

Conversational AI Chatbot for Tele-communication.

School of Continuing Studies-3760,

University of Toronto.

Jaspreet Bhatia.



AGENDA

- ❖ Business Case.
- ❖ Proposed Architecture.
- ❖ Screenshots
- ❖ Advantages/Limitations.

Business Case

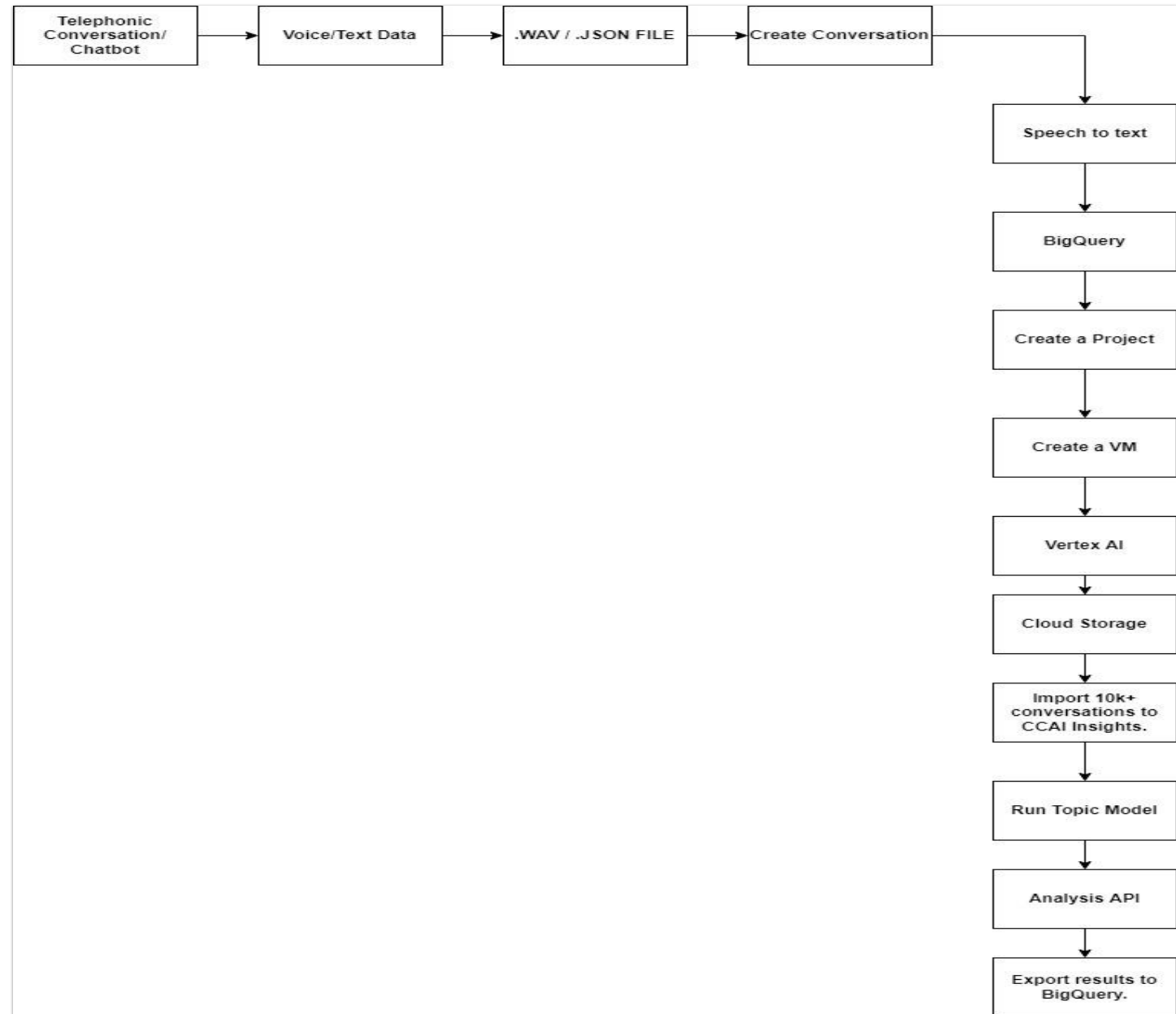
Current scenario:


- ◆ Huge amount of data flowing in daily.
- ◆ Storing the voice/text transcripts.
- ◆ Analysis of the conversations spoken in different languages (English , French etc.) for business cases (Internet, billing etc.)
- ◆ Context-based Topic modeling for identifying the major issues in the conversations.
- ◆ Storing the results of the topic model.

