



Katholieke  
Universiteit  
Leuven

**Faculty of Engineering Science**  
*Master of Artificial Intelligence*

# **Privacy Impact Assessment for Gmail**

## **Authors**

Reniflal Ebenezer Sundaralal(r0659916)  
Savithri Radhakrishnan(r0826319)  
Vaijeyanthi Venkateswaran(r0877769)

Academic year 2022–2023

## Contents

1. Introduction.....	3
2. Product Description .....	3
2.1 Functionality .....	3
2.2 Stakeholders .....	4
2.3 Data Collection .....	5
2.4 Design Architecture .....	8
3. Privacy Impact assessment.....	12
3.1. Technical aspects .....	15
3.2. Legal aspects:.....	15
3.3. Ethical concerns .....	16
3.4. Market.....	17
3.5. Compliance assessment.....	17
3.6. LINDDUN: privacy threat modelling .....	21
4. Recommendations.....	24
5. Bibliography .....	25

# 1. Introduction

“Digital freedom stops where that of users begins...” (- Stephane Nappo).

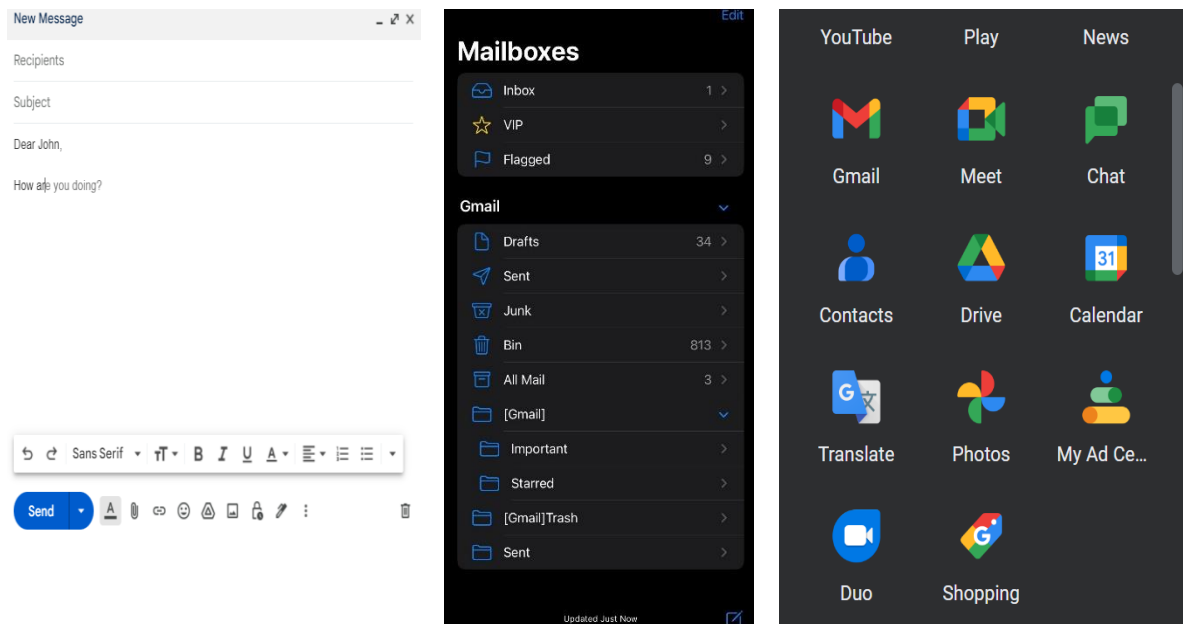
When we started using the internet, most of us signed up for one of the email services existed at that time, to define our identity in the digital world. Gmail has always been famously liked by most people for its easy-to-use interface, storage etc. Whilst it may make our life easier, we must also recognise how much we share about our life through Gmail. As much as privacy means a great deal to us, how much does it mean to the company Google which owns Gmail. Therefore, we are going to perform a privacy impact assessment on Gmail which explores to what extent Gmail preserves its user’s right to privacy in the light of existing laws and ethical principles.

Gmail is a free email communication service can be accessed on any web browser and through Gmail mobile app. The minimum age requirement to create a Gmail account is 13 in most of the countries, while in some countries it is 14 years<sup>[1]</sup>. “Gmail is by far the most popular email service, with more than 1.5 billion active users, compared to Microsoft Outlook (400 million) and Yahoo Mail (225 million)”<sup>[2]</sup>. The scope of Gmail extends far beyond the e-mail app, because we login to all google services through the Gmail account. Our digital lifestyle is carried out with several google services giving access to all our information to Google in the end. Hence while discussing how Gmail works, it's also important to understand how Gmail links to other Google services and third-party applications providing a gateway to all our data.

## 2. Product Description

### 2.1 Functionality

Gmail offers a wide range of user-friendly features. The principal feature is that users can compose an e-mail and send it to anyone anywhere with a valid e-mail address. What makes Gmail more attractive is that it gives free storage 15GB to all its users for personal use. As shown in Figure 1, Smart compose is one of the features of Gmail, uses ML algorithm, which allows users to write emails faster by suggesting the expected words; Gmail also provides Spam, phishing & malware protection. By signing up for Gmail, the users are given access to use Google services such as chat, Google meet, Google Calendar, Google drive and so on. Therefore, Gmail account can be considered as the token to experience Google world.

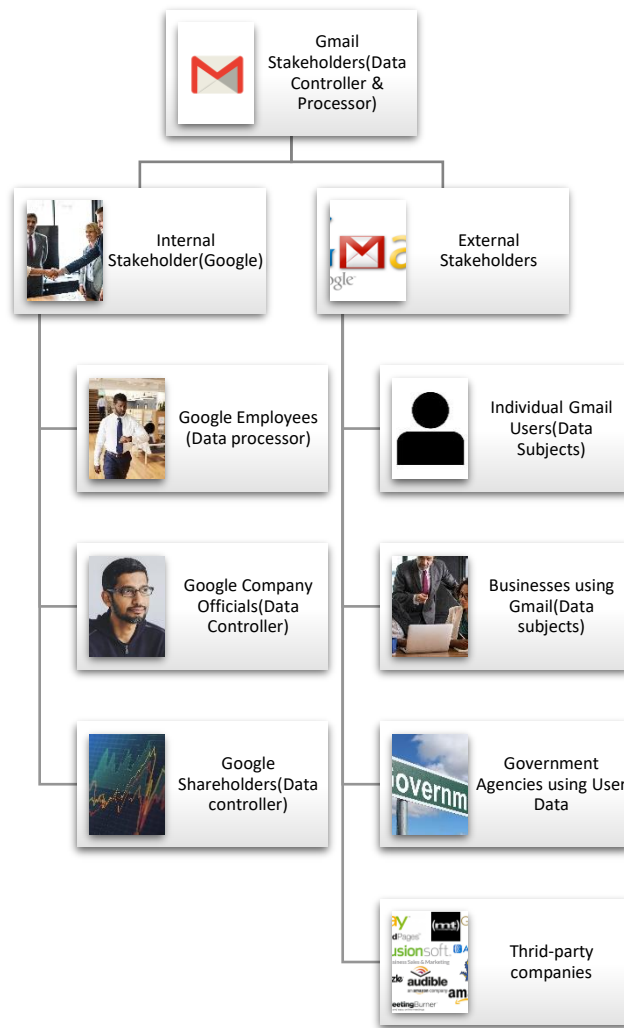


*Figure 1: Smart Compose, email categories, Google services*

## 2.2 Stakeholders

In the context of data privacy, the data subjects of Gmail are individual users and businesses and are considered be the most important stakeholders (Figure 2) and should usually put on a pedestal when it comes to how their data is being treated. The data controller is Google who gets data through Gmail app, and on an organisation level it defines the collection of data, the amount of data being retained and who are all given access to the data. When it comes to the organisation, Google's board of directors, investors should also comply to data privacy rules and regulations since they are involved in making decisions as to how user's data is being controlled. The first party within the organisation is Google employees who are responsible for developing and maintaining Google services. Hence, they are called as processor who is responsible for processing the data on behalf of Google.

