

# TRANSCRIBED AUDIO AUTOMATIC CHECKLIST BASED ON GEN AI

---

Ana Paula Oliveira Bertholdo  
June 10-13, 2024



# About me

- PhD in Computer Science
  - University of Sao Paulo (USP)
- Information Technologist
  - Campinas State University (UNICAMP)
- Over 17 years of work experience
- Data Scientist
  - Brasilprev



[ana.bertholdo@brasilprev.com.br](mailto:ana.bertholdo@brasilprev.com.br)

[www.brasilprev.com.br](http://www.brasilprev.com.br)

# Brasilprev

The leading private pension company in Brazil



# Scenario

## Customer Relationship Center (CRC)

- The CRC receives an average of 1000 audio files per day.
- Each audio has an average duration of 10 minutes.
- Evaluators listen to the entire audio to identify items related to the customer service protocol.



# Problem

- Evaluators conduct random audio checks
  - Most audios are not evaluated
- Human evaluators listen to complete audios
  - Audios are reviewed by evaluators who listen to the entire call in a non-automated manner and without integration into databases



# Goal

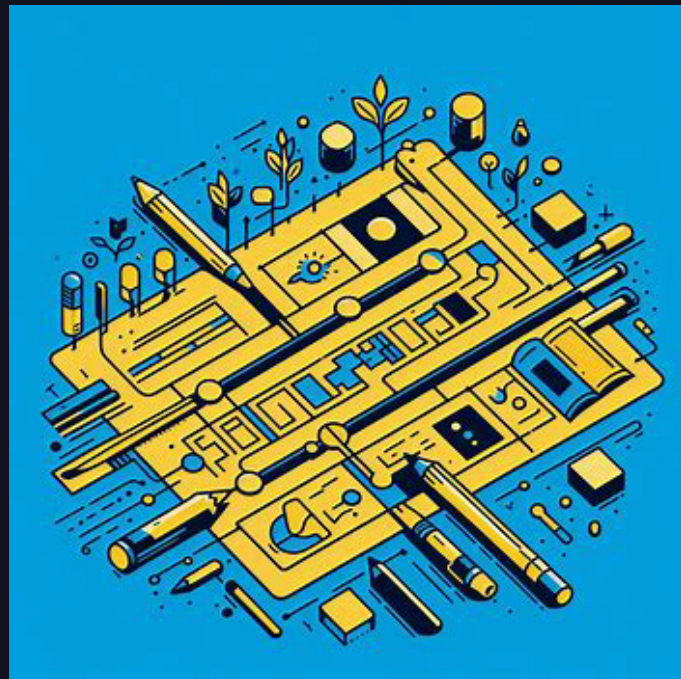
## Automated checklist for transcribed audio based on Generative AI

- Evaluation of all audio recordings from the CRC, assessing adherence to the service quality protocol for each call.
- There is no need to listen to the entire audio.



