

Supplementary materials

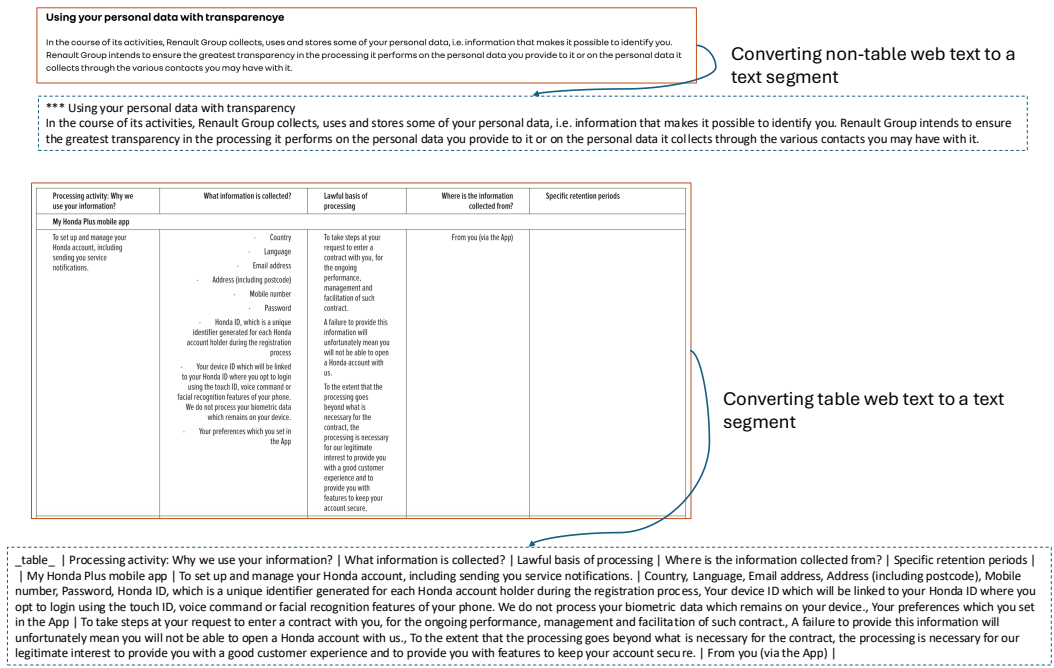


Fig. 1. Example of converting non-table and table text to text segments

Prompt

role: system

content:

- [1] You are an expert to analyse the **TEXT SEGMENT** to extract data flows.
- [2] If **TEXT SEGMENT** starts with `_table_:` Treat `|` as separators; Treat first line as the table heading; Treat second line as the table content.
- [3] Read and understand the **TEXT SEGMENT** and then strictly follow the below rules to produce your responses:

- (a) If the **TEXT SEGMENT** at least talk about one party collects data or personal information from another party, or a party shares data or personal information to another party, OUTPUT the extracted data flows in multiple JSON objects.
- (b) The JSON objects must use the format:

```
{ "Output": [{
  "data_sender": "",
  "data_type": [],
  "data_receiver": []
}] }
```

- (c) Respond only with valid JSON.
- (d) Each data flow represents one party (i.e., `data_receiver`) collects personal data (i.e., `data_type`) from another party (i.e., `data_sender`), or a party (i.e., `data_sender`) shares personal data (i.e., `data_type`) to another party (i.e., `data_receiver`).
- (e) For `data_types`: extract all atomic personal data (i.e., `data_type`) following these rules: (1) when dealing with sentences that have combined `data_types`, split them into individual `data_types` for a clearer representation; (2) each `data_type` MUST appear in **TEXT SEGMENT**, and do not change the cases; (3) DO NOT INCLUDE any other text in the answer such as input or query text or your deduction or your explanation; (4) remove Pronouns in the identified strings; (5) DO NOT INCLUDE specific addresses, postcodes, email addresses, companies, organisations, or geographical information; (6) if you can not identify a `data_type`, leave it empty.
- (f) For `data_sender/data_receiver`: (1) when dealing with sentences that have combined `data_receivers` or `data_senders`, split them into individual `data_receiver` or `data_sender` for a clearer representation; (2) each `data_sender` or `data_receiver` string MUST appear in **TEXT SEGMENT**, do not change the cases; (3) if no `data_sender` is explicitly stated in the **TEXT SEGMENT**, leave `data_sender` empty; (4) if no `data_receiver` is explicitly stated in the **TEXT SEGMENT**, leave `data_receiver` empty.
- (g) OUTPUT only "None" for other unrelated scenarios

role: user

content:

TEXT SEGMENT: One Privacy Policy Text Segment

Fig. 2. Prompt template for extracting data flows following the Groq Chat Completions API format

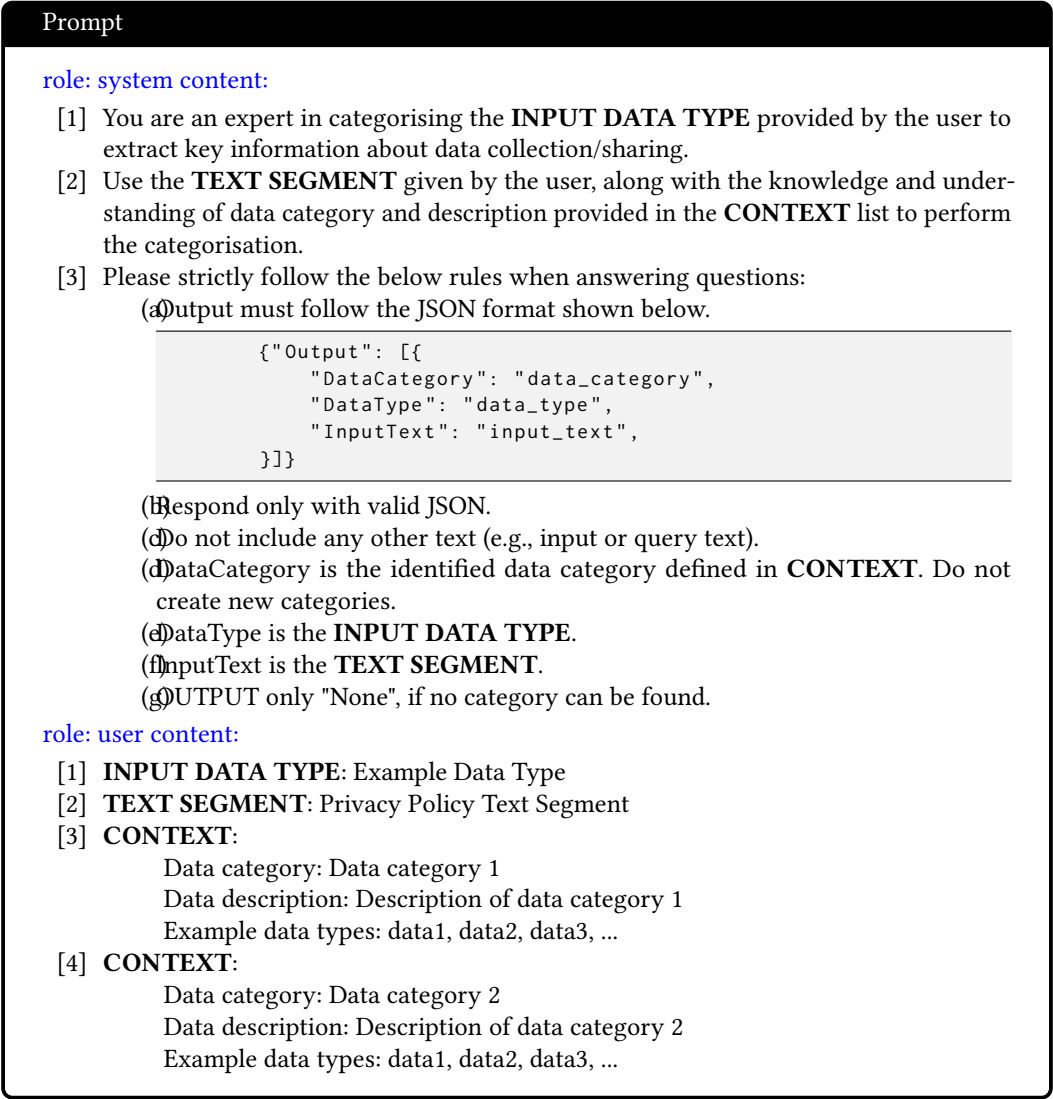


Table 1. Data collection and sharing practices stated in Google app safety section