The external stakeholders are Government agencies to not only ensure Google comply with data privacy regulations but also in case of Federal investigation. When a user login to a third-party application with Gmail id, they are prompted about the privacy policy of the external service provider. From that moment on, their data is shared with these third-party applications to use their extended functionalities and for advertising purposes.

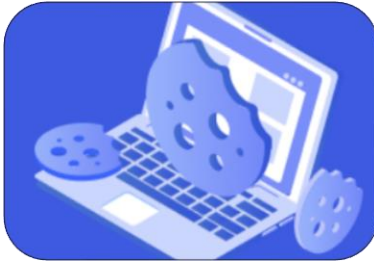


*Figure 2: Gmail Stakeholders*

## 2.3 Data Collection

To start with, we will see how data is being collected from a user and what type data ideally being provided to Google to use Gmail which uses different technologies to collect data: cookies, pixel tags, local storage, such as browser web storage or application data caches, databases, and server logs. The overview of these technologies is shown in Figure 3 and 4.

We will now see the type of data that is being collected which has been consolidated into Table 1. It is common that we must at least provide our basic information for identification purposes. To our knowledge now, based on Google's privacy policy, the company does collect more information than what we see on the screen when we first signed or want to sign up for a Gmail/Google account. Overall, the data being collected are classified into two broad categories: Things users create or provide to Google and Information collected by Google as we use their services <sup>[4]</sup>.



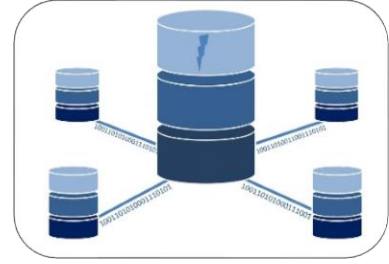
### Cookies

- Cookies are identifiers placed on a browser to allow users to access Gmail's functionality
- Essential cookies: to ensure that the primary functionality of Gmail application works fine
- Non-essential: used to tailor ads, search items etc based on user preferences



### Pixel Tags

- This technology is used to track users' activity and normally placed on a website or within a body of email. For instance, this is used to track how often the user has opened a particular email.



### Local Storage

- Browser web storage: stores data in a browser. This is used for data retrieval purposes when browser has been closed and reopened. Ex: HTML 5
- Application data cache: allows web application to run without internet
- Databases/Google cloud servers
- Server logs: records page requests when user visits Gmail application.

Figure 3: Different technologies used to collect data<sup>[4]</sup>

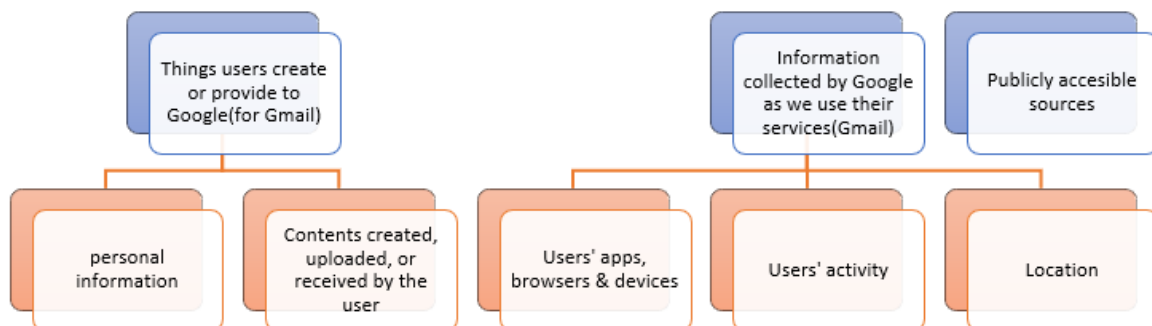


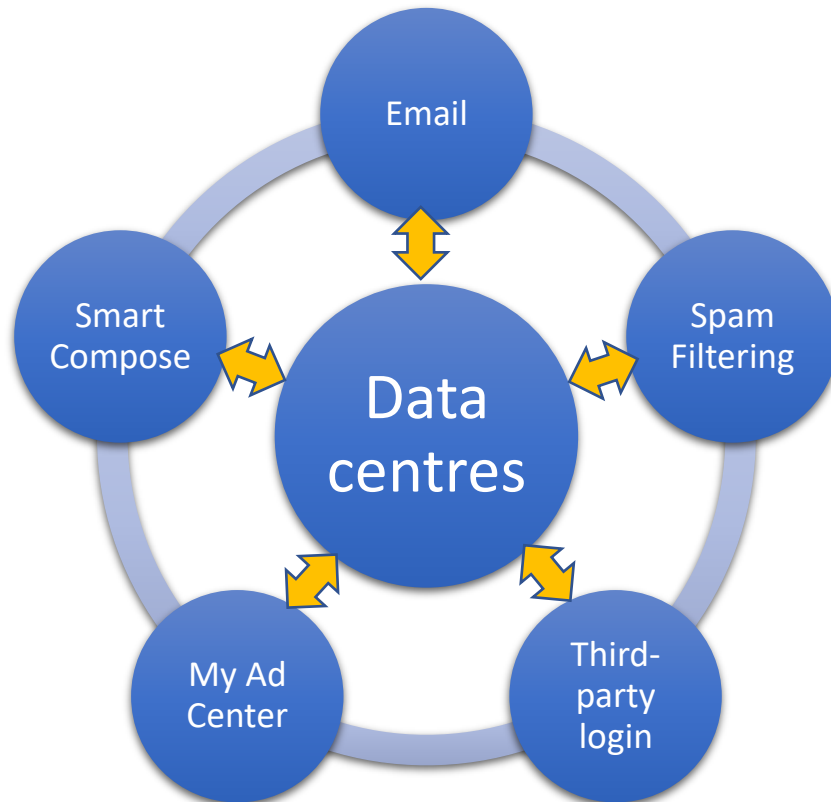
Figure 4: Data categories

Table 1: Type of Data collected and purpose

Category	Example Data	Purpose
User's personal information	Name, password, email, address, phone number, billing information	for personal identification, payment service
Contents created, uploaded, or received by the user	Emails, photos, videos, docs, spreadsheets	To provide, maintain, improve services, and develop (new) services
Users' apps, browsers & devices	unique identifier, device type and settings, operating system, mobile network information, IP address, system activity, and the date, time	uniquely identify a browser, app, or device, including security and fraud detection, syncing services such as your email inbox, remembering users' preferences, and providing personalized advertising
Web & App Activity	Battery level, how frequently the app is being used, Quality and length of users' network connections (like mobile, Wi-Fi, and Bluetooth)	To check what's using the most battery on a user's device to help make common features to use less battery; crash reports help make the Android OS more reliable.
Users' activity	Terms searched, Videos watched, Views and interactions with content and ads, Voice and audio information, Purchase activity, People with whom user communicate or share content, Activity on third-party sites and apps that use Google services, sender, and recipient email address	for recommendation features such as suggesting YouTube videos based on users' likes and dislikes
Location	GPS and other sensor data, IP address, places labeled as Home or Work, information about things near users' device, such as Wi-Fi access points, cell towers, and Bluetooth-enabled devices	To offer features such as driving directions, search results for things near user, and ads based on user's general location.
Publicly accessible sources	users'/business name in local newspaper	The data retrieved from public resources are used to generate more helpful Google's language models (Ex: Gmail's smart compose)

## 2.4 Design Architecture

The Gmail application comprises of five important features that are interconnected: Email, Third-party login, Google AdSense, Smart Compose, Spam Filtering and uses Data centres to access data (Figure 5).