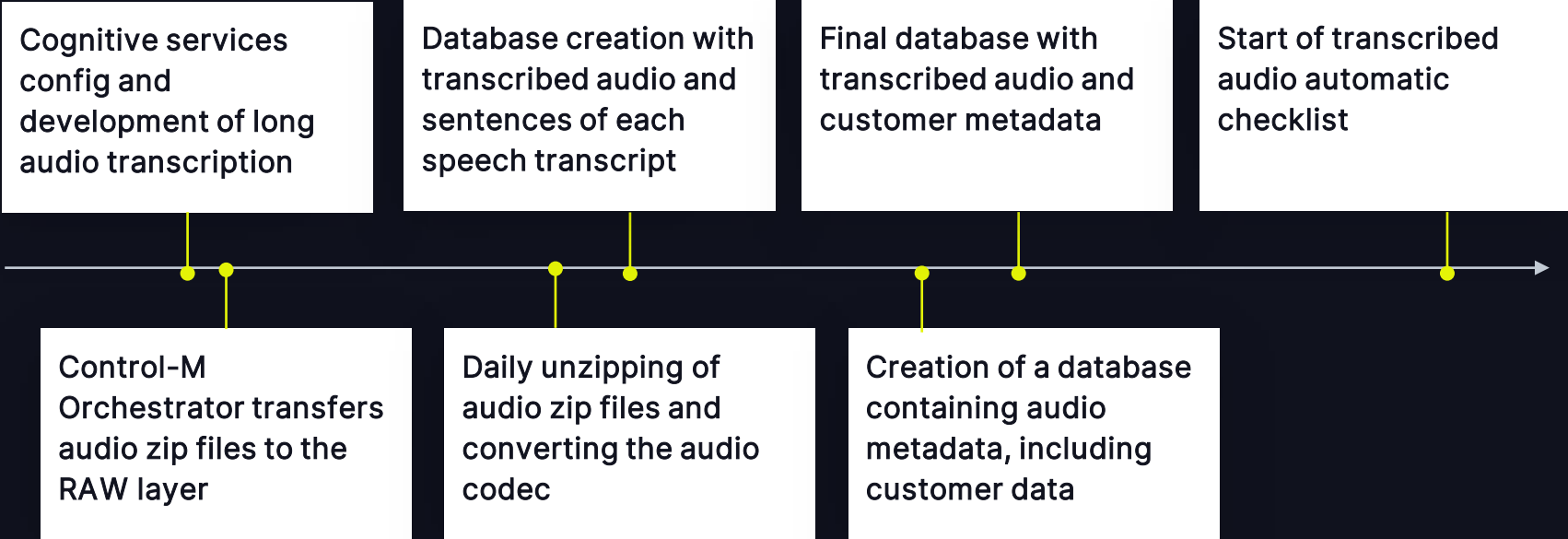
# Method

- Long audio transcription pipeline in Azure Databricks
- Automated checklist for transcribed audio
  - Performance/cost comparison of LLMs with and without Vector Search (BGE)



# Method

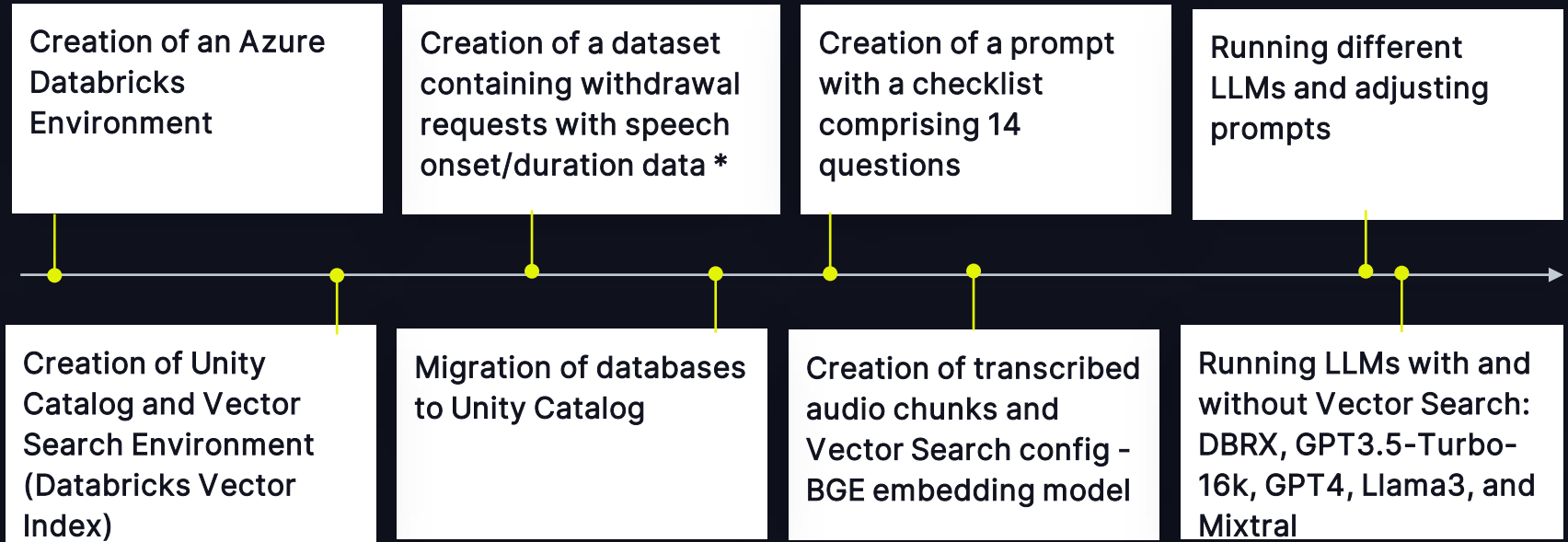
## Long audio transcription pipeline in Azure Databricks





