Studio Ghibli AI trend is a privacy nightmare in disguise, warn experts

Studio Ghibli: Experts caution that while many platforms claim to either not store images, the definitions of terms like "deletion" remain vague.

The latest viral trend on the internet is the Studio Ghibli-style which has captured the imagination of everyone on the internet. From politicians to influencers, everyone has been using ChatGPT to give their photos a whimsical makeover. While users are captivated by the charm of this new digital experience, cybersecurity experts are sounding alarms about the potential privacy risks that come with sharing personal photos on these AI platforms.

The trend took off after OpenAI's GPT-4o model, which enables users to recreate personal images in the artistic style of the beloved Japanese animation studio, Studio Ghibli. While the tools behind the transformation are user-friendly and the results often magical, experts warn that the fine print in terms of service agreements and the way data is handled could have serious implications for users' privacy.

Experts warn against Ghibli makeover

Cybersecurity professionals caution that while many platforms claim to either not store images or delete them after a single use, the definitions of terms like "deletion" remain vague and open to interpretation.

In the absence of clear policies, it’s unclear whether user photos are truly erased instantly or if they linger in the system longer than expected.

Vishal Salvi, CEO of Quick Heal Technologies, explains that photos hold more than just facial data. They contain metadata such as location coordinates, timestamps, and device information, which can inadvertently expose sensitive personal details.

The AI tools powering these transformations utilise neural style transfer (NST) algorithms, which separate content from artistic styles and blend the user's image with reference artwork. While this process appears harmless, vulnerabilities exist.

Salvi points out that model inversion attacks could allow adversaries to reconstruct the original photos from the stylized Ghibli images.
"Even if companies claim they don't store your photos, fragments of your data might still end up in their systems," he says. "Uploaded images can definitely be repurposed for unintended uses, like training AI models for surveillance or targeted advertising."

The ease and fun of transforming personal photos into art can obscure the hidden risks, warns Pratim Mukherjee, Senior Director of Engineering at McAfee.
"Eye-catching results and viral filters create an experience that feels light and effortless, but it often comes with hidden privacy risks," Mukherjee explains.

He further highlights the normalization of casual data sharing, which users may not fully comprehend, especially when platforms silently collect data in the background.

Deepfake and identity theft threat
Mukherjee also cautions that this seemingly harmless trend could lead to more severe consequences. With the rapid evolution of AI technology, stolen personal photos could fuel the creation of deepfakes or identity fraud. "When access to something as personal as a camera roll is granted without a second thought, it's not always accidental," Mukherjee says. "These platforms often encourage quick engagement, which distracts users from the potential long-term implications of sharing sensitive data."

Vladislav Tushkanov, Group Manager at Kaspersky AI Technology Research Centre, adds that while some companies prioritise the security of the data they collect, no protection system is entirely foolproof.

"Due to technical issues or malicious activity, data can leak, become public, or even appear for sale on underground websites," he warns. "Stolen user data, including images, is often trafficked on dark web forums, making it especially important for users to be cautious about what they share."

Tushkanov highlights another critical concern: "The hard part is, you can't change your face the way you can reset a password. Once a photo is out there, it's out there." This makes the potential for identity theft or exploitation even more troubling.

Vague service agreements

Another pressing issue is the opacity of terms of service agreements. These policies are often lengthy, filled with jargon, and difficult for users to fully understand. Mukherjee notes that many platforms bury crucial information regarding data usage in the fine print.

"Just because someone clicks 'accept' doesn't mean they're giving truly informed consent," he explains. "If it's not clear how your image will be used, or whether it's deleted at all, it's worth asking if the fun is really worth the risk."

Governments around the world are beginning to take notice of these privacy concerns. Some have already enacted regulations requiring clearer disclosures of how user data is handled, while others are still considering similar steps.

Experts suggest that more transparent policies and simplified disclosures could go a long way in helping users make informed decisions about the AI tools they choose to use.

What can you do?

To mitigate risks, experts recommend that users exercise caution when engaging with such platforms. Tushkanov advises combining strong security practices - such as using unique, complex passwords and enabling two-factor authentication - with common sense.

Additionally, Salvi recommends stripping metadata from photos before uploading them to AI tools. On the regulatory front, he advocates for mandatory differential privacy certification and standardized audits to close compliance gaps.

Mukherjee calls for governments to mandate that platforms provide upfront, clear disclosures regarding data usage before users grant access to their photos. "These disclosures should be simple, concise, and easy to understand, helping users make more informed decisions about their privacy."

While the technology behind these AI tools is fascinating, it’s essential to remain aware of the privacy risks lurking beneath the surface. As always, when it comes to sharing personal data online, being informed is key.