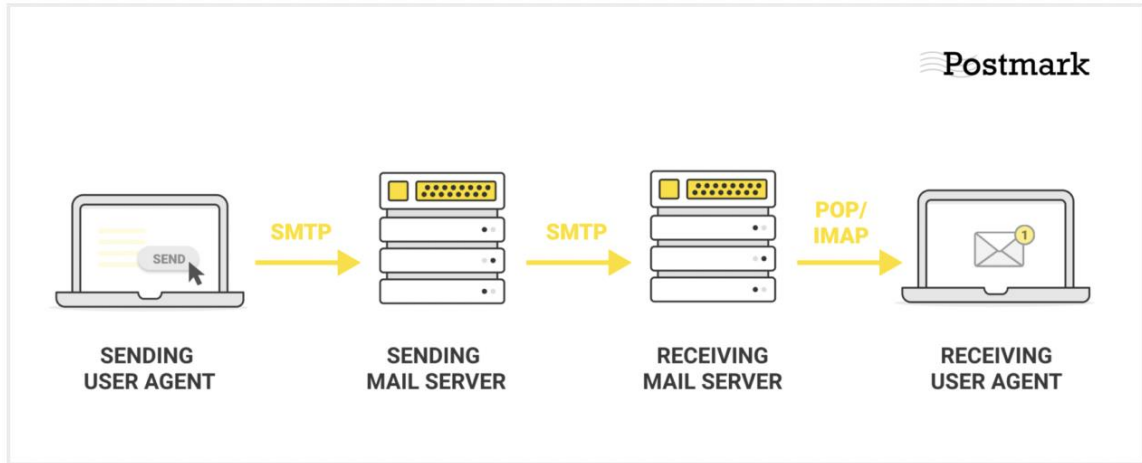


*Figure 5: Gmail in a nutshell*

### 2.4.1 E-mail Architecture:

Gmail uses SMTP (Simple Mail Transfer Protocol) to send emails, the process is illustrated in Figure 6. SMTP servers handle sending, receiving, and relaying of emails. Gmail primarily uses IMAP (Internet Message Access Protocol) to receive emails but supports POP (Post office Protocol) email access as well from other mailbox client applications.





Different protocols in the sending process: SMTP is used to send email, POP and IMAP to receive mail

Figure 6: SMTP protocol <sup>[14]</sup>

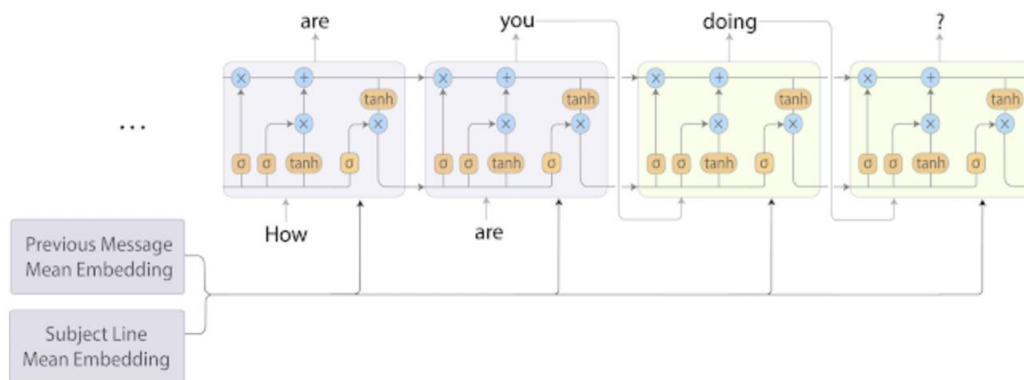
## 2.4.2 Data Centres

Gmail uses ISO certified <sup>[15]</sup> Data Centres to store its massive data and are distributed across the globe <sup>[39]</sup>.

- 🌐 Operating System: a customized version of Debian Linux <sup>[17]</sup>
- 🌐 Network Infrastructure: Clos network topology <sup>[18]</sup> to handle data traffic between data centres

## 2.4.3 Smart Compose:

As explained before, Smart Compose is a feature which gives autocomplete suggestions. As shown Figure 7, Gmail combines Bag-of-Words model with an RNN-LM (Recurrent Neural Network – Language Model) for Smart Compose. It uses previous emails and the email subject to identify context. Google researchers have no access to the actual email content <sup>[34]</sup>, instead they work on simulated data to develop and train the model.

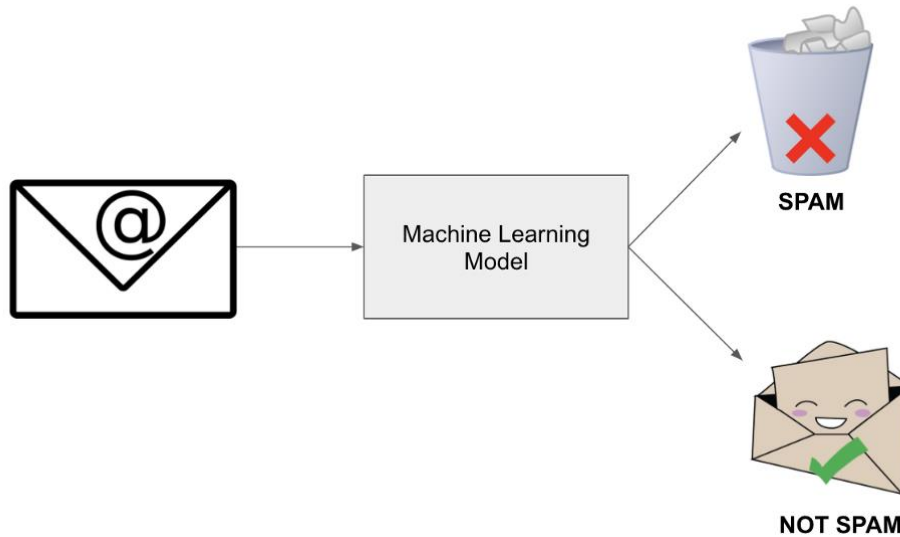


Smart Compose RNN-LM model architecture. Subject and previous email message are encoded by averaging the word embeddings in each field. The averaged embeddings are then fed to the RNN-LM at each decoding step.

Figure 7: Smart Compose Architecture <sup>[34]</sup>

#### 2.4.4 Gmail Spam Filter

Gmail uses artificial neural networks for spam filtering. The artificial neural network uses “spam”, “not spam” selected by the user as well as user behaviour while using the application as input for training the neural network <sup>[35]</sup>. Since this is trained using the specific behaviour of the user, the spam filtering works differently from person to person (Figure 8).



*Figure 8: Spam Filter working model <sup>[40]</sup>*

#### 2.4.5 Third-Party Web logins:

With Gmail account, users can login to third-party applications. Google uses OAuth (Open Authorization) for this purpose. Many of these third-party websites have two options to login to their account: a regular sign-up or use OAuth login <sup>[23]</sup>. When a user logs in using Gmail OAuth, the next window will show what information will be shared with the third-party website. It is usually information like email, phone, contacts, etc. Once agreed, Gmail returns to the third-party application and gives an authentication token. The third-party website will then use this token to access the limited information <sup>[23]</sup>. The authentication process happens as shown in Figure 9.