# Method

## Transcribed audio automatic checklist



\* 300 audios

# Method





## Code Overview

Shared Structure		With Vector Search		Without Vector Search	
Name	Type	Name	Type	Name	Type
<a href="#">Chunk_Size_Exploration</a>	Folder	<a href="#">2-ChunkingPreVectorSearch - 300Por100</a>	Notebook	<a href="#">5-V1.3_LLMs_Without_VectorSearch_Audio</a>	Notebook
<a href="#">Final_Experiment</a>	Folder	<a href="#">3-PK_For_Chunks</a>	Notebook		
Name	Type	<a href="#">4-ConfigVectorSearch</a>	Notebook		
<a href="#">With_Vector_Search</a>	Folder	<a href="#">5-V1.3_LLMs_With_VectorSearch_1ChunkForQuestion</a>	Notebook		
<a href="#">Without_Vector_Search</a>	Folder				
<a href="#">1-Database_Creation</a>	Notebook				
<a href="#">6-LLMs_Cost_Analysis</a>	Notebook				

# Method

## Inputs

These parameters control the generation process of the model, affecting the diversity, randomness, and focus of the output.

- Temperature  0.01
- Top P  0.95
- Top K  1
- Max tokens  2000

### Time markers in sentences (without Vector Search)

Frase 25 - Speaker 1 - OFFSET PT3M21.44S - DURATION PT10.6S - Tá aqui pra gente poder processar o resgate é o motivo que você está solicitando, é compra de imóvel, de veículo, outros pagamentos. ;

Frase 26 - Speaker 2 - OFFSET PT3M32.04S - DURATION PT4.24S - É de imóvel. Eu expliquei Pra Ela que assim quem está mais em dinheiro, depois eu retorno. ;

Frase 27 - Speaker 1 - OFFSET PT3M37S - DURATION PT4.12S - Né? Ela já ela mencionou que já estava prestando a consultoria pra você, né? ;

Frase 28 - Speaker 2 - OFFSET PT3M41.4S - DURATION PT0.44S - Isso. ;

Frase 29 - Speaker 1 - OFFSET PT3M43.69S - DURATION PT1.36S - O imóvel é pro moradia ou. ;

Frase 30 - Speaker 2 - OFFSET PT3M45.05S - DURATION PT2.2S - Investimento, moradia. ;

Frase 31 - Speaker 1 - OFFSET PT3M48.53S - DURATION PT0.79S - Ai, que legal. ;

Frase 32 - Speaker 1 - OFFSET PT3M49.36S - DURATION PT0.92S - Ai, que legal. ;

### Time markers in chunks (with Vector Search)

> ; Frase 45 - Speaker 2 - OFFSET PT5M24.79S - DURATION PT0.6S - Ma...

> ; Frase 77 - Speaker 1 - OFFSET PT9M59.92S - DURATION PT3.44S - P...

> O pagamento vai ser no Banco do Brasil mesmo. ; Frase 34 - Speaker ...

> ; Frase 69 - Speaker 1 - OFFSET PT9M29.43S - DURATION PT9.56S - A...

> ; Frase 41 - Speaker 2 - OFFSET PT5M2.79S - DURATION PT0.72S - Ok...

> ; Frase 61 - Speaker 1 - OFFSET PT8M25.88S - DURATION PT29.56S - ...

> ; Frase 63 - Speaker 1 - OFFSET PT8M58.21S - DURATION PT3.36S - V...

∨ ;  
Frase 26 - Speaker 2 - OFFSET PT3M32.04S - DURATION PT4.24S - É de imóvel. Eu expliquei Pra Ela que assim quem está mais em dinheiro, depois eu retorno. ;

Frase 27 - Speaker 1 - OFFSET PT3M37S - DURATION PT4.12S - Né? Ela já ela mencionou que já estava prestando a consultoria pra você, né? ;

Frase 28 - Speaker 2 - OFFSET PT3M41.4S - DURATION PT0.44S - Isso.

