifically, limited research evaluates the nexus between actual user sharing behaviors, protective digital hygiene practices (such as metadata stripping and facial blurring), platform policy transparency, and the behavioral "Privacy Paradox" unique to visual data submissions in tools like ChatGPT, Google Gemini, and Microsoft Copilot. This study addresses this critical gap by providing a targeted empirical investigation ($N = 130$) into the awareness, threat perceptions, and protective precautions surrounding personal visual media shared with generative AI systems.

1.3 Significance of the study

The rapid expansion of multimodal generative AI platforms such as ChatGPT, Google Gemini, and Microsoft Copilot has fundamentally transformed how digital users interact with artificial intelligence, shifting from text-only prompts to complex visual data exchanges. While existing literature has extensively evaluated privacy concerns in social media and text-based AI systems, empirical research addressing the specific privacy and cybersecurity vulnerabilities of multimodal visual media uploads remains limited.

1.4 Objectives

1. To examine the level of awareness among young adults aged 18–30 regarding the privacy and cybersecurity risks associated with sharing personal photos and videos with generative AI platforms.
2. To analyse users' sharing behaviour by examining the frequency, purpose, and types of personal visual content shared with generative AI platforms.
3. To identify potential privacy and cybersecurity risks associated with submitting personal photos and videos to generative AI platforms, including exposure of personally identifiable information, identity-related risks, unauthorized profiling, social engineering, and misuse of visual information.
4. To examine the privacy and data-handling practices disclosed by selected generative AI platforms, namely ChatGPT [1], Google Gemini [2], and Microsoft Copilot [3], with particular attention to user controls, data retention, data usage, and privacy-related safeguards.
5. To assess the protective practices adopted by users when sharing personal visual information with generative AI platforms and identify areas where improved awareness or safer practices may be required.
6. To propose practical recommendations for users, platform developers, and policymakers to reduce privacy and cybersecurity risks when interacting with generative AI platforms using personal photos and videos.

1.5 Research Questions

RQ1. What is the level of awareness among young adults aged 18–30 regarding the privacy and cybersecurity risks associated with sharing personal photos and videos with generative AI platforms?

RQ2. What are the common patterns of personal photo and video sharing among young adults when using generative AI platforms?

RQ3. What privacy and cybersecurity risks may arise from sharing personal photos and videos with generative AI platforms?

RQ4. How do the privacy and data-handling practices of ChatGPT, Google Gemini, and Microsoft Copilot address the protection and management of user-submitted visual information?

RQ5. What protective practices do users adopt before or while sharing personal photos and videos with generative AI platforms?

RQ6. What measures can be recommended to improve user awareness and promote safer sharing of personal visual information with generative AI platforms?

1.6 Proposed Hypotheses

H1: There is a significant behavioral gap (the Privacy Paradox) among young adults, wherein high stated baseline concern regarding visual privacy does not significantly deter users from actively uploading personal photos and videos to generative AI platforms.

H2: Higher levels of user awareness regarding hidden image data (such as EXIF metadata and OCR text extraction) are positively associated with a greater frequency of adopting pre-upload protective practices (such as facial blurring, cropping, and metadata stripping).

H3: Greater clarity and user confidence in understanding platform privacy disclosures and data retention terms are positively associated with overall user trust in generative AI platforms.

H4: Academic specialization (such as postgraduate versus undergraduate student status) and active daily platform usage frequency significantly influence the adoption of proactive risk mitigation and privacy control settings.

2. REVIEW OF LITERATURE

2.1 Multimodal Generative AI and Visual Data Exposure

Unlike text prompts, digital photographs and videos encode multi-layered information that extends beyond what is immediately visible.

2.2 The "Privacy Paradox" in AI Ecosystems

A well-established concept in digital privacy literature is the Privacy Paradox, originally formulated by Barnes [4] to describe the contradiction between users' expressed privacy concerns and their actual online sharing behavior.