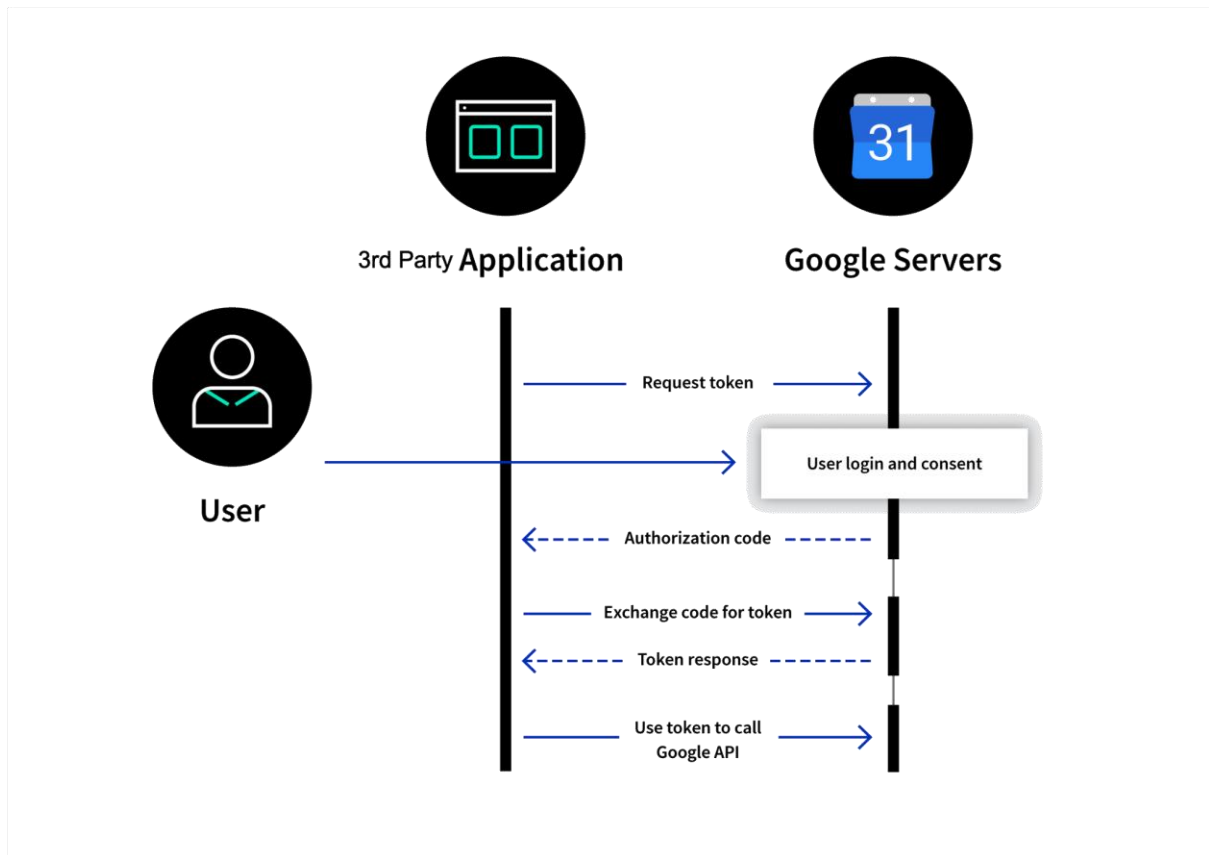


Figure 9: OAUTH flow diagram <sup>[24]</sup>

#### 2.4.6 My Ad Center:

Figure 10: Gmail's My Ad Center allows users to personalize their ad settings. For this it uses data collected through Cookies, user's activity (Table 1: Search terms, YouTube videos watched etc).

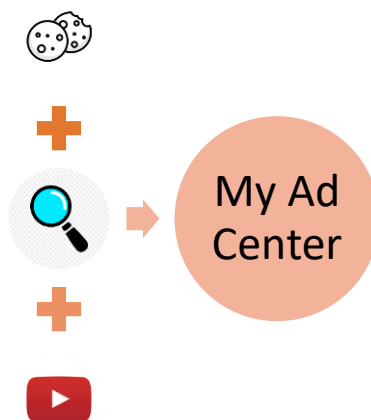


Figure 10: My Ad Center diagram





### 3. Privacy Impact assessment






*“Privacy is not just about learning your name. What fundamentally for me privacy is about inferences made about you that reduce personal self-determination. i.e., It’s when an inference about you is used to make decisions about your life or maybe to nudge you into making decisions about your life.” – Dr Joss Wright (from Lecture 02 of Privacy and Big Data)*






Taking inspiration from one of the lectures of Privacy and Big Data, we understand that the real harm in privacy is not just about collecting data but also inferences drawn based on data and the way in which it affects an individual’s life. Hence, it is extremely crucial that the data subjects(users/businesses) should be able to practice their right to privacy when they use Gmail.

Table 2 consolidates different types of data collected, the associated risks and GDPR requirements which will be discussed in later section in detail.

*Table 2: Gmail Privacy Impact Assessment overview*

Type of data collected	Source of information	User Trust assumption	Necessary/ Unnecessary	Privacy impact
Name, Login password, age	During creation of Gmail account	Data is not used for profiling and is not shared with external service providers.	Necessary	 Risk of profiling and discrimination (Ref: Art. 5 and Art. 6, GDPR)
Gender	During creation of Gmail account	Data is not used for profiling and is not shared with external service providers.	Unnecessary	 Risk of profiling and discrimination (Ref: Art. 5 and Art. 6, GDPR)
Scanning email content, recipient email address	User’s activity	Email content and contact information is not scanned/shared.	Unnecessary	 (Ref: Art. 5 and Art. 6, GDPR)
Browser connected to Gmail account	User’s activity with Gmail-connected devices	Search history is not scanned/shared.	Necessary	

				<p>Users have control over what data is shared.</p> <p></p> <p>But there is a risk of user identification (See LINDDUN analysis)</p>
Search terms	Information collected by Google as we use their services <sup>34</sup>	Search history is not scanned/shared.	Necessary for advertisements	<p></p> <p>Users have control over what data is shared.</p>
Search history	Cookies	Search history is not scanned/shared.	Unnecessary	<p></p> <p><u>Risk of profiling and discrimination.</u> (Ref: Articles 5 and 6, GDPR for processing of personal data)</p>
Location (especially labelled locations like 'Home' and 'Work')	User's activity with Gmail-connected devices	Users are not tracked.	Necessary, if and only if users willingly agreed to share such details.	<p></p> <p>Users have control over what data is shared. (Ref: Articles 5 and 6, GDPR)</p>
Bluetooth devices, Cell tower, WiFi points near the Gmail-connected device	Google Fast Pair Service ( <b>GFPS</b> )	Users are not tracked.	Unnecessary	<p></p> <p>Risk of tracking and profiling users and their recipients.</p>

				(Ref: Art.5 and Art.6, GDPR)
Voice/Audio information- especially “few seconds before users say ‘Hey Google’ “	Voice action services	Audio is not collected before the activation command and users are not re-identified based on their voice profiles.	Storage of voice data that is few seconds before activation command ‘Hey Google’ is <u>unnecessary</u>	 <p>Although Google encrypts the audio transcripts during transit, collection of audio data prior to activation command is a clear infringement of user privacy.</p>
Message and call logs, duration of calls, frequency of calls	User’s activity with Gmail-connected devices	Data is not transparent to the backend.	Unnecessary	 <p>Risk of tracking and profiling users and their recipients.</p>
Videos watched on YouTube, through Gmail account	User’s activity with Gmail-connected devices	Profiling and discrimination do not occur.	Necessary for advertisements	 <p>Users have control over what data is shared.</p>
Purchase activity	User’s activity with Gmail-connected devices	Data is not used for profiling and is not shared with external service providers.	Unnecessary	 <p>Risk of tracking and profiling users.</p> <p>(Ref: Art. 17 GDPR-Right to Erasure or the ‘Right to be forgotten’)</p>
Data retention after deletion (for a period of 180 days)	Not clear	Data is deleted as soon as users choose to delete	Unnecessary	

		all information linked with Gmail and thereby other Google apps.		<u>This is a clear case of privacy infringement. (Ref: Art. 17 GDPR-Right to Erasure or the ‘Right to be forgotten’)</u>
--	--	--	--	--

In the following section we discuss the privacy impact assessment of Gmail through the lens of Lessig’s Modalities of Regulation-namely: Law, Technology, Norms/Ethics and Market.