CLOUD TECHNOLOGY

- ◇ Google Cloud Platform.
- ◇ Google CCAI Insights – Topic Model.
- ◇ Speech to text API
- ◇ Custom Highlights.

Proposed Architecture




 Google Cloud Platform

psso-sbx-tel-dev ▼

DASHBOARD

ACTIVITY

RECOMMENDATIONS

 Project info

Project name

psso-sbx-tel-dev


Project number

411216105654

Project ID

psso-sbx-tel-dev

[ADD PEOPLE TO THIS PROJECT](#)

 Billing

Estimated charges

For the billing period Apr 1 – 12, 2022

Take a tour of billing

USD \$53.64

PROJECT

Google Cloud Platform

ps0-sbx-tel-dev

Search Products, resources, docs (/)

13

J

IAM & Admin

IAM

Identity & Organization

Policy Troubleshooter

Policy Analyzer

Organization Policies

Service Accounts

Workload Identity Federat...

Labels

Tags

Settings

Privacy & Security

IAM

+ ADD

- REMOVE

HELP ASSISTANT

PERMISSIONS

RECOMMENDATIONS HISTORY

View By:

PRINCIPALS

ROLES

☐ Include Google-provided role grants

Filter

Enter property name or value

<input type="checkbox"/>	Type	Principal	Name	Role	Security insights	Inheritance
<input type="checkbox"/>		411216105654-compute@developer.gserviceaccount.com	Compute Engine default service account	Owner	5339/5340 excess permissions	
<input type="checkbox"/>		411216105654@cloudservices.gserviceaccount.com	Google APIs Service Agent	Compute Admin	588/588 excess permissions	
				GCE Storage Bucket Admin	16/16 excess permissions	
				GCS Storage Bucket Owner	5/5 excess permissions	
				Notebooks Legacy Admin	642/642 excess permissions	
				Owner	5340/5340 excess permissions	
				Storage Admin	25/25 excess permissions	

Project IAM Roles

Filter

Filter

Name	↓ Requests	Errors (%)	Latency, median (ms)	Latency, 95% (ms)
Notebooks API	30,203	0	394	514
Compute Engine API	29,006	0	190	258
Contact Center AI Insights API	10,160	3	120	254
Cloud Monitoring API	4,844	86	336	507
Cloud Speech-to-Text API	335	0	99	593
Vertex AI API	146	0	88	249
Cloud Logging API	13	100	280	499
Data Labeling API	8	0	786	1,022

API

← Bucket details

REFRESHHELP ASSISTANTLEARN

sample_test_ccai_stry

Location

Storage class

Public access

Protection

us (multiple regions in United States)

Standard

Not public

None

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

LIFECYCLE

Public access

Not public

This bucket is not publicly accessible. If you know objects should never be exposed on the public internet, you should also prevent public access to this bucket. [Learn more](#)

PREVENT PUBLIC ACCESS

Access control

Uniform: No object-level ACLs enabled

36 days left to change this setting

All object access is controlled by bucket permissions and objects cannot have their own access control lists (ACLs). To allow per-object access, you can switch to fine-grained access within 90 days. [Learn more](#)

SWITCH TO FINE-GRAINED

Google Cloud Platform

psb-sbx-tel-dev

Search Products, resources, docs (/)

13

J

Cloud Storage

← Bucket details

REFRESHHELP ASSISTANTLEARN

Browser

sample_test_ccai_stry

Location

Storage class

Public access

Protection

us (multiple regions in United States)

Standard

Not public

None

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

LIFECYCLE

Buckets > sample_test_ccai_stry

UPLOAD FILES

UPLOAD FOLDER

CREATE FOLDER

MANAGE HOLDS

DOWNLOAD

DELETE

Filter by name prefix only

Filter objects and folders

Show deleted data

<input type="checkbox"/>	Name	Size	Type	Created	Storage class	Last modified	Public access	Version history	Encryption	Retention expiration
<input type="checkbox"/>	ABCD/	-	Folder	-	-	-	-	-	-	-

CLOUD STORAGE

VM INSTANCE

python-20220...