2.3 Platform Privacy Policies, Data Ingestion, and Governance

The governance of user-submitted visual media depends heavily on platform privacy policies and terms of service. However, academic evaluations of privacy documentation consistently highlight significant barriers to user comprehension

2.4 Cybersecurity Threat Vectors: Deepfakes, Biometric Profiling, and Data Leakage

From a cybersecurity perspective, the submission of personal photos and videos to external cloud servers introduces severe threat vectors:

- Synthetic Media and Deepfake Exploitation: High-resolution facial photographs and personal video clips serve as optimal training inputs for deepfake

synthesis algorithms. Exposed facial media can be exploited for unauthorized voice/image cloning, digital impersonation, and social engineering attacks. Standard frameworks such as C2PA [7] advocate for content provenance and cryptographic watermarking to mitigate deepfake proliferation.

- **Biometric Feature Extraction:** Uploaded images processed by facial recognition systems allow for the extraction of biometric templates. Without clear retention boundaries, biometric data can be permanently indexed across database breaches.
- **Training Data Memorization and Leakage:** Research in deep learning security demonstrates that large generative models can memorize unique training samples. If private visual uploads are ingested into training sets, there is a technical risk of training data extraction, where visual inputs are accidentally reproduced in response to external prompts.

2.5 Synthesis and Research Gap

While existing literature extensively examines privacy risks on social media platforms [4] and text-based AI models, there is a notable empirical gap regarding how young adult users (18 – 30 years old) interact specifically with multimodal generative AI visual upload features. Furthermore, limited research evaluates the alignment between users' protective practices (such as metadata stripping or facial blurring), their policy reading habits, and actual platform-level safeguards provided by platforms like ChatGPT [1], Google Gemini [2], and Microsoft Copilot [3]. This study fills this gap by conducting a structured empirical assessment ($N = 130$) of user awareness, sharing behaviors, platform policies, and protective practices within this emerging domain.

3. METHODOLOGY

3.1 Research Design

This study adopts a descriptive, cross-sectional research design to examine the privacy and cybersecurity risks associated with sharing personal photos and videos with generative Artificial Intelligence (AI) platforms. The descriptive design is appropriate because the study focuses on systematically examining and presenting the existing patterns of AI usage, personal visual-data sharing behaviour, privacy awareness, perceived cybersecurity risks, and protective practices among the selected respondents.

The study does not manipulate the behaviour of participants or introduce an experimental intervention. Instead, data are collected through a structured questionnaire and analyzed to describe the respondents' existing experiences, perceptions, awareness, and practices regarding the sharing of personal visual

information with generative AI platforms. The findings are used to identify patterns and relationships within the collected responses and to provide a clearer understanding of the privacy and cybersecurity considerations associated with the use of such platforms.

3.2 Target Population

The target population for this study comprises young adults aged 18–30 who actively use digital technologies and generative AI platforms such as ChatGPT, Google Gemini and Microsoft Copilot. This demographic is particularly relevant as young adults form a substantial proportion of early adopters of AI tools for educational, professional, and personal purposes. The primary focus was on student populations (undergraduate and postgraduate) as well as young working professionals. Out of the total $N = 130$ survey respondents, 99 respondents (76.15%) fell within the primary young adult age bracket (18–30 years), with 40.77% being postgraduate students ($n = 53$), 28.46% undergraduate students ($n = 37$), and 25.38% working professionals ($n = 33$). Gender distribution across the sample was balanced, consisting of 51.54% female ($n = 67$) and 48.46% male ($n = 63$) participants.

3.3 Sampling

A non-probability convenience sampling method combined with snowball sampling was utilized to recruit participants. This approach was selected due to its practicality and efficiency in reaching active digital users across academic institutions and professional networks. The structured survey link was distributed digitally via academic communication channels, instant messaging platforms (such as WhatsApp student groups), and social media networks. Participation was entirely voluntary, and respondents were encouraged to share the survey link with peers within the target age demographic. A total of 130 completed and valid responses were collected and incorporated into the primary dataset for analysis ($N=130$).

3.4 Data Collection

Primary data was collected through an online cross-sectional survey administered using Google Forms. The survey remained active through September 2026. Prior to filling out the questionnaire, respondents were provided with a detailed informed consent statement explaining the academic purpose of the research, ensuring strict confidentiality, and confirming that no personally identifiable visual media or sensitive credentials would be stored or shared. All 130 participants explicitly agreed to the participation terms before accessing the survey questions.

3.5 Survey Instrument

The data collection tool consisted of a structured questionnaire containing 44 items, comprising single-choice, multiple-choice, and 5-point Likert scale questions. The instrument was structured into six distinct functional sections:

1. Demographics: Age group, gender, and current occupation/status.
2. AI Platform Usage & Sharing Behaviour: Frequency of generative AI usage, platforms utilized, frequency of sharing personal photos/videos, visual content types (e.g., selfies, ID documents, background details), and underlying reasons for uploading media.
3. Privacy Concerns & Risk Perception: User concern levels regarding unauthorized image use, prolonged data retention, biometric/facial data processing, model training, third-party access, and deepfake synthesis.
4. Policy Awareness & Platform Trust: Reading habits regarding platform privacy policies, level of confidence in data handling, and factors influencing platform trust.
5. Protective Practices & Precautions: Self-reported safeguards adopted before uploading visual media (e.g., face blurring, cropping sensitive areas, avoiding document uploads).
6. User Recommendations & Awareness Needs: Opinions on required user education methods and platform-level safeguards.

3.6 Policy Analysis Framework

To complement the primary survey data, this study incorporates a comparative policy analysis framework examining the publicly available privacy documentation and data-handling policies of leading generative AI platforms: ChatGPT (OpenAI), Google Gemini (Google) and Microsoft Copilot (Microsoft). The evaluation focuses on five key analytical dimensions:

1. Data Retention & Storage: Policies governing how long uploaded visual media (photos, videos, screenshots) and associated metadata are stored on platform servers.
2. Model Training & Data Ingestion: Platform disclosures regarding whether user-submitted visual content is automatically utilized to train, fine-tune, or improve underlying multimodal models.
3. User Opt-Out & Deletion Controls: Availability, clarity, and accessibility of user mechanisms to disable model training, opt out of data logging, or permanently delete uploaded visual media.

4. Third-Party Access & Sharing: Express terms regarding third-party vendor access, human reviewer access, and data transfers.
5. Technical Safeguards: Disclosed security controls, including encryption standard specifications in transit and at rest.

3.7 Threat Analysis

A conceptual cybersecurity and privacy threat framework was developed to categorize the potential vectors of risk associated with uploading personal visual content to generative AI platforms. The analysis evaluates five primary threat categories:

1. Personally Identifiable Information (PII) Exposure: Accidental disclosure of sensitive visual elements embedded within uploaded images, including national ID cards, driver's licenses, contact details, financial credentials, or residential backgrounds.
2. Biometric & Facial Identification: Extraction or processing of facial geometry and biometric features without explicit user consent or clear retention boundaries.
3. Metadata & Environmental Tracking: Exposure of Exchangeable Image File Format (EXIF) metadata, including geographic coordinates (GPS), device timestamps, camera specifications, and location contexts.
4. Synthetic Media & Deepfake Exploitation: Downstream misuse of uploaded personal facial media for unauthorized synthetic media generation, identity impersonation, or social engineering.
5. Unauthorized Model Ingestion: Involuntary inclusion of private visual media into public model training corpora, potentially resulting in persistent visual data memorization or accidental generation to external users.

3.8 Data Analysis

Quantitative survey data collected from the 130 respondents were processed and analyzed using Python data science libraries (pandas, numpy, and matplotlib/seaborn). The analysis primary focus includes:

- Descriptive Statistical Analysis: Frequencies, percentages, means, and standard deviations to summarize user demographic profiles, platform usage frequency, and types of uploaded visual media.
- Perception & Likert-Scale Categorization: Distribution analysis of 5-point Likert scale items measuring user privacy concerns, policy reading habits, and risk perceptions across specific threat categories.

- Cross-Tabulation & Trend Analysis: Comparative analysis examining differences in risk perception and protective practices across age brackets, educational levels, and usage frequencies.

3.9 Ethical Considerations

The study was conducted in accordance with standard academic research ethics guidelines:

- Informed Consent: Prior to survey participation, all respondents were presented with a clear consent form outlining the research objectives, voluntary nature of participation, and right to withdraw at any time.
- Anonymity & Confidentiality: No personal identifying visual media, original photographs, or credentials were requested or collected during the study. Email addresses collected during survey administration were strictly anonymized and segregated from response records.
- Data Protection & Storage: Survey response datasets were stored securely in password-protected, encrypted cloud storage accessible exclusively to the primary research authors.

4. RESULTS

4.1 Demographic Profile of Respondents

A total of N=130 valid survey responses were analyzed. The sample demonstrated a strong representation of the primary target demographic of young adults aged 18–30 (76.15% n=99), with the largest single age cohort being 21–23 years (39.23% n=51). The gender ratio was highly balanced, comprising 51.54% female (n=67) and 48.46% male (n=63) participants. In terms of occupational status, students formed the predominant majority (69.23% n=90), split between postgraduate (40.77% n=53) and undergraduate students (28.46% n=37), followed by working professionals (25.38% n=33).

Demographic Variable	Category	Frequency (n)	Percentage (%)
Age Group	Below 18	5	3.85%
	18–20	34	26.15%
	21–23	51	39.23%
	24–26	14	10.77%
	Above 30	26	20.00%
Gender	Female	67	51.54%
	Male	63	48.46%
Occupation Status	Postgraduate student	53	40.77%
	Undergraduate student	37	28.46%
	Working	33	25.38%

	professional		
	Homemaker	7	5.38%

4.2 Generative AI Usage Frequency and Platform Adoption

Generative AI platforms are deeply integrated into the daily digital routines of the surveyed sample. Nearly three-quarters of respondents (74.62%, n=97) report using generative AI tools on an active basis, with 43.08% (n=56) utilizing them daily and 31.54% (n=41) several times a week.

Regarding platform adoption, ChatGPT emerged as the most widely adopted platform (88.46%, n=115), followed closely by Google Gemini (70.77%, n=92), Claude (43.08%, n=56), Meta AI (39.23%, n=51), and Microsoft Copilot (27.69%, n=36). Specialized visual AI models such as Adobe Firefly, ComfyUI, Runway, and Grok were utilized by smaller subsets of respondents (3.85% to 5.38%).

Usage Parameter	Category	Frequency (n)	Percentage (%)
Overall Usage Frequency	Daily	56	43.08%
	Several times a week	41	31.54%
	Rarely	15	11.54%
	Occasionally	10	7.69%
	Never	8	6.15%
Platforms Used (Multiple Select)	ChatGPT	115	88.46%
	Google Gemini	92	70.77%
	Claude	56	43.08%
	Meta AI	51	39.23%
	Microsoft Copilot	36	27.69%

4.3 Personal Visual Media Sharing Behavior

When asked whether they had ever uploaded a personal photo or video to a generative AI platform, 50.77% (n=66) of respondents confirmed doing so, while 38.46% (n=50) reported never uploading personal visual media, and 10.77% (n=14) were uncertain or could not recall.

In terms of upload frequency, 16.15% (n=21) upload personal visual media very frequently, 16.92% (n=22) upload sometimes, 23.08% (n=30) upload rarely, and 33.85% (n=44) reported never uploading visual content.

Types of Visual Content Shared Among visual content types, Educational/work-related images were the most frequently uploaded (52.31%,n=68), followed by Selfies / Personal photographs (44.62%,n=58) and Photos for editing or enhancement (38.46%,n=50). Crucially from a cybersecurity perspective, a notable proportion of users

admitted to uploading highly sensitive media containing personally identifiable information (PII):

- Videos containing other people: 28.46% (n=37)
- Photographs with friends: 25.38% (n=33)
- Family photographs: 20.77% (n=27)
- Videos containing oneself: 19.23% (n=25)
- Identity documents (e.g., ID cards, licenses): 11.54% (n=15)
- Screenshots containing personal information: 7.69% (n=10)
- Primary Motivations for Uploading Visual Content

The leading functional reasons for sharing photos and videos with AI platforms included:

- Creating AI-generated images: 16.15% (n=21)
- Removing or changing image backgrounds: 10.00% (n=13)
- Educational tasks: 9.23% (n=12)
- Image editing or enhancement: 8.46% (n=11)
- Professional/work purposes: 7.69% (n=10)
- Image analysis and visual query: 4.62% (n=6)

4.4 General Privacy Concerns & Risk Beliefs

User perception regarding general privacy and cybersecurity risks associated with generative AI visual uploads demonstrates high baseline concern:

- Baseline Privacy Concern: Over two-thirds of respondents (67.70%, n=88) expressed high privacy concern prior to uploading personal visual media (53.08%, n=69 rating 5/5 and 14.62%, n=19 rating 4/5).
- Information Exposure Awareness: 73.08% (n=95) of respondents were aware that image content can reveal hidden personal details (58.46%, n=76 rating 5/5 and 14.62%, n=19 rating 4/5).
- Cybersecurity Risk Belief: 66.92% (n=87) explicitly believe that uploading personal photos/videos to generative AI platforms creates tangible cybersecurity risks, while 24.62% (n=32) reported being unsure.
- Overall Safety Perception: Only 14.62% (n=19) view uploading personal visual media to AI platforms as safe (11.54% safe, 3.08% very safe). In contrast, 52.31% (n=68) view it as unsafe or very unsafe (31.54% very unsafe, 20.77% unsafe), with 33.08% (n=43) remaining neutral.

Privacy & Cybersecurity Perception Item	Metric Category /	Response Percentage (%)	Frequency (n)
Pre-Upload Privacy Concern Level	High Concern (4-5/5)	67.70%	88
	Moderate Concern (3/5)	26.92%	35
	Low Concern (1-2/5)	5.38%	7
Awareness of Hidden Data in Images	High Awareness (4-5/5)	73.08%	95
	Moderate Awareness (3/5)	11.54%	15
	Low Awareness (1-2/5)	15.38%	20
Belief in Created Cybersecurity Risks	Yes	66.92%	87
	Not Sure	24.62%	32
	Maybe	8.46%	11

4.5 Specific Cybersecurity & Privacy Threat Concerns

When evaluated against specific threat categories, respondents highlighted acute concern over synthetic media misuse, model ingestion, and personal data exposure:

- Deepfake & Synthetic Media Misuse: 53.08% (n=69) expressed high concern (50.00%, n=65 extremely concerned; 3.08%, n=4 very concerned).
- Model Improvement & Ingestion: 44.62% (n=58) expressed high concern regarding their data being used to train AI models (40.77%, n=53 extremely concerned).
- Personal Information Exposure: 60.00% (n=78) expressed high concern over personal details being exposed (40.00%, n=52 extremely concerned; 20.00%, n=26 very concerned).
- Data Storage Duration: 54.62% (n=71) reported concern over platforms storing visual media longer than expected (30.00%, n=39 extremely concerned; 24.62%, n=32 very concerned).
- Unauthorized Access & Loss of Control: 46.92% (n=61) were concerned about unauthorized party access (36.15%, n=47 extremely concerned), and 43.08% (n=56) expressed high concern regarding total loss of control over uploaded media (31.54%, n=41 extremely concerned).

Specific Threat Category	Extremely Concerned (n %)	Very / Moderately Concerned (n %)	Slightly Concerned / Not Sure (n %)
Deepfake / Synthetic Media Misuse	65 (50.00%)	18 (13.85%)	42 (32.31%)
AI Model Improvement / Training Ingestion	53 (40.77%)	32 (24.62%)	45 (34.62%)
Personal Information Exposure	52 (40.00%)	32 (24.62%)	41 (31.54%)
Extended Data Storage / Retention	39 (30.00%)	46 (35.38%)	33 (25.38%)
Unauthorized Third-Party Access	47 (36.15%)	19 (14.62%)	64 (49.23%)
Loss of Content Control	41 (31.54%)	24 (18.46%)	65 (50.00%)

4.6 Policy Awareness, Understanding, and Platform Trust

A notable paradox exists between user desire for transparency and actual policy comprehension:

- Policy Reading Habits: 41.54% (n=54) state they always check privacy policies before uploading media, 15.38% (n=20) check often, 13.85% (n=18) rarely, 10.77% (n=14) sometimes, 8.46% (n=11) never, and 10.00% (n=13) do not know where to find policy details.
- Platform Policy Exposure: ChatGPT privacy documentation has been read by 90.00% (n=117) of users, followed by Google Gemini (58.46%,n=76) and Microsoft Copilot (23.08%,n=30).
- Understanding Confidence: 30.77% (n=40) rated their comprehension of how platforms handle visual data at the absolute lowest score (1/5), indicating severe ambiguity surrounding platform terms.
- Importance of Clear Disclosures: 53.84% (n=70) rated clear platform explanations as vital (4–5/5).
- Reaction to Retention Disclosures: If informed explicitly that visual uploads are retained or repurposed, 54.62% (n=71) would refuse to upload, 15.38% (n=20) would switch platforms, and 12.31% (n=16) would strip sensitive info first. Only 3.08% (n=4) would upload regardless.

Factors Increasing Platform Trust

The top factors capable of enhancing user trust in generative AI visual processing include:

- Clear privacy policy: 94.62% (n=123)
- Clear explanation of data storage: 87.69% (n=114)
- Option to delete uploaded data: 76.15% (n=99)
- Explicit consent before re-purposing: 75.38% (n=98)
- Strong encryption/security measures: 73.08% (n=95)
- Transparency regarding third-party access: 66.92% (n=87)
- Clear retention period details: 53.08% (n=69)
- Option to opt out of model training: 43.85% (n=57)

4.7 Protective Precautions & Future Behavioral Intentions

User Precautions Taken Before Uploading Visual Content
To mitigate risks, respondents adopt proactive visual safeguards:

- Remove unnecessary personal details: 74.62% (n=97)
- Avoid uploading identity documents: 72.31% (n=94)
- Remove metadata/location details: 53.08% (n=69)
- Blur faces: 50.77% (n=66)
- Crop sensitive portions: 47.69% (n=62)
- Verify model training usage: 44.62% (n=58)
- Review platform privacy policy: 42.31% (n=55)
- Use temporary/non-sensitive images: 40.77% (n=53)
- Obtain consent from co-subjects: 29.23% (n=38)
- Take no precautions: 16.15% (n=21)

Behavioral Avoidance & Future Change

- Prior Avoidance: 60.77% (n=79) of respondents have previously avoided uploading specific photos/videos due to privacy concerns. Primary triggers for avoidance included: documents/IDs (68.46%), sensitive content (67.69%), facial visibility (62.31%), location indicators (62.31%), and presence of other individuals (60.00%).
- Willingness to Change Behavior: After completing the survey evaluation, 52.31% (n=68) confirmed they would alter their visual media sharing habits on AI platforms, with an additional 30.00% (n=39) stating they would consider doing so.
- Need for User Awareness: 80.77% (n=105) agreed or strongly agreed that users urgently require heightened awareness regarding AI visual risks. The most recommended educational

channels were College/University workshops (75.38%,n=98), Social media campaigns (69.23%,n=90), and In-platform privacy warnings (66.92%,n=87).

5. DISCUSSION

The rapid integration of multimodal generative AI platforms into daily academic, professional, and creative workflows has introduced unprecedented convenience alongside complex privacy and cybersecurity challenges. By synthesizing the empirical survey data collected from 130 respondents with the structural analysis of platform data-handling policies, this discussion contextualizes key behavioral patterns, risk perceptions, policy disconnects, and protective measures among young adult users.

5.1 The "Privacy Paradox" in AI Visual Data Sharing

A central finding of this study is the prominent emergence of the Privacy Paradox a well-documented behavioral phenomenon where users express acute concern regarding their privacy, yet engage in online sharing behaviors that directly contradict those stated concerns. The empirical data illustrates this tension clearly:

- **High Risk Perception vs. Active Sharing:** While 67.70% (n=88) of respondents reported high privacy concern before uploading visual content and 66.92% (n=87) acknowledged that uploading personal media introduces cybersecurity risks, more than half of the sample (50.77%,n=66) actively upload personal photos and videos to generative AI platforms.
- **Exposure of Highly Sensitive Content:** Despite 73.08% (n=95) recognizing that digital images contain hidden personal details, respondents frequently upload media containing personally identifiable information (PII). This includes photographs with friends (25.38%), family photos (20.77%), videos containing themselves (19.23%), identity documents (11.54%), and screenshots containing personal info (7.69%).
- **This behavioral gap is largely driven by the high functional utility offered by multimodal AI models** such as automated background removal, image analysis, and synthetic media generation. The perceived immediate benefit of obtaining an instant AI-generated output often outweighs the abstract, deferred risks associated with data retention or unauthorized secondary usage.

6. RECOMMENDATIONS

Based on the empirical findings (N=130), policy gap analysis, and threat matrix, a multi-tiered recommendation framework is proposed to establish safer

visual interaction standards in generative AI environments.

6.1 Recommendations for Individual Users (Personal Visual Hygiene)

To minimize personal data exposure while benefiting from multimodal AI capabilities, users should adopt a proactive "Pre-Upload Visual Hygiene" workflow:

1. Practice Client-Side Media Sanitization:

- **Strip Metadata:** Remove EXIF geolocation tags, timestamps, and device details prior to uploading photographs using mobile or desktop metadata-cleaning tools.
- **Redact Visual PII:** Crop out or blur background visual elements such as vehicle license plates, institutional identity badges, house numbers, or financial documents before processing.
- **Obtain Consent for Co-Subjects:** Seek explicit approval from family members, friends, or colleagues appearing in images/videos prior to submitting them to third-party AI platforms.

2. Enforce Strict Exclusion Categories:

- **Zero-Trust Document Rule:** Never upload original official identity documents (e.g., Aadhaar cards, driver's licenses, passports) or screenshots containing unencrypted passwords or bank details into public AI prompts.

3. Configure Privacy Settings & Opt-Outs:

- **Disable Model Training:** Actively navigate to platform account settings (e.g., ChatGPT Data Controls or Google Gemini Activity) to toggle off options that allow user-submitted media to be used for model training and feature fine-tuning.
- **Utilize Temporary Sessions:** Use temporary or non-logged session modes when analyzing non-sensitive images to prevent persistent cloud prompt history retention.

6.2 Recommendations for Generative AI Platform Developers

AI platform providers must move beyond passive legal disclosures and integrate Privacy-by-Design (PbD) principles directly into user interface (UI) and image processing pipelines:

1. Implement Automated In-Platform Pre-Upload Scanners:

- **Real-Time PII & Face Detection:** Deploy client-side computer vision models that scan uploaded images in real time and trigger contextual warnings when sensitive identity documents, facial geometry, or geographic metadata are detected.
- **One-Click Redaction Tools:** Provide integrated editing features (e.g., automated face blurring or document redact buttons) directly within the AI upload prompt box.

2. Provide Granular and Transparent Data Retention Controls:

- **Explicit Opt-In Defaults:** Transition from default opt-in model training to explicit opt-in mechanisms for visual media ingestion.
- **Guaranteed Server Purging:** Establish verifiable server-side data purging schedules (e.g., 24-hour automated deletion for transient visual queries) and provide users with a "Delete All Uploaded Media" action button.

3. Adopt Provenance & Synthetic Watermarking Standards:

- **Implement C2PA (Coalition for Content Provenance and Authenticity)** digital watermarking and machine-readable metadata standards for all AI-generated or enhanced visual outputs to prevent synthetic deepfake misuse and support source traceability.

6.3 Recommendations for Policy Makers and Educational Institutions

Regulatory bodies and educational institutions play a pivotal role in bridging the gap between platform disclosures and user comprehension:

1. Institutional Digital Literacy Modules:

- **University & College Workshops:** As requested by 75.38% (n=98) of survey respondents, academic institutions should integrate practical cybersecurity modules covering AI visual privacy, deepfake risks, and prompt hygiene into student orientation programs.

2. Standardized Regulatory Disclosure Frameworks:

- **Mandatory AI Visual Warnings:** Regulatory authorities (such as CERT-In or MEITY in India) should mandate standard, iconography-based privacy warnings on all AI platforms operating within the jurisdiction, explicitly stating retention

periods and model ingestion terms prior to image submission.

- **Data Minimization Mandates:** Require AI platforms to enforce strict data minimization for multimodal prompts, ensuring visual uploads are retained solely for the duration required to generate the immediate query output.

7. FUTURE SCOPE

To extend the findings of this research, several directions for future empirical and technical investigation are proposed:

- **Longitudinal Behavioral Studies:** Future research should implement longitudinal tracking to observe how user sharing habits, privacy concerns, and safeguard adoption evolve over time as multimodal AI tools become deeper integrated into personal devices.
- **Cross-Demographic and Regional Sampling:** Expanding future survey populations across broader geographical regions, age groups (e.g., adolescents and older adults), and professional domains will allow for comparative cross-demographic risk assessments.
- **Technical Forensic and Interception Auditing:** Technical cybersecurity researchers can perform network traffic analysis, client-side memory inspection, and forensic metadata tracking to empirically verify platform claims regarding EXIF stripping, encryption standards, and server-side data purging.
- **Empirical Testing of In-Platform Privacy Prompts:** Controlled experimental studies can test the effectiveness of different in-platform warning mechanisms (e.g., real-time facial detection pop-ups vs. passive policy links) in deterring users from uploading sensitive visual data.
- **Regulatory Compliance Frameworks:** Investigating the impact of emerging AI governance regulations (such as India's DPDP Act 2023 and the EU AI Act) on platform default privacy settings and mandatory visual data minimization standards.

8. CONCLUSION

- This study investigated the privacy and cybersecurity risks associated with sharing personal photos and videos with generative AI platforms, focusing on user awareness, sharing behavior, platform policies, and protective practices among young adults aged 18–30. The empirical findings reveal a pervasive Privacy Paradox: while a clear majority of users express high baseline concern regarding visual privacy (67.70%) and recognize that uploading personal media introduces cybersecurity threats (66.92%), over half (50.77%) actively upload personal

photos and videos to multimodal AI tools. Crucially, users frequently share media containing sensitive personally identifiable information (PII)—including family photographs (20.77%), videos of co-subjects (28.46%), identity documents (11.54%), and screenshots containing personal details (7.69%).

- While users exhibit high concern over synthetic media exploitation (50.00% extremely concerned) and involuntary model training ingestion (40.77%), platform policy comprehension remains severely lacking. Despite high self-reported exposure to platform terms, nearly one-third of users (30.77%) rate their understanding of how platforms store and process visual data at the lowest possible confidence level, underscoring the ambiguity of current data-handling disclosures. Although respondents actively employ individual safeguards such as removing personal information (74.62%) and avoiding identity document uploads (72.31%) individual digital hygiene alone cannot mitigate systemic infrastructure risks.
- Addressing these challenges requires a synchronized effort between developers, educators, and regulators. AI platform providers must implement Privacy-by-Design principles, including real-time client-side PII scanners, automated face-blurring tools, transparent data retention schedules, and clear opt-out mechanisms for model training. Simultaneously, academic institutions must deploy practical AI literacy and cybersecurity workshops (75.38% user preference) to bridge the gap between risk awareness and actual sharing practices. By establishing transparent platform safeguards alongside proactive user hygiene, young adults can safely leverage the transformative capabilities of generative AI without compromising their personal digital privacy.

(DPDP Act)," The Gazette of India, Extraordinary, Part II, Sec. 1, Aug. 2023.

- European Parliament and Council of the European Union, "Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)," Official Journal of the European Union, L series, May 2024.
- Coalition for Content Provenance and Authenticity (C2PA), "C2PA Technical Specification for Digital Content Provenance," C2PA Governance, Ver. 1.3, 2023.
- P. Mehta, K. Musale, and A. Rajput, "Privacy and Cybersecurity Risks of Sharing Personal Photos and Videos with Generative AI Platforms: A Study of User Awareness and Protective Practices," Primary Survey Dataset (N = 130), D Y Patil Deemed to be University School of Humanities and Sciences, Navi Mumbai, 2026.
- P. Mehta, D. Khatri, A. Dhopawkar, and A. Naidu, "The Dark Side of Digital India: An Investigation into UPI Fraud, Financial Cybercrime, and Payment Security," International Research Journal of Engineering and Technology (IRJET), 2026.
- D. Mungara, H. Sri Ramulu, and Y. Acar, "Security and Privacy Advice for Digital Users in India," 34th USENIX Security Symposium, pp. 6085–6103, 2025.
- Reserve Bank of India, "BE(A)WARE: Booklet on Modus Operandi of Financial and Digital Frauds," RBI Consumer Education Release, 2022.
- Indian Cybercrime Coordination Centre (I4C), "National Cybercrime Reporting Portal Guidance," Ministry of Home Affairs, Government of India, 2026.

8. REFERENCES

- OpenAI, "ChatGPT Privacy Policy and Data Controls," OpenAI Help Center, 2026.
- Google, "Google Gemini Apps Privacy Hub & Data Handling Notice," Google Privacy & Terms, 2026.
- Microsoft, "Microsoft Privacy Statement: Copilot and Multimodal Data Protection," Microsoft Security & Privacy, 2026.
- S. B. Barnes, "A privacy paradox: Social networking in the United States," First Monday, vol. 11, no. 9, 2006.
- Ministry of Law and Justice, Government of India, "The Digital Personal Data Protection Act, 2023