### 3.1. Technical aspects

**Lack of strong data encryption:** Google automatically encrypts data transmitted by users via Gmail, using Transport Layer Security (TLS). TLS encryption is a widely used data encryption method which makes sure that data cannot be tampered during transmission. But this comes with a drawback: TLS only works if the receiver has the same kind of encryption. Even though TLS ensures connection security between two devices or servers, some versions of TLS are prone to Man-in-the-Middle (MiM) attacks. MiM is a type of ‘eavesdropping in which the attacker intercepts and then controls the entire conversation.’ <sup>[28]</sup>

### 3.2. Legal aspects:

The General Data Protection Regulation (GDPR) enlists legal requirements to be complied by the data controller (in this case, Gmail) or service provider, towards the protection of people’s privacy rights (applicable for EU citizens and residents of the EU). It is to be noted that the data protection rights however apply only to data that can be identifiable, not to anonymized data, as mentioned in Article 2 GDPR <sup>[31]</sup>. Furthermore, Article 4 GDPR clearly defines the terms data, processing, profiling etc that are henceforth used to assess the privacy impact for Gmail users <sup>[32]</sup>.

Users of Gmail expose a lot of personal data including names, recipient email addresses, attachments and conversations which can be identifiable by hackers in a classic Man-in-the-Middle (MiM) attack. The following section discusses the relevant legal requirements that help in assessing the privacy impact for the primary app users.

1. **Lawful processing of personal data:** Article 5 and 6 of GDPR insists on the necessity for lawful, fair, and transparent processing of sensitive data (identifiable data such as name, location, race, age, ethnicity) “Requires that personal data shall be processed lawfully, fairly and in a transparent manner.” <sup>[31], [33]</sup>. As per the data collection policy of Gmail, user data is collected both for actual functioning and for extended features like Autofill, text classification which poses threat to privacy of users. In particular, the classification of Gmail content into “Primary”, “Spam” and other custom-made categories require collection and processing of personal data which need to be strictly encrypted and

pseudonymized according to GDPR regulation. Despite various data anonymization strategies adopted by Google Cloud services, user data is at risk for hackers and in case of a data leak.

2. **Data minimization:** Article 5 of GDPR <sup>[9]</sup> insists on collection of personal data only to an extent that is deemed necessary for the direct functionality of the service: “Personal data shall be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed (‘data minimization’)”. With respect to the auto-fill feature of Gmail, it is not clearly stated if the data collected for this feature is being stored in the backend server of Gmail. This feature clearly doesn’t comply with the data minimization requirement of GDPR.
3. **Data retention:** With reference to Table 2, we see that Gmail takes a period of 180 days just to delete users’ data connected to a Gmail account, following the users’ request for deletion. Following a user’s request for deletion, Google Cloud follows a four-stage process (Figure 11) to delete all the data, it is not clear what data remains in the cloud for such a long period of time. This is a clear violation of the GDPR requirement of the data subjects’ Right to be forgotten.

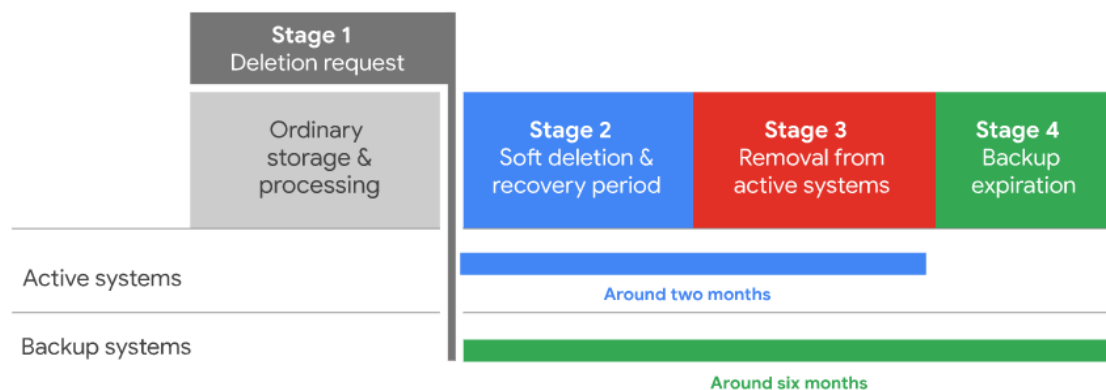


Figure 11: The four-stage data deletion process adopted by Google cloud services.

### 3.3. Ethical concerns

- **Transparency** - “Transparency is a property of a system that makes it possible to get certain information regarding a system’s inner workings” <sup>[26]</sup>. In this context, Transparency entails the users’ rights to know what type of data is being collected, how the data is being processed, along with the right to know if/when they are being tracked. The major concern with usage of Gmail is the humongous amount of data that is being collected from every user, apart from the fact that data retention-deletion policies are not obvious for naive as well as experienced users. From Table 2, we see that data collection includes Bluetooth devices, cell towers and Wi-Fi points near the Gmail-connected devices, which is a violation of Data Minimization requirement of GDPR, and Gmail thereby is not transparent about why such data is collected and what inference is being made.
- **Risk of user profiling and discrimination** - Collection of user data like language preferences, searched items can lead to user profiling, as can be seen in targeted



advertisements. Further, the storage and access of this information in the backend can be seen as a threat in case of employers trying to extract information about their employees or potential candidates. From the privacy policy, it is evident that Google Cloud uses the collected information for security/fraud detection purposes <sup>[5]</sup>, and this might inadvertently label some users as criminals or frauds and thus lead to unforeseen circumstances. According to Recital 71 of GDPR, “Data subjects have the right to obtain human intervention and to challenge the decision” <sup>[29]</sup>. In addition, this Recital also states that data subjects (users) have the right to not be subject to any kind of profiling that is entirely automated.

### 3.4. Market

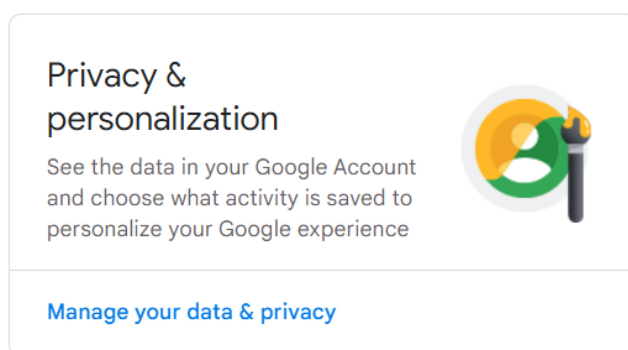
Gmail predominantly generates revenue through e-mail marketing, personalised ads using AdSense, and GSuite subscription which business and professionals to create their own domain emails. Gmail can make market decision, perform research, refine its features based on inference drawn from the collected data.

One of the four modalities of regulation is Market, if Gmail can identify market trends at an early stage, it can react to such trends effectively based on what users prefer. Although, it is a free supply and demand, we can see that the demand is only being “predicted” using big data analytics which helps large companies to identify potential customers and make more profit, but not the data subjects in general. The research work performed on collected data (after getting users’ consent) leads to new technological innovations, for instance, Smart Compose.