EDIT

RESET

CREATE MACHINE

DETAILS

OBSERVABILITY

OS INFO

SCREENSHOT

SHOW MORE

Basic information

Name	python-20220303-211024
Instance Id	8437445934552276891
Description	None
Type	Instance
Status	Running
Creation time	Mar 3, 2022, 10:40:36 AM UTC-05:00
Zone	us-west1-b
Instance template	None
In use by	None
Reservations	Automatically choose (default)

VM instances

CREATE INSTANCE

IMPORT VM

REFRESH

OPERATIONS

HELP ASSISTANT

LEARN

SHOW INFO PANEL

2 instances could be resized to save you up to an estimated \$71 per month. [Learn more](#)

DISMISS ALL

INSTANCES

INSTANCE SCHEDULE

VM instances are highly configurable virtual machines for running workloads on Google infrastructure. [Learn more](#)

Filter

Enter property name or value

?

⋮

	Status	Name ↑	Zone	Recommendations	In use by	Internal IP	External IP	Connect
<input type="checkbox"/>	✓	python-20220303-211024	us-west1-b	💡 Save \$32 / mo 💡 Save \$38 / mo		10.138.0.3 (nic0)	35.233.135.84	SSH ▾ ⋮

VERTEX AI WORKBENCH

✓ python-20220303-211024

BASIC INFO	HEALTH	MONITORING	LOGS
Region	us-west1 (Oregon)		
Zone	us-west1-b		
Environment ?	Python 3 (with Intel® MKL)		
Environment version	M90		
Machine type ?	n1-standard-4 (4 vCPUs, 15 GB RAM)		
GPU ?	None		
Boot disk	100 GB disk		
Data disk	100 GB disk		
Created	Mar 3, 2022, 10:40:33 AM		
Last modified	Apr 7, 2022, 9:50:46 AM		
Backup	Not specified		
Subnetwork	default		
Service account	411216105654-compute@developer.gserviceaccount.com		
Permission mode	Service account		
Sudo access	Enabled		
File downloads	Enabled		
nbconvert	Enabled		
Realtime Collaboration	Disabled		
Notebook health	System health report enabled Custom metrics reporting not enabled Cloud Monitoring agent not installed		

Workbench

NEW NOTEBOOK

REFRESH

START

STOP

RESET

DELETE

SHOW INFO PANEL

MANAGED NOTEBOOKS

USER-MANAGED NOTEBOOKS

EXECUTIONS

SCHEDULES

Notebooks have JupyterLab 3 pre-installed and are configured with GPU-enabled machine learning frameworks. [Learn more](#)

Filter

Enter property name or value

<input type="checkbox"/>	<input checked="" type="checkbox"/>	Notebook name ↑		Zone	Auto-upgrade	Environment	Machine type	GPUs	Permission	Last modified
<input type="checkbox"/>	<input checked="" type="checkbox"/>	python-20220303-211024	OPEN JUPYTERLAB	us-west1-b	—	NumPy/SciPy/scikit-learn	4 vCPUs, 15 GB RAM ▼	None ▼	Service account	Apr 7, 2022, 9:50:46 AM

NOTEBOOK DETAILS

```
[1]: import pandas as pd
    from tqdm import tqdm
    import json
    import re
    import random
    from pathlib import Path
    from pandas.io.json import json_normalize
```

```
[2]: # _df=pd.read_json("Data/abcd_v1_1.json", "r", encoding = "utf8", lines=True)
    # _df=pd.read_json(open("Data/abcd_v1.1.json", "r"), encoding="utf8", lines=True)
    _df=pd.read_json("Data/abcd_v1.1.json", "r", encoding = "utf8", lines=True)
```

