

AURA

AI-POWERED UNIFIED RESPONSE AND ANALYSIS

Advisor: Charnon Pattiyanon

Kasidis Manasurangkul
Niracha Janavatara
Supakorn Etitum
Pon Yimcharoen
Shine Min Kha



WHO ARE THE NHSO?



สปสช.

สายด่วน สปสช.

1 3 3 0

www.nhso.go.th

30 Baht Healthcare Universal Coverage Scheme

Thai government agency dedicated to public health and wellness

1330 Call Center

Handles public inquiries, processes claims, provides guidance on health insurance procedures



เรื่อง การนำเข้าข้อมูล 16 ไฟฟ์ เลขที่ KB 1220

การนำเข้าข้อมูล 16 ไฟฟ์

ขั้นตอนการนำเข้า 16 ไฟฟ์

- 1) เลือกเมนู บริการ
- 2) เลือกเมนู นำเข้าข้อมูล 16 ไฟฟ์
- 3) เลือกชนิดของไฟล์ที่ต้องการนำเข้า รูปแบบ DBF file หรือ Text file
- 4) เลือกสิทธิประโยชน์ที่ต้องนำเข้า ประเภท OP Refer/ OP/ IP
- 5) เลือกประเภทข้อมูลที่ต้องการนำเข้า 譬如 OP Refer/ OP/ IP
- 6) กดปุ่ม แนบไฟล์

หมายเหตุ กรณีนำเข้าประเภท OP Refer ต้องเลือกสิทธิประโยชน์เป็น UCS จึงจะสามารถนำเข้าได้

เรื่อง Lab Catalogue KB1262
เรื่อง NON UCEP (มาตรฐาน) KB66
เรื่อง OP Anywhere KB1355
เรื่อง Telemedicine คณฑ์ไทยในต่างประเทศ KB1356
เรื่อง UCEP COVID KB1068
เรื่อง UCEP PLUS COVID สิทธิ์ครม... KB1093
เรื่อง UCEP PLUS COVID สิทธิ์บัตร KB1094
เรื่อง UCEP PLUS COVID สิทธิ์ห้อง KB1109
เรื่อง UCEP สิทธิ์ยาสีฟัน KB1273
เรื่อง กลุ่มคนไทยไร้สิทธิ KB1039
เรื่อง กลุ่มเด็กตั้งแต่ 0-5 ปี PP KB1039
เรื่อง กลุ่มผู้สูงอายุ 60 ปี KB1039

NHSO CRM SYSTEM Call_ID: 00058 00:29 จบการโทร

ข้อมูลการติดต่อ

ชื่อ-นามสกุล
หมายเลขบัตรประชาชน
อีเมล
หมายเลขโทรศัพท์

ประเภทการโทร

ประเภทการโทร

หัวข้อฐานความรู้

แนะนำหัวข้อฐานความรู้

KB1303 X KB1387 X KB1297 X

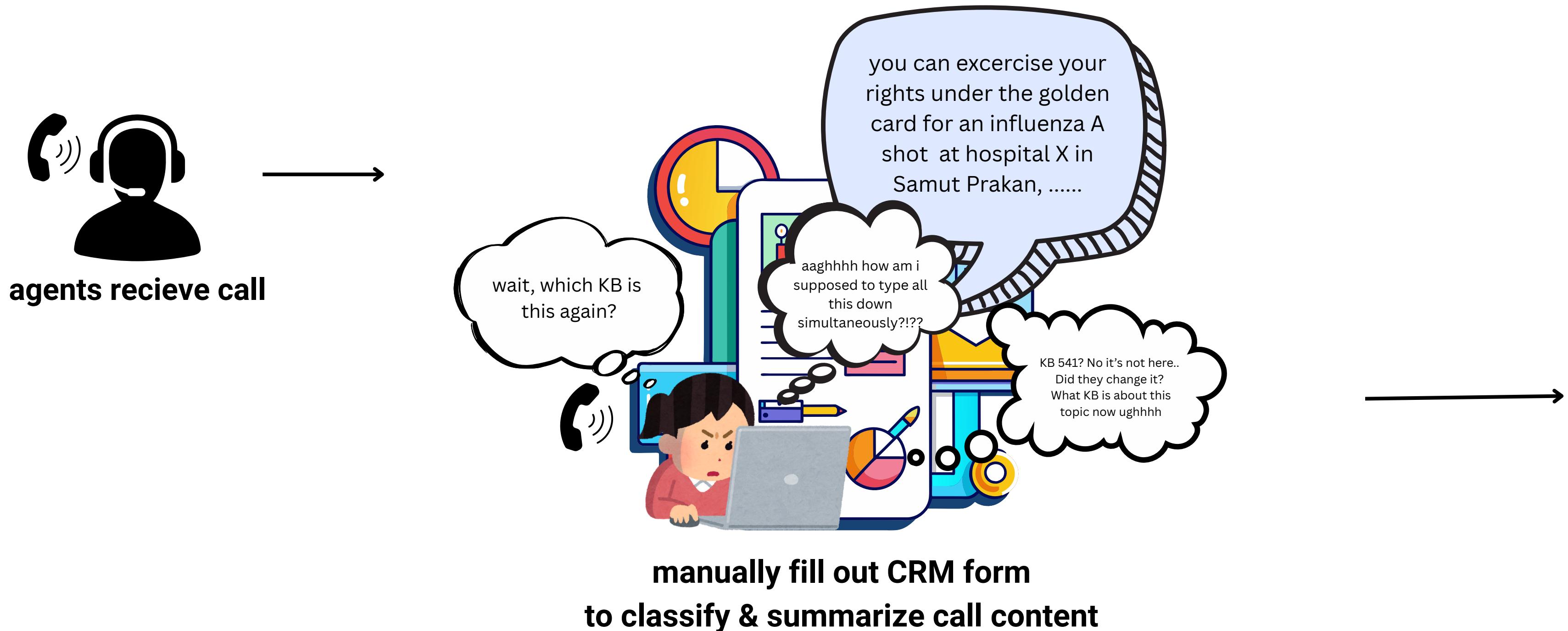
หัวข้อฐานความรู้

แนะนำหัวข้อฐานความรู้

รายละเอียดการโทร

Call Details

CURRENT METHODS:



PROBLEM:

Agents struggle to categorize fields like
Knowledge Base classification and
Call Content Summaries
accurately and efficiently.

time-consuming

prone to human error

delays crucial decision-making

missing data

OUR SOLUTION:



Automated Call Analysis:

Leveraging AI to transcribe, summarize, and classify calls in real-time when a call ends.

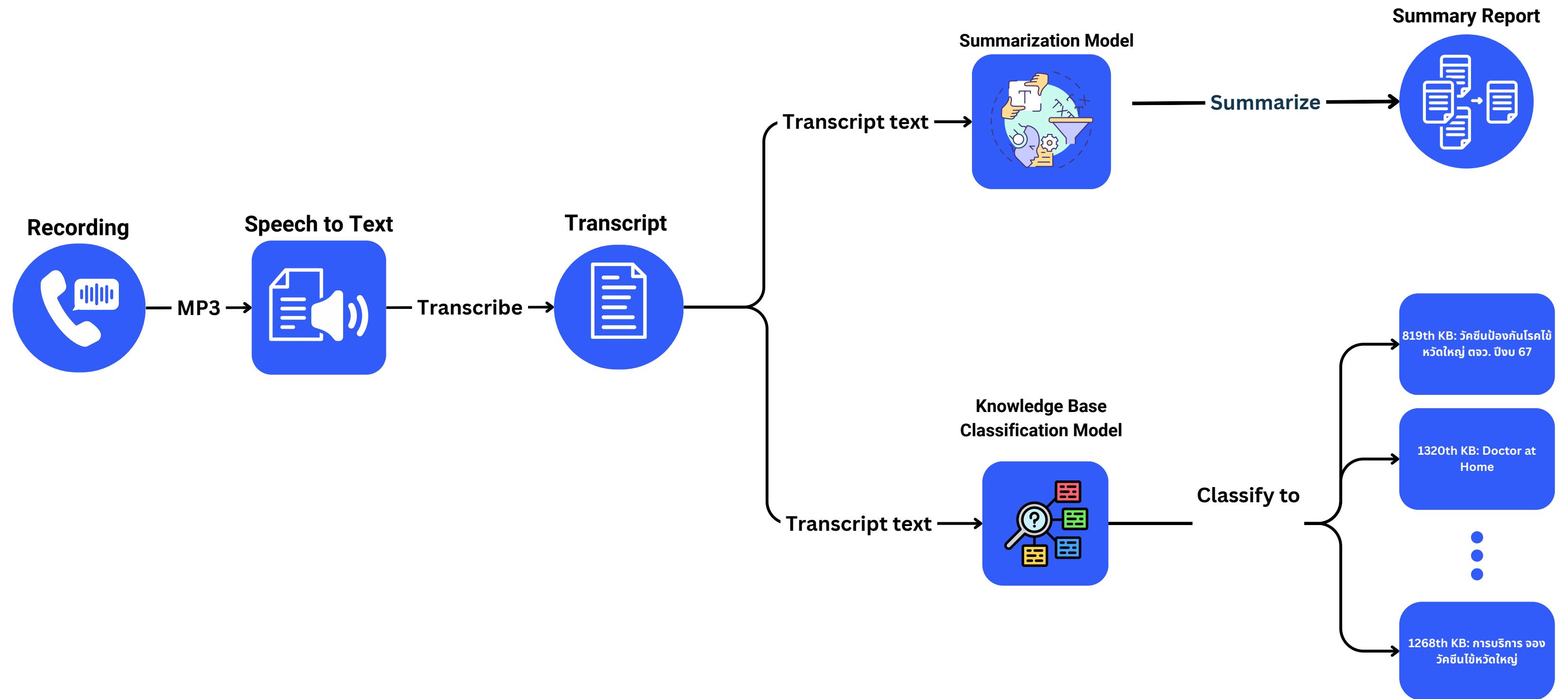
CRM Integration

Models populate structured fields (eg. dropdown classifications, disease fields) in the NHSO CRM automatically.

Scalable and Precise:

Handles large call volumes with high accuracy and integrate seamlessly into NHSO workflows.

SYSTEM WORKFLOW

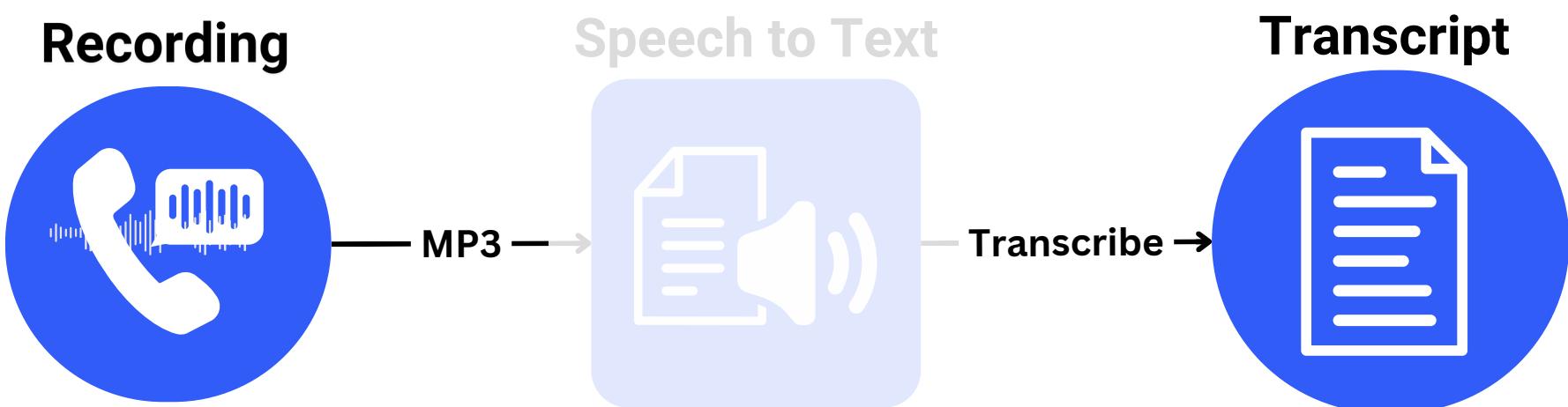




GOALS AND PROGRESS



DATA SYNTHESIS



No call recording received.

Processed
Knowledge Base Documents

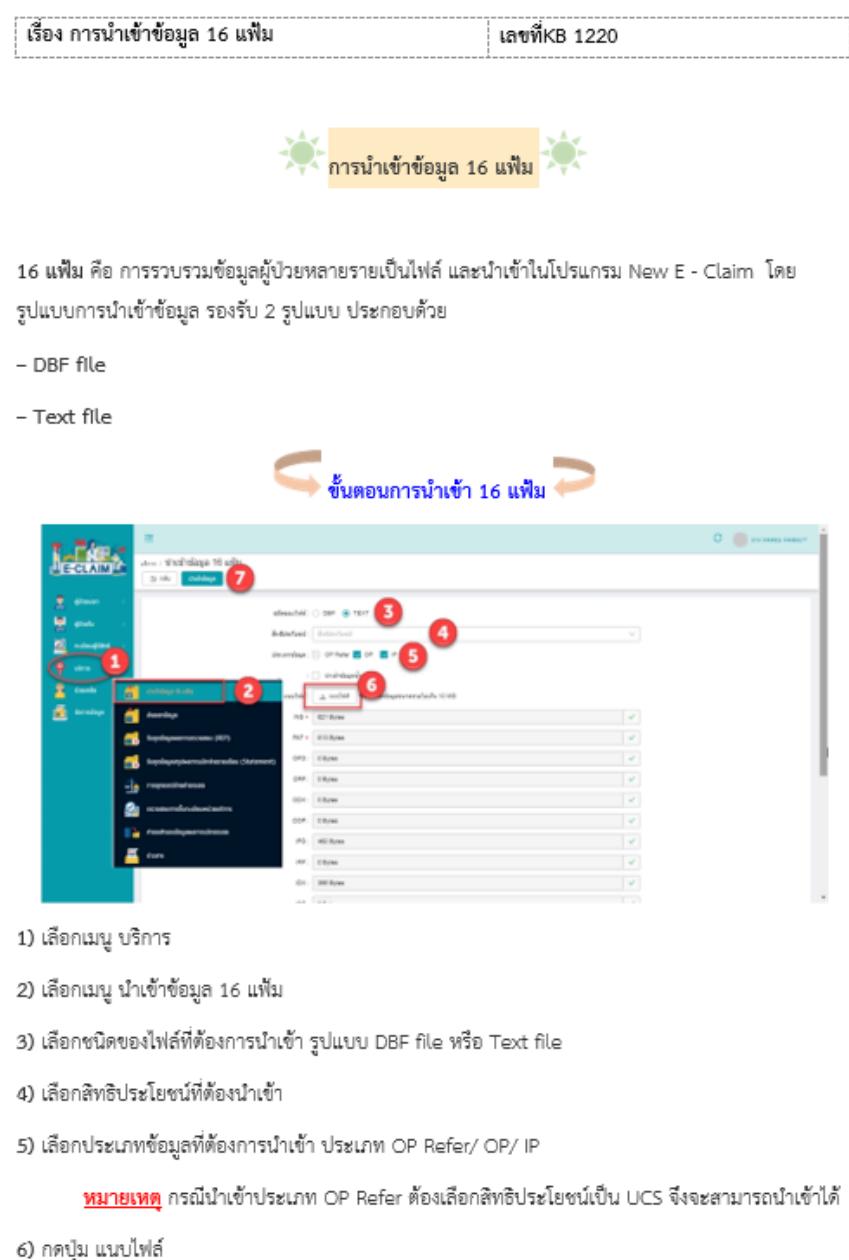
Synthesised Call Transcripts



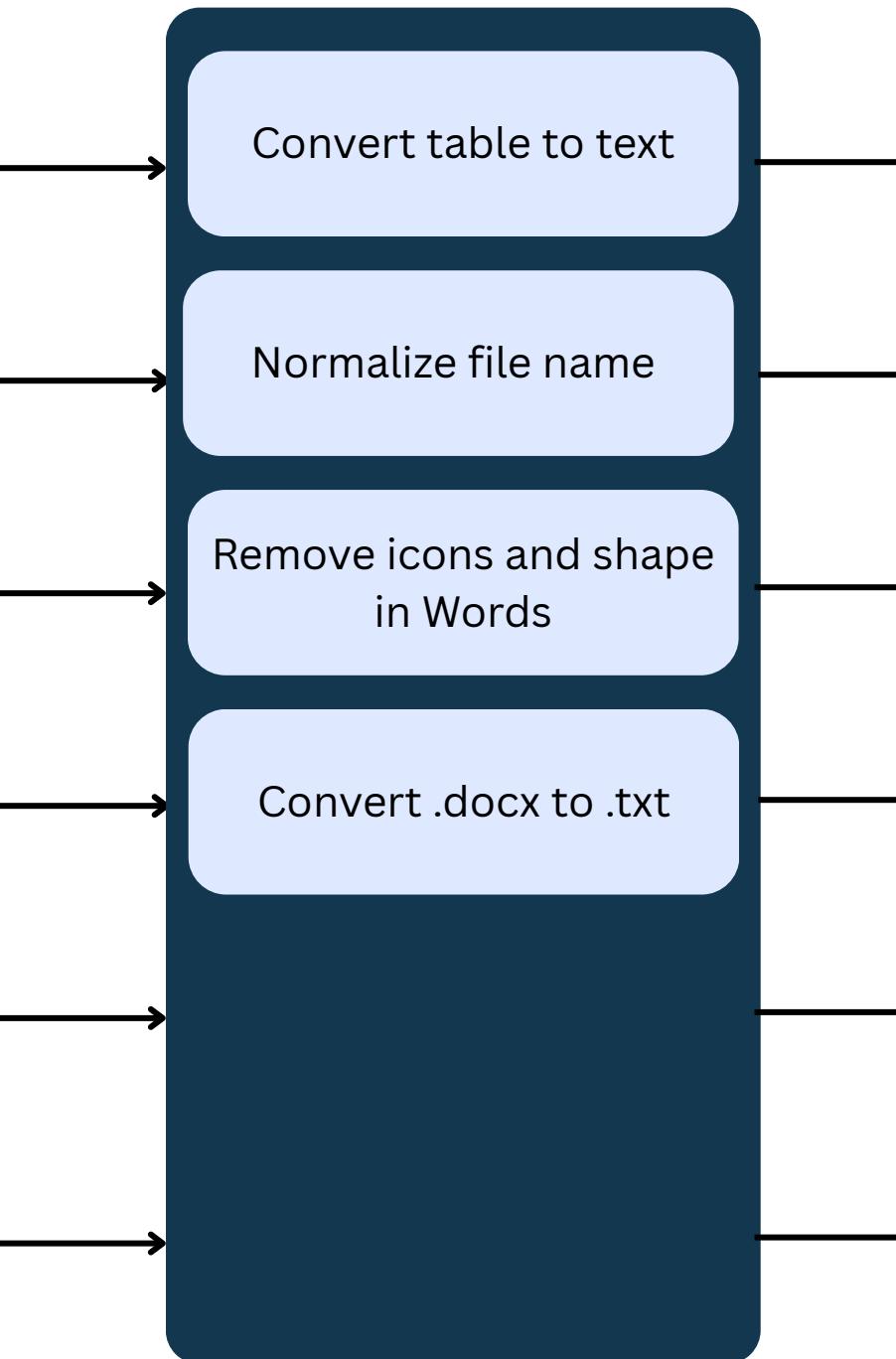
KB PROCESSING

Raw KB files

- เรื่อง Lab Catalogue KB1262
- เรื่อง NON UCEP (มาตรา7) KB66
- เรื่อง OP Anywhere KB1355
- เรื่อง Telemedicine ตนไทยในต่างป...
- เรื่อง UCEP COVID KB1068
- เรื่อง UCEP PLUS COVID สิทธิค...
- เรื่อง UCEP PLUS COVID สิทธิป...
- เรื่อง UCEP PLUS COVID สิทธิส...
- เรื่อง UCEP PLUS COVID สิทธิอป...
- เรื่อง ucep โรคทั่วไป KB1352
- เรื่อง UCEP สิทธิข้าราชการกรมบัญ...
- เรื่อง UCEP สิทธิครุภัณ KB1120
- เรื่อง UCEP สิทธิประกันสุขภาพ KB1094
- เรื่อง UCEP สิทธิรัฐวิสาหกิจ KB1109
- เรื่อง UCEP สิทธิหลักประกันสุขภาพ...
- เรื่อง UCEP สิทธิอปท. KB120
- เรื่อง กรณีผ่าตัดด้วยหุ่นยนต์ (Robo...
- เรื่อง กรณีสิทธิอยู่หมัดจาก KB1093
- เรื่อง กลุ่มคนที่เป็นสิทธิ KB1273
- เรื่อง กลุ่มเด็กปีแรกวัยรุ่น อายุ 6-24 ...
- เรื่อง กลุ่มเด็กเล็กอายุ 0-5 ปี PP KB...
- เรื่อง กลุ่มผู้สูงอายุ 60 ปี KB1039



Text Cleaning



Cleaned .txt file

[TABLE]
เรื่อง การนำเข้าข้อมูล 16 ไฟล์ | เลขที่ KB 1220
[/TABLE]

การนำเข้าข้อมูล 16 ไฟล์

ไฟล์ คือ การรวบรวมข้อมูลผู้ป่วยหลายรายเป็นไฟล์ และนำเข้าในโปรแกรม New E - Claim โดยรูปแบบการนำเข้าข้อมูล รองรับ 2 รูปแบบ ประกอบด้วย

- DBF file
- Text file

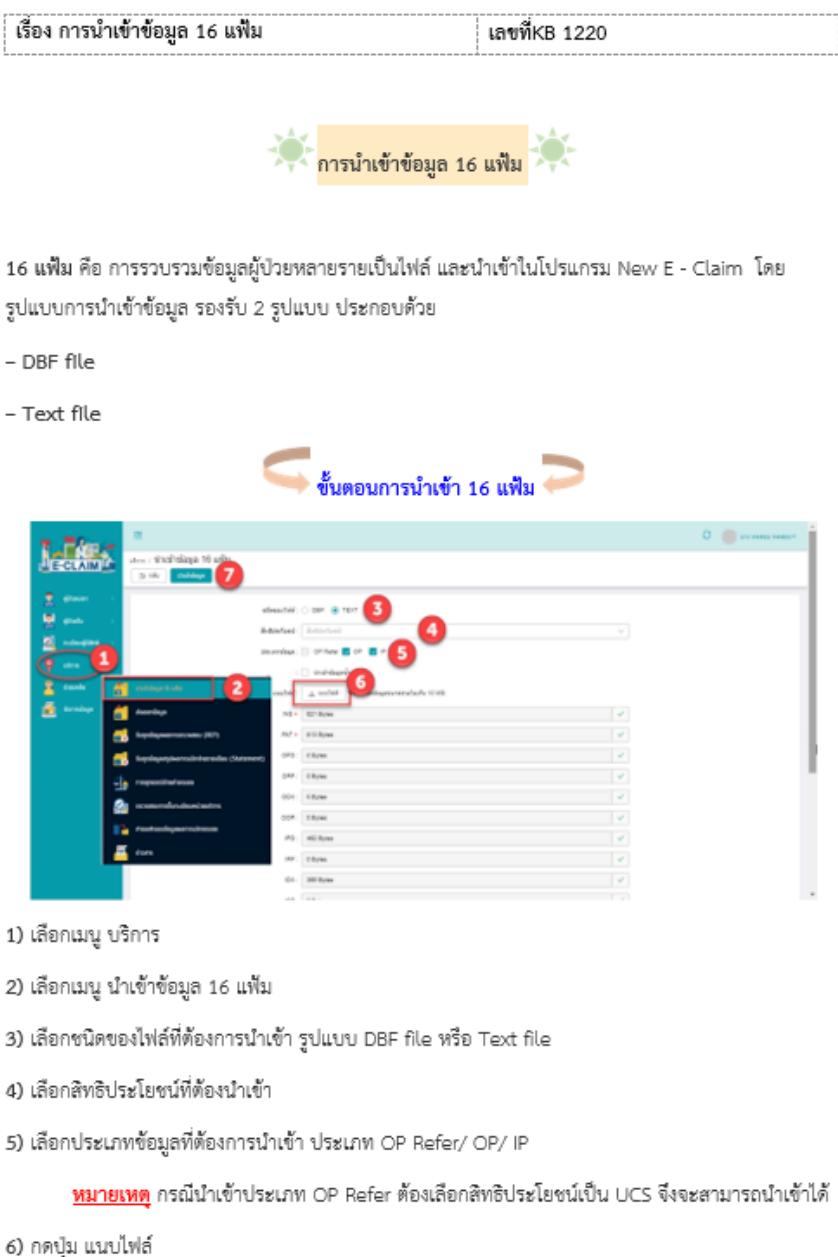
ขั้นตอนการนำเข้า 16 ไฟล์

[รูปภาพ] [/รูปภาพ]

KB PROCESSING

Raw KB files

- เรื่อง Lab Catalogue KB1262
- เรื่อง NON UCEP (มาตรา7) KB66
- เรื่อง OP Anywhere KB1355
- เรื่อง Telemedicine คนไทยในต่างป...
- เรื่อง UCEP COVID KB1068
- เรื่อง UCEP PLUS COVID สิทธิค...
- เรื่อง UCEP PLUS COVID สิทธิป...
- เรื่อง UCEP PLUS COVID สิทธิหน...
- เรื่อง UCEP PLUS COVID สิทธิอป...
- เรื่อง ucep โรคทั่วไป KB1352
- เรื่อง UCEP สิทธิเข้าร้านการกรมบัญ...
- เรื่อง UCEP สิทธิครุภัณฑ์ KB1120
- เรื่อง UCEP สิทธิประกันสุขภาพ KB1094
- เรื่อง UCEP สิทธิรัฐวิสาหกิจ KB1109
- เรื่อง UCEP สิทธิหลักประกันสุขภาพ...
- เรื่อง UCEP สิทธิอปท. KB120
- เรื่อง กรณีผ่าตัดด้วยหุ่นยนต์ (Robo...
- เรื่อง กรณีสิทธิอยู่หมู่เดียว KB1093
- เรื่อง กลุ่มคนไทยไร้สิทธิ KB1273
- เรื่อง กลุ่มเด็กโภคและวัยรุ่น อายุ 6-24 ...
- เรื่อง กลุ่มเด็กเล็กอายุ 0-5 ปี PP KB...
- เรื่อง กลุ่มผู้สูงอายุ 60 ปี KB1039



Text Cleaning

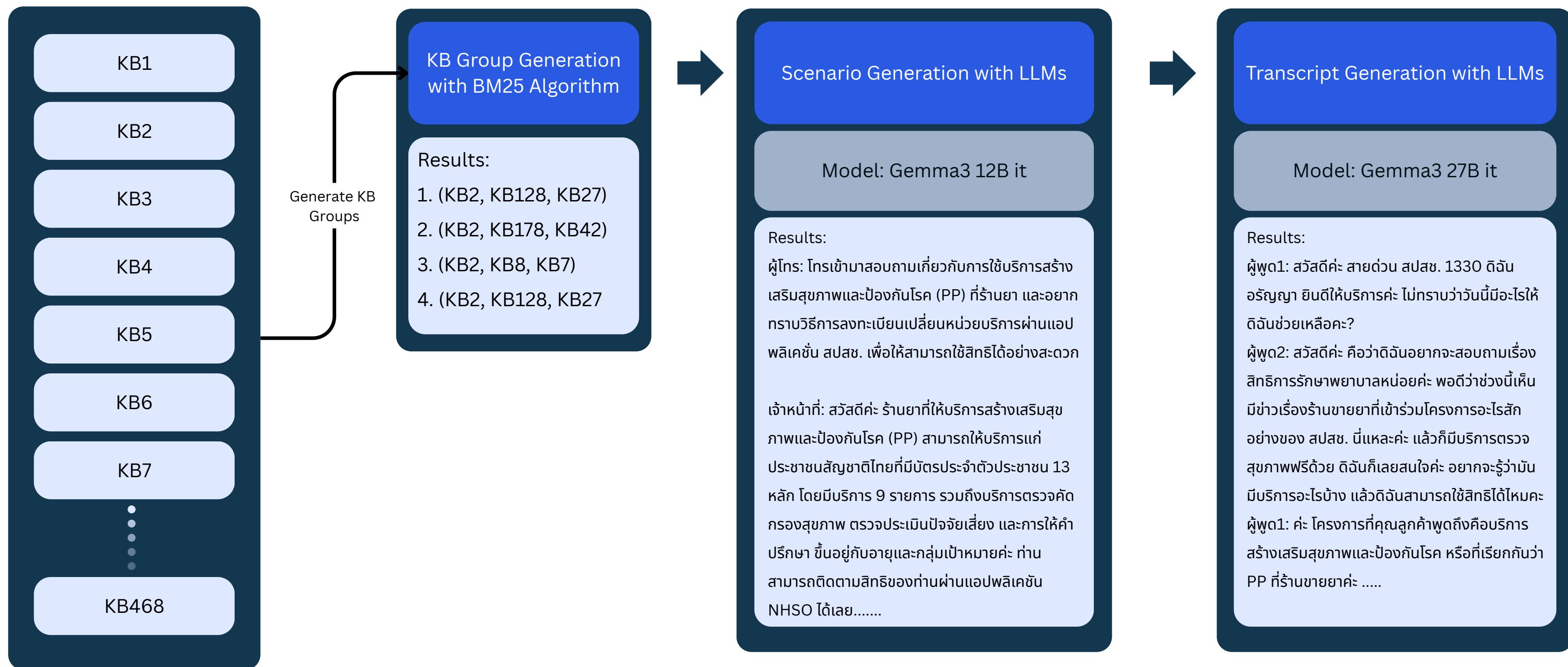
- Convert table to text
- Normalize file name
- Remove icons and shape in Words
- Convert .docx to .txt
- Image Explanation with Gemini 2.0 API / Qwen2.5 VL 3B Instruct

Cleaned .txt file

- [รูปภาพ]นี่คือคำอธิบายโครงสร้างและเนื้อหาข้อความจากภาพในภาษาไทย:
- ภาพนี้เป็นส่วนหนึ่งของระบบ E-CLAIM ของ NHSO และแสดงหน้าจอสำหรับนำเข้าข้อมูล 16 ไฟล์
- 1. เมบุหลัก: การด้านซ้ายมือ มีเมบุหลักซึ่งมีรายการต่างๆ ได้แก่ "บริการ" (เน้นด้วยวงกลมสีแดง) "ช่วยเหลือ" และ "จัดการข้อมูล"
- 2. เมบุย่อย: เมื่อคลิกที่ "บริการ" จะปรากฏเมบุย่อย ซึ่งมีรายการ "นำเข้าข้อมูล 16 ไฟล์" (เน้นด้วยกรอบสีแดง) พร้อมกับตัวเลือกอื่นๆ เช่น "ส่งออกข้อมูล," "รับชุดข้อมูลผลการตรวจสอบ (REP)," "รับชุดข้อมูลสรุปผลการเบิกจ่ายรายเดือน (Statement)," "การอุทธรณ์จ่ายค่าชดเชย," "ตรวจสอบการรับประทานยา," "การอุทธรณ์จ่ายค่าชดเชย," "สำรวจข้อมูลผลการเบิกจ่าย," และ "ข่าวสาร"
- [/รูปภาพ]

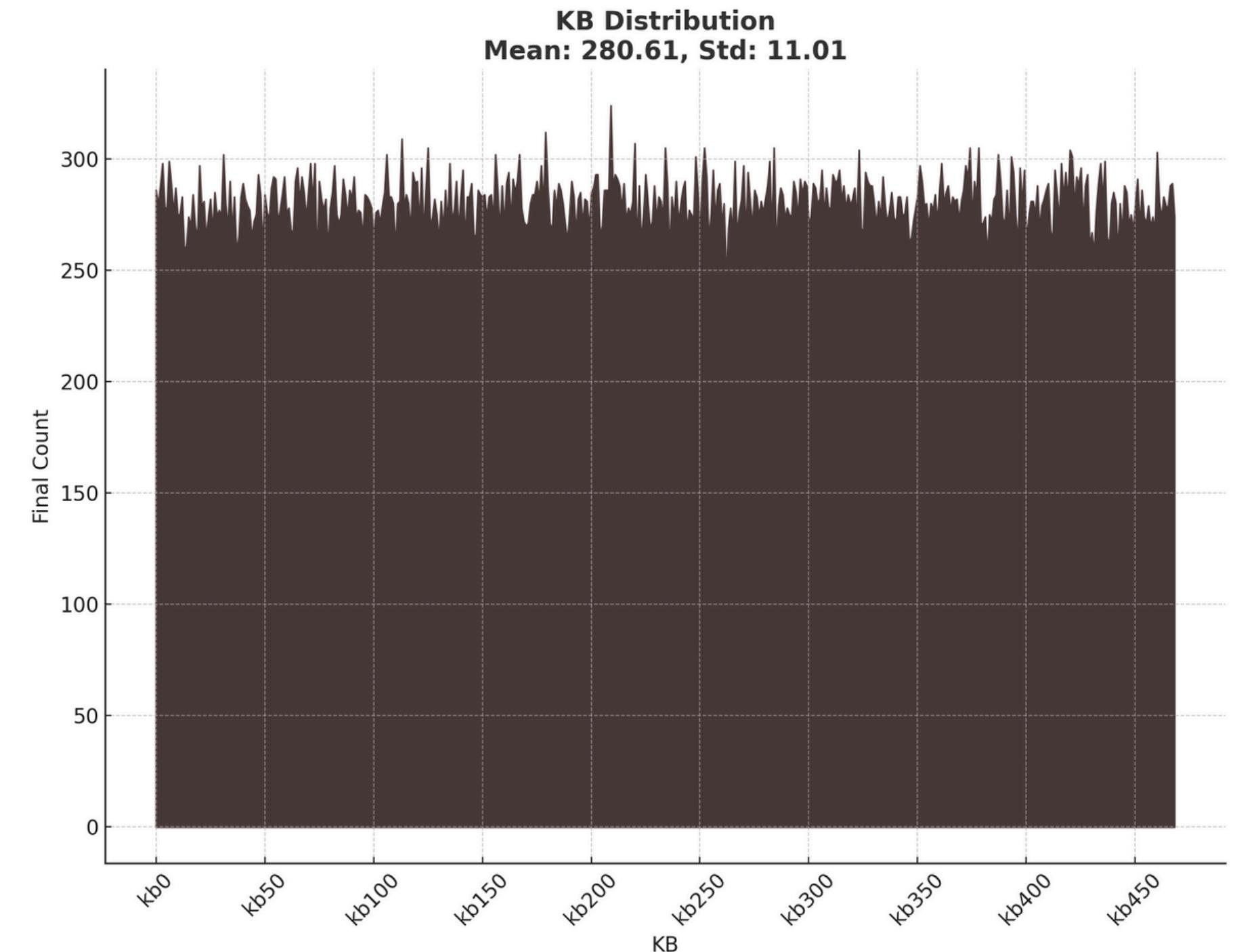
DATA SYNTHESIS

All KBs



DATA SYNTHESIS

Transcript Distribution



SPEECH TO TEXT



Fine-tuned Monsoon Whisper,
reducing the word error rate

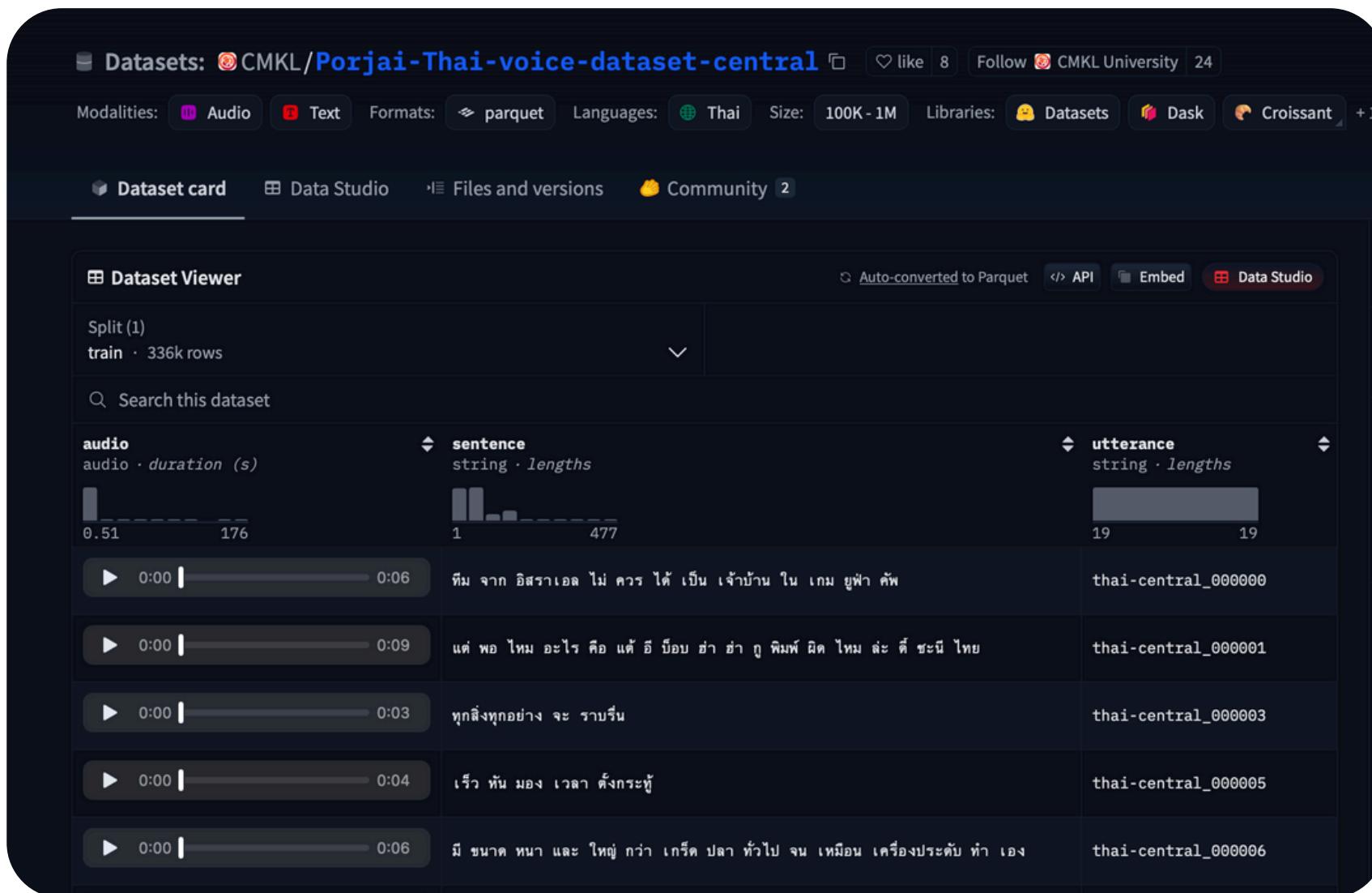
We segment audio files into smaller chunks and use
batch processing to transcribe them efficiently.

Use PyAnnote for speaker diarization and integrate its
output into the transcription pipeline to improve multi-
speakers transcript accuracy

FINETUNING STT MODEL

Trained using CMKL/Porjai-
Thai-voice-dataset-central

336k entries



Distilled Whisper TH	Monsoon	Fine Tuned Monsoon
0.3822	0.3337	0.3222
0.3005	0.2989	0.3217
0.2488	0.283	0.2146
0.3157	0.3337	0.3333
0.1299	0.0909	0.1299
0.2222	0.1852	0.2037
0.1447	0.1447	0.0263
0.1818	0.0364	0.0
0.3611	0.0972	0.25
0.1047	0.1395	0.0465
0.1042	0.1042	0.0625
0.0678	0.2373	0.0847
0.0676	0.1081	0.0676
0.122	0.0854	0.061
0.2824	0.3294	0.3294
0.2613	0.1532	0.1441
0.1471	0.1275	0.1373
0.4478	0.194	0.2687
0.4074	0.6852	0.2778
0.2274	0.2033	0.1602

Word Error Rate Comparison

SPEAKER DIARIZATION

Preprocessing:

Resample, mono conversion

Voice Activity Detection:

Detects speech vs. silence

Embedding:

Extracts speaker voice features

Segmentation:

Chunks audio for analysis

Clustering:

Groups similar voices and assigns speaker labels

Post-processing:

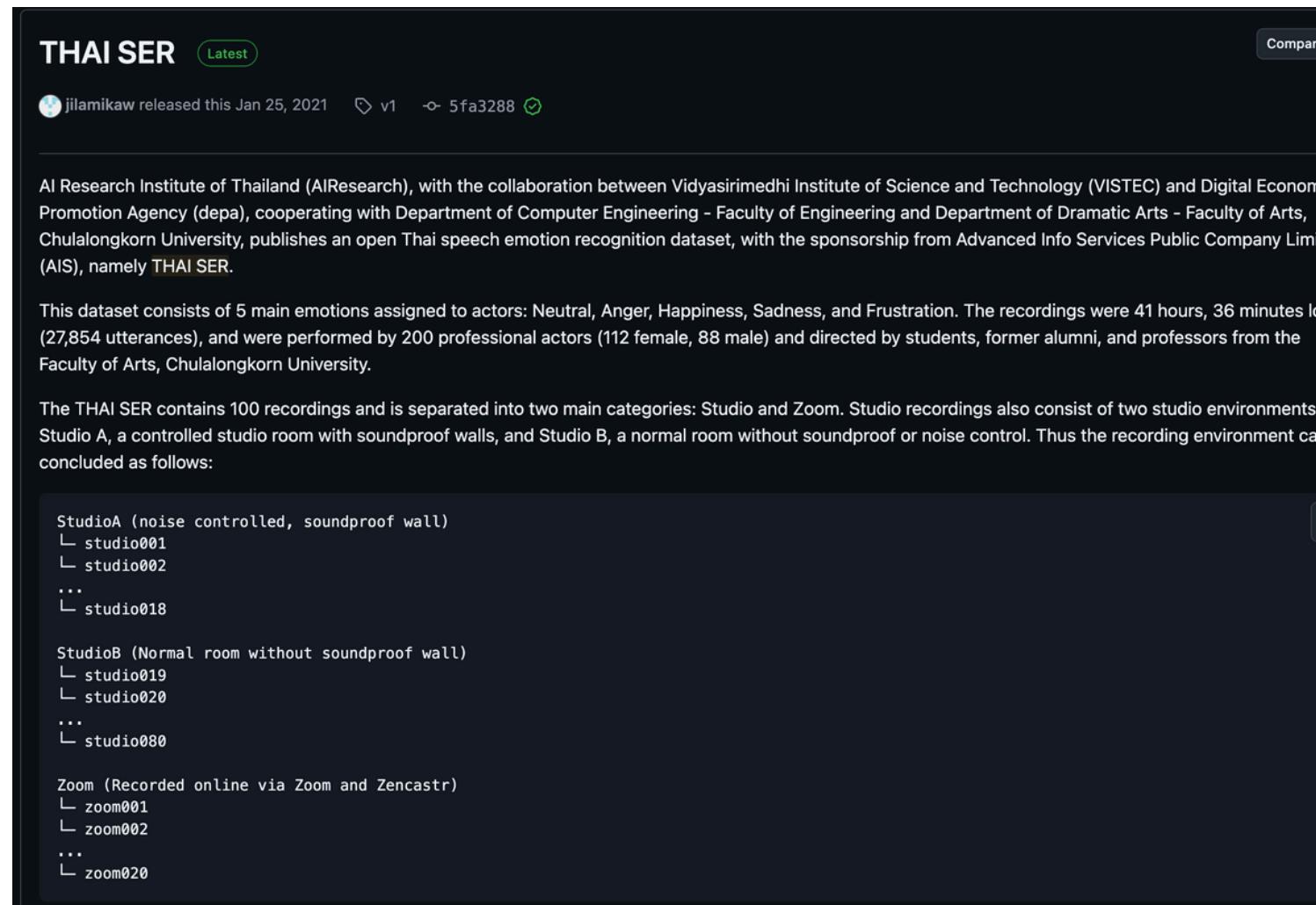
Merges short/overlapping segments



Speakers	DER
Male and Female	1.72%
Male and Male	30.24%
Female and Female	46.73%
Total	10.63%

FINETUNING SPEAKER DIARIZATION

THAI SER 200 Male and Female Actors



Speaker Diarization Finetuning Results

Epoch	Diarization Error Rate	Confusion
0	0.1439	0.1404
1	0.0889	0.0854
2	0.0255	0.0190
3	0.0118	0.0084
4	0.0118	0.0080

KB CLASSIFICATION



Researched **embedding models**, **chunking** and **retrieval methods**, **vector databases**, and their combinations



Implement Final Pipeline
of best performing components



Implement
KB retrieval/classification
and updates as **API**



MODELS

WangchanBERTa

PhayathaiBERT

BGE-M3

**SCT-Distil-model-
phayathai-bert-bge-m3**

CHUNKING METHODS

recursive

semantic

ClusterSemanticChunker

<https://research.trychroma.com/evaluating-chunking>

BGE_confidence

RETRIEVAL METHODS

BM25

(Lexical Search)

Direct Semantic Search

(Dense Retrieval)

Hybrid

(BM25 lexical + Vector Search)



MODELS

WangchanBERTa

PhayathaiBERT

BGE-M3

**SCT-Distil-model-
phayathai bert-bge-m3**

INDEXING METHODS

IndexFlatL2

IndexIVFlat

IndexHNSWFlat

RETRIEVAL METHODS

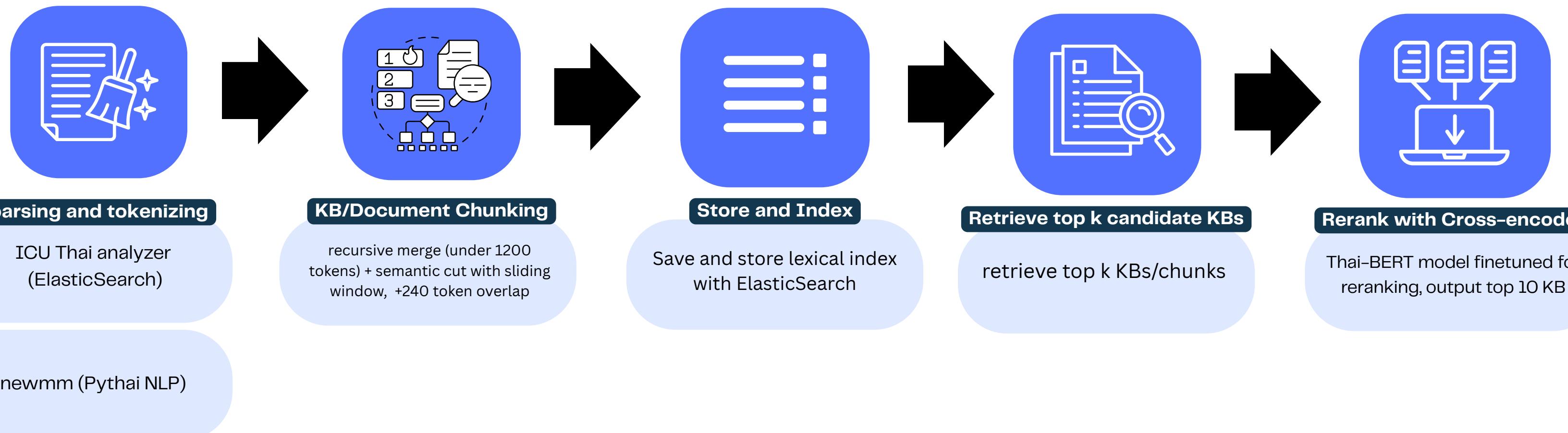
BM25
(Lexical Search)

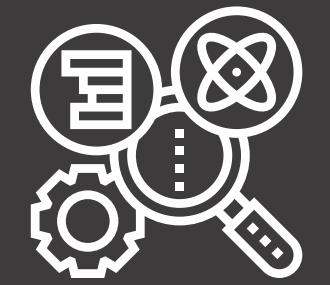
Direct Semantic Search
(Dense Retrieval)

Hybrid
(BM25 lexical + Vector Search)



BM25 + THAI BERT RERANKER

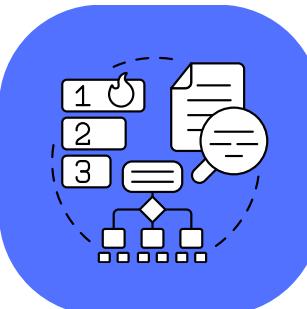




DIRECT SEMANTIC SEARCH + RERANKER



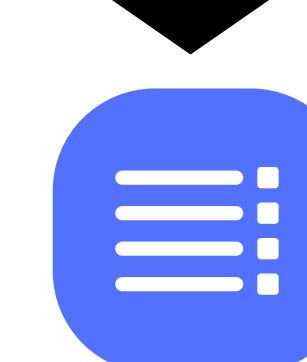
parsing and tokenizing
SentencePiece 50 k
(automatic with BGEM3)



KB/Document Chunking
recursive merge (under 1200 tokens) + semantic cut with sliding window, +240 token overlap



Embed Chunks
Ecnepcde chunks or KB and Transcript with Thai embedding model



Store and Index
Store generated embeddings in a vector DB

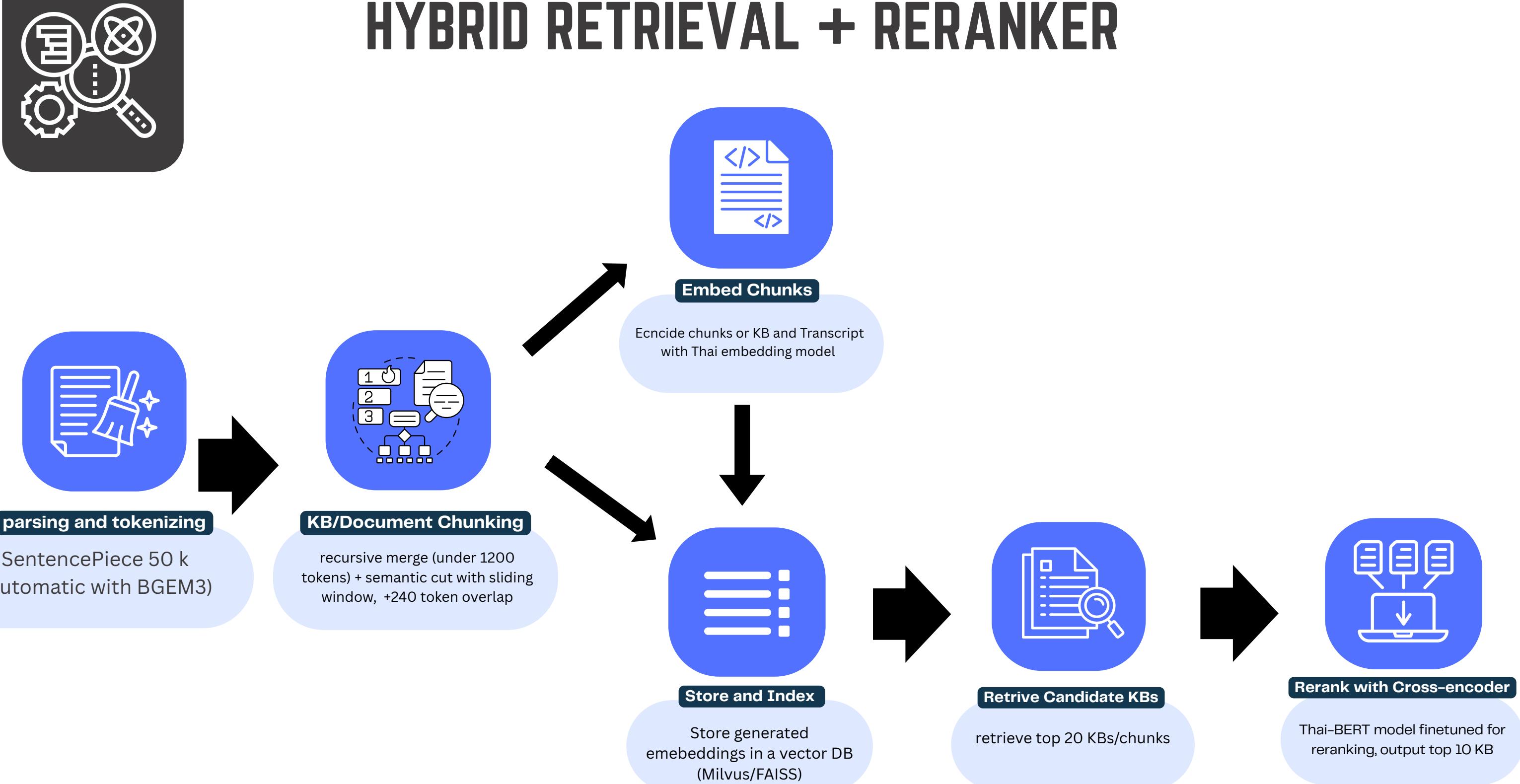


Retrieve Candidate KBs
retrieve top k KBs/chunks

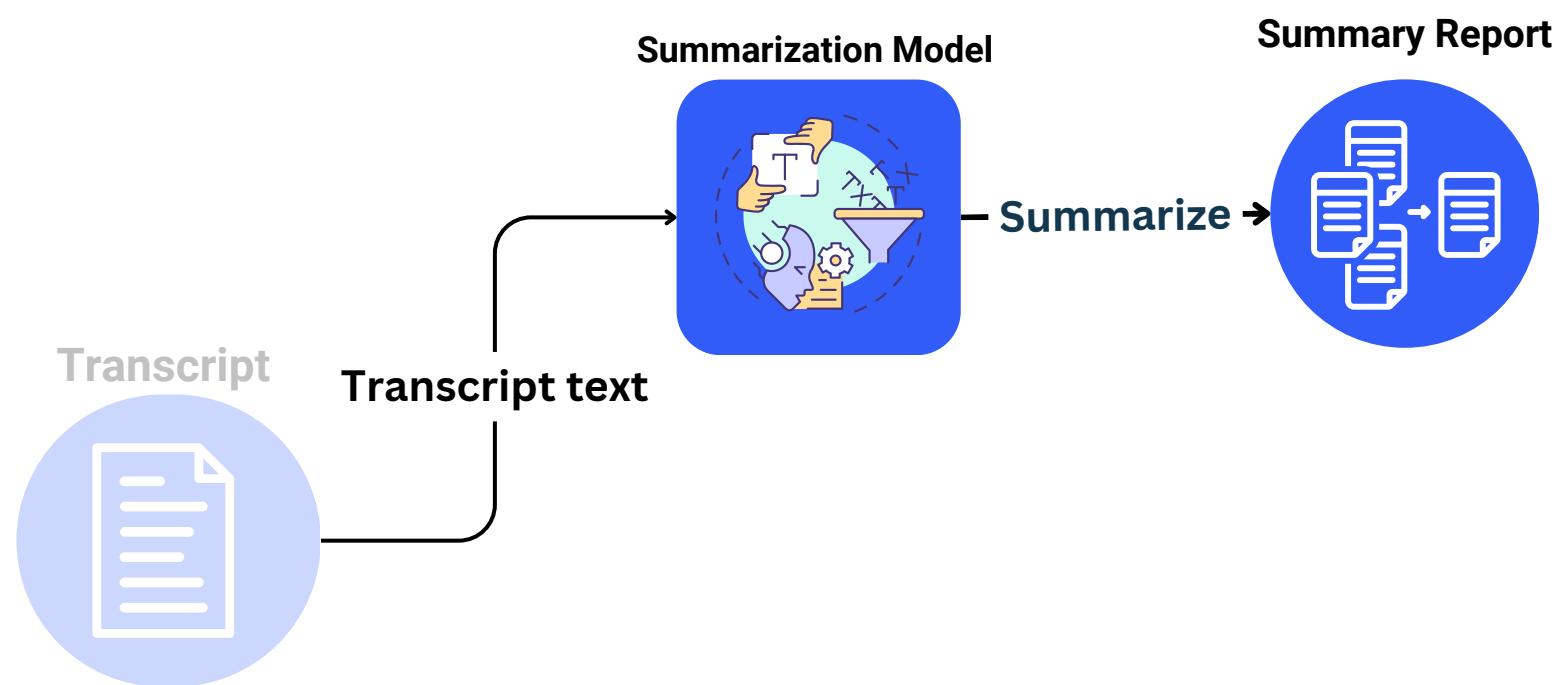


Rerank with Cross-encoder
Thai-BERT model finetuned for reranking, output top 10 KB

HYBRID RETRIEVAL + RERANKER



SUMMARISATION MODEL



Distill Gemma 3 27B to generate high-quality answers.

Explore suitable pre-trained models. Train Llama 3.2 1B using the distilled outputs.

Train Qwen 2.5 0.5B using the same distilled outputs.

FINE-TUNED MODELS

Qwen 2.5 0.5B Instruct



Llama 3.2 1B Instruct



QWEN 2.5 0.5B INSTRUCT

Evaluation Results

Model	Pre-train	Fine-tune
Training Data Size	-	51.3 K
Testing Data Size		5.7 K
ROUGE-1	0.28	0.70
ROUGE-2	0.15	0.52
ROUGE-L	0.27	0.67

Result Example

Generated Text

ໂກຣແຈ້ງວ່າຮູ້ສັກໄມ່ສະບາຍຕົວມາສອງສານວັນ ມີອາການເຈັບຄອ ບ້ານຸກໄໂລ ໂອ ແລະຮູ້ສັກອ່ອນພຶ່ຢ ໂດຍໄມ່ແປ່ໃຈວ່າເປັນອາການຂອງໂຄວິດຮ້ອງໃໝ່ຫວັດໃໝ່ ເຈົ້າໜ້າກໍໄດ້ເປີຍບເຖິງອາການເບື້ອງຕົ້ນຂອງກົ່ງສອງໂຮງ ແລະແນະນຳໃຫ້ຜູ້ໂກຣໄປຕຽວຈາກເຊື້ອໂຄວິດທີ່ໂຮງພຍາບາລຮ້ອສຄານພຍາບາລໃກ້ລ້ານ ເນື່ອງຈາກຜູ້ໂກຣມີຄວາມກັງວລເຮື່ອການຕິດເຊື້ອຈາກກໍ່ອື່ນເນື່ອງຈາກໄມ່ສະດວກໃນການເດີນກາງໄປໂຮງພຍາບາລ ເຈົ້າໜ້າກໍໄດ້ໃຫ້ຂ້ອມູລເກື່ຽວກັບຊຸດຕຽວ ATK (Antigen Rapid Test) ທີ່ສາມາດກຳໄດ້ເອງກໍ່ບ້ານ ແລະແນະນຳໃຫ້ພັກຝ່ອນ ດື່ນນ້ຳ ແລະຫລັກເລື່ອງການສັນພັສໃກ້ລ້ັສິດກັບຜູ້ອື່ນ ມີຄວາມກຳໄດ້ເອງກໍ່ບ້ານ ແລະໄປພັບແພທຍໍ່ຫາກອາການໄມ່ດີຂຶ້ນກາຍໃນ 2-3 ວັນ ຜູ້ໂກຣຮັບການຄຳແນະນຳແລະຂອບຄຸນເຈົ້າໜ້າກໍ່

Reference Text:

ຜູ້ປ່ວຍໂກຣສັພກເຂົ້າມາຍັງສາຍດ່ວນ ສປປຊ. 1330 ດ້ວຍອາການເຈັບຄອ ບ້ານຸກໄໂລ ໂອ ແລະຮູ້ສັກອ່ອນພຶ່ຢ ໂດຍມີຄວາມກັງວລ ວ່າອາຈາເປັນໂຮງໂຄວິດ-19 ເນື່ອງຈາກມີອາການສຸຄຸມເສີຍການຮັບຮສແລະກລິ່ນຮ່ວມດ້ວຍ ເຈົ້າໜ້າກໍໄດ້ໃຫ້ຂ້ອມູລເປີຍບເຖິງອາການຂອງໂຄວິດ-19 ແລະໄໝ້ຫວັດໃໝ່ ແລະແນະນຳໃຫ້ຜູ້ປ່ວຍໄປຕຽວຈາກເຊື້ອໂຄວິດ-19 ເພື່ອຄວາມແປ່ໃຈ ເນື່ອງຈາກຜູ້ປ່ວຍໄມ່ສະດວກໄປໂຮງພຍາບາລ ເຈົ້າໜ້າກໍຈຶ່ງແນະນຳກາງເລືອກໃນການໃໝ່ຊຸດຕຽວ ATK (Antigen Rapid Test) ດ້ວຍຕົນເອງກໍ່ບ້ານ ແລະໃຫ້ຄຳແນະນຳເກື່ຽວກັບວິທີການໃໝ່ແລະພົກພະວານຕຽວຈາກພັກຝ່ອນ ມີຄວາມອູ່ ຄວາມພັກຝ່ອນແລະຫລັກເລື່ອງການສັນພັສໃກ້ລ້ັສິດກັບຜູ້ອື່ນ ມີຄວາມກຳໄດ້ເອງກໍ່ບ້ານ ແລະໄປພັບແພທຍໍ່ຫາກອາການໄມ່ດີຂຶ້ນກາຍໃນ 2-3 ວັນ ມີຄວາມກຳໄດ້ເອງກໍ່ບ້ານ ແລະໄປພັບແພທຍໍ່ຫາກອາການໄມ່ດີຂຶ້ນກາຍໃນ 2-3 ວັນ ມີຄວາມກຳໄດ້ເອງກໍ່ບ້ານ

META-LLAMA/LLAMA-3.2-1B-INSTRUCT

Evaluation Results

Model	Pre-train	Fine-tune
Training Data Size	-	51.3 K
Testing Data Size		5.7 K
ROUGE-1	0.34	0.75
ROUGE-2	0.19	0.58
ROUGE-L	0.32	0.72

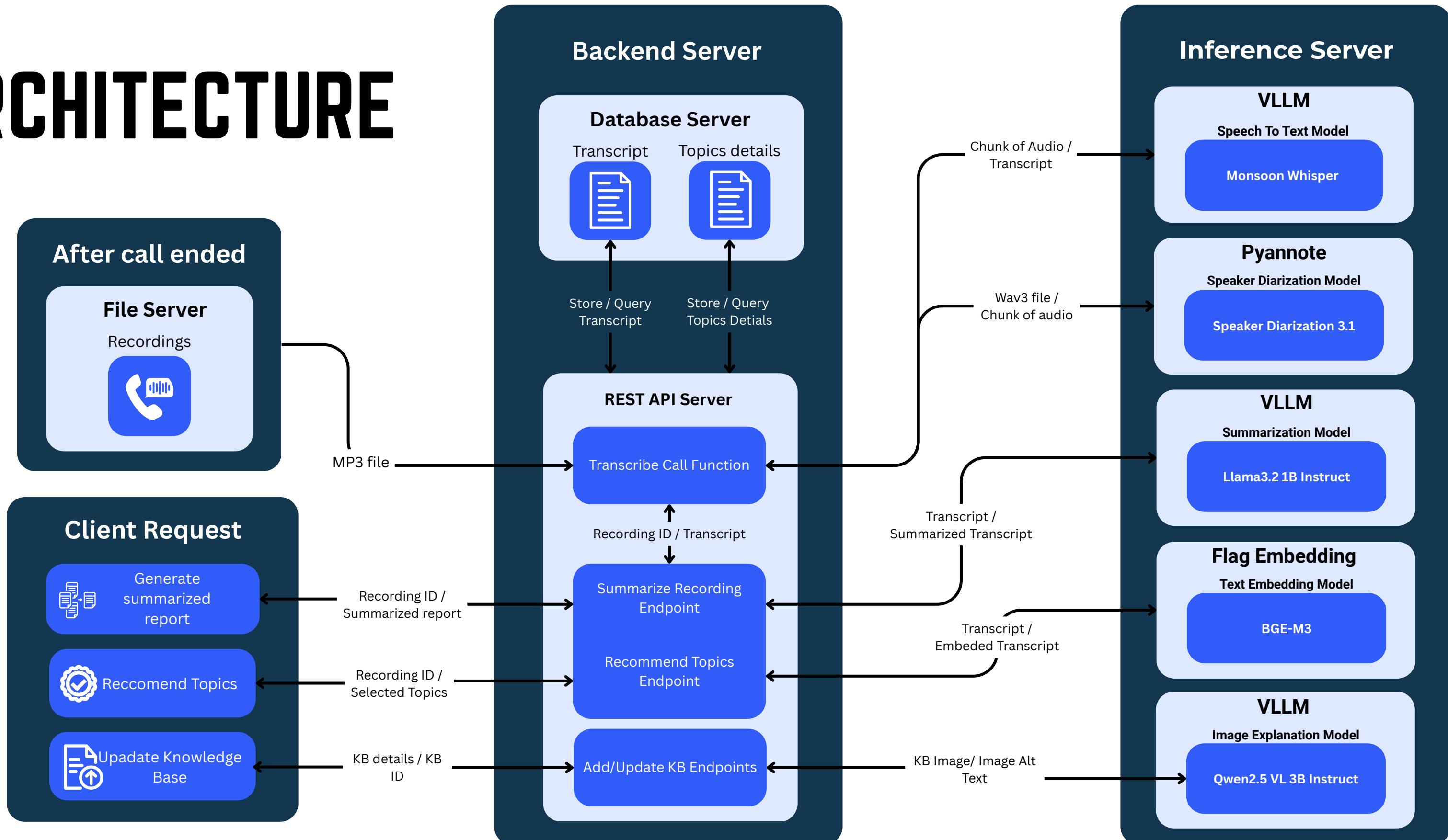
Generated Text

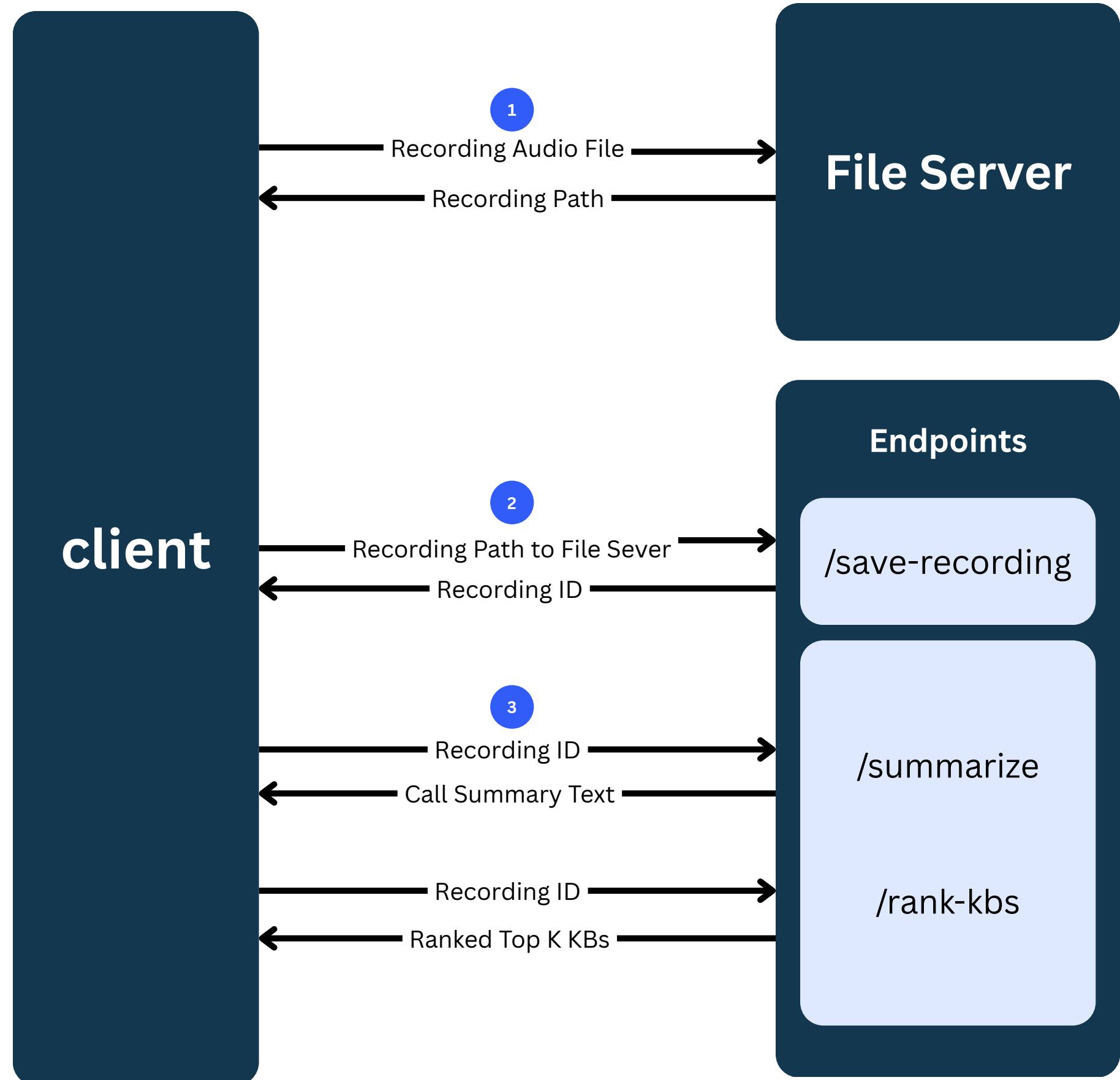
ผู้โดยสารสามารถข้อมูลเกี่ยวกับการใช้บริการที่คลินิกพยาบาลชุมชนอบอุ่นและการเบิกจ่ายผ่านระบบ KTB โดยเฉพาะรายการบริการสร้างเสริมสุขภาพและป้องกันโรค (PP) ผู้โดยสารมีความกังวลเกี่ยวกับการจัดการข้อมูลการเบิกจ่ายของคลินิก หากข้อมูลผิดพลาดหรือต้องแก้ไข คลินิกสามารถแก้ไขข้อมูลในระบบ KTB ได้ภายใน 72 ชั่วโมงหลังจากการส่งข้อมูล หากเกิดกำหนดจัดตั้งรอบการประเมินผลของ สปสช. ซึ่งมีสองรอบต่อเดือน คือวันที่ 15 และวันสิบเดือน หากสถานะยังคงเป็น \"รออนุมัติ\" แม้ข้อมูลถูกต้องแล้ว ผู้โดยสารสามารถตรวจสอบเหตุผลในระบบ KTB หรือติดต่อ Call Center 1330 หรือ Line OpenChat ของ สปสช. เพื่อขอความช่วยเหลือเพิ่มเติม นอกจากนี้ ผู้โดยสารยังสามารถเกี่ยวกับขั้นตอนการเพิ่มรายการบริการใหม่ เช่น การตรวจ HPV DNA Test ซึ่งคลินิกจะต้องขออนุมัติศักยภาพจากเขต สปสช. พร้อมส่งเอกสารประกอบการพิจารณา เจ้าหน้าที่ได้ให้ข้อมูลเกี่ยวกับขั้นตอนการเบิกจ่ายผ่าน KTB และแนะนำช่องทางการติดต่อเพิ่มเติมหากมีข้อสงสัยเพิ่มเติม

Reference Text :

ผู้โดยสารสามารถเกี่ยวกับวิธีการใช้บริการที่คลินิกพยาบาลชุมชนอบอุ่น และการเบิกจ่ายผ่าน KTB โดยมีข้อสงสัยเกี่ยวกับการดำเนินงานของคลินิกในการเบิกจ่ายบริการสร้างเสริมสุขภาพและป้องกันโรค (PP) เจ้าหน้าที่ได้อธิบายว่าคลินิกสามารถแก้ไขข้อมูลการเบิกจ่ายในระบบ KTB ได้ภายใน 72 ชั่วโมง และ สปสช. จะมีรอบการประเมินผลข้อมูลทุกวันที่ 15 และสิบเดือน หากข้อมูลถูกต้องแต่สถานะยัง \"รออนุมัติ\" ให้ตรวจสอบเหตุผลในระบบ KTB หรือติดต่อ Call Center 1330 หรือ Line OpenChat ของ สปสช. เพื่อขอความช่วยเหลือเพิ่มเติม นอกจากนี้ ผู้โดยสารสามารถเกี่ยวกับขั้นตอนการเบิกจ่ายผ่าน KTB ซึ่งเจ้าหน้าที่แจ้งว่า หน่วยบริการต้องบันทึกข้อมูลในระบบ KTB หลังให้บริการเสร็จสิ้น และ สปสช. จะอนุมัติการเบิกจ่ายทุก 15 วัน หากคลินิกต้องการเพิ่มรายการบริการ เช่น การตรวจ HPV DNA Test ต้องขออนุมัติจากเขต สปสช. ก่อน

ARCHITECTURE





TECHNOLOGY STACK

Backend/API Server



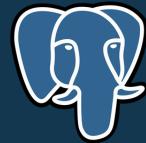
FastAPI

Model Inference Engine



FlagEmbedding

Database



PostgreSQL



elasticsearch



Deployment





NHSO CRM SYSTEM

Call_ID: 00058

00:33

งานการไฟฟ้า



ข้อมูลการติดต่อ

ชื่อ-นามสกุล

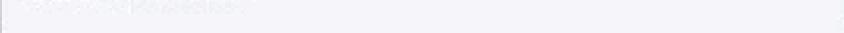
หมายเลขบัตรประชาชน

อีเมล

หมายเลขโทรศัพท์

ประเภทการโทร

ประเภทการโทร



หัวข้อฐานความรู้

Q_ เมืองจากท่องเที่ยว

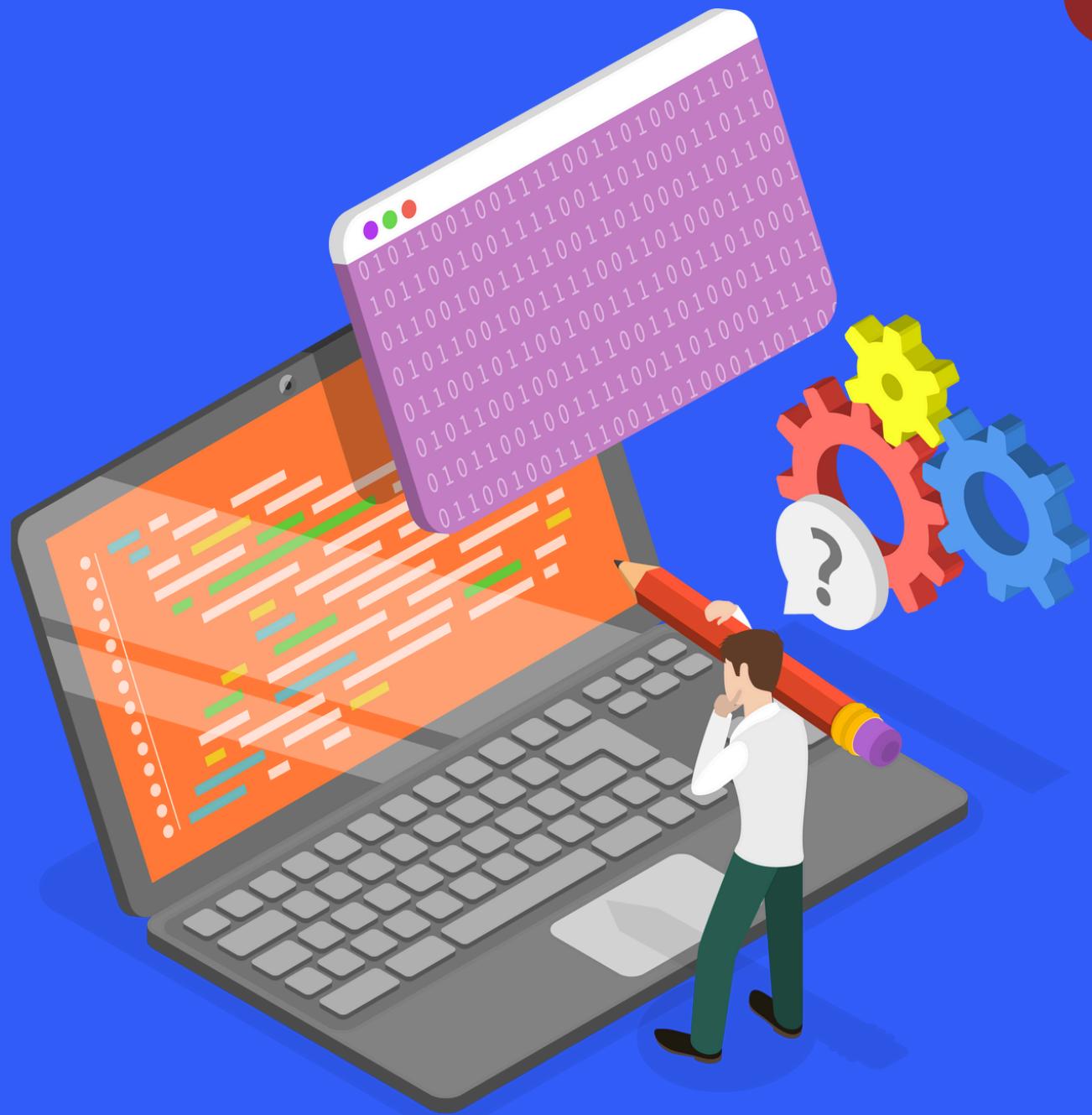
แนะนำหัวข้อฐานความรู้

รายละเอียดการโทร

รายละเอียดการโทร

ISSUES/ CHALLENGES

NO DATA PROVIDED



**unable to finetune models
with real data**

**unable to test pipeline
with real data**

AURA

AI-POWERED UNIFIED RESPONSE AND ANALYSIS

OUR TEAM



PON YIMCHAROEN



SHINE MINKHA



KASIDIS MANASURANGKUL



NIRACHA JANAVATARA



SUPAKORN ETITUM

THANK YOU!