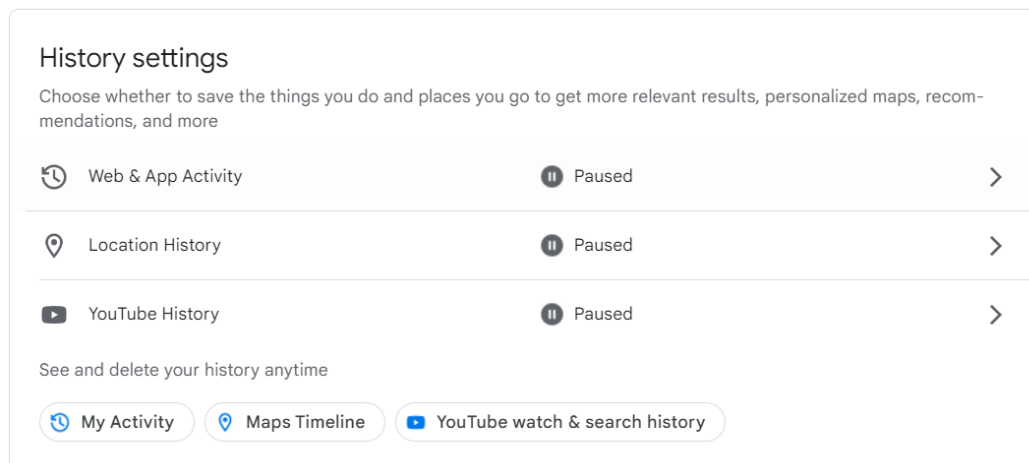
### 3.5. Compliance assessment

In this section we discuss the legal, ethical compliance strategies adopted by Gmail in accordance with GDPR checklist <sup>[38]</sup> and Article 32 <sup>[30]</sup>

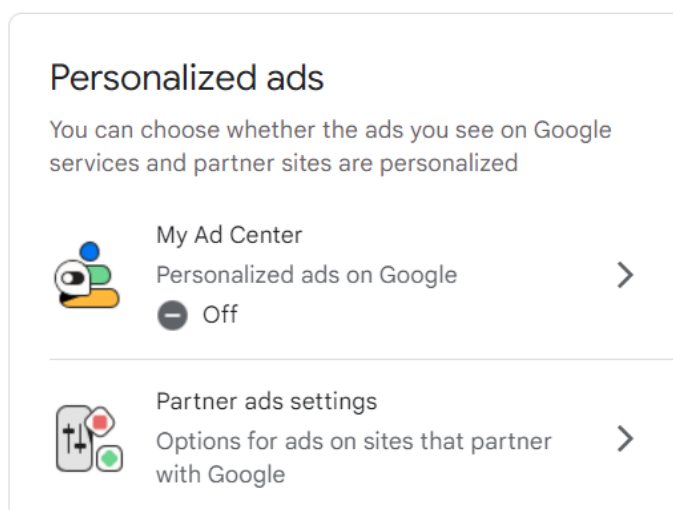
- Gmail provides Privacy recommendations for its users.



- Users can view and change their web activity.



- Users are given control over personalized advertisements

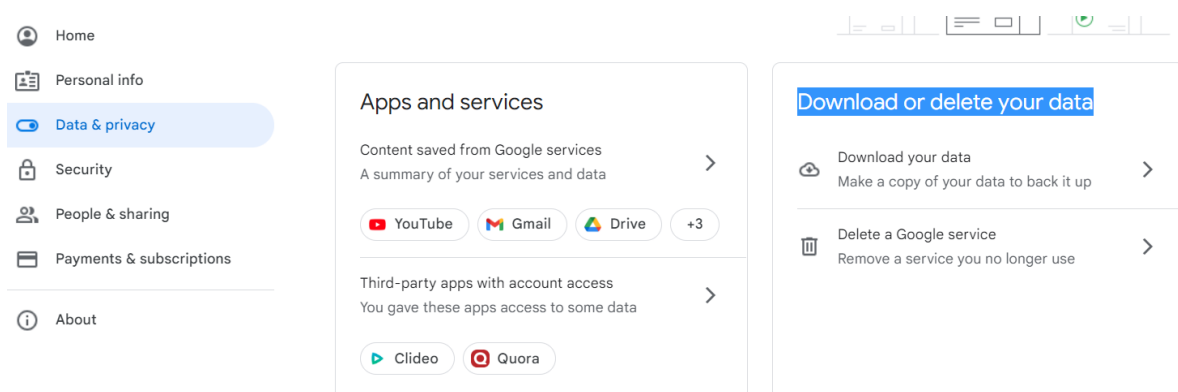


Gmail's Data Anonymization Methods	Example
<b>K-anonymity</b>	Sensitive data such as pin code and telephone number are replaced by the same set of sequential numbers
<b>L-diversity</b>	To ensure users cannot be identified based on the search topics
<b>Differential privacy</b> <sup>[13], [27]</sup>	Addition of noise to sensitive data to make it unidentifiable

- Google Takeout: An easy-to-use interface to view/download users' data. The screenshots below are taken in real time to demonstrate how Gmail allows users to view/download shared data.

### Step 1:

Go to My account → Data & Privacy → scroll to Download or delete your data



### Step 2:

In Google takeout screen: select the data items/type you would like to download/delete → Data & Privacy → scroll to Download or delete your data (the user can also to delete data shared for research purposes)

## ← Google Takeout

Your account, your data.  
Export a copy of content in your Google Account to back it up  
or use it with a service outside of Google.

### YOUR EXPORTS

#### Your latest export

47 products on December 4, 2022

Manage exports

### CREATE A NEW EXPORT

#### 1 Select data to include

47 of 48 selected

Products

Deselect all

#### 1 Select data to include Photos you added yourself, as well as contacts from your interactions in Google products like Gmail. [More info](#) 47 of 48 selected

vCard format



#### Crisis User Reports

Information provided to help others during crises



CSV format



#### Data Shared for Research

Responses saved with your Google Account from your participation in Google research studies and projects.



Multiple formats



#### Drive

Files you own that have been stored in your [My Drive](#) and [Computers](#). [More info](#)



Multiple formats

Advanced settings

All Drive data included

## Step 3:

Select frequency and file type → click on Export.

#### 2 Choose file type, frequency & destination

Transfer to:

Send download link via email

When your files are ready, you'll get an email with a download link. You'll have one week to download your files.

#### Frequency



Export once

1 export



Export every 2 months for 1 year

6 exports

#### File type & size

File type:

.zip

Zip files can be opened on almost any computer.

File size:

2 GB

## ← Google Takeout

### CREATE A NEW EXPORT



Select data to include

47 of 48 selected



Choose file type, frequency & destination

#### Export progress



Google is creating a copy of files from 47 products

This process can take a long time (possibly hours or days) to complete. You'll receive an email when your export is done.

