University of Pisa


Business and Project Management

Project

# Business Meeting Notes generation and evaluation using Generative AI

Marco Bologna

A.Y. 2022-2023

# Contents

# 1 Introduction to the problem

In the dynamic landscape of modern business operations, meetings are indispensable for collaboration, decision-making, and strategic planning. However, the effectiveness of meetings is often hampered by the challenges of accurately capturing and summarizing the discussions, decisions, and action items that arise during these sessions.
Traditional note-taking methods, whether manual or digital, require significant time and effort and are prone to human error, leading to potential misunderstandings or loss of critical information.

The advent of *Artificial Intelligence* (AI) technologies presents a novel solution to these challenges.
At the heart of this project lies this fundamental question: *Can generative AI create comprehensive and informative Business Meeting Notes using only the recording of the meeting?*
To answer to this question, this document explores the development of an AI-based system designed to transcribe recorded meetings and generate concise, coherent, and comprehensive meeting notes.
The system aims to address the following key challenges:

- **Transcription Accuracy**: Converting spoken language into written text with high accuracy, including the recognition of technical jargon, names, and industry-specific terminology.
- **Speaker Identification**: Distinguishing between different speakers in the meeting and accurately associating statements and ideas with the respective individuals.
- **Content Summarization**: Identifying and summarizing key points, decisions, and action items from lengthy discussions, ensuring that the essence of the meeting is captured without loss of critical information.
- **Formatting and presentation**: Creating a template to show Business Meeting Notes in the most complete way.

The following chapters will detail the methodologies employed in developing this AI-based meeting notes generation system, the technical and operational challenges encountered and the solutions devised to overcome these challenges.

# 2 The Dataset

In the quest to develop an AI-driven solution for generating business meeting notes, the choice of an appropriate dataset is pivotal.
The ideal dataset would encompass a broad spectrum of meeting scenarios, discussions, and decision-making processes typical of a business environment.
However, a significant challenge in this endeavor is the inherent confidentiality of business meetings, which makes accessing genuine corporate meeting recordings and transcripts a complex issue due to privacy and proprietary concerns.

To navigate these constraints, the *Meeting Bank* dataset was chosen as the foundation for this project.
While the dataset does not originate from a corporate setting, it comprises comprehensive records of city council meetings from six major U.S. cities, amounting to over 3,579 hours of video and audio content.
The public domain nature of the "MeetingBank" dataset offers a unique advantage. The meetings it contains, being of a public and official nature, are not encumbered by the confidentiality issues that beset corporate meetings. This accessibility allows for a robust analysis and application of AI technologies in transcribing, understanding, and summarizing complex meeting discussions without breaching privacy norms.

The dataset provides a rich tapestry of dialogues, covering a wide array of topics from municipal governance, public policy discussions, budget hearings, to community issues, mirroring the multifaceted nature of conversations that occur in business settings. This variety ensures that the AI model trained on this dataset can handle a broad spectrum of discussion themes, terminologies, and interaction dynamics.

# 3 Used Tools

## 3.1 Audio to Text Transformation

The transcription of audio recordings into accurate text is a critical step in the process of generating business meeting notes. This chapter delves into the rationale behind the selection of *Speechmatics* for speech-to-text conversion and elucidates the pivotal role of speaker diarization in enhancing the quality and utility of the transcribed text.

**Speechmatics**, a leading provider in the field of automatic speech recognition (ASR), was chosen for its advanced capabilities and robust performance across diverse audio qualities and linguistic nuances.
A distinguishing feature that sets Speechmatics apart is its proficiency in speaker diarization, a crucial factor for this project given the multiplicity of participants typical in meeting scenarios.
**Speaker diarization** is a process that involves partitioning an audio stream into homogeneous segments according to the speaker's identity. Essentially, it answers the question "*Who spoke when?*". This feature is instrumental in attributing dialogue to the correct participants within a meeting, thereby preserving the conversational context and dynamics essential for accurate note-taking.
Speechmatics' diarization technology employs sophisticated machine learning algorithms to analyze the audio stream, differentiate between speakers, and accurately label the segments. This capability not only enhances the clarity of the transcribed text but also lays a solid foundation for generating structured and comprehensible meeting notes.

## 3.2 Text Generation

Following the transcription of meeting audio into text, the next pivotal step in the creation of business meeting notes is the generation of coherent, concise, and relevant summaries.
This process leverages the cutting-edge capabilities of **OpenAI**'s *GPT-3.5* and *GPT-4* models, accessed via API, to transform the raw transcriptions into polished meeting notes.
OpenAI's Generative Pre-trained Transformer models, particularly *GPT-3.5* and *GPT-4*, represent the forefront of natural language processing technology. These models have been trained on a diverse range of internet text, enabling them to generate human-like text based on the input

they receive. For this project, the choice to utilize *GPT-3.5* and *GPT-4* was driven by their unparalleled ability to understand context, generate coherent narratives, and maintain consistency in text generation.
The initial phase of the project utilized *GPT-3.5* for the generation of meeting notes.
While *GPT-3.5* offered impressive capabilities in terms of understanding context and generating coherent text, certain limitations became apparent as the project evolved. The need for handling more extensive conversations, producing higher quality summaries, and accurately recognizing and attributing speakers led to the decision to upgrade to *GPT-4*.

- One of the pivotal reasons for this transition was *GPT-4*'s **expanded token limit** for both input and output. This enhancement allowed for the processing of longer segments of transcribed text in a single prompt, enabling more comprehensive summaries of extended discussions. The increased token limit facilitated a more seamless integration of entire meeting transcripts, reducing the need to segment the text into smaller parts and thus preserving the contextual continuity.

- *GPT-4*'s advanced algorithms and larger training dataset have significantly **improved the quality** of text generation. This model exhibits a superior ability to distill complex discussions into concise, clear, and relevant summaries. The enhanced quality of *GPT-4*'s summaries ensures that the essence of the meetings is captured more effectively, with critical decisions, action items, and discussion points accurately reflected in the notes.

- A notable advancement with *GPT-4* is its refined capability to **understand and attribute speakers** within the conversation. Utilizing the speaker labels provided by the *Speechmatics* transcription (e.g., Speaker 1, Speaker 2), *GPT-4* adeptly associates these labels with the actual names of the speakers if mentioned within the conversation.

For businesses looking to implement this AI-driven meeting note generation system, ensuring that each participant introduces themselves at the beginning of the conversation can significantly enhance the accuracy of speaker attribution. This practice, coupled with *GPT-4*'s advanced speaker recognition, can streamline the process of creating structured and personalized meeting notes, making the content more accessible and useful for all stakeholders.

## 3.3 PDF Generation

In the endeavor to streamline the process of converting meeting notes into professional and structured PDF documents, this project harnessed the power of PDFGeneratorAPI. This platform stands out for its ability to generate PDFs from customizable templates using JSON data, marrying flexibility with automation in document creation.

The initial step involved designing a template that would universally fit the format of business meeting notes. Using *PDFGeneratorAPI*'s intuitive browser-based template editor, I crafted a layout that includes specific fields tailored to capture the essence of any meeting.
These fields are:

- **Title**: Captures the primary focus or subject of the meeting, providing a clear and concise reference point.
- **Location**: Details the venue of the meeting, whether a physical location or a virtual meeting space.
- **Date**: Records the date the meeting took place, crucial for historical context and chronological organization.
- **Attendees**: Lists the participants of the meeting, acknowledging their presence and contributions.
- **Agenda Items Discussed**: Enumerates the key topics addressed during the meeting, offering a snapshot of the discussion points.
- **Summary**: Provides a succinct overview of the meeting's