Data Category	Audi		Ford		Honda		Hyundai		Kia		Lexus		Nissan		Polestar		Renault		Vauxhall	
	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C
Location																				
Approximate location																	Y		Y	
Precise location		Y															Y		Y	
Personal Info																				
Name		Y		Y	Y	Y		Y		Y	Y	Y				Y	Y	Y	Y	Y
Email address		Y		Y	Y	Y		Y		Y	Y	Y		Y		Y	Y	Y	Y	Y
User IDs	Y	Y		Y	Y	Y						Y			Y		Y	Y	Y	
Address		Y		Y	Y	Y					Y	Y					Y	Y		Y
Phone number		Y		Y				Y		Y	Y	Y			Y		Y	Y	Y	Y
Race and ethnicity																				
Political or religious beliefs																				
Sexual orientation																				
Other info						Y											Y	Y	Y	Y
Financial Info									Y											
User payment info									Y											
Purchase history																				
Credit score																				
Other financial info																				
Health & Fitness																				
Health info																				
Fitness info																				
Messages																				
Emails																				
SMS or MMS																				
Other in-app messages																				
Photos and Videos									Y	Y					Y					
Photos									Y	Y					Y					
Videos																				
Audio files																				
Voice or sound recordings																				
Music files																				
Other audio files																				
Files and docs							Y		Y										Y	
Files and docs							Y		Y										Y	
Calendar																				
Calendar events		Y																		
Contacts							Y		Y											
Contacts							Y		Y											
App activity																				
App interactions		Y		Y	Y	Y		Y									Y		Y	Y
In-app search history																				
Installed apps																				
Other user-generated content					Y	Y				Y					Y					
Other actions																			Y	Y
Web browsing																				
Web browsing history																				
App info and performance																				
Crash logs				Y	Y	Y		Y		Y					Y		Y		Y	
Diagnostics				Y	Y	Y		Y		Y							Y		Y	
Other app performance data											Y						Y			
Device or other IDs																				
Device or other IDs		Y		Y	Y	Y		Y		Y	Y	Y			Y		Y	Y	Y	Y

S: Data shared
C: Data collected

Table 2. Data collection and sharing practices represented using Apple privacy labels

Data Category	Audi			Ford			Honda			Hyundai			Kia			Lexus			Nissan			Polestar			Renault			Vauxhall		
	T	L	N	T	L	N	T	L	N	T	L	N	T	L	N	T	L	N	T	L	N	T	L	N	T	L	N			
Location																														
Precise location													Y		Y			Y		Y			Y				Y			
Coarse location										Y								Y		Y			Y							
Contact info																														
Name				Y			Y			Y			Y					Y		Y			Y				Y			
Email address				Y			Y			Y			Y		Y			Y		Y			Y				Y			
Phone number				Y			Y			Y			Y		Y			Y		Y			Y				Y			
Physical address				Y			Y						Y		Y			Y		Y							Y			
Other user contact info													Y		Y															
Health and fitness																														
Health																														
Fitness																														
Financial info																														
Payment info																					Y									
Credit info																					Y									
Other financial info																														
Sensitive info																														
Contacts																														
User content																														
Emails or text messages																														
Photos or videos															Y												Y			
Audio data																														
Gameplay content																														
Customer support				Y																	Y					Y	Y			
Other user content																					Y									
Browsing history																														
Search history													Y														Y			
Identifiers																														
User ID				Y			Y						Y		Y				Y	Y		Y	Y		Y		Y			
Device ID				Y			Y						Y					Y	Y	Y					Y		Y			
Purchase history																					Y									
Usage data																														
Product interaction				Y			Y								Y	Y	Y		Y			Y				Y	Y			
Advertising data																						Y					Y			
Other usage data																					Y									
Diagnostics																														
Crash data				Y			Y								Y		Y		Y			Y				Y	Y			
Performance data				Y			Y								Y				Y			Y				Y	Y			
Other diagnostic data							Y																							
Surroundings																														
Environment scanning																														
Body																														
Hands																														
Head																														
Other data types							Y						Y																	

T: “Data used to track you”
L: “Data linked to you”
N: “Data not linked to you”