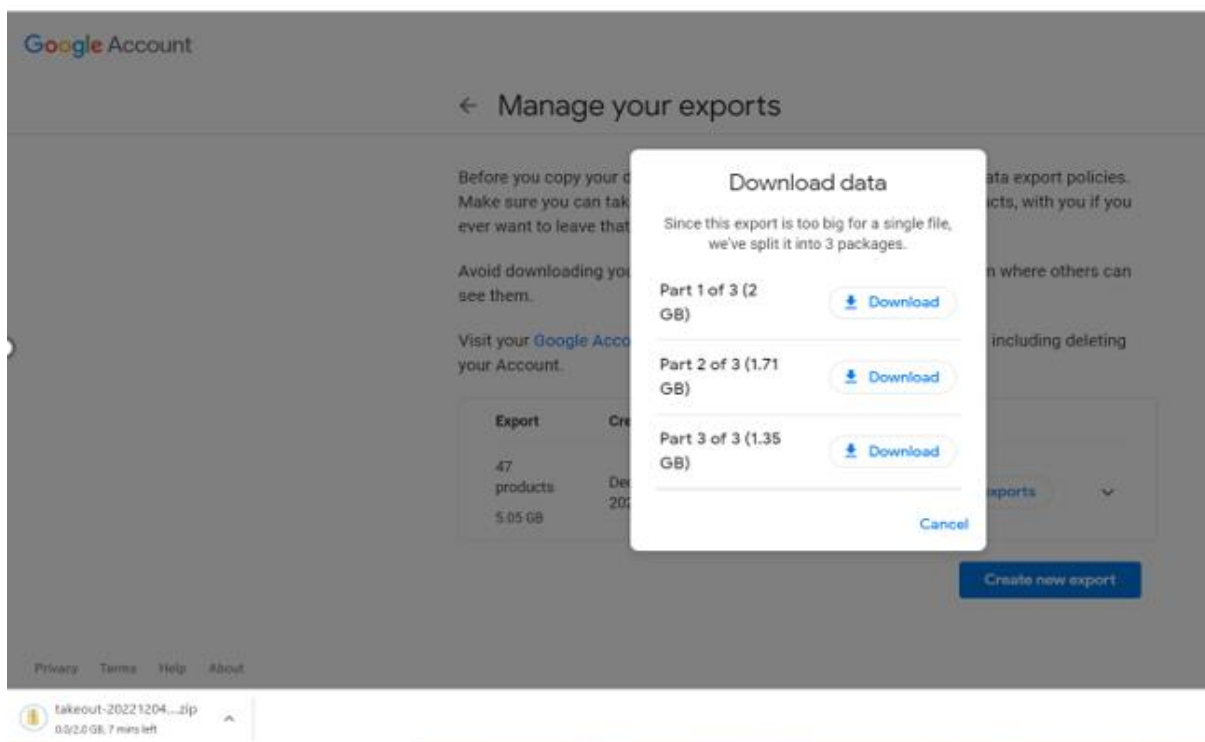
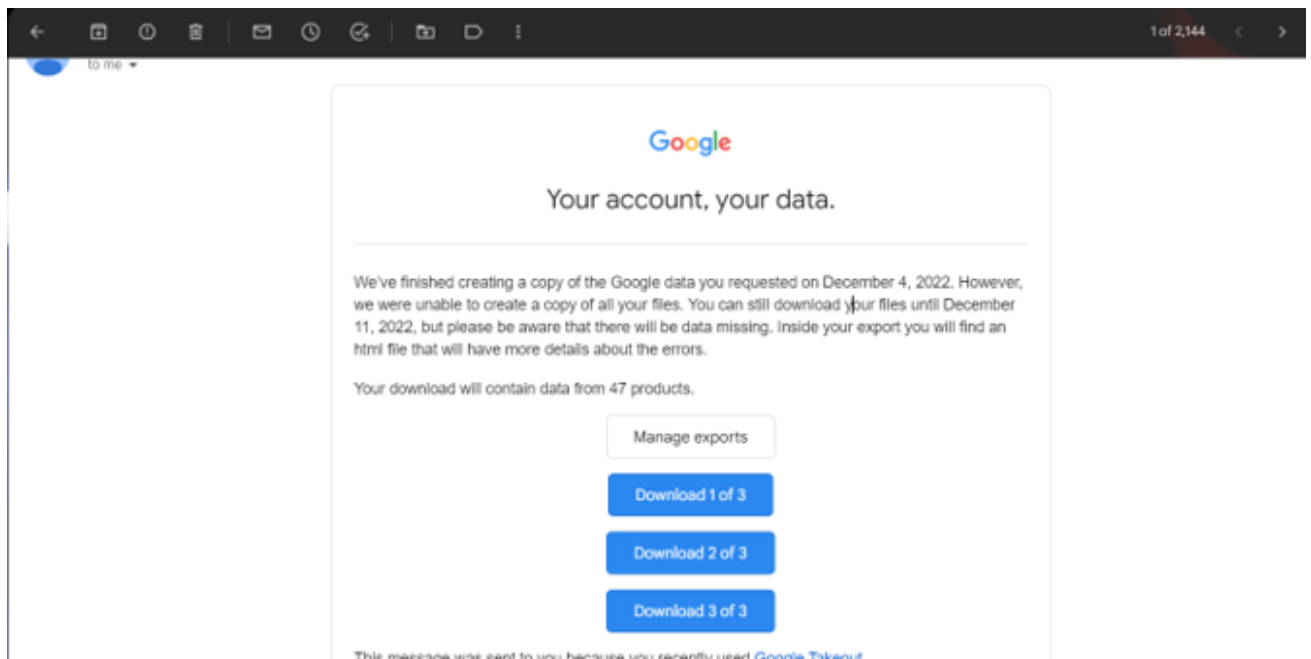
Created: December 13, 2022, 9:36 PM

Cancel export

Create another export

#### Step 4:

Once the user receives the email (took 2hrs) → click on Download.



### 3.6. LINDDUN: privacy threat modelling

We first need to identify potential privacy threats of Gmail so that we can come with recommendations. We used LINDDUN for privacy threat modelling<sup>[37]</sup>, a proactive approach of analysing from the attacker's perspective. It allows analysts to elicit threats and come up with mitigating strategies in software architecture. The model has 7 threat categories:

Linkability, Identifiability, Non-repudiation, Detectability, Disclosure of Information, Unawareness, and Non-compliance.

**Step 1:** LINDDUN uses a Data Flow Diagram (DFD) as graphical model of the system-under-analysis <sup>[37]</sup>: Gmail (Figure 12)

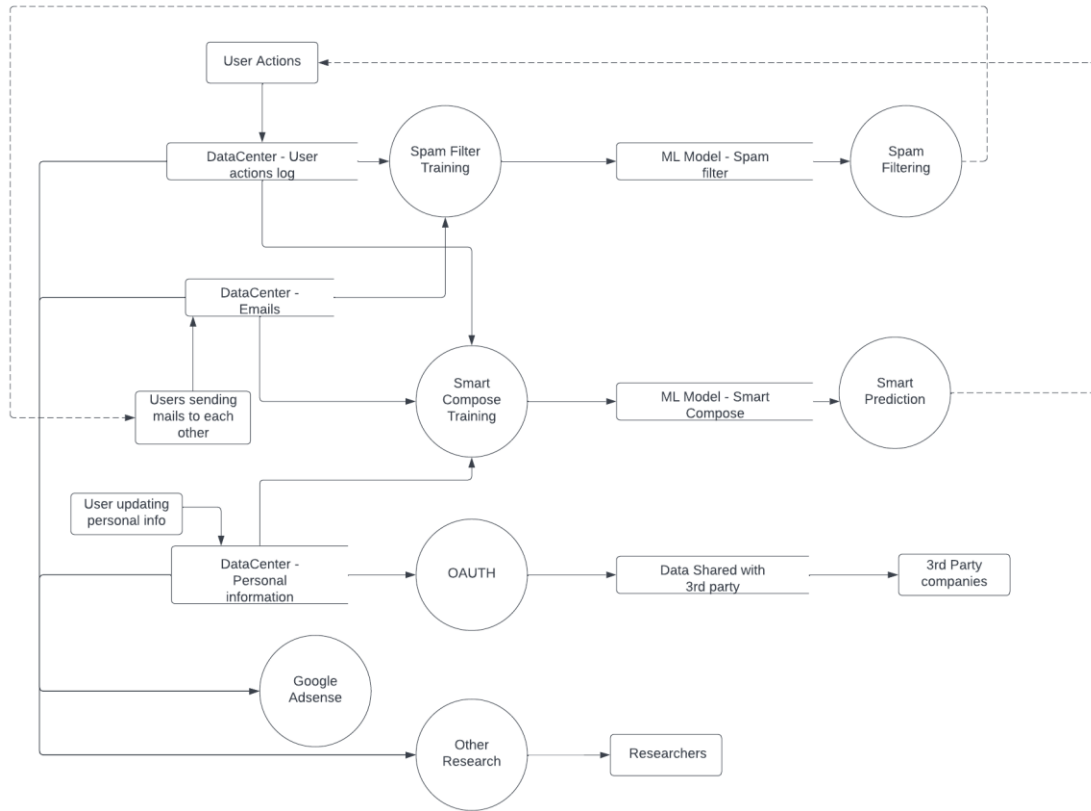


Figure 12: Data Flow Diagram (DFD)