/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:3: FutureWarning: Starting with pandas version 2.0 all arguments of read_json except for the argument 'path_or_buf' will be keyword-only
This is separate from the ipykernel package so we can avoid doing imports until

```
[3]: _df
```

```
[3]:
```

	train	dev	test
0	{'convo_id': 3592, 'scenario': {'personal': {...	{'convo_id': 3532, 'scenario': {'personal': {...	{'convo_id': 4989, 'scenario': {'personal': {...

```
[5]: _df['train'].head(1)
```

NOTEBOOK

Google Cloud Platform

pso-sbx-tel-dev

Search speed

13

J

Speech-to-Text

Overview

Transcriptions

Model Adaptations

Transcriptions

+ NEW TRANSCRIPTION

Your transcriptions are listed below. You can switch workspaces to find other transcriptions that are stored in a different workspace.

Workspace
sample_test_ccai_stry/m... ?

Filter Enter property name or value ?

<input type="checkbox"/>	Name	Status	Audio file	Encoding	Sample rate	Channels	Language
<input type="checkbox"/>	cm50LW1xanoteGFklCgyMDIyLTAzLTE3IDEyXzU...	Success	gs://sample_test_ccai_stry/mp3_2/audio-files/c...	MP3	48,000	2	en-US

SPEECH-TO-TEXT

[7]: df

[7]:	convo_id	scenario	original	delexed
0	3592	{'personal': {'customer_name': 'crystal minh',...	[[agent, Hi!], [agent, How can I help you?], [...	[{'speaker': 'agent', 'text': 'hi!', 'turn_cou...
1	9489	{'personal': {'customer_name': 'alessandro pho...	[[agent, good afternoon, how can I help you?],...	[{'speaker': 'agent', 'text': 'good afternoon,...
2	3695	{'personal': {'customer_name': 'joyce wu', 'me...	[[customer, HEY HO!], [agent, good afternoon, ...	[{'speaker': 'customer', 'text': 'hey ho!', 't...
3	5798	{'personal': {'account_id': 'xjiw8ufp3b', 'cus...	[[agent, Welcome to AcmeBrands! How can I help...	[{'speaker': 'agent', 'text': 'welcome to acme...
4	3647	{'personal': {'customer_name': 'norman bouchar...	[[agent, Hello, how can i help you], [customer...	[{'speaker': 'agent', 'text': 'hello, how can ...
...
8029	2229	{'personal': {'customer_name': 'chloe zhang', ...	[[agent, Welcome to AcmeBrands. How may I help...	[{'speaker': 'agent', 'text': 'welcome to acme...
8030	6109	{'personal': {'customer_name': 'crystal minh',...	[[agent, Hello, how can I help you today?], [c...	[{'speaker': 'agent', 'text': 'hello, how can ...
8031	9892	{'personal': {'customer_name': 'crystal minh',...	[[agent, Hello, how may I help you?], [custome...	[{'speaker': 'agent', 'text': 'hello, how may ...
8032	4036	{'personal': {'customer_name': 'chloe zhang', ...	[[customer, Hi], [agent, Hello, what can I do ...	[{'speaker': 'customer', 'text': 'hi', 'turn_c...
8033	5882	{'personal': {'customer_name': 'david williams...	[[agent, how may i help you], [customer, Hello...	[{'speaker': 'agent', 'text': 'how may i help ...

8034 rows × 4 columns

Action Based Conversation Dialogue (ABCD) Dataset.


[56]: final


[56]:	index	role	text	Count_id	Time	clean_text	transcript
	0	14	AGENT	ok, unfortunately because it has been more tha...	0 2270422	ok, unfortunately because it has been more tha...	ok, unfortunately because it has been more tha...
	1	21	AGENT	OK, I have let my manager know, they will give...	0 2976195	OK, I have let my manager know, they will give...	OK, I have let my manager know, they will give...
	2	35	AGENT	It is currently in progress and the payment me...	1 6810274	It is currently in progress and the payment me...	It is currently in progress and the payment me...
	3	57	END_USER	Hello, I would like to change my shipping deat...	3 6532489	Hello, I would like to change my shipping deat...	Hello, I would like to change my shipping deat...
	4	74	END_USER	Hello. I have a really cool party coming up. A...	4 3519225	Hello. I have a really cool party coming up. A...	Hello. I have a really cool party coming up. A...