> ; Frase 73 - Speaker 2 - OFFSET PT9M54.53S - DURATION PT0.48S - T...

### No time markers in chunks (with Vector Search)

> ; - Mateus. ; - Vai. ; - Oh, meu Deus, hã, já. ; - Choveu. ; - Alô?...

> ; - Perfeito, então. Então, muito obrigada. Desejo um ótimo ...

> O pagamento vai ser no Banco do Brasil mesmo. ; - Isso na ...

> ; - Amanhã já conta como o primeiro dia útil de análise tá? E...

> ; - Okay? ; - A sua renda mensal, só uma média. ; - Tá, 4. ; - 1...

> ; - O valor ele está aplicado no fundo fixo. 2 tá que é o fund...








> ; - Você faz a sua declaração de imposto aqui no Brasil mes...




∨ ;  
- É de imóvel. Eu expliquei Pra Ela que assim quem está mais em dinheiro, depois eu retorno. ;  
- Né? Ela já ela mencionou que já estava prestando a consultoria pra você, né? ;  
- Isso. ;  
- O imóvel é pro moradia ou. ;  
- Investimento, moradia.

> ; - Tá bom? ; - Tá bom, ficou com alguma? ; - Dúvida, não. ; - ...

# Sample Results from LLMs

Question: Has the purpose for the money withdrawal been requested? If yes, what was the reason for the withdrawal?

Input	DBRX	GPT3.5-Turbo-16K	GPT4	Llama3	Mixtral
Audio with speech onset and duration (Without VS)	Yes, the reason mentioned was "property purchase". 	Yes, the reason for the money withdrawal was the purchase of a property. (Phrase 25 - OFFSET PT3M21.44S - DURATION PT10.6S)	Yes, the reason for the money withdrawal was the purchase of a property. (Phrase 26 - Speaker 2 - OFFSET PT3M32.04S - DURATION PT4.24S)	Yes, the reason for the money withdrawal was the purchase of a property. (Phrase 26 - OFFSET PT3M32.04S - DURATION PT4.24S)	Answer: Yes, the reason for withdrawing money is the purchase of a property, as Speaker 1 mentioned in Phrase 25: It's here so we can process the withdrawal. The reason you are requesting is the purchase of property, vehicle, or other payments. 
Chunk with speech onset and duration (With VS)	The reason for the withdrawal was not requested. 	A: There is no request for the reason for withdrawal. 	No, the reason for the withdrawal was not asked. (Context)	The reason for the withdrawal was not requested. 	The reason for the withdrawal was not requested in this section. 
Chunk without speech onset and duration (With VS)	 The reason for the withdrawal is the purchase of property.	A: Yes, the reason for the withdrawal was the purchase of a property.	Yes, the reason for the withdrawal is the purchase of a property.	Yes, the reason for the withdrawal is the purchase of a property.	Yes, the reason for the withdrawal is the purchase of a property.

-  Correct answer
-  Missing information
-  Incorrect answer

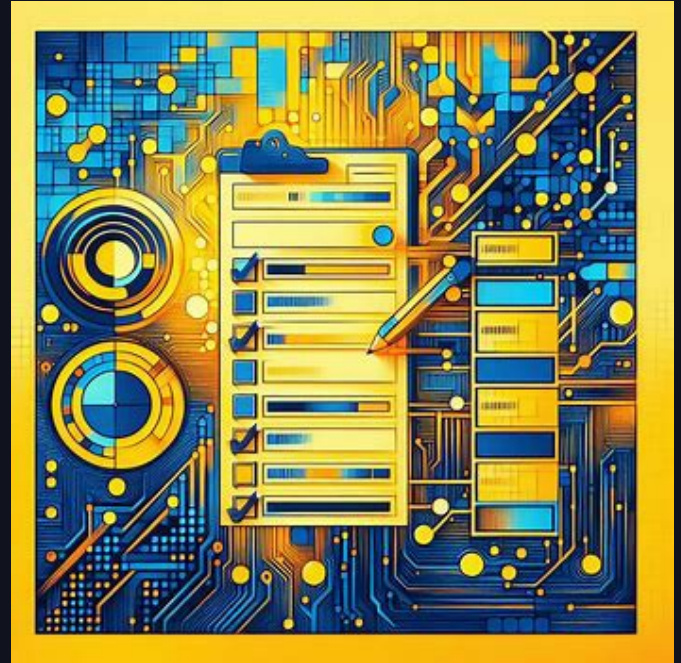
# Results

## Scheme for comparing LLMs with and without Vector Search

LLM	Audio or Chunk	With or Without Time	Embedding Tokens Cost	Foundation Input Tokens	Foundation Output Tokens	Duration	Language	Accuracy	Context	Foundation Tokens Cost	Total Cost	Cost Ranking	Total Performance	Performance Ranking	% Cost Diff	Final Ranking
DBRX	Audio	With	0	606.34	520	4.97	1	13	4	0.00162475	0.001625	1	18	6	0%	6
DBRX	Chunk	With	0.00138	368.28	299	2.97	1	11	6	0.00094896	0.002329	6	18	6	43%	6
DBRX	Chunk	Without	0.000845	383.16	321	3.72	1	9	9	0.00100962	0.001855	2	19	5	14%	2
Mixtral	Audio	With	0	1570.7	1500	13.9	1	11	3	0.0022854	0.002285	5	15	9	0%	10
Mixtral	Chunk	With	0.00138	489.7	427	4.15	1	6	6	0.00067186	0.002052	3	13	10	-10%	9
Mixtral	Chunk	Without	0.000845	937.72	881	7.88	1	9	9	0.00134989	0.002195	4	19	5	-4%	3
Llama3	Audio	With	0	2059.2	2000	31.68	1	8	8	0.00805922	0.008059	12	17	7	0%	8
Llama3	Chunk	With	0.00138	856.32	817	13.38	1	12	11	0.00330733	0.004687	11	24	3	-42%	7
Llama3	Chunk	Without	0.000845	670.08	643	10.47	1	13	13	0.00259909	0.003444	8	27	1	-57%	1
GPT3.5	Audio	With	0	4035	964	10	1	12	10	0.0034635	0.003464	9	23	4	0%	5
GPT3.5	Chunk	With	0.00138	3295	806	11	1	10	5	0.0028565	0.004237	10	16	8	22%	11
GPT3.5	Chunk	Without	0.000845	1990	749	10	1	11	11	0.0021185	0.002964	7	23	4	-14%	4
GPT4	Audio	With	0	4035	1029	56	1	13	13	0.36558	0.36558	15	27	1	0%	14
GPT4	Chunk	With	0.00138	3295	957	60	1	10	12	0.31254	0.31392	14	23	4	-14%	13
GPT4	Chunk	Without	0.000845	1990	725	33	1	12	12	0.2064	0.207245	13	25	2	-43%	12

# Results

- Reduction in the number of input tokens with VS
- VS requires the choice of “good” embedding
- The results showed improvement after removing the speech onset and duration data.
- Biggest cost reductions for GPTs/Llama3
- The average execution time with Vector Search has been reduced to 18 seconds per audio, with a median of 10 seconds.



# Conclusions

- Next steps:
  - Assessment of other embedding models
  - Human curation for the Top 3 scenarios (user feedback to build an evaluation dataset)
  - Human-verified LLM eval
- Application in other use cases: legal documents
- Innovation Process as a part of the Data Science Cycle

# DATA+AI SUMMIT

THANK YOU!

Ana Paula Oliveira Bertholdo

[ana.bertholdo@brasilprev.com.br](mailto:ana.bertholdo@brasilprev.com.br)

[www.brasilprev.com.br](http://www.brasilprev.com.br)



Special Thanks to:



Angela Assis  
Bruno Venceslau  
Arley Mendonça

Brasilprev Team:

Ana Alice Pilon  
Gabriel Caballeria de Oliveira  
Emerson Fernandes da Silva

Databricks Team:

Ana Caroline Sanchez Silva  
Alan Silva  
Claudio Seidi Takamiya  
Eduardo Bevilaqua