**Step 2:** Systematically iterate over the DFD elements to identify privacy threats <sup>[37]</sup> (Tables 3 & 4)

Table 3: LINDUNN threat analysis

	Threat Target	L	I	N	D	D	U	N
Dataflow	DataCenter -> DataCenter							
	Browser -> DataCenter	A1	A1		A1	A1		
	DataCenter -> Browser	A1	A1		A1	A1		
	DataCenter -> 3rd Party	A2	A2		A2	A2		
	DataCenter -> Research	A8						
Entity	Users	A1	A2				A3	
	Gmail employees and							

	researchers							
	3rd Party							
Process	Training Spam Filter Model							
	Training Smart Compose Model							
	Spam Filtering							
	Smart Predictions							
	Data sharing with 3rd Party	A2	A2	A2	A2	A2		
	Other Research	A8						A6
	Google AdSense			A7				
Datastore	DataCenter	A5						A4
	3rd Party DataBase	A2	A2		A2	A2		A2

Table 4: Privacy threats identified

A1	TLS certification is not fully secure if sender and receiver of mail are not both encrypted. Also, need to minimize data
A2	No clear understanding of 3rd party security and privacy policies. Google shares the information after consent from data subject to any website which uses OAUTH.
A3	User is unaware of in which research their data is used and in which location and how their data is stored
A4	Google stores data for up to 180 days after user requests to delete it
A5	Violates Data minimization principle
A6	Google gets a blanket consent for the collected data to use in research. But according to GDPR, consent should be for specific purpose and research areas
A7	User actions are tracked by both the AdSense and the 3rd party for payment and success rate calculations
A8	Risk of Linkability: Large amounts of data are collected and used in Research

## 4. Recommendations

Based on the identified threats, we have come up with recommendations consolidated into Table 5. These improvements will help Gmail to better protect the privacy of the users and improve the trust users have on Gmail and Google as whole.

*Table 5: Recommendations*

<b>Aspects (Technical/Legal/Ethical/Market)</b>	<b>Risks</b>	<b>Recommendations and Mitigation strategies</b>	<b>User Impact (How does the recommendation positively impact users?)</b>
Ethical & Legal	Linkability, Non-Compliance, Identifiability	Collect only necessary data. For instance, e-mail communication with non-Gmail contact (Ex: abc@yahoo.com) should not be used for purposes other than intended purposes.	Ensures trust, complies with GDPR data minimization standard
Legal	Non-Compliance	Get consent from the user before using their data in each research work.	Gives users more control over their data
Legal	Non-Compliance	Gmail shouldn't take more than 30 days to delete the data. The processes taken by Gmail for data deletion at every stage should be notified to the users explicitly to ensure transparency and trust.	Compliance to GDPR - right to be forgotten.
Market	Linkability and Identifiability	Share data with third-party only if they comply with the legal and ethical	Makes sure that data is shared only with websites which are safe.



		requirements	
Technical	Identifiability	TLS is good only when both sending and receiving mail servers use TLS. Hence an additional encryption would help or if not, inform the user that the receiving server is not TLS encrypted.	Makes email communication more secure.

## 5. Bibliography

1. Age limit: <https://support.google.com/accounts/answer/1350409?hl=en#zipy=%2Ceurope>
2. <https://www.theguardian.com/technology/2021/may/09/how-private-is-your-gmail-and-should-you-switch>
3. <https://policies.google.com/privacy#footnote-cookies>
4. <https://policies.google.com/privacy#infocollect>
5. Type of data collected: [Privacy Policy – Privacy & Terms – Google](#)
6. Purpose of data collection: <https://policies.google.com/privacy#whycollect>
7. Data retention: <https://policies.google.com/privacy#inforetaining>
8. Data retention policy: <https://policies.google.com/technologies/retention?hl=en>
9. Article 5 GDPR: <https://gdpr-info.eu/art-5-gdpr/>
10. Data Protection Impact Assessment (DPIA): <https://cloud.google.com/privacy/data-protection-impact-assessment>
11. <https://www.theguardian.com/technology/2021/may/09/how-private-is-your-gmail-and-should-you-switch>
12. <https://www.gov.uk/government/news/cma-to-investigate-google-s-privacy-sandbox-browser-changes>
13. Google Cloud GDPR Compliance: <https://cloud.google.com/privacy/gdpr>
14. <https://postmarkapp.com/guides/everything-you-need-to-know-about-smtp>
15. <https://www.google.com/about/datacenters/>
16. [https://en.wikipedia.org/wiki/Google\\_data\\_centers](https://en.wikipedia.org/wiki/Google_data_centers)
17. [https://events.static.linuxfound.org/sites/events/files/lcjp13\\_merlin.pdf](https://events.static.linuxfound.org/sites/events/files/lcjp13_merlin.pdf)
18. <https://storage.googleapis.com/pub-tools-public-publication-data/pdf/43837.pdf>
19. [https://www.washingtonpost.com/business/technology/google-encrypts-data-amid-backlash-against-nsa-spying/2013/09/06/9acc3c20-1722-11e3-a2ec-b47e45e6f8ef\\_story.html?tid=a\\_inl\\_manual](https://www.washingtonpost.com/business/technology/google-encrypts-data-amid-backlash-against-nsa-spying/2013/09/06/9acc3c20-1722-11e3-a2ec-b47e45e6f8ef_story.html?tid=a_inl_manual)
20. Google privacy policy: <https://policies.google.com/privacy>
21. <https://cryptome.org/2012/12/google-cloud-sec.pdf>

22. <https://www.theguardian.com/technology/2021/may/09/how-private-is-your-gmail-and-should-you-switch>
23. <https://www.scienceabc.com/innovation/oauth-how-does-login-with-facebook-google-work.html>
24. <https://www.nylas.com/blog/integrate-google-oauth>
25. <https://policies.google.com/technologies/cookies?hl=en-US>
26. <https://ethics-of-ai.mooc.fi/chapter-4/1-transparency-in-ai>
27. Anonymization <https://policies.google.com/technologies/anonymization?hl=en-US>
28. <https://www.techtarget.com/iotagenda/definition/man-in-the-middle-attack-MitM>
29. <https://gdpr-info.eu/recitals/no-71/>
30. <https://gdpr-info.eu/art-32-gdpr/>
31. <https://gdpr-info.eu/art-2-gdpr/>
32. <https://gdpr-info.eu/art-4-gdpr/>
33. <https://gdpr-info.eu/art-6-gdpr/>
34. <https://ai.googleblog.com/2018/05/smart-compose-using-neural-networks-to.html>
35. <https://techtheday.com/google-makes-use-of-machine-learning-to-help-gmail-users-deal-with-spam/>
36. Jiow, H. J. (2013). Cyber crime in Singapore: An analysis of regulation based on Lessig's four modalities of Constraint1. *International Journal of Cyber Criminology*, 7(1), 18-27.
37. LINDDUN: <https://www.linddun.org>
38. <https://gdpr.eu/checklist/>
39. <https://www.google.com/about/datacenters/locations/>
40. <https://towardsdatascience.com/logic-and-implementation-of-a-spam-filter-machine-learning-algorithm-a508fb9547bd>