	17538	130881	END_USER	How often should I wash them. I do like doing...	8031 4368986	How often should I wash them. I do like doing...	How often should I wash them. I do like doing...
	17539	130884	AGENT	However if you are heavy duty user or play spo...	8031 4062570	However if you are heavy duty user or play spo...	However if you are heavy duty user or play spo...
	17540	130915	END_USER	Hello, Every time I go to search on the websit...	8033 1443527	Hello, Every time I go to search on the websit...	Hello, Every time I go to search on the websit...
	17541	130916	AGENT	i understand you are having issues with our si...	8033 6205350	i understand you are having issues with our si...	i understand you are having issues with our si...
	17542	130921	AGENT	first, try logging in and out of your account,...	8033 2476343	first, try logging in and out of your account,...	first, try logging in and out of your account,...


17543 rows × 7 columns

DATA CLEAN-UP

Monitoring

Settings

Marketplace

Release Notes

Location

us (multiple regions in United States)

Storage class

Standard

Public access

Not public

Protection

None

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

LIFECYCLE

Buckets > sample_test_ccai_stry > ABCD > train_clean

UPLOAD FILES

UPLOAD FOLDER

CREATE FOLDER

MANAGE HOLDS

DOWNLOAD

DELETE

Filter by name prefix only

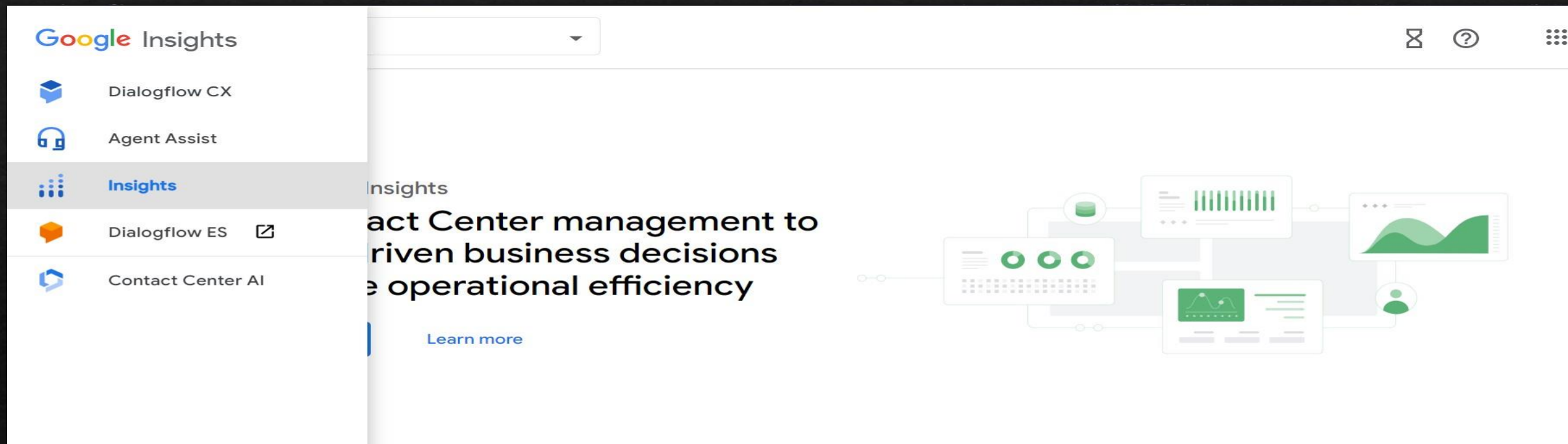
Filter

Filter objects and folders

Show deleted data

<input type="checkbox"/>	Name	Size	Type	Created	Storage class	Last modified	Public access	Version history	Encryption	
<input type="checkbox"/>	.ipynb_checkpoints/	—	Folder	—	—	—	—	—	—	
<input type="checkbox"/>	0.json	545 B	application/json	Apr 7, 202...	Standard	Apr 7, 202...	Not public	—	Google-managed key	<div><div>↓</div><div></div></div>
<input type="checkbox"/>	1.json	288 B	application/json	Apr 7, 202...	Standard	Apr 7, 202...	Not public	—	Google-managed key	<div><div>↓</div><div></div></div>
<input type="checkbox"/>	10.json	1.1 KB	application/json	Apr 7, 202...	Standard	Apr 7, 202...	Not public	—	Google-managed key	<div><div>↓</div><div></div></div>
<input type="checkbox"/>	100.json	935 B	application/json	Apr 7, 202...	Standard	Apr 7, 202...	Not public	—	Google-managed key	<div><div>↓</div><div></div></div>
<input type="checkbox"/>	1000.json	561 B	application/json	Apr 7, 202...	Standard	Apr 7, 202...	Not public	—	Google-managed key	<div><div>↓</div><div></div></div>
<input type="checkbox"/>	1001.json	726 B	application/json	Apr 7, 202...	Standard	Apr 7, 202...	Not public	—	Google-managed key	<div><div>↓</div><div></div></div>
<input type="checkbox"/>	1002.json	581 B	application/json	Apr 7, 202...	Standard	Apr 7, 202...	Not public	—	Google-managed key	<div><div>↓</div><div></div></div>

OUTPUT FILES TO CLOUD STORAGE



GOOGLE CCAI INSIGHTS



Conversation history

Export

