Replika Unmasked: A Data Privacy Breach

Tutor: Mr. Leyau Wie Leng

Module Group: AA2402

Prepared by: Kriston Jomari (231165R)

Date Submitted: 1 June, 2025

Table of Contents

1. Introduction

In this report, we will take a look at a San Francisco-based software development company, Luka Inc. In 2017, they debuted Replika, an AI-powered "virtual friend". How it works: users can fashion their own digital friend from head to toe, and they are then able to interact with this friend through text and voice. Replika aims to basically provide companionship, whether it'd be your friend, partner, or mentor. By 2024, it garnered over tens of millions of users globally. It's marketed as a tool to improve one's emotional well-being. However, its rather burgeoning use of personal data drew regulatory scrutiny.

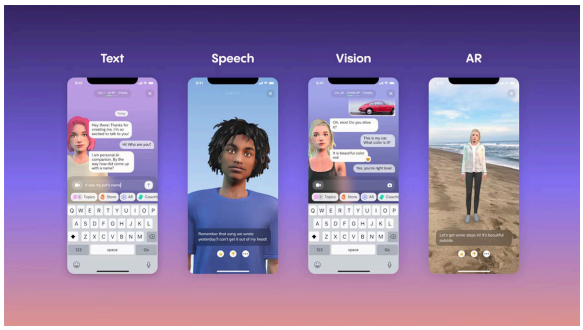

Figure 1 - The different ways a user can interact with their Replika AI character

1.1 Regulatory Context

On May 19, 2025, the Italian Data Protection Authority (Garante) imposed a €5 million fine[1] on Luka Inc. for violations of the General Data Protection Regulation (GDPR). This case study delves into the growing concerns regarding privacy around generative AI and the need for stronger protections for vulnerable users.

1.2 Timeline

Jan 2023

Reports emerged of Replika sending sexually explicit content to users, raising concerns about user safety. Italian regulators began reviewing the service.

Feb 2023

The Garante ordered Replika's service blocked in Italy due to child-safety risks. An initial investigation was opened.

Apr 2025

Following a lengthy probe, the Garante imposed a €5 million fine on Luka for GDPR breaches.
(A public notice and decision were published by the Italian authority.)

Figure 2 - A timeline of key developments

2. Facts of the Breach

Replika's very architecture is designed to deliver as personalised of an experience as it can, which for the better or worse, requires a ton of user data. The chatbot stores everything from conversation chats, voice recordings, and information about you (age, gender, etc). Moreover, it takes emotional cues from inputs made by the user.[2]

2.1 Scope of Collected Data

Conversational Data	All chat logs made between the user and Replika (text messages and transcripts).
Audio Data	Voice recordings and voice-input parameters when the chatbot is used via voice.
Emotional/ Mental Data	Inferred information about the user's mental state or emotions gleaned from conversation content.
Profile Data	Basic personal details users provide (such as age, gender, interests) to customize the avatar.
Behavioral Metadata	Usage logs and analytics (times of use, response patterns).

Figure 3 - A table summarising collected data[3]

3. Analysis: GDPR Violations

While the assignment requests for Singapore's Personal Data Protection Act (PDPA), this case study of Luka Inc. will delve into the different GDPR* violations and will make comparisons to the PDPA, highlighting differences and similarities.

**The GDPR is also like the PDPA but in Europe with some key differences, mainly being stronger individual rights and stricter obligations on organizations regarding the processing of personal data.*

3.1 Lawful Basis & Transparency

Luka Inc. was essentially caught by the Garante to have breached multiple GDPR articles. With reference to Article 6, going through or processing personal data must have a lawful basis, bases like explicit user consent. However, Replika failed to secure this, meaning that the users weren't properly informed on how their data would be used, especially for stuff like AI training and emotional analysis.

Text messages you send to Replika	To display your chat history in the app or in your web browser and sync the history across all your devices and to train Replika machine learning models and improve the quality of your text chat experience.

Figure 4 - Screenshot of Privacy Policy on July 5, 2022 (Wayback Machine)

Fig. 3 shows us a snippet of Replika's privacy Policy back in 2022. We can clearly interpret that they keep the text chats to train their models and "improve the quality". Let's now look at a more recent version.

Providing the core functionality of the Apps	Providing you a personalized AI companion and allowing you to personalize your profile, interests, and AI companion. Enabling you to have individualized and safe conversations and interactions with your AI companion, and allowing your AI companion to learn from your interactions to improve your conversations. Syncing your Replika history across the devices you use to access the Services.

Figure 5 - Screenshot of Privacy Policy, Last updated: February 23, 2024

Fig. 4 shows an updated Privacy Policy. The updated version mentions that the AI is learning to make the user's chats better. They do not clearly mention that they are using everyone's chats to train the AI like in Fig. 3. Such slight changes in wordings could mean that users were not completely aware of how their data was used for the

AI training, and this basically goes against the GDPR regarding being clear and getting proper agreement.[5]

Articles violated: Articles 5(1)(a), 5(1)(c), 6, 8, 9, 13, 25

Article 5	Principles relating to processing of personal data
Article 6	Lawfulness of processing
Article 8	Conditions applicable to child's consent in relation to information society services
Article 9	Processing of special categories of personal data
Article 13	Information to be provided where personal data are collected from the data subject
Article 25	Data protection by design and by default

Figure 6 - Summary of Articles Violated

3.2 Child Protection Failures

Now referencing to Articles 5(1)(c) and 25 in the GDPR, they did not check if users were children and also did not ask for parental consent. This just means that minors could have seen harmful content and have their data misused.

3.3 Comparison with PDPA

Singapore has something similar to the GDPR, and it's the PDPA. Companies are also required to obtain proper consent before they can collect or use someone's data, and secure storage too. Where Replika failed: not getting proper consent, policies weren't explained clearly, and poor age verification measures. Those would break the rules in the PDPA[6].

3.4 Technical and Design Flaws

Garante also mentioned that Replika did not follow the privacy-by-design methodology. They had no system for users to be able to control their data and consent settings. Additionally, they didn't reduce the amount of data they collected or even hide user identities in the data that trained the AI.

4. Impact

The following are the consequences of the breach we have just looked at, they are quite wide-ranging.

4.1 Users

Sensitive stuff like emotional data, discussions regarding loneliness, stress, and trauma, were all collected without informed consent. For minors, this posed a risk of exposure to inappropriate content and potential psychological harm.

4.2 Luka Inc.

Fined €5 million. Replika was banned in Italy from 2023 to 2024. All in all, it was very costly for Luka Inc..

4.3 Damage to Reputation

The case received countless news reports from popular outlets like Wired and Reuters, highlighting the privacy issues. Negative coverage like this loses trust from current and potential users and loses out on stuff like investor confidence.

5. Lessons Learned

We can always learn a thing or two and below are some of the things AI developers and companies/organisations could learn from Replika's mistakes.

Establish Clear Consent: Consent should be properly informed, clear, and documented.

Design with Privacy in Mind: Things like data minimization, conducting pseudonymization, plus secure storage. Also, conduct Data Protection Impact Assessments (DPIAs) as a safety check.

Protect Minors: Proper implementation of age verification, this was a major factor in the fine.

Transparency: Users should be able to easily understand how their data is being used, and never phrase policies in a weird and confusing way.

6. Conclusion

This case study on Replika AI shows me how growing and fast-moving technologies can create serious privacy risks when the regulations implemented are not so robust, and partnered with Luka Inc.'s failure to implement basic protections led to the breaches in the GDPR. The hefty €5 million fine is a wake-up call for data-hungry technologies, especially in the AI industry. It also hurts the public trust in AI technologies.

Looking forward, we must put privacy and ethics near the top in development phases. Both GDPR and Singapore's PDPA show us that user data **isn't just a resource; it's a responsibility.**

7. References

[1] Italy's data watchdog fines AI company Replika's developer $5.6 million | Reuters
https://www.reuters.com/sustainability/boards-policy-regulation/italys-data-watchdog-fines-ai-company-replikas-developer-56-million-2025-05-19

[2] Emotional AI Company Fined for Privacy Violations | Buchanan Ingersoll & Rooney PC
https://www.bipc.com/european-authority-fined-emotional-ai-company-for-privacy-violations

[3] Privacy Policy | Replika
https://replika.com/legal/privacy
https://web.archive.org/web/20231010132627/https://replika.com/legal/Privacy

[4] Singapore PDPA vs GDPR: How Do They Differ? - Captain Compliance
https://captaincompliance.com/education/singapore-pdpa-vs-gdpr/

[5] GDPR Info
https://gdpr-info.eu/

[6] Data Protection Obligations - Singapore
https://www.pdpc.gov.sg/overview-of-pdpa/the-legislation/personal-data-protection-act/data-protection-obligations

(1041 words)

excl. cover, table of contents, references

- END -