Understanding User Needs and Attitudes for Privacy Protection Tools in Online Visual Content Sharing

CHUN-WEI CHIANG, Purdue University, USA
HARRY YIZHOU TIAN, Purdue University, USA
MING YIN, Purdue University, USA

Visual content shared on social media often includes sensitive elements that can threaten personal privacy. While privacy protection tools—some of which are powered by the state-of-the-art generative AI (Gen-AI) technologies—have been increasingly developed to address such visual privacy concerns by identifying sensitive elements in visual content and suggesting or applying modifications to process the visual content, the success of these tools depends on how well they meet users' nuanced needs and preferences. In this study, we conducted semi-structured interviews with 18 individuals who have either experienced or caused privacy violations in shared visual content in the past to gather first-hand perspectives on stakeholders' privacy concerns, their preferences for how to address these concerns, and their attitude toward the use of generative AI for privacy protection. Our findings highlight that sensitive elements are often not limited to direct identifiers but include contextual combinations and external information that can lead to unintended inferences. Decisions about whether and what to modify are shaped by concerns about privacy effectiveness, content value, content meaning, and emotional or social relevance, while choices around how to modify are influenced by recognition difficulty, visual content integrity, contextual consistency, atmosphere, and usability of modification methods. Participants saw Gen-AI as a promising tool for lowering editing barriers and enhancing creative control but also raised concerns about data usage, manipulation, and transparency. Importantly, we identify tensions between uploaders and depicted individuals, emphasizing the need for shared consent mechanisms and user-centered design in privacy protection. We conclude by discussing design implications for context-aware, flexible, and ethically responsible privacy tools.

CCS Concepts: • **Human-centered computing** → **User studies**; • **Security and privacy** → **Human and societal aspects of security and privacy**.

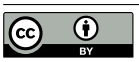
Additional Key Words and Phrases: Visual Privacy, Social Network, Anonymization, Generative AI

ACM Reference Format:
Chun-Wei Chiang, Harry Yizhou Tian, and Ming Yin. 2025. Understanding User Needs and Attitudes for Privacy Protection Tools in Online Visual Content Sharing. *Proc. ACM Hum.-Comput. Interact.* 9, 7, Article CSCW514 (November 2025), 31 pages. https://doi.org/10.1145/3757695

1 Introduction

Visual content—such as photographs and videos—shared on social media plays a powerful role in personal expression and social connection but frequently contains sensitive elements that people do not want to share with the public. Disclosures of sensitive visual content can lead to privacy violations, not only for those who post but also for individuals depicted in the content. Uploaders may be unaware of the risks they impose on others [80, 97], and the rapid dissemination of content across social media platforms amplifies the consequences. This underscores the need for developing

Authors' Contact Information: Chun-Wei Chiang, chiang80@purdue.edu, Purdue University, West Lafayette, Indiana, USA; Harry Yizhou Tian, tian253@purdue.edu, Purdue University, West Lafayette, Indiana, USA; Ming Yin, mingyin@purdue.edu, Purdue University, West Lafayette, Indiana, USA.



ACM 2573-0142/2025/11-ARTCSCW514
https://doi.org/10.1145/3757695

privacy protection tools to assist people in mitigating privacy violation concerns in online visual content sharing.

Over the years, researchers and practitioners have developed many tools to help mitigate privacy concerns in online photo/video sharing on social media. For example, machine learning techniques have been used to analyze visual content and detect sensitive elements that can potentially lead to privacy violations [101, 110]. Once sensitive elements are detected, these tools can provide a wide range of solutions for users to mitigate their privacy concerns. Mechanisms have been designed to dissuade photo uploaders from sharing the photo without communicating with other stakeholders of the photo [20], facilitate multi-party privacy conflict resolutions [97, 107], or allow different parties involved to set their individual privacy settings [52, 64, 86]. When uploaders of the visual content are primarily responsible for addressing privacy concerns, diverse modification methods are provided to them to edit the sensitive elements in photos/videos (e.g., obfuscate the sensitive elements) to decrease the risk of privacy violations while maintaining viewer satisfaction [13, 43, 55, 65].

Most recently, with the advent of generative AI (Gen-AI) technologies, efforts have been made to incorporate Gen-AI into privacy protection tools to enable both the detection of privacy-violating elements in visual content and a seamless modification of visual content. For example, Xu et al. introduced a generative content replacement method to replace privacy-threatening components in a photo with similar and realistic substitutes produced by generative AI [117]. Monteiro et al. developed a tool, allowing users to express their privacy concerns in natural language. They then leveraged generative AI (e.g., multimodal large language models like GPT-4o) to detect sensitive content in photos that relate to users' concerns and recommend obfuscation techniques, which can be powered by generative AI technologies (e.g., diffusion models) [79].

Despite these advancement, the effectiveness of the privacy protection tools—especially the emerging solutions powered by Gen-AI technologies—depends on how well they align with the need and preference of key stakeholders in online visual content sharing, including both the content uploaders and content co-owners (e.g., the individuals depicted in the visual content). To inform human-centered designs of these tools, prior research has collected information on what *third parties* (e.g., viewers of the visual content) consider as sensitive elements within visual content [87, 127, 129], and assessed the effectiveness of different visual content modification methods by asking *third parties* to judge the privacy levels in both original and modified images [43, 44, 66, 67, 117]. However, privacy perceptions are highly subjective [48], and third parties' judgments may differ significantly from those of the *primary stakeholders*, such as the content uploaders or subjects depicted in the images. Moreover, despite growing enthusiasm for integrating the state-of-the-art AI technologies like Gen-AI into privacy protection tools, primary stakeholder's perceptions of these technologies in real-world visual content sharing workflows remain largely unexplored, making it difficult to assess their real-world relevance and to design tools that are genuinely usable and useful for the stakeholders. To foster more trustworthy and human-centered privacy solutions, it is essential to foreground the viewpoints of those directly involved in content creation and sharing. By doing so, we can develop a more nuanced understanding of stakeholder needs and attitudes—including the tensions between content co-owners and uploaders, and the trade-offs they must navigate when adopting advanced technologies like Gen-AI for privacy protection.

Therefore, to deepen our understanding of primary stakeholders' needs and attitudes toward privacy protection tools, we adopt a qualitative approach by conducting semi-structured interviews with 18 individuals who have either caused or suffered from privacy violations previously when sharing visual content online. This allows us to gather first-hand accounts that reflect the nuanced, and sometimes conflicting, roles individuals play in privacy-related decisions. By examining these dual perspectives, we explore not only how people define and respond to privacy threats, but also how they prefer such violations to be addressed. To support the human-centered design and

responsible deployment of privacy protection tools, our study is guided by the following three research questions:

- **RQ1**: What elements in online visual content are considered "sensitive," and how do different stakeholders (e.g., visual content uploader vs. content co-owners who are depicted in the content) differ in defining and identifying these sensitive elements?
- **RQ2**: What factors influence how uploaders and co-owners decide whether and how to modify sensitive elements in visual content to address privacy concerns?
- **RQ3**: What are people's perceptions of the use of generative AI in protecting privacy in visual content?

Our findings highlight that sensitive elements in visual content extend beyond clearly identifiable components (e.g., human faces, identifiable markers, unique appearance of location) to include subtle contextual cues. These include combinations of co-occurring elements that suggest sensitive associations, as well as external information that, when aggregated across posts, captions, or platforms, can unintentionally reveal identities or situations. We also identified a tension between perspectives: uploaders often overlook subtle privacy risks and assess harm based on intent or overt identifiers (e.g., faces), whereas individuals depicted in the content are more attuned to contextual cues and concerned with emotional or social consequences (see §4.1). We also identified differing perspectives in modification preferences, with uploaders focusing on maintaining the consistency of meaning, emotional tone, visual appeal, and the usability of the modification method, whereas depicted individuals emphasized the importance of effective concealment to ensure their privacy is protected (see §4.2). Finally, we found that while users recognize the potential for generative AI for privacy protection in visual content sharing, they also have deep concerns regarding the ethical risks, lack of control over data, and potential misuse, and consistently called for greater transparency in disclosing the use of generative AI should it be adopted for privacy protection purposes (see §4.3).

In summary, our research offers a qualitative, stakeholder-centered investigation into how users perceive and manage visual privacy risks on social media in the era of generative AI. We conclude by outlining design implications for developing human-centered privacy protection tools that are sensitive to subtle contextual cues, reconcile divergent stakeholder priorities, and incorporate transparency and control in the application of generative AI for real-world visual content sharing.

2 Related Work

2.1 Visual Content Privacy Concerns on Social Media

Privacy concerns related to visual content shared on social media often revolve around four primary aspects: "*who*" is depicted, "*where*" the content was captured, "*when*" it occurred, and "*what*" is shown and conveyed. Each of these elements can contribute to privacy violations, making them crucial considerations in understanding the broader implications of sharing visual content online.

Who is depicted in a photo/video can become a primary privacy concern, particularly when the individuals involved have not consented to their likeness being shared publicly [3, 41, 49, 59]. Privacy in this context extends beyond the individual uploading the photo [11]; it also encompasses the privacy of friends [106, 123], significant others [22], family members [2, 78], and bystanders [12, 26, 30, 48, 49, 80, 116]. Moreover, the privacy of vulnerable or marginalized groups, including children [5], individuals with disabilities [4], and members of the LGBTQ+ community [27], requires particular attention. Additionally, identification can occur directly through facial recognition [84, 130], or even indirectly through distinctive features like tattoos [7], body shape and posture [43], or unique accessories [128].

Where a photo/video is taken can reveal sensitive information about a person's life. Location information such as house number [129], private place [47], visited places [36, 45, 48, 120], or places that imply personal information or activity, such as hospitals [64, 90], funeral [30], and street protests [30], can expose someone's life that they may wish to keep private. Additionally, locations that influence social impressions, such as bars [18, 64] and Alcoholics Anonymous meetings [30], can lead to unintended judgments or social stigmatization.

When a photo/video is taken adds another layer of sensitivity, especially when it captures moments intended to remain private. This aspect refers not only to the specific time the photograph was taken but also to the broader temporal context it reveals. For instance, images captured during holidays or weekends—times generally considered more personal—are often seen as more sensitive [64]. Additionally, metadata associated with digital photos, such as timestamps, allows inference about a person's activities and movements [35, 68, 80].

Finally, *what* is depicted and conveyed in the photo or video can encompass a wide range of potentially sensitive material, from the activities being performed to the objects and contextual clues present. Photos capturing embarrassing moments [112] or potentially controversial activities, such as consuming alcohol or using a stigmatized product [10, 11, 94], can be particularly sensitive when shared publicly. Moreover, images containing nudity, sexual content, or other mitigated material are frequent sources of concern [66, 87, 122, 129]. Also, photos that include personal documents, financial information, or other confidential materials can lead to severe privacy breaches, including identity theft or financial fraud [54, 60]. The content with entertaining values would also trade off others' privacy, resulting in a high likelihood of sharing private content publicly [6].

These direct indicators of sensitive content offer a starting point for automated detection, but they fail to capture the nuanced and context-dependent nature of perceived sensitivity. Moreover, what counts as sensitive varies across individuals and situations [1, 8], suggesting that rigid, predefined lists of sensitive attributes are inadequate. These limitations underscore the need for us to obtain in-depth understandings of how different stakeholders define and identify sensitive elements in visual content to inform better privacy violation detections in privacy-preserving tools.

2.2 Social Contexts and Interdependent Privacy in Visual Content Sharing

Privacy concerns around visual content on social media are shaped not only by image content but also by external, contextual factors. *Communication Privacy Management* (CPM) theory [91] explains that individuals regulate disclosures based on privacy rules shaped by demographics, personality traits, and social context. Prior research shows gender differences in privacy attitudes, with women generally more cautious than men [33, 46, 63, 104, 109]. Age also influences privacy orientations. Older adults are more attentive to institutional data handling and often take additional protective measures [126, 131], while younger users prioritize social privacy [92]. Education and geographic location also matter. For instance, those with higher education worry less about information privacy [131], while rural users are more restrictive on privacy practice [37]. Personality traits, like narcissism [73], humor, or dark triad characteristics [6, 42], affect willingness to share content despite privacy risks. Social factors such as peer influence, online communities, and loneliness also shape privacy attitudes [73].

Building on CPM theory, the relationship between photo uploaders and others depicted in or affected by shared visual content introduces complex dynamics of co-ownership and boundary coordination. These dynamics of multiparty co-ownership and conflicting privacy expectations have been a longstanding focus in the CSCW community, particularly in studies of photo-sharing practices and collaborative content management [69, 70, 99, 103]. Conflicts often arise when uploaders share content without fully considering the privacy expectations of co-owners, leading to privacy violations [48]. The subjective nature of privacy perceptions, which can vary widely

between stakeholders, adds another layer of complexity [25, 127]. Additionally, stakeholders may not always realize they have been recorded in shared media [26, 75], exacerbating the potential for privacy breaches. Consequently, multiparty photos often result in conflicting privacy interests among the stakeholders [52, 81, 106, 107]. Similarly, users tend to make careful, thoughtful decisions when selecting recipients for sensitive visual content, balancing privacy concerns with the potential social, professional, and personal consequences of sharing [66].

While this body of work has illuminated many factors influencing visual privacy decisions, there is limited understanding of how individuals, both content posters and those affected, identify which visual elements require protection and how they decide whether or how to modify them. Our study builds on this foundation by investigating users' lived experiences and decision-making processes when employing privacy protection tools for visual content.

2.3 Methods for Privacy Preservation

In response to these visual privacy concerns, researchers and technologists have developed a variety of solutions and tools aimed at preserving user privacy. One common strategy for visual privacy is editing images to obscure sensitive content and reduce identifiability. Early methods like cropping [77, 113] were followed by masking techniques using filters of various shapes (e.g., boxes [57, 58], thin bars [124], silhouettes [58, 67, 89]), edges [58], and point-light [17], and styles (e.g., solid colors [57, 58], avatars [65, 89, 125], pixelation [57], and blurring [14, 34, 43, 57, 67, 108].) These techniques are widely available on social media platforms like Facebook, Instagram, and X, where users can easily add stickers or blurring masks to their photos with a single click. Removing sensitive elements and filling the gaps with plausible content is another privacy protection technique [40, 58]. More recently, generative AI has provided new opportunities for privacy protection by synthesizing plausible replacements that maintain the overall coherence and aesthetic of the image [13, 23, 56, 61, 65, 82, 117]. Another emerging strategy for visual privacy is multi-party privacy control, which acknowledges the presence of multiple stakeholders in shared visual content. While most social media platforms provide basic access controls, they often prioritize the uploader's preferences and offer limited support for negotiation among stakeholders [10]. In contrast, research has explored more collaborative approaches, such as systems that notify identified individuals through automatic recognition [119] or infers appropriate access based on social collective preferences [50, 52, 105]. These efforts aim to shift privacy decision-making from being solely uploader-driven to more participatory and context-aware models.

While technical solutions are promising for privacy protection on social media, understanding users' perceptions and requirements for privacy protection tools and techniques is essential to ensure they effectively address privacy concerns without compromising user experience. Previous research has demonstrated a privacy-utility tradeoff between protecting privacy and maintaining the value or shareability of the content [43, 66, 67, 108]. More recently, Xu et al. investigated the user perception of AI-generated content (AIGC), offering new insights into how generative AI may impact perceptions of privacy and content quality [117]. These works have mainly concentrated on evaluating privacy protection tools and techniques from the perspectives of the viewers of the visual content rather than the key stakeholders of the visual content, such as the creator (and uploader) of the photo/video and people who are captured in the photo/video. However, the stakeholders often attach deeper emotional and contextual significance to visual content, which may lead to differences in their privacy concerns, preferences, and expectations than those of the third-party viewers [102]. This leaves a critical gap in understanding how these key stakeholders perceive, negotiate, and enact privacy within the emotional and contextual richness of their lived experiences. The way decisions around visual privacy are shaped by a mix of personal, relational, and situational factors highlights the importance of addressing this gap. To understand the complexities surrounding

privacy violations in visual content and understand how tools could be best designed to reduce such violations, in this study, we interviewed individuals who had experienced these violations firsthand, whether they were the ones who violated others' privacy or were victims themselves. This dual perspective provides valuable insights into the nuanced emotions, expectations, and ethical dilemmas involved in sharing visual content while preserving privacy.

3 Methods

This study investigates people's needs and attitudes toward privacy protection tools when sharing visual content (e.g., photos, videos) online. These tools serve two main goals: (1) identifying sensitive elements in visual content that may cause privacy violation and (2) enabling users to apply modification techniques, potentially powered by generative AI, to help mitigate those privacy risks. To examine how these goals should be achieved from the perspective of individuals affected by visual privacy breaches, we conducted a semi-structured interview study with participants who had previously experienced privacy violations resulting from visual content shared online.

3.1 Participant Recruitment

To identify potential interview participants, in July 2024 to March 2025, we first distributed a survey to 266 participants, all aged 18 or older and residing in the United States, through Prolific[1]. The survey first collected participants' basic demographic information (e.g., gender, age range, ethnicity). Then, participants were asked to reflect that if they had any previous first-hand experience of privacy violation when sharing photos or videos online, either as their own privacy got violated by others (i.e., they are the "victims" of privacy violation) or they uploaded visual content that violated others' privacy (i.e., they are the "uploaders" of privacy-violating content)[2]. If so, participants were prompted to briefly describe those experiences. At the end of the survey, participants indicated if they would be willing to participate in an interview study later about their privacy violation experience and ways to mitigate similar situations. All respondents of the survey were compensated with 30 cents.

In this survey, 100 respondents indicated that they had been involved in privacy violations due to online visual content sharing. Specifically, 33 respondents had both experienced violations of their privacy and caused privacy violations to others, 40 only had their own privacy violated by others, and 26 had only caused violations to others. We invited all the respondents who reported to be involved in privacy violations and indicated that they were willing to participate in the interview study to take part in a semi-structured interview. In total, 18 participants (9 female, 9 male) agreed to participate. Among them, 7 participants had both experienced and caused privacy violations, 6 had only experienced such violations themselves, and 5 had only caused violations without experiencing them personally. Table 1 provides an overview of the participants' demographic background. While the sample size may appear modest, it aligns with similar other interview-based studies probing into people's privacy-related experience [112, 121]. Recruiting participants for interviews on sensitive topics, especially those involving both personal experiences and actions that may have affected others, poses significant challenges. As we acknowledge the potential limitations on generalizability due to sample size (see more in Section 5.5), we also note that this sample size reflects not only the number of eligible participants who consented to the interview but also the inherent difficulty of obtaining candid and thoughtful reflections on ethically complex situations.

[1]https://www.prolific.com/

[2]Privacy violation was defined as cases where personal information or activities that stakeholders do not want to be disclosed are unauthorizedly revealed due to the visual content sharing.

ID	Gender	Age	Ethnicity	Experienced	Caused
P1	Female	25-34	White	Yes	Yes
P2	Male	18-24	White	Yes	No
P3	Female	35-44	White	Yes	No
P4	Male	35-44	White	No	Yes
P5	Female	25-34	White	Yes	No
P6	Male	35-44	White	Yes	Yes
P7	Male	45-54	Hispanic	Yes	No
P8	Female	35-44	White	Yes	No
P9	Female	25-34	White	Yes	Yes
P10	Male	45-54	White	No	Yes
P11	Male	55-64	Black	Yes	Yes
P12	Female	25-34	White	Yes	Yes
P13	Male	25-34	Black	No	Yes
P14	Female	45-54	White	No	Yes
P15	Male	45-54	Black	No	Yes
P16	Female	65+	Black	Yes	Yes
P17	Male	25-34	Asian	Yes	Yes
P18	Female	25-34	Black	Yes	No

Table 1. The demographics and privacy violation experience in online visual content sharing for the 18 participants of our semi-structured interviews.

3.2 Interview Procedure

Semi-structured interviews were conducted via Webex by the first author. The procedure of the study was approved by the IRB of the authors' institution.

Each interview consisted of three phases. The first phase was designed to explore what elements should visual content privacy protection tools identify as "sensitive" based on participants' real-world experiences of privacy violations. Since all participants of our interview study indicated in the survey that they had privacy violation experiences before due to photos or videos shared online, in the first phase, participants were asked to recall these experiences. They were prompted to elaborate on the content of the photo or video that led to privacy violations, the stakeholders involved, the context of the violation (e.g., the location and time that the photo or video was taken), and the information that was unauthorizedly disclosed. For each privacy violation experience that a participant shared, they were asked to reflect on what they believe as the key sensitive elements in the visual content that caused privacy violations.

In the second phase of our study, we examined participants' perspectives on how they would modify the photos or videos to avoid privacy violations, should they be given the chance to do so in their past privacy violation experiences. This includes reflections on which elements of the visual content they would modify and how they would modify them. Specifically, most of the participants' privacy violation experiences involved the unauthorized disclosure of a person's information due to the presence of the person's face in the visual content (i.e., either the participant's own face or others' faces that showed up in the participant's photo/video). Thus, we usually started this phase by prompting the participant to consider if they would modify the face for privacy protection and then proceeded on to other elements of the visual content that participants would like to modify in order to mitigate privacy concerns. For each element that the participant chose to modify, they were asked to express their preferences on various modification methods, including but not limited

to blurring, removal, replacement, and applying stickers. They were encouraged to evaluate the advantages and limitations of each method. Moreover, we also asked the participant to elaborate on that after the modification of the selected element, whether the resulting photo/video was free of privacy violation concerns, how well it aligned with their intention to share it online (when the participant was the photo/video uploader), and how comfortable they were if others sharing it online (when the participant was the victim of privacy violation). These questions aimed to reveal factors that influence people's decisions on "*what to modify*" and "*how to modify*", which then inform how privacy protection tools should process the visual content to mitigate privacy concerns.

In the final phase, we aimed to understand participants' attitudes toward the use of generative AI in privacy protection tools for addressing privacy concerns in visual content. We first asked participants if they had heard of the generative AI technology, and if so, explain what it is. If the participant indicated that they did not know what generative AI is, or their explanations suggested their understanding was not accurate, we provided the participant with further explanations and clarifications. After that, participants were asked to reflect on their perceived benefits and drawbacks of using generative AI technologies for privacy protection in visual content sharing. Participants were prompted to discuss whether they would use generative AI for modifying photos/videos that have privacy violation concerns, and whether they would feel comfortable allowing others to use these technologies when facing similar privacy concerns. Finally, we also asked about their thoughts on how the visual content should be presented if generative AI technologies were used to modify the content for privacy protection purposes (e.g., whether and how it should be explicitly labeled).

Each of our interviews lasted between 30 and 70 minutes. We audio-recorded all interviews for further analysis, which produced a total of 13 hours and 56 minutes of audio data. Participants were compensated at a rate of $4 per 15-minute interval, with total payments ranging from $8 to $20, depending on the interview duration. Given the sensitive nature of the interview topic, we employed stringent anonymization protocols to protect participants' identities throughout the data collection and analysis process. These included removing all personally identifiable information from interview transcripts. Additionally, any references to specific events or details that could potentially lead to the identification of participants were carefully redacted during the analysis phase.

3.3 Data Analysis

Using TurboScribe, an online transcribing tool[3], we first generated the initial transcripts of all interviews based on the audio recordings. The first author then cross-referenced the transcripts with the original audio recordings to correct errors in the transcripts. Next, we operated reflexive thematic analysis [15, 16] on the interview transcripts, where we adopted an inductive coding approach [96], chosen for its flexibility in identifying emergent themes [16]. Reflexive thematic analysis recognizes that researchers shape how data is interpreted, bringing their scholarly expertise to the analytical process. This approach emphasizes deep researcher engagement with data and reflective analytical discussions, rather than prioritizing inter-rater reliability metrics or consensus-driven reliability metrics [16]. Each transcript was coded by two researchers independently, where they employed both descriptive and in vivo coding [96] to preserve participants' voices while identifying recurring ideas and concepts. Upon the completion of coding of all transcripts, the two researchers collaboratively reviewed and discussed their codes, grouping codes based on similarities and patterns within the data [31, 115]. The two researchers then iteratively refined the themes through ongoing analytic discussions, focusing on thematic clarity, internal consistency, and distinctiveness. Themes were

[3]https://turboscribe.ai

critically reviewed to ensure they reflected participants' experiences and perspectives, captured the complexities of multi-stakeholder privacy, and contributed interpretively to our research goals.

4 Results

In the interviews, 18 participants shared 25 privacy violation experiences due to visual content shared online, either as the person whose privacy was violated or as the one unintentionally violating others' privacy. Through analyzing participants' experiences and perspectives, we identify a few key themes that reveal people's needs and attitudes toward privacy protection tools, including which elements of the visual content should be deemed as sensitive, how to process these elements appropriately, and the perceptions towards using generative AI technologies for privacy protection purposes. Below, we will explore these themes in greater detail.

4.1 RQ1: What Elements are Considered Sensitive in Visual Content Sharing?

In the first phase of our interviews, we asked participants to recall their privacy violation experiences involving privacy violations. Our goal was not to create an exhaustive list of sensitive features but to uncover how stakeholders interpret sensitivity based on their roles and relationships to the shared content.

4.1.1 Sensitive Features from Direct Identifiers. In our interviews, 8 participants shared privacy violation experiences that they believed the key reason for the violation was that someone's face was revealed in the visual content without permission, among which 4 of them involved faces of children or teenagers. However, perceptions of sensitivity varied significantly depending on participants' roles in the content. From the perspective of content co-owners, the visibility of their faces in publicly shared images without consent constituted a key sensitive element, often leading to discomfort and privacy concerns. They associated facial exposure with a loss of control over their online identity and a sense of vulnerability. As P3 explained, "*I just like to have control over my information and my photos. And so I didn't like that people had the power to post things without my permission.*" In contrast, many content uploaders did not perceive faces as sensitive elements. This disconnect was often rooted in a lack of awareness, e.g., uploaders were unaware that others might consider being depicted in the image a privacy concern. As P12 noted, "*I knew that they [P12's friends] personally didn't post pictures, but I didn't really know that they considered it a privacy issue.*" In a few other cases, the content uploaders did not even realize that co-owners were present in the content they posted. As P1 expressed, "*I didn't even realize it at the time, but there was a woman who had a baby stroller in the background.*" This contrast illustrates how privacy boundaries are shaped not just by the content itself but by one's position in the content-sharing dynamic. These gaps in various stakeholders' perceptions and attention illustrate how privacy violations can occur unintentionally, especially when uploaders are focused on their own perspective or goals in sharing the content.

Concerns around images of minors further revealed the complexity of these dynamics. In some cases, parents expressed stronger concern for their children's privacy than for their own, choosing not to share photos of their kids at all. As P10 expressed, "*They [P10's friends] were at an event with their child, and I took photos of them and their child. [...] I never noticed that they didn't post photos of their child.*" These dynamics extended into familial relationships as well. In several cases, parents acting as uploaders overlooked the privacy concerns of their children, assuming they had the right to share family photos. The conflicts emerged between parents and their children, especially as children grew older and developed their own sense of privacy. For example, P2 shared the following experience: "*There would be just a lot of times that she [P2's mother] would insist upon a photo, and I wouldn't want a photo taken. And then, I would relent and I would say that she could take one, but don't*

upload it. And then, she'd upload it. So, I personally feel like it's kind of privacy-violating because I said no, and I was a minor at the time." These cases further underscore the complexity of determining whose privacy preferences should take precedence, particularly in familial relationships.

4.1.2 Contextually Sensitive Features from Internal Co-Occurrence. While individual elements, such as a person's face or a specific object, may not always be considered sensitive on their own, the co-occurrence of multiple elements within a single visual context (e.g., a single photo) can produce unintended inferences that compromise privacy. In 14 cases, participants described how the combination of people, objects, locations, and timing created a narrative that exposed information they or others preferred to keep private.

These co-occurrences often invoked broader social meanings tied to cultural, professional, or legal norms. Participants highlighted concerns when individuals were photographed alongside alcohol, cigarettes, or in nightlife venues, contexts that might conflict with professional roles or public personas. For example, P14 explained how her photo unintentionally violated her friend's privacy: "*There's alcohol in the picture. So, he was upset that he was drinking in the picture because he's a teacher, too.*" Similarly, P4 described how party photos of a friend receiving workers' compensation were later used against them in a legal dispute, despite the friend being genuinely injured. These examples underscore how the interplay between visual elements can lead to misinterpretation, resulting in reputational harm or institutional consequences.

Co-occurrence also impacted perceptions of property ownership and personal safety. P8 explained how being photographed in front of her residence, alongside identifiable landmarks, revealed her home location: "*There wouldn't be a problem [if P8 was not in the photo] because no one will really know what that was.*" Additionally, some participants considered the combination of a person's location and the timing for when the visual content was taken to be sensitive. For example, posting a photo in real time, especially with cues from public spaces, could inadvertently disclose someone's whereabouts. As P9 recalled, "*She [P9's female friend] got really upset with me in turn because he [the boyfriend of P9's friend] was upset that she had lied to him about where she was going to be.*"

In several cases, participants described privacy violations that arose not from explicit personal information being shown but from how the visual content implied social affiliations, intentions, or emotional stances, particularly in sensitive interpersonal contexts. For instance, P17 recalled one photo that included a friend's ex-partner, noting, "*We [P17 and her friend's ex-partner] took selfies.*" The friend saw the co-occurrence of P17 and the ex-partner as symbolizing an alliance with him, and took it as a breach of social and emotional privacy norms: "*It was foreseen as support for him [the ex-partner] ... And she was already struggling with depression... So it was a friendship that has ended.*" This case illustrates how the co-occurrence of individuals in a photo can carry inferred social meanings. Rather than any single visible element, the relational context and the viewer's personal knowledge led to a deeply personal sense of privacy violation.

Co-owners were often aware of these nuanced risks, recognizing how combinations of people, locations, or timing might inadvertently disclose their sensitive information about relationships, behavior, or identity. In contrast, uploaders frequently overlook these subtleties. Several uploaders noted they were unaware of the potential for harm until the affected individuals expressed concern. P1 had shared a photo that unintentionally included a woman with a stroller, revealing her parental status, which she had chosen to keep private. As P1 admitted, "*I did not realize that that was a big privacy breach beforehand because I was inexperienced. But afterwards, I could definitely see the logic. I was very respectful in taking it down.*" This highlights an asymmetry in privacy perception: While co-owners often interpret shared visuals through the lens of personal context and lived experience, uploaders may lack the background knowledge needed to anticipate these risks, resulting in privacy violations that are unintentional but deeply consequential.

4.1.3 Contextual Sensitivity from External Information, Cross-Photo Inferences, and Repurposing. In some cases, contextual sensitivity often emerged not from what a visual explicitly revealed, but from how it could be interpreted or combined with external information. Participants shared incidents in which partial features, though not overtly identifying, enabled others to infer personal information by cross-referencing with other publicly available data. For example, one participant described an experience involving an explicit selfie shared on Discord. She intended to obscure her identity by covering her face, yet the photo included a tattoo and family pets. A third party was able to link the photo to her Reddit account, where her real name had once been shared, thereby compromising her privacy. As she explained: "*I had my phone like this [using her phone to cover her face]. But I do have an identifying mark on my shoulder. I have a tattoo. So, that was in the photograph. Somebody had found my information, my real name from Reddit that I had posted years ago. ... My animals [also appeared in the photo], which in hindsight, might be how they found my Reddit.*" This case underscores that even seemingly minor visual cues, such as tattoos, body shape, or personal belongings, can carry identifying value when aggregated with external data. It suggests that users may underestimate these risks and that privacy tools should account for features beyond facial recognition.

In other cases, privacy sensitivity emerged from the combination of multiple posts or external contextual clues, rather than any single image. For example, P12 recounted an instance where her privacy was compromised due to two consecutive photo posts. The first image showed the uploader standing outside her residence, while the second depicted her and the uploader inside the residence. P12 believed that while viewers of the first photo itself may not know the photo captured the neighborhood of her house, such a connection could be built when they saw both photos. As she stated, "*I felt like the context, people might have already seen the front of my house. So I just didn't want to be associated with it [P12's residence].*"

Participants also expressed concern over how post-processing actions, such as cropping, captioning, or re-uploading, could shift a photo's perceived meaning or draw attention to unintended details. P18, for example, recounted how a casual photo taken by her hairstylist, intended to showcase a hairstyle transformation, spread across multiple platforms and was eventually used without her consent to promote a product: "*They [the wholesale site] were using the picture to sell a type of brush. ... which was a lie. But it was really weird to me to see how fast that picture had spread.*" Although the original photo was not perceived as sensitive, its repurposing in a misleading commercial context resulted in reputational discomfort. Such cases often arise not from accidental sharing but from the deliberate actions of third-party uploaders or platforms seeking to exploit visual content for promotional or misleading purposes.

In more severe cases, participants described malicious repurposing that resulted in direct harm. For example, one participant had shared a photo of herself at the beach on social media platforms, which by itself did not compromise her privacy. However, a third party (who knew her personally) later reposted the photo along with her contact information to a compensated dating platform. The photo itself did not reveal sensitive information, but its new context resulted in harassment and unwanted attention. These examples highlight that privacy violations are not always rooted in what a visual shows explicitly. Instead, they may emerge through reinterpretation, sequencing, cross-referencing, or malicious repurposing, often in ways unintended by the original uploader. Effective privacy-preserving tools must account for these broader risks and support users in anticipating how content may evolve or be exploited after publication.

4.2 RQ2: What Factors Influence People's Attitudes Towards How to Process Sensitive Visual Content?

In the second phase of our interviews, participants shared their perspectives on how they would process the visual content to avoid privacy violations. Based on their first-hand privacy violation experiences, we identified key factors that influence decisions regarding which elements in visual content to modify and how such modifications are carried out. These decisions were shaped by a combination of personal, contextual, and practical considerations. The findings helped us to understand how privacy protection tools could be designed to process visual content to improve their usability and user experience.

4.2.1 Factors Influencing Whether and What to Modify. We identified a few factors that influence people's decisions on whether to modify the visual content to mitigate privacy concerns, and if yes, how to select which sensitive elements of the visual content to modify.

Effectiveness of privacy protection after the modification. A key factor that influences people's decisions on whether and what to modify in the visual content is whether the modification (on the selected element) can effectively mitigate the privacy violation concerns. As such, most of the uploaders and co-owners in our interviews agreed that directly modifying the sensitive element(s) was an effective approach for processing the visual content to mitigate privacy concerns. For example, P13, as an uploader, considered a case where the privacy violation was caused by the presence of a person's face and noted "*I think [face modification] should solve the problem because no face [and] no kids [in the photo]. So, you shouldn't be stressing if your face is not showing.*" However, participants who experienced privacy violations due to the unauthorized disclosure of their sensitive element(s) shared mixed views. For instance, P12, whose private property was disclosed due to the visibility of their house number, noted "*I think that [house address modification] actually would have probably solved my issue with it. If it [house address modification] could have covered up where the location was exactly, I think that would have been good.*" In contrast, P3 expressed, "*I feel like it's still mine [after face modification]. Like, I want to have some sort of control and ownership over it.*"

Meaning consistency of the visual content after the modification. When deciding which sensitive elements in the visual content to modify, participants considered not just the presence of individual sensitive features, but how those features contributed to the focal point or overall meaning of the photo. In many cases, the focal point was defined by a combination of elements, such as people, objects, location, and time, that together told a particular story. Modifying the core elements risked undermining the intended context or emotional significance of the image. As P4, a photo uploader, noted, "*I think that [modifying the location] might change the overall purpose of the photo [party] too much. Especially if you're using it to kind of document that you and somebody else were doing something, and you place that in a different context, kind of destroys the initial intent there.*" Similarly, P2, whose parents frequently documented him and shared the images without his consent, reflected, "*[If my face is modified], what would be the point of the picture then?*"

In contrast, when the sensitive elements are not the focal point, uploaders preferred to modify less central elements to protect privacy while keeping the focal narrative intact. For instance, when a photo shows a person alongside a sensitive object, such as alcohol, modifying the sensitive object was often a preferred solution as it addresses the root cause of the privacy concern without changing the overall meaning of the visual content. As a victim, P7 described, "*Probably the biggest thing is the alcohol. Anytime you throw in any type of alcohol or substance [to an event], it just makes it seem unethical.*" P14, who uploaded a photo that depicted a friend having alcohol, stated "*I think that [modifying the sensitive object] is an interesting concept because it seemed like to me, what he*

was upset about was obviously the alcohol. ... And a good time does not always mean you have to have alcohol. That is for sure. So that [alcohol] is not the point of the picture."

Perspective-taking and recognition of others' privacy needs. Participants' decisions about whether to modify the visual content and which elements to modify were sometimes shaped by their ability to empathize with those depicted in the image. This perspective-taking often emerged from their own life experiences. P1, who had previously shared a photo containing a child's face, now preferred not to upload photos containing children's faces as she put herself into the parent's shoes, "*I probably would have now being older and having been pregnant myself. I understand her [the child's mother] thought process [of taking down the photo] from that aspect.*" This shift in perspective led some participants to reevaluate past decisions or adopt a more cautious approach to sharing.

Additionally, some uploaders described how recognizing the co-owners' privacy needs influenced their decisions to modify visual content in ways that respected those individuals' preferences. As a photo uploader, P4 values not only the effectiveness of modifying the victim's face but also empowering victims by giving them editing control: "*That [modifying face features] could be helpful. ... If they [victims] had an option to remove themselves from a picture, that would also be helpful.*"

Tradeoffs between content value and modification effort. We found that when the elements in photos or videos are highly sensitive or difficult to modify, people may choose not to modify the visual content at all and simply not share the content. In cases where the image was not particularly meaningful, they preferred to avoid posting rather than investing effort in editing. P10 reflected this view, stating, "*At that point, why go through the extra work of editing the person out, and just not posting a photo? ... I think that [modifying the photo] is a solution, but I just don't know that it will be a common solution.*" This illustrates a cost-benefit logic, where the perceived emotional or social value from posting the visual content had to outweigh the burden of modification.

4.2.2 Factors Influencing Decisions on How to Modify. After selecting the elements they prefer to modify in the visual content, participants also reflected on how they would like to modify them. Based on their reflections, we found a few key factors that influence people's selection of modification methods.

Recognition difficulty. The primary goal of visual content modification is to protect individuals' privacy by concealing sensitive elements, including people, locations, time, and objects, in visual content. Naturally, after the modification, how easy or difficult it is for viewers of the visual content to recognize or infer the sensitive elements originally presented in it is a critical factor that influences participants' choice of modification methods.

Participants worried that subtle modifications like blurring might still allow viewers to infer the hidden information. For example, P7 indicated "*I would be more comfortable if it was just removed. Um, if I went the edit route, then some type of blurred out. ... Like in my example, people know that these are my friends or coworkers and they see one person blurred out, they're going to be able to kind of tell, "Hey, who's missing?" "Oh, it's XX [P7's name]." "Why is he blurred out?" And it leads to more questions.*"

Another important consideration was how robust the modification technique is in resisting reverse engineering attempts. To increase the difficulty of recognition, participants combined multiple techniques. As P1 described her strategy, "*I like to blur it first ... and then put the overlay of the emoji on top.*" This layered strategy adds complexity, making it harder to uncover or reverse the modification and reveal the concealed content.

In addition to removing sensitive visual elements, participants proposed another approach of adding new elements, which are commonly perceived as identifiable and unchangeable, such as tattoos and buildings, to mislead viewers. For instance, P11 suggested a way to protect other's

privacy: "*She didn't have any tattoos, so if you could even add tattoos to the hands, I think that will work.*"

Integrity of visual content. Participants' choice of modification methods is significantly influenced by their perceptions of whether the modification can preserve the integrity of the visual content, including aspects such as quality, aesthetics, size, and shape. Several participants noted that clunky modification techniques can divert audience attention, thus diminishing the overall quality of the image. As P6 explained, "*I feel like putting a sticker on a photo deteriorates the photo overall. If you have a sticker on top of the photo, I think that's the most obvious and eye-catching part of the photo now.*"

Participants remarked that although removing sensitive elements by cropping may not drastically impact a photo's quality or visual appeal, it does alter the photo's dimensions, which could compromise its integrity. P1 expressed, "*I wouldn't have cropped the image for the integrity of the photo.*" To mitigate the negative effects of cropping, P4, an amateur photographer, offered an alternative approach: "*Let's say there's a group of people and one person doesn't want to be in the video. You could always have the camera kind of pan in and move out, like, a rack focus or whatnot, and avoid that person being in there without just a giant black screen.*"

On the other hand, some participants suggested that as public awareness of privacy concerns increases, society shows greater tolerance for the quality reduction of visual content resulting from modifications. P4 noted this shift in social attitudes, remarking, "*These days, everybody knows that people have reasons to not want to be on social media, or not want to be tagged for something. ... It [blurring individuals] is certainly a more acceptable thing today.*"

Consistency with the context. Participants expressed preferences for modification methods that keep the consistency between the modified photo/video and the original one, as well as between the modified part of the photo/video and the rest of it. For instance, blurring sensitive elements was often considered as allowing much of the context in the original photo/video to be preserved. P4 noted, "*I prefer a blur because it still shows that a person was there.*"

On the other hand, participants noted that the effectiveness of covering sensitive elements with stickers or AI-generated content in maintaining context consistency largely depends on implementation. For instance, P1 described adding a puppy emoji over people in the background of a dog park photo to protect their privacy, explaining, "*It would make sense for me to put a puppy emoji over top of the woman's face because the content is about dogs. But I would think that if the emoji were not relevant to the content posted, there could be potential for some questions about what people are trying to convey.*"

Furthermore, participants expressed concern that poor implementation could distort original cultural messages in photos and videos, particularly those related to race, gender, and religion. For instance, P13 highlighted the potential misalignment caused by changing an individual's ethnicity in an advertisement, explaining, "*If the video is supposed to be portraying a message based on ethnicity, I think it [covering individuals with stickers or AI-generated content] would rather cause a problem because it means that the message you're trying to push out there would not be pushed out. So if trying to promote an Asian business, an Asian video with an Asian man in there. The Asian man in there would sell more to the Asian crowd. If you end up changing the ethnicity, it means you might not end up reaching the Asian market you're trying to reach.*"

Atmosphere. Participants employ modification techniques not only to protect privacy but also to enhance or adjust the atmosphere of visual content, allowing photo and video uploaders to express themselves further through the editing process. As P13 described, "*I am able to create whatever I want to create. I think it would depend on my mood. When I'm looking at a video picture, if it's making*

me feel happy, I would want to put some happy face on there. Be it a sticker, be it someone else's face, a meme face, or something."

In addition, different modification methods evoke distinct atmospheres. For example, the application of stickers is often perceived as a casual and informal approach. As P12 noted, "*I personally prefer to put little stickers over my face. Sometimes on Instagram, I'll post a picture when I'm not wearing makeup or I don't look great, but I think the rest of the picture is cute.*" In contrast, blurring is viewed as a more professional and formal method. As P3 illustrated, "*If I was using a sticker, it would be in more of a casual way, or like fun, or not as serious, but was blurring it in some sort of professional environment.*" Moreover, removing sensitive elements without visible signs of alteration is often considered a more refined technique. For instance, P10, a professional visual editor, elaborated, "*Even if I'm not a big fan of AI, but if it were AI-driven, it could run through a set of steps and programmatically remove the entire person and that would be the most elegant method.*"

Usability. Participants emphasized the need for simple and straightforward modification techniques that do not require extensive technical expertise or effort to use. Techniques such as blurring or applying stickers were frequently preferred for privacy protection due to their usability. As P11 noted, "*I think blur or emoji. I think those are better. It's simple and straightforward.*" Importantly, 8 participants explicitly highlighted the need for greater control over the modification process. For instance, P13 noted, "*The ideal way would be if I'm able to change exactly what I want to change.*" P10 highlighted the necessity of human supervision, indicating that users should retain final decision-making authority to verify that outcomes align with their expected results. Participants also believed the affordance of modification methods is important and suggested privacy protection tools could follow design patterns used in existing photo/video editing tools. For instance, P9 mentioned, "*[The modification methods should be] like Photoshop, the way that they have to just edit pictures and the basic apps to edit pictures and stuff is how I'd expect to use it.*"

4.3 RQ3: What are People's Perceptions of the Use of Generative AI in Visual Content Privacy Protection?

Finally, in the third phase of our interviews, participants shared a range of perceptions about using generative AI for visual privacy protection. Their reflections captured both enthusiasm for its creative and efficient capabilities and concern about its ethical, practical, and social implications. Below, we present themes grounded in these perspectives.

4.3.1 Perceived Benefits of Generative AI for Privacy Protection. Participants identified several potential benefits of using generative AI for visual content privacy. These benefits primarily centered on how AI can empower users with limited technical skills to take control of their visual data and how it can help clarify the intended message of an image while removing or masking sensitive elements. The following themes illustrate how participants viewed generative AI as a promising tool for enhancing both privacy and expressive control.

Gen-AI as a support for user autonomy and accessibility. For participants with limited technical skills, Gen-AI offered a way to engage with complex editing tasks without needing prior experience. For example, P3 stated that, "*I like the ability of AI, for those of us without a lot of skills*", framing Gen-AI as a bridge for users with limited technical experience. In addition, participants noted that generative AI could streamline the traditional labor-intensive process of high-quality photo editing for privacy protection. For instance, P10, who had experience in Generative AI-powered editing tools and traditional editing tools, mentioned, "*It [Generative AI-powered editing tools] would save time for the image editor.*"

However, we note the efficiency benefit of Gen-AI in visual privacy protection was not universally perceived. Some participants perceived AI-based editing to be time-consuming or cumbersome

when simpler methods sufficed. P5 expressed, "*I like the idea of being able to easily do that [by] myself. And maybe the downside might be time. It sounds like it could be time-consuming.*" These responses reflect that while Gen-AI can reduce technical barriers, it does not always align with participants' expectations for convenience.

Enhancing privacy and intent through generative AI mediated content modification. Many visual content uploaders described generative AI as a helpful tool for enhancing privacy and emphasizing the intended message of their media. Some participants saw it as a means of removing distracting or sensitive information while sharpening the viewer's attention on the original subject. As P7 mentioned, "*The pros are obviously the ability to remediate an issue [about privacy] that you may have with a picture being posted.*" Others emphasized the technology's role in refining communicative intent. As P4 remarked, "*The pros of some generative AI [are] that they can definitely help you better express the initial intent and idea of the photo.*" Beyond privacy, participants highlighted generative AI's creative potential for expressing one's vision more freely. For instance, P1 said, "*If you always wanted to see yourself in front of Niagara Falls, you can do that now. Or if you wanted to create a short little video of your dog running in front of the Sahara Desert, you can do that [with generative AI].*" These findings show that some users view generative AI not just as a corrective tool, but as a creative aid that enhances both privacy and communicative intent.

4.3.2 Perceived Concerns and Trade-Offs of Using Generative AI for Privacy Protection. Despite recognizing its potential, many participants expressed concerns about the limitations and risks of using generative AI in privacy-related image editing. These concerns spanned technical, ethical, and social dimensions, including authenticity, manipulation, unintended exposure, and even malicious misuse. This section highlights the tensions and uncertainties participants felt, revealing the complex trade-offs they perceived in adopting generative AI for privacy protection.

Tensions between authenticity and AI-generated edits. Participants expressed concerns that AI-generated modifications risk compromising the authenticity of visual content. As P3 questioned, "*Is it considered art or reality?*" Similarly, P4 pointed out, "*A photo is supposed to be some of the most solid documentation of how things were at any given point of time and space, and manipulation of those can bring some of that into question.*" Due to concerns about authenticity, some participants expressed strong opposition to using generative AI for privacy protection. For example, P2 claimed using generative AI for privacy protection was worse than revealing sensitive elements, describing, "*I'd rather just be censored entirely than be edited into someone else.*" Likewise, P1 expressed a strong preference for authenticity, noting, "*[I will] not edit myself because I'm a huge proponent of being authentic. ... I like people portraying themselves as they are, and I would not edit other people unless it was blurring or cropping out.* These views reflect a discomfort with AI-based edits that cross into manipulation or misrepresentation, particularly when authenticity is core to a participant's self-image or ethical beliefs.

Concerns about imperfection in AI-generated edits. Another major concern that participants shared on the use of generative AI for privacy protection is that AI-generated content can be potentially inaccurate and imperfect. Participants frequently noted that generative AI could produce flawed or unnatural visual outputs, which may compromise the credibility or aesthetic integrity of the edited content. As P9 noted, "*You can still spot different flaws, like fingers and stuff like that.*" The capacity to identify AI-generated content and recognize its errors varies significantly among individuals. P14 described, "*My son likes to point things out that I don't even recognize as being AI-generated with photos, and he'll point things out to me that he finds to be really glaringly obvious.*", illustrating how some viewers may be more aware of the inaccuracies of AI-generated edits than others. These imperfections can also increase the burden on the user. As P12 observed the need for repeated attempts to achieve satisfactory outcomes: "*Sometimes you get slightly wonky and strange*

results, so you have to do it a few times.", highlighting the additional time and effort required to achieve acceptable edits. These findings underscore how users' perceptions of Gen-AI for privacy protection are shaped not only by its conceptual promise but also by their direct experiences with technical imperfections. These experiences influence their trust in its reliability and suitability for privacy-sensitive use cases.

Perceived risks of malicious and misleading use of generative AI. While generative AI presents potential opportunities for privacy protection, many participants, particularly women, voiced concerns about its potential for misuse. These concerns centered not on the tool itself but on how it might be repurposed by others with harmful intent. As P5 stated, "*If they change the picture to protect my privacy, that would be fine with me. If they had some sort of bad intentions behind editing the video or the picture, that would bother me.*" Participants described a range of possible harms, from personal violations to broader societal consequences. For example, P7 warned of the reputational dangers of synthetic misplacement, stating, "*[With generative AI,] you can put people in the wrong place at the wrong time.*" Also, P9 illustrated her concerns: "*You could take a picture of anybody and [edit] out their bathing suit. This was a huge issue not only with people's privacy but with children having pictures taken of them.*"

These risks were seen as magnified in online environments where visual content spreads rapidly and is rarely fact-checked. Participants expressed concern about the broader implications for public discourse and truthfulness in digital media, stating, "*It can create a false narrative about someone or an event that actually didn't happen. And especially online nowadays, [when] people sharing things, most of whom don't [take] great research into information about what is factual and what is not factual.*" These comments reflect a growing anxiety that generative AI may not only endanger individual privacy but also compromise societal trust in visual documentation. As AI-generated images become more convincing, participants worried that the line between fact and fabrication may blur, leading to misinformed judgments, reputational damage, or public confusion. This concern highlights that the perceived risks of generative AI extend beyond personal harm to include broader epistemic and social consequences.

Concerns about data use in generative AI training. Participants with technical backgrounds raised significant concerns that using these tools could introduce new forms of privacy leakage. P10 described unease over how training data is acquired, noting, "*Photos are being used without the permission of the originators to train the generative AI to generate these images. There are additional ethical concerns there, too.*" Participants worried not only about past data usage but also about what might happen to images uploaded during the editing process. P9 further explained, "*If I'm putting my pictures into that [AI-powered] software, is it still going to be just my picture? Can it be used to train the AI to generate other pictures? Will my face end up coming up in some generated AI picture?*" These reflections highlight a layered tension in participants' perceptions. Although generative AI offers mechanisms for protecting visible identities, it simultaneously prompts anxiety over invisible data flows. As such, privacy protection through generative AI was not viewed as inherently safe; instead, it was seen as a potential trade-off between visible privacy (e.g., obscuring features) and infrastructural privacy (e.g., control over data ownership and usage).

4.3.3 Normative Expectations for AI-Modified Visual Content. In addition to functional and ethical concerns, participants articulated expectations for how generative AI should be used, disclosed, and governed, both at the platform level and in interpersonal contexts. These expectations reflected values around transparency, control, and consent, especially when AI-modified images were shared publicly or involved other people. The following themes illustrate how participants envisioned more respectful, accountable uses of generative AI in visual content.

Balancing automation and manual precision in generative AI tools. While participants recognized the potential of AI for automating privacy protection, many emphasized the importance of maintaining control over the editing process. Rather than fully delegating visual modifications to AI, they expressed a preference for hybrid systems that allow for human oversight and fine-tuning. For instance, P16 stated, "*I'm not an advocate of anything automatically. I think that checks and balances needs to be there... as long as there is a quick recovery mode.*" Others envisioned interactive, iterative workflows in which users act as decision-makers rather than passive recipients of AI output. P10 described a preferred system where "*It [AI] goes through like 20 or 30 different iterations ... and presents 2 or 3 options that are tasteful... The person running the software can still say, "Oh, no, that's not an iteration that I want. That causes a problem". *" This reflects a desire for preview, selection, recovery, and control, reinforcing the idea that even well-functioning AI should operate under human guidance.

Consent and control in AI-modified visual content. Participants expressed a strong desire for agency over how their images are edited using generative AI, particularly when they are the co-owners of the visual content. A recurring concern was the need for prior approval before AI-modified photos or videos were posted publicly. For example, P6 shared, "*I may ask to do it myself first because I'd be more comfortable in being able to change it myself to the way I like it and then let them [photo takers] share it.*" P12 echoed this emphasis on control, stating, "*I wouldn't use it on other people, probably, unless they specifically said it was okay. [If not,] I would rather use a sticker than generate a fake face for them.*" Participants were not opposed to AI modifications in principle but stressed that transparency, preview, and mutual agreement were necessary conditions for ethical use. These expectations reflect a broader shift toward shared authorship and negotiated control in AI-mediated edits, especially in privacy-sensitive contexts.

AI modification disclosure on social media as a double-edged sword in privacy protection. Across the board, participants emphasized the importance of transparency in how generative AI-modified images and videos are presented on social media, even when used for privacy protection. This concern was voiced consistently, regardless of whether individuals had experienced privacy violations themselves or had previously caused them. As P3 said, "*I think if AI is used to modify a photo in any way, there should be some sort of disclaimer associated with the photo.*"

However, participants' opinions varied on the extent to which transparency should be enforced. Several participants advocated for explicit labels or access to the original content to prevent misinterpretation. For instance, P7 underscored the need for access to unmodified images, noting,: "*[If] there's a picture of you doing something wrong, but I don't know that it's modified. Then, an original photo that could be shown that "I wasn't doing that" or "I wasn't there".*" Others suggested that platform should record all the changes made by generative AI editing tools. P3 noted, "*The metadata of the photo has to indicate that there have been all of these changes.*" Conversely, some participants felt that disclosing every detail of an AI edit could undermine the privacy protections those modifications were intended to provide. P12 expressed the tension, stating, "*I think it's probably good to label it overall. I don't think you should have to say what you changed with the AI because then that kind of defeats the purpose.*" This illustrates a key dilemma in participants' perceptions. While labeling fosters accountability and helps audiences interpret visual content, excessive transparency may erode the protective function of privacy-enhancing edits.

Finally, some participants extended the discourse beyond image/video modification transparency to encompass the transparency of the entities providing AI-powered editing tools. P10 voiced that ethical considerations and transparency in AI development and deployment were center to his decision-making process as a consumer, and he had a preference for companies that openly communicate their values and practices.

5 Discussion

Based on the findings of our interview study, in this section, we reflect on the implications for and challenges in designing privacy protection tools to satisfy user needs and preferences, summarize the lessons learned on how generative AI technologies could be appropriately used for privacy protection purposes, and discuss the limitations of our study.

5.1 Contextual Challenges in Privacy Protection

Our findings highlight how visual privacy concerns are deeply shaped by context, revealing a potential disconnect between the perspectives of stakeholders who are directly involved in privacy violation instances and those of external evaluators. Participants described privacy not as a static checklist of sensitive elements but as a fluid judgment shaped by situational factors, emotional associations, and interpersonal dynamics.

This complexity aligns with the theory of *Contextual Integrity* [83], which conceptualizes privacy as the appropriateness of information flows within specific social contexts. According to this framework, privacy norms are defined by who is sharing information, with whom, what kind of information is being shared, and under what conditions. Our data supports this view. For example, one participant in our study intentionally shared explicit content publicly that external evaluators might classify as highly sensitive [66, 87, 129]. Yet, for the participant, the sharing was appropriate given their self-defined norms and intended audience. These cases illustrate how standard AI-based approaches that rely on fixed taxonomies of sensitive content may overlook the subtle but crucial contextual factors that shape privacy boundaries.

Beyond individual content, our findings further underscore the role of external information and temporal sequencing in shaping contextual sensitivity. Participants noted that the combinations of multiple information, such as location indicators, timestamps, and identifying features across posts, can lead to unintended inferences about an individual's whereabouts or routines. Importantly, participants described how content that appeared harmless in isolation could become privacy-invasive when combined with other posts or viewed over time. Additionally, repurposing previously shared content, such as re-uploading an old image in a new context, was perceived as a violation of contextual expectations. These findings suggest that privacy risks are not just about what is visible in a photo but how meaning evolves across time, platforms, and social relationships. These phenomena reflect the contextual integrity framework, which emphasizes that informational norms are contingent on what, when, where, and how information is disclosed. Moreover, it reveals the limitations of privacy violation detection systems that focus solely on discrete elements rather than examining the broader patterns of content dissemination.

5.2 Power Imbalance and Decision-Making Authority in Visual Content Privacy Protection

Our findings illuminate a recurring tension between content uploaders and co-owners who are depicted in the content. Participants described situations where uploaders made unilateral decisions about whether and how content was shared, often without consulting the individuals depicted. This power imbalance significantly shapes privacy outcomes. In some cases, this led to discomfort or conflict, particularly when the shared content revealed personal or sensitive aspects of someone's life. This dynamic is captured by the *Communication Privacy Management (CPM)* theory [91], which frames privacy as a process of negotiating boundaries around shared information. According to CPM, when multiple people are involved in a piece of information, such as visual content, each has a stake in how it is disclosed. When only one party (typically the uploader) exercises control,

the absence of mutual agreement can result in "boundary turbulence," where differing privacy expectations lead to misunderstandings or harm.

Several participants voiced a desire for more collaborative control, where individuals depicted in a photo could request its removal or approve modifications before it is shared. For instance, participants noted that they would prefer to modify content themselves before allowing someone else to post it, highlighting the importance of agency in how one's image is managed. These views reflect a broader push for systems that support co-ownership and consensual decision-making. Notably, most of the uploaders we interviewed expressed a willingness to take down or modify visual content if the stakeholders asked them to do so. However, these uploaders also acknowledged that they often failed to recognize the sensitive nature of certain elements, underscoring how privacy is highly contextual and subjective. This finding complements our earlier discussion on contextual challenges in identifying sensitive elements and further illustrates the difficulty of ensuring informed consent when privacy perceptions are not mutually understood or communicated.

Despite the interest in collaborative control over visual content, such features are rarely supported on mainstream social media platforms. This gap reflects broader structural and cultural factors. First, platform incentives are often misaligned with privacy protections. Most social platforms are optimized for rapid sharing [19, 88], user engagement [28, 29], and content virality [39]. These goals may be hindered by mechanisms that require waiting for approval or negotiating consent among co-present individuals. Second, the prevailing social norm on social media platforms is always uploader-centric [9]. Social media usage typically assumes that uploaders manage their own boundaries through tagging, untagging, or private messaging. These practices place the burden of privacy protection on those affected, rather than building in proactive, consent-based safeguards. Finally, there is a lack of regulatory or legal requirements mandating shared decision-making over visual content [93]. In many jurisdictions, responsibility for managing privacy risks falls entirely on users, and platforms face minimal liability for content that depicts others without their consent. These structural and cultural constraints help explain why, despite their ethical importance, co-consent tools remain rare in current platform designs.

Together, these findings highlight the limitations of uploader-centric norms and the urgent need to reimagine visual content sharing as a more equitable and collaborative process. Addressing the power imbalance between uploaders and those depicted in visual content requires more than individual goodwill. It also demands structural support from platforms and intentional design interventions.

5.3 Designing Context-Aware and Consent-Based Visual Privacy Tools

Building on the challenges and tensions identified in our findings, we outline design implications for developing future privacy protection tools in ways that are context-sensitive, ethically considerate, and supportive of both uploaders and those depicted in visual content.

5.3.1 Support for Contextual Sensitivity Awareness. A central challenge identified in our findings is that privacy violations often emerge from internal or external contextual cues rather than static or predefined categories of sensitive components. Participants described privacy violations arising not only from clearly identifiable features but also from subtle, context-dependent cues within the visual content, as well as from external factors such as metadata, sequencing, and content repurposing. These observations call for privacy protection systems that are capable of reasoning across multiple levels of context. Internally, systems should consider how co-located elements can collectively lead to unintended disclosure. Externally, detection must account for how content can be repurposed or inferred using surrounding metadata, previous uploads, or public information. This aligns with participants' concerns about how others might "piece together" identities or

situations from fragmented yet interconnected content. To address these complexities, future privacy tools could consider including an analysis to evaluate the "identifiability" of different components presented in the visual content, potentially through automated cross-referencing with different informational sources. For instance, tools might allow users to explore the sensitivities of different elements within an image when viewed in relation to previous posts or public data. With this analysis, users can be prompted to process those components that turn out to be highly identifiable (or the privacy protection tools can directly process them) to avoid unwanted inference of users' private information.

5.3.2 Support Uploaders in Recognizing Elements Sensitive for Depicted Individuals. Our findings suggest that uploaders often lack awareness of potential privacy leakage, particularly regarding how others depicted in an image may perceive and experience exposure. Recent research has shown the potential of large language models (LLMs) to raise uploaders' awareness of their own location privacy risks by analyzing photos for location-based cues [74]. Yet, our findings point to a broader need: privacy tools should be designed to surface potential sensitivities not just from the uploader's perspective but also from the viewpoint of others in the image. One potential approach is to incorporate AI-driven sensitivity detection systems that highlight elements commonly flagged in past cases (e.g., faces, uniforms, identifiable locations) while encouraging the uploader to reflect on how those depicted might perceive the visibility of these elements. Additionally, systems could provide examples or questions that prompt uploaders to consider the preferences of those depicted through LLM-based agents. Furthermore, generative AI could simulate the viewpoint of individuals depicted in the image, highlighting elements that might compromise their privacy based on their role or context (e.g., personal artifacts, clothing, background cues). Rather than imposing these assessments, these tools could allow users to toggle between perspectives, explore inferred interpretations, and ultimately decide how to edit or share the content. In this framing, generative AI functions not as an arbiter of privacy but as a supportive lens for users navigating complex, context-driven disclosure risks. Its potential lies in enhancing user awareness, enabling flexible content review, and reinforcing human-centered decision-making in privacy-sensitive environments.

5.3.3 Support for Co-ownership and Shared Consent Mechanisms. Our findings revealed a recurring tension between uploaders and individuals depicted in shared visual content. To mitigate this imbalance, privacy tools should move beyond uploader-centric controls and support co-ownership and shared consent practices. A foundational step is to cultivate a norm of "asking first" before capturing or sharing photos and videos involving others. This shift in social expectation can be facilitated through design. For example, platforms could embed lightweight prompts or reminders, such as "Did you get consent to share this?", during the uploading process when faces or multiple individuals are detected. Educational pop-ups or onboarding messages could also model real-life scenarios where asking first is appropriate, particularly in private or emotionally sensitive contexts. By reinforcing the idea that visibility should not be assumed as consent, tools can help establish a culture of shared responsibility for privacy protection in everyday visual sharing.

In addition to these social nudges, systems could introduce automated consent request mechanisms that formally involve depicted individuals in the sharing decision. Prior research has demonstrated that such co-consent workflows are technically feasible and actionable [21, 64, 71, 118]. Importantly, these consent processes should extend beyond the moment of posting to also include downstream sharing and visual editing. These mechanisms should not be limited to obtaining one-time consent for posting alone. Instead, they could be extended to support layered consent, including consent to redistribute content across platforms or audiences (e.g., from private chats to public timelines) and consent to apply visual edits, such as anonymization, enhancement, or

stylization, before or after content is shared. A layered consent model would allow individuals to set boundaries on how their representations are used and transformed, reinforcing transparency and trust. For instance, consent prompts could notify individuals if their face will be blurred or if an AI-based filter will be applied, allowing them to opt in or decline. Supporting this type of granular control respects the agency of individuals featured in visual media and aligns with the broader principle of shared responsibility in privacy protection.

5.3.4 Modification Methods Should Also Be Context-Aware. Effective privacy protection tools must navigate the trade-off between safeguarding sensitive information and preserving the usability, meaning, and aesthetic integrity of visual content. Our findings reveal that users do not evaluate modification methods solely based on how well they conceal identities or objects but also on how these methods impact the overall context and emotional tone of the image. Participants emphasized that modifications need to be selective to maintain contextual consistency. When privacy risks stem from the co-occurring visual elements, obscuring just one of these elements can break the link while keeping the image's message intact. For example, they preferred hiding others' faces to highlight a location or event, or obscuring stigmatized objects like alcohol to preserve social context. These insights highlight the need for flexible, fine-grained editing tools over one-size-fits-all methods like face blurring. Moreover, modification methods themselves carry significant social and emotional implications. Participants favored methods that could reduce the visibility of obscuring targets and do not disrupt the essential narrative or aesthetics of the shared content. For many social media users, especially the young generation, the visual style of privacy edits must align with their online identity or "personal brand." Even technically effective visual anonymization methods were frequently dismissed if they disrupted the intended tone or atmosphere of the content. These findings suggest the need for privacy tools that offer stylistically diverse, content-aware options. Allowing users to preview and select edits and personalize styles based on prior choices can better support both privacy and self-expression.

5.3.5 Users as Final Decision Makers. While automated privacy protection tools show potential for detecting sensitive elements and implementing privacy safeguards, most participants expressed a strong preference for maintaining control over the process. They emphasized the importance of detailed control over which elements are modified and how these modifications are applied. While participants acknowledged the potential of AI to assist with anonymization or risk detection, they consistently rejected full automation. Instead, they described a preference for hybrid systems in which the AI generates suggestions, but the human makes the final call. This preference aligns with findings in prior research [95], which indicate that users often prioritize controllability over the precision of automation tools. To address this need, automated privacy protection tools should prioritize low-friction interaction design. Features such as fast previews, side-by-side comparisons, and undo/redo controls can help users quickly test different options without losing control. These designs ensure that users remain active participants in the process while also making it practically feasible to apply privacy protections in everyday settings. This balanced approach combines the efficiency of automation with users' desire for agency in protecting both their privacy and that of others depicted in the content.

5.4 Sociotechnical Pathways Toward Responsible Generative AI Use in Privacy Protection

5.4.1 Ethical Data Practices in Generative AI. Participants' concerns underscore that privacy risks in Gen-AI-powered privacy protection tools extend beyond visible outputs to include hidden data practices. Participants expressed deep unease over the lack of clarity about how their data is handled, especially in terms of whether uploaded content might be reused for model training or appear

in future AI-generated outputs. To address these concerns, developers of privacy-preserving AI tools must prioritize transparency and ethical data stewardship as core design goals. First, systems should provide clear disclosures about how user-uploaded images are processed, stored and whether they contribute to future model training. For instance, AI tools could include opt-in or opt-out mechanisms for data retention and model improvement, along with real-time indicators or logs showing where and how data is used. Second, designers should consider implementing explainable AI (XAI) features [98, 111] that allow users to understand what kinds of transformations were applied to their content and what training data might inform those outputs. This could involve surfacing example inputs that resemble the user's image or generating brief justifications for automated decisions, such as why certain features were targeted for modification. Lastly, it is critical to integrate privacy-preserving model architectures such as federated learning [76] or differential privacy techniques [38], especially for systems designed to evolve with user feedback. These approaches can help mitigate concerns about unintentional data exposure while still enabling iterative improvement. By making data flows visible and controllable and by embedding ethical constraints into model development, privacy tools can better align with users' expectations of safety, autonomy, and informed participation.

5.4.2 Making AI-Modified Visual Content Recognizable. Beyond concerns about how generative models are trained, participants also questioned the trustworthiness of the visual content they encountered online. Our findings reveal a complex landscape surrounding the use of generative AI in privacy protection on social media. Participants acknowledged the potential benefits of generative AI in safeguarding personal information but also expressed significant ethical concerns. A primary concern among participants was the potential for generative AI to be used to create fake images for malicious purposes, such as privacy violations, harassment, and disinformation. The rise of Deepfake in recent years exemplifies the risk [32, 62, 72, 85]. Notably, female participants voiced heightened concerns about the misuse of AIGC in sexual contexts, highlighting how generative AI can exacerbate existing power imbalances and gender-based harassment in digital spaces.

These discussions revealed a nuanced tension around transparency. However, participants' perspectives on transparency are diverse and, at times, conflicting. While some advocated for full disclosure to prevent the spread of misinformation, others worried that such disclosures could undermine the privacy protections that generative AI aims to enhance. For example, if an image is modified to obscure someone's identity for privacy reasons, logging the modified location for the AI-altered version might unintentionally highlight the modification and invite unwanted speculation. In this way, excessive transparency could work against the very goals of protection.

These divergent views suggest that transparency in generative AI content should not be approached as a binary but as a design space requiring careful calibration. To support user trust without compromising privacy, future systems could offer optional and context-sensitive visual cues, such as subtle watermarks [51], blur patterns, or editable metadata, that allow content to be recognizable as AI-modified without necessarily revealing what was changed or why. In addition, generative AI visual modification systems could be designed to reject user requests that conflict with the intended ethical purpose of the tool, similar to how large language models currently decline to answer harmful prompts [114]. These approaches reflect a need for transparency strategies that are not only technically feasible but socially responsive to the varied expectations and risks faced by different user groups.

5.4.3 Distributing Responsibility for Transparency in Generative AI Use. Another critical issue around transparency is determining who holds responsibility for ensuring transparency disclosures. Transparency, as a collective concern, also raised questions of responsibility and governance. Some participants voluntarily disclosed their use of generative AI on visual content when sharing it on

social media. However, such voluntary disclosure relies heavily on the discretion and honesty of individual users and are not always feasible or consistent. These findings highlight the limitations of relying on end-user action alone, and instead point to the need for shared accountability across stakeholders. In line with CSCW's long-standing interest in infrastructure and cooperation, we argue that governments, AI companies, and social media platforms should work together to establish robust guidelines and enforcement mechanisms that ensure transparency in the use of generative AI, supporting both accountability and trust in digital spaces [53]. Governments and policymakers should mandate AI companies to incorporate transparency features into their products, require platforms to clearly indicate when content is modified by AI, and hinder the spread of malicious generative AI-generated content. For instance, California recently enacted Assembly Bill No. 602, which increases penalties for distributing unauthorized AI-generated explicit content.

From a platform perspective, system-level mechanisms, such as automated detection and labeling of AI-generated content, can reduce the burden on users and promote consistent transparency. For instance, Facebook and Instagram have begun implementing such features by watermarking the AI-modified altered content [24]. However, these efforts still largely depend on user self-disclosure and remain limited in scope. Finally, our results also highlight participants' need for transparency regarding companies developing these AI tools. This underscores the importance of collaborative efforts among governments, social media platforms, and AI tool providers to build a stronger framework for managing AI-generated content misuse, enhancing transparency, and safeguarding privacy without stifling innovation. Future research could explore how such collaborations might evolve to address emerging challenges, particularly within global regulatory landscapes and through cross-platform cooperation.

5.5 Limitations

Our study includes interviews with eighteen U.S.-based individuals who had experienced or caused privacy violation incidents in online visual content sharing. While the relatively small sample size may limit the breadth of perspectives captured, it is consistent with qualitative research practices that prioritize depth over breadth [100]. Rather than statistical generalizability, we emphasize analytical generalizability, where insights contribute to conceptual understanding and inform future research and design. We acknowledge that the participant sample lacks demographic diversity, particularly with respect to ethnicity, cultural background, geography, and physical abilities. The majority of our participants self-identified as White and were based in the United States, which may constrain the applicability of our findings to other cultural contexts. Cultural norms and values significantly shape how individuals perceive privacy and visual content sharing. Hence, future research should explore how these dynamics vary across regions, communities, and social media ecosystems. While some core concerns around visual privacy may transcend cultural boundaries, our findings can serve as an initial step that lays the groundwork for future cross-cultural and intersectional investigations.

Additionally, our data collection relied on participants' retrospective accounts of privacy violations. This approach is susceptible to recall bias and selective memory. To mitigate these effects, we deployed the semi-structured interview format, which allowed us to probe deeper into the reason behind participants' preferences, encouraging reflection and helping to mitigate some potential biases. Finally, this study was conducted between 2024 and 2025, thus some findings may be limited to the sociotechnical landscape of that time. While specific technologies may evolve, the underlying privacy concerns and human behaviors we identified are likely to remain relevant.

6 Conclusion

Through semi-structured interviews with individuals who had experienced or caused privacy violations, our study sheds light on the nuanced needs and concerns surrounding privacy protection for visual content shared on social media and attitudes toward generative AI in privacy protection. Our findings reveal that privacy risks encompass more than directly identifiable elements like faces. They also extend to internal context-dependent cues, such as co-occurring elements, to external cues, such as content sequencing and external knowledge that can be used for inference. These insights underscore the limitations of prescriptive or static approaches to privacy protection. These nuanced understandings of sensitivity informed participants' decisions about whether and how to modify content, which were shaped by a careful balancing of recognition difficulty, integrity of visual content, consistency with the context, atmosphere after modification, and usability of the modification method. Finally, we find that generative AI based privacy preserving tools were often viewed as enabling more seamless and aesthetically coherent modifications; however, this perceived utility was accompanied by concerns about imperfection and malicious use, hidden data usage, loss of content authenticity, and a lack of transparency regarding how changes were made. Importantly, the study surfaced a fundamental tension between the autonomy of uploaders and the privacy expectations of individuals depicted in shared content, highlighting asymmetries in power and control over visibility and consent. Together, these findings suggest a need for privacy tools that are context-aware, flexible, and designed to support negotiation between stakeholders with potentially competing interests. Moreover, the ethical complexities introduced by AI-driven modification, particularly regarding manipulation, consent, and accountability, call for transparency-enhancing features and the consideration of advocacy mechanisms to protect the rights and agency of all parties involved.

Acknowledgments

We are grateful to the anonymous reviewers who provided many helpful comments. We thank the support of the National Science Foundation under grant CNS-2114123 on this work. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors alone.

References

[1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (2015), 509–514.

[2] Anne Adams, Sally Jo Cunningham, and Masood Masoodian. 2007. Sharing, privacy and trust issues for photo collections. (2007).

[3] Shane Ahern, Dean Eckles, Nathaniel S Good, Simon King, Mor Naaman, and Rahul Nair. 2007. Over-exposed? Privacy patterns and considerations in online and mobile photo sharing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 357–366.

[4] Tousif Ahmed, Apu Kapadia, Venkatesh Potluri, and Manohar Swaminathan. 2018. Up to a limit? privacy concerns of bystanders and their willingness to share additional information with visually impaired users of assistive technologies. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–27.

[5] Mary Jean Amon, Nika Kartvelishvili, Bennett I Bertenthal, Kurt Hugenberg, and Apu Kapadia. 2022. Sharenting and children's privacy in the united states: Parenting style, practices, and perspectives on sharing young children's photos on social media. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–30.

[6] Mary Jean Amon, Aaron Necaise, Nika Kartvelishvili, Aneka Williams, Yan Solihin, and Apu Kapadia. 2023. Modeling user characteristics associated with interdependent privacy perceptions on social media. *ACM Transactions on Computer-Human Interaction* 30, 3 (2023), 1–32.

[7] Fabio Bacchini and Ludovica Lorusso. 2018. A tattoo is not a face. Ethical aspects of tattoo-based biometrics. *Journal of Information, Communication and Ethics in Society* 16, 2 (2018), 110–122.

[8] Lemi Baruh, Ekin Secinti, and Zeynep Cemalcilar. 2017. Online privacy concerns and privacy management: A meta-analytical review. *Journal of Communication* 67, 1 (2017), 26–53.

[9] Anja Bechmann and Stine Lomborg. 2013. Mapping actor roles in social media: Different perspectives on value creation in theories of user participation. *New media & society* 15, 5 (2013), 765–781.

[10] Andrew Besmer and Heather Lipford. 2009. Tagged photos: concerns, perceptions, and protections. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. 4585–4590.

[11] Andrew Besmer and Heather Richter Lipford. 2010. Moving beyond untagging: photo privacy in a tagged world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1563–1572.

[12] Divyanshu Bhardwaj, Alexander Ponticello, Shreya Tomar, Adrian Dabrowski, and Katharina Krombholz. 2024. In Focus, Out of Privacy: The Wearer's Perspective on the Privacy Dilemma of Camera Glasses. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[13] Dmitri Bitouk, Neeraj Kumar, Samreen Dhillon, Peter Belhumeur, and Shree K. Nayar. 2008. Face swapping: automatically replacing faces in photographs. *ACM Trans. Graph.* 27, 3 (aug 2008), 1–8. doi:10.1145/1360612.1360638

[14] Michael Boyle, Christopher Edwards, and Saul Greenberg. 2000. The effects of filtered video on awareness and privacy. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. 1–10.

[15] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association.

[16] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology* 18, 3 (2021), 328–352.

[17] Kelly E Caine, Wendy A Rogers, and Arthur D Fisk. 2005. Privacy perceptions of an aware home with visual sensing devices. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 49. SAGE Publications Sage CA: Los Angeles, CA, 1856–1858.

[18] Nicholas Carah. 2014. Watching nightlife: Affective labor, social media, and surveillance. *Television & New Media* 15, 3 (2014), 250–265.

[19] Albert KM Chan, Chris P Nickson, Jenny W Rudolph, Anna Lee, and Gavin M Joynt. 2020. Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic. *Anaesthesia* 75, 12 (2020), 1579.

[20] Mauro Cherubini, Kavous Salehzadeh Niksirat, Marc-Olivier Boldi, Henri Keopraseuth, Jose M Such, and Kévin Huguenin. 2021. When forcing collaboration is the most sensible choice: Desirability of precautionary and dissuasive mechanisms to manage multiparty privacy conflicts. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–36.

[21] Umur A Ciftci, Gokturk Yuksek, and Ilke Demir. 2023. My face my choice: Privacy enhancing deepfakes for social media anonymization. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1369–1379.

[22] Kathryn D Coduto and Allison McDonald. 2024. Delete it and move on": Digital management of shared sexual content after a breakup. *Proceedings of the ACM on Human Factors in Computing Systems (CHI '24), Honolulu, HI, USA* (2024), 11–16.

[23] Daniel Da Silva Costa, Pedro Nuno Moura, and Ana Cristina Bicharra Garcia. 2021. Improving human perception of GAN generated facial image synthesis by filtering the training set considering facial attributes. In *2021 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, 100–106.

[24] Emilia David and Alex Heath. 2024. Meta says you better disclose your AI fakes or it might just pull them. https://www.theverge.com/2024/2/6/24062388/meta-ai-photo-watermark-facebook-instagram-threads

[25] Jasmine DeHart, Makya Stell, and Christan Grant. 2020. Social media and the scourge of visual privacy. *Information* 11, 2 (2020), 57.

[26] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. 2014. In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2377–2386.

[27] Michael A DeVito, Ashley Marie Walker, and Jeremy Birnholtz. 2018. 'Too Gay for Facebook' Presenting LGBTQ+ Identity Throughout the Personal Social Media Ecosystem. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–23.

[28] Paul M Di Gangi and Molly M Wasko. 2016. Social media engagement theory: Exploring the influence of user engagement on social media usage. *Journal of Organizational and End User Computing (JOEUC)* 28, 2 (2016), 53–73.

[29] Rebecca Dolan, Jodie Conduit, John Fahy, and Steve Goodman. 2016. Social media engagement behaviour: a uses and gratifications perspective. *Journal of strategic marketing* 24, 3-4 (2016), 261–277.

[30] Cori Faklaris, Francesco Cafaro, Asa Blevins, Matthew A O'Haver, and Neha Singhal. 2020. A snapshot of bystander attitudes about mobile live-streaming video in public settings. In *Informatics*, Vol. 7. MDPI, 10.

[31] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods* 5, 1 (2006), 80–92.

[32] Asher Flynn, Anastasia Powell, Adrian J Scott, and Elena Cama. 2022. Deepfakes and digitally altered imagery abuse: A cross-country exploration of an emerging form of image-based sexual abuse. *The British Journal of Criminology* 62, 6 (2022), 1341–1358.

[33] Joshua Fogel and Elham Nehmad. 2009. Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in human behavior* 25, 1 (2009), 153–160.

[34] Andrea Frome, German Cheung, Ahmad Abdulkader, Marco Zennaro, Bo Wu, Alessandro Bissacco, Hartwig Adam, Hartmut Neven, and Luc Vincent. 2009. Large-scale privacy protection in google street view. In *2009 IEEE 12th international conference on computer vision*. IEEE, 2373–2380.

[35] Diane Gan and Lily R Jenkins. 2015. Social networking privacy—Who's stalking you? *Future Internet* 7, 1 (2015), 67–93.

[36] Kambiz Ghazinour and John Ponchak. 2017. Hidden privacy risks in sharing pictures on social media. *Procedia computer science* 113 (2017), 267–272.

[37] Eric Gilbert, Karrie Karahalios, and Christian Sandvig. 2008. The network in the garden: an empirical analysis of social media in rural life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1603–1612.

[38] Maoguo Gong, Yu Xie, Ke Pan, Kaiyuan Feng, and Alex Kai Qin. 2020. A survey on differentially private machine learning. *IEEE computational intelligence magazine* 15, 2 (2020), 49–64.

[39] Marco Guerini, Carlo Strapparava, and Gozde Ozbal. 2011. Exploring text virality in social networks. In *proceedings of the international AAAI conference on web and social media*, Vol. 5. 506–509.

[40] Grant S Hamilton. 2010. Photoshop tips and tricks every facial plastic surgeon should know. *Facial Plastic Surgery Clinics* 18, 2 (2010), 283–328.

[41] Eszter Hargittai et al. 2010. Facebook privacy settings: Who cares? *First Monday* (2010).

[42] Rakibul Hasan, Bennett I Bertenthal, Kurt Hugenberg, and Apu Kapadia. 2021. Your photo is so funny that i don't mind violating your privacy by sharing it: effects of individual humor styles on online photo-sharing behaviors. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[43] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. 2018. Viewer experience of obscuring scene elements in photos to enhance privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[44] Rakibul Hasan, Yifang Li, Eman Hassan, Kelly Caine, David J. Crandall, Roberto Hoyle, and Apu Kapadia. 2019. Can Privacy Be Satisfying? On Improving Viewer Satisfaction for Privacy-Enhanced Photos Using Aesthetic Transforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300597

[45] Benjamin Henne, Christian Szongott, and Matthew Smith. 2013. SnapMe if you can: Privacy threats of other peoples' geo-tagged media and what we can do about it. In *Proceedings of the sixth ACM conference on Security and privacy in wireless and mobile networks*. 95–106.

[46] Mariea Grubbs Hoy and George Milne. 2010. Gender differences in privacy-related measures for young adult Facebook users. *Journal of interactive advertising* 10, 2 (2010), 28–45.

[47] Roberto Hoyle, Luke Stark, Qatrunnada Ismail, David Crandall, Apu Kapadia, and Denise Anthony. 2020. Privacy norms and preferences for photos posted online. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 4 (2020), 1–27.

[48] Roberto Hoyle, Robert Templeman, Denise Anthony, David Crandall, and Apu Kapadia. 2015. Sensitive lifelogs: A privacy analysis of photos from wearable cameras. In *Proceedings of the 33rd Annual ACM conference on human factors in computing systems*. 1645–1648.

[49] Roberto Hoyle, Robert Templeman, Steven Armes, Denise Anthony, David Crandall, and Apu Kapadia. 2014. Privacy behaviors of lifeloggers using wearable cameras. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. 571–582.

[50] Hongxin Hu, Gail-Joon Ahn, and Jan Jorgensen. 2011. Detecting and resolving privacy conflicts for collaborative data sharing in online social networks. In *Proceedings of the 27th Annual Computer Security Applications Conference* (Orlando, Florida, USA) *(ACSAC '11)*. Association for Computing Machinery, New York, NY, USA, 103–112. doi:10.1145/2076732.2076747

[51] JaeYoung Hwang and SangHoon Oh. 2023. A Brief Survey of Watermarks in Generative AI. In *2023 14th International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 1157–1160.

[52] Panagiotis Ilia, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi, and Sotiris Ioannidis. 2015. Face/off: Preventing privacy leakage from photos in social networks. In *Proceedings of the 22nd ACM SIGSAC Conference on computer and communications security*. 781–792.

[53] Basileal Imana, Aleksandra Korolova, and John Heidemann. 2023. Having your Privacy Cake and Eating it Too: Platform-supported Auditing of Social Media Algorithms for Public Interest. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–33.

[54] Shareen Irshad and Tariq Rahim Soomro. 2018. Identity theft and social media. *International Journal of Computer Science and Network Security* 18, 1 (2018), 43–55.

[55] Pavel Korshunov and Touradj Ebrahimi. 2013. Using face morphing to protect privacy. In *2013 10th IEEE international conference on advanced video and signal based surveillance*. IEEE, 208–213.
[56] Pavel Korshunov and Touradj Ebrahimi. 2013. Using face morphing to protect privacy. In *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*. 208–213. doi:10.1109/AVSS.2013.6636641
[57] Pavel Korshunov, Andrea Melle, Jean-Luc Dugelay, and Touradj Ebrahimi. 2013. Framework for objective evaluation of privacy filters. In *Applications of Digital Image Processing XXXVI*, Vol. 8856. SPIE, 265–276.
[58] Takashi Koshimizu, Tomoji Toriyama, and Noboru Babaguchi. 2006. Factors on the sense of privacy in video surveillance. In *Proceedings of the 3rd ACM workshop on Continuous archival and retrival of personal experences*. 35–44.
[59] Hanna Krasnova, Oliver Günther, Sarah Spiekermann, and Ksenia Koroleva. 2009. Privacy concerns and identity in online social networks. *Identity in the Information Society* 2 (2009), 39–63.
[60] Matthew B Kugler. 2019. From identification to identity theft: Public perceptions of biometric privacy harms. *UC Irvine L. Rev.* 10 (2019), 107.
[61] Andrei OJ Kwok and Sharon GM Koh. 2021. Deepfake: a social construction of technology perspective. *Current Issues in Tourism* 24, 13 (2021), 1798–1802.
[62] Hao-Ping Lee, Yu-Ju Yang, Thomas Serban Von Davier, Jodi Forlizzi, and Sauvik Das. 2024. Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.
[63] Kevin Lewis, Jason Kaufman, and Nicholas Christakis. 2008. The taste for privacy: An analysis of college student privacy settings in an online social network. *Journal of computer-mediated communication* 14, 1 (2008), 79–100.
[64] Fenghua Li, Zhe Sun, Ang Li, Ben Niu, Hui Li, and Guohong Cao. 2019. Hideme: Privacy-preserving photo sharing on social networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 154–162.
[65] Yifang Li and Kelly Caine. 2022. Obfuscation Remedies Harms Arising from Content Flagging of Photos. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 35, 25 pages. doi:10.1145/3491102.3517520
[66] Yifang Li, Nishant Vishwamitra, Hongxin Hu, and Kelly Caine. 2020. Towards a taxonomy of content sensitivity and sharing preferences for photos. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
[67] Yifang Li, Nishant Vishwamitra, Bart P Knijnenburg, Hongxin Hu, and Kelly Caine. 2017. Effectiveness and users' experience of obfuscation as a privacy-enhancing technology for sharing photos. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–24.
[68] Chi Liu, Tianqing Zhu, Jun Zhang, and Wanlei Zhou. 2022. Privacy intelligence: A survey on image privacy in online social networks. *Comput. Surveys* 55, 8 (2022), 1–35.
[69] Jacob Logas, Poojita Garg, Rosa I Arriaga, and Sauvik Das. 2024. The Subversive AI Acceptance Scale (SAIA-8): A Scale to Measure User Acceptance of AI-Generated, Privacy-Enhancing Image Modifications. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–43.
[70] Jacob Logas, Ari Schlesinger, Zhouyu Li, and Sauvik Das. 2022. Image DePO: towards gradual decentralization of online social networks using decentralized privacy overlays. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–28.
[71] Juniper Lovato, Antoine Allard, Randall Harp, and Laurent Hébert-Dufresne. 2020. Distributed consent and its impact on privacy and observability in social networks. *arXiv preprint arXiv:2006.16140* (2020).
[72] Kweilin T Lucas. 2022. Deepfakes and domestic violence: perpetrating intimate partner abuse using video technology. *Victims & Offenders* 17, 5 (2022), 647–659.
[73] Christoph Lutz and Giulia Ranzini. 2017. Where dating meets data: Investigating social and institutional privacy concerns on Tinder. *Social Media+ Society* 3, 1 (2017), 2056305117697735.
[74] Ying Ma, Shiquan Zhang, Dongju Yang, Zhanna Sarsenbayeva, Jarrod Knibbe, and Jorge Goncalves. 2025. Raising Awareness of Location Information Vulnerabilities in Social Media Photos using LLMs. *arXiv preprint arXiv:2503.20226* (2025).
[75] Sophie Maddocks. 2018. From non-consensual pornography to image-based sexual abuse: Charting the course of a problem with many names. *Australian Feminist Studies* 33, 97 (2018), 345–361.
[76] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.
[77] Andrew D Miller and W Keith Edwards. 2007. Give and take: a study of consumer photo-sharing culture and practice. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 347–356.
[78] Tehila Minkus, Kelvin Liu, and Keith W Ross. 2015. Children seen but not heard: When parents compromise children's online privacy. In *Proceedings of the 24th international conference on World Wide Web*. 776–786.
[79] Kyzyl Monteiro, Yuchen Wu, and Sauvik Das. 2024. Manipulate to Obfuscate: A Privacy-Focused Intelligent Image Manipulation Tool for End-Users. In *Adjunct Proceedings of the 37th Annual ACM Symposium on User Interface Software*

and Technology. 1–3.

[80] Joshua Morris, Sara Newman, Kannappan Palaniappan, Jianping Fan, and Dan Lin. 2021. "Do you know you are tracked by photos that you didn't take": large-scale location-aware multi-party image privacy protection. *IEEE Transactions on Dependable and Secure Computing* 20, 1 (2021), 301–312.

[81] Tamara Mujirishvili, Anton Fedosov, Kooshan Hashemifard, Pau Climent-Pérez, and Francisco Florez-Revuelta. 2024. "I Don't Want to Become a Number": Examining Different Stakeholder Perspectives on a Video-Based Monitoring System for Senior Care with Inherent Privacy Protection (by Design).. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 774, 19 pages. doi:10.1145/3613904.3642164

[82] Yuta Nakashima, Tatsuya Koyama, Naokazu Yokoya, and Noboru Babaguchi. 2015. Facial expression preserving privacy protection using image melding. In *2015 IEEE International Conference on Multimedia and Expo (ICME)*. 1–6. doi:10.1109/ICME.2015.7177394

[83] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.

[84] Seong Joon Oh, Rodrigo Benenson, Mario Fritz, and Bernt Schiele. 2016. Faceless person recognition: Privacy implications in social media. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*. Springer, 19–35.

[85] Chidera Okolie. 2023. Artificial intelligence-altered videos (deepfakes), image-based sexual abuse, and data privacy concerns. *Journal of International Women's Studies* 25, 2 (2023), 11.

[86] Alexandra-Mihaela Olteanu, Kévin Huguenin, Italo Dacosta, and J-P Hubaux. 2018. Consensual and privacy-preserving sharing of multi-subject and interdependent data. In *Proceedings of the 25th network and distributed system security symposium (NDSS)*. Internet Society, 1–16.

[87] Tribhuvanesh Orekondy, Bernt Schiele, and Mario Fritz. 2017. Towards a visual privacy advisor: Understanding and predicting privacy risks in images. In *Proceedings of the IEEE international conference on computer vision*. 3686–3695.

[88] Babajide Osatuyi. 2013. Information sharing on social media sites. *Computers in human behavior* 29, 6 (2013), 2622–2631.

[89] José Ramón Padilla-López, Alexandros Andre Chaaraoui, Feng Gu, and Francisco Flórez-Revuelta. 2015. Visual privacy by context: proposal and evaluation of a level-based visualisation scheme. *Sensors* 15, 6 (2015), 12959–12982.

[90] César Palacios-González. 2015. The ethics of clinical photography and social media. *Medicine, Health Care and Philosophy* 18, 1 (2015), 63–70.

[91] Sandra Petronio. 2002. *Boundaries of privacy: Dialectics of disclosure*. Suny Press.

[92] Kate Raynes-Goldie. 2010. Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook. *First Monday* (2010).

[93] Alex Rochefort. 2020. Regulating social media platforms: A comparative policy analysis. *Communication Law and Policy* 25, 2 (2020), 225–260.

[94] Lynsey K Romo, Charee M Thompson, and Erin E Donovan. 2017. College drinkers' privacy management of alcohol content on social-networking sites. *Communication Studies* 68, 2 (2017), 173–189.

[95] Quentin Roy, Futian Zhang, and Daniel Vogel. 2019. Automation accuracy is good, but high controllability may be better. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.

[96] Johnny Saldaña. 2021. The coding manual for qualitative researchers. (2021).

[97] Kavous Salehzadeh Niksirat, Diana Korka, Hamza Harkous, Kévin Huguenin, and Mauro Cherubini. 2023. On the potential of mediation chatbots for mitigating multiparty privacy conflicts-a wizard-of-Oz study. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–33.

[98] Wojciech Samek and Klaus-Robert Müller. 2019. Towards explainable artificial intelligence. In *Explainable AI: interpreting, explaining and visualizing deep learning*. Springer, 5–22.

[99] Morgan Klaus Scheuerman, Kandrea Wade, Caitlin Lustig, and Jed R Brubaker. 2020. How we've taught algorithms to see identity: Constructing race and gender in image databases for facial analysis. *Proceedings of the ACM on Human-computer Interaction* 4, CSCW1 (2020), 1–35.

[100] Brett Smith. 2018. Generalizability in qualitative research: Misunderstandings, opportunities and recommendations for the sport and exercise sciences. *Qualitative research in sport, exercise and health* 10, 1 (2018), 137–149.

[101] Anna Squicciarini, Cornelia Caragea, and Rahul Balakavi. 2017. Toward automated online photo privacy. *ACM Transactions on the Web (TWEB)* 11, 1 (2017), 1–29.

[102] Luke Stark. 2016. The emotional context of information privacy. *The Information Society* 32, 1 (2016), 14–27.

[103] Frederic Stutzman and Woodrow Hartzog. 2012. Boundary regulation in social media. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 769–778.

[104] Fred Stutzman and Jacob Kramer-Duffield. 2010. Friends only: examining a privacy-enhancing behavior in facebook. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1553–1562.

[105] Jose M. Such and Natalia Criado. 2016. Resolving Multi-Party Privacy Conflicts in Social Media. *IEEE Transactions on Knowledge and Data Engineering* 28, 7 (2016), 1851–1863. doi:10.1109/TKDE.2016.2539165

[106] Jose M Such and Natalia Criado. 2018. Multiparty privacy in social media. *Commun. ACM* 61, 8 (2018), 74–81.

[107] Jose M Such, Joel Porter, Sören Preibusch, and Adam Joinson. 2017. Photo privacy conflicts in social media: A large-scale empirical study. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 3821–3832.

[108] Yasuhiro Tanaka, Akihisa Kodate, Yu Ichifuji, and Noboru Sonehara. 2015. Relationship between willingness to share photos and preferred level of photo blurring for privacy protection. In *Proceedings of the ASE BigData & SocialInformatics 2015*. 1–5.

[109] Sigal Tifferet. 2019. Gender differences in privacy tendencies on social network sites: A meta-analysis. *Computers in Human Behavior* 93 (2019), 1–12.

[110] Lam Tran, Deguang Kong, Hongxia Jin, and Ji Liu. 2016. Privacy-cnh: A framework to detect photo privacy with convolutional neural network using hierarchical features. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 30.

[111] Xinru Wang and Ming Yin. 2021. Are explanations helpful? a comparative study of the effects of explanations in ai-assisted decision-making. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*. 318–328.

[112] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. 2011. " I regretted the minute I pressed share" a qualitative study of regrets on Facebook. In *Proceedings of the seventh symposium on usable privacy and security*. 1–16.

[113] Samantha Warren. 2002. Show me how it feels to work here: Using photography to research organizational aesthetics. *ephemera* 2, 3 (2002), 224–245.

[114] Joel Wester, Tim Schrills, Henning Pohl, and Niels van Berkel. 2024. "As an AI language model, I cannot": Investigating LLM Denials of User Requests. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–14.

[115] Michael Williams and Tami Moser. 2019. The art of coding and thematic exploration in qualitative research. *International management review* 15, 1 (2019), 45–55.

[116] Yanlai Wu, Xinning Gui, Pamela J. Wisniewski, and Yao Li. 2023. Do Streamers Care about Bystanders' Privacy? An Examination of Live Streamers' Considerations and Strategies for Bystanders' Privacy Management. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 127 (apr 2023), 29 pages. doi:10.1145/3579603

[117] Anran Xu, Shitao Fang, Huan Yang, Simo Hosio, and Koji Yatani. 2024. Examining Human Perception of Generative Content Replacement in Image Privacy Protection. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.

[118] Kaihe Xu, Yuanxiong Guo, Linke Guo, Yuguang Fang, and Xiaolin Li. 2015. My privacy my decision: Control of photo sharing on online social networks. *IEEE Transactions on Dependable and Secure Computing* 14, 2 (2015), 199–210.

[119] Kaihe Xu, Yuanxiong Guo, Linke Guo, Yuguang Fang, and Xiaolin Li. 2017. My Privacy My Decision: Control of Photo Sharing on Online Social Networks. *IEEE Transactions on Dependable and Secure Computing* 14, 2 (2017), 199–210. doi:10.1109/TDSC.2015.2443795

[120] Jinghan Yang, Ayan Chakrabarti, and Yevgeniy Vorobeychik. 2020. Protecting geolocation privacy of photo collections. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 524–531.

[121] Alyson L Young and Anabel Quan-Haase. 2009. Information revelation and internet privacy concerns on social network sites: a case study of facebook. In *Proceedings of the fourth international conference on Communities and technologies*. 265–274.

[122] Jun Yu, Zhenzhong Kuang, Baopeng Zhang, Wei Zhang, Dan Lin, and Jianping Fan. 2018. Leveraging content sensitiveness and user trustworthiness to recommend fine-grained privacy settings for social image sharing. *IEEE transactions on information forensics and security* 13, 5 (2018), 1317–1332.

[123] Lingjing Yu, Sri Mounica Motipalli, Dongwon Lee, Peng Liu, Heng Xu, Qingyun Liu, Jianlong Tan, and Bo Luo. 2018. My friend leaks my privacy: Modeling and analyzing privacy in social networks. In *Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies*. 93–104.

[124] Xiaoyi Yu, Kenta Chinomi, Takashi Koshimizu, Naoko Nitta, Yoshimichi Ito, and Noboru Babaguchi. 2008. Privacy protecting visual processing for secure video surveillance. In *2008 15th IEEE International Conference on Image Processing*. IEEE, 1672–1675.

[125] Lin Yuan and Touradj Ebrahimi. 2017. Image privacy protection with secure JPEG transmorphing. *IET Signal Processing* 11, 9 (2017), 1031–1038.

[126] Eva-Maria Zeissig, Chantal Lidynia, Luisa Vervier, Andera Gadeib, and Martina Ziefle. 2017. Online privacy perceptions of older adults. In *Human Aspects of IT for the Aged Population. Applications, Services and Contexts: Third International Conference, ITAP 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part II 3*. Springer, 181–200.

[127] Sergej Zerr, Stefan Siersdorfer, Jonathon Hare, and Elena Demidova. 2012. Privacy-aware image classification and search. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. 35–44.

[128] Lan Zhang, Xiang-Yang Li, Kebin Liu, Cihang Liu, Xuan Ding, and Yunhao Liu. 2018. Cloak of invisibility: Privacy-friendly photo capturing and sharing system. *IEEE Transactions on Mobile Computing* 18, 11 (2018), 2488–2501.

[129] Chenye Zhao, Jasmine Mangat, Sujay Koujalgi, Anna Squicciarini, and Cornelia Caragea. 2022. Privacyalert: A dataset for image privacy prediction. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 1352–1361.

[130] Wenyi Zhao, Rama Chellappa, P Jonathon Phillips, and Azriel Rosenfeld. 2003. Face recognition: A literature survey. *ACM computing surveys (CSUR)* 35, 4 (2003), 399–458.

[131] Tomasz Zukowski and Irwin Brown. 2007. Examining the influence of demographic factors on internet users' information privacy concerns. In *Proceedings of the 2007 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*. 197–204.

Received October 2024; revised April 2025; accepted August 2025