☰ Conversation filters

Total conversations
5309



Average duration
🕒 0m 0s

Average turns ?
1




Top topics
No topic data for this set of conversations.

Agent ID	Duration	Turns	Primary Topic	Channel	Import date (GMT-4)	Analysis date (GMT-4)	Analysis status	
—	0m 0s	3	—	Chat	Today, 2:19 PM	—	— Not analyzed	⋮
—	0m 2s	1	—	Chat	Today, 2:19 PM	—	— Not analyzed	⋮
—	-1m 56s	3	—	Chat	Today, 2:19 PM	—	— Not analyzed	⋮
—	0m 1s	1	—	Chat	Today, 2:19 PM	—	— Not analyzed	⋮
—	-1m 57s	2	—	Chat	Today, 2:19 PM	—	— Not analyzed	⋮
—	—	1	—	Chat	Today, 2:19 PM	—	— Not analyzed	⋮

CONVERSATION HUB


 Insights


pso-sbx-tel-dev




Understand major call drivers with Insights Topic models.

We analyze and bring the information that you need. Learn more about our topic modeling and how you can best leverage us. [View documentation](#)

**Import conversations**
Bring 10k conversations for best accuracy.

**Train the topic model**
Training is as simple as one-click

**Review and deploy your model**
Edit and deploy the model to use across your data.

Id	Name	Status	Topics
----	------	--------	--------

Build a topic model to know what's driving your calls

An Insights topic model uses Google's natural language processing to generate primary topics for each conversation in your dataset. You can then deploy the model to analyze future conversations as they're imported. [Learn More](#)

Here's how it works

✓

Select at least 10K conversations for training.

For better accuracy we need at least 10,000 conversations to train your model.

Conversation channel
Chat

Total conversations selected: **5,373**

☒ Use all conversations

☐ Filter conversations

✓



Name your training model (Optional)

Model name

TOPIC MODEL

 This conversation hasn't been analyzed. Please start analysis to see more details.

Start analysis

Agent Name

Duration
-1m 59s

Turn Count
1

Language Code
-

Topics

Date & Time
Apr 12, 2022 - 3:33 PM GMT-4

Labels
None

Customer sentiment



Conversation



I'll need bit of information from you so I can verify your order and validate the purchase as well. Thank you for providing your Name for account look up.



Thank you, I was able to verify your identity. Next could you provide me with your Username and Email address so I can check on the order?



ANALYSIS API

LIMITATIONS

Conversations may contain personal details like Credit card, Pin number etc.

Speech to text should correctly highlight more important phrases after conversion.

Data Privacy to be maintained.

Huge data storage, storage time, data analysis required.

Updating the input data as more data flows in from different business use-cases.

Topic Model to be updated over time.

REFERENCES



UOFT CLASS
LECTURES/NOTES/ASSIGN
MENTS.



GOOGLE CLOUD
DOCUMENTATION.



GOOGLE CCAI
DOCUMENTATION.



NLP TEXTBOOKS.



ABCD OPEN-SOURCE
DATASET.



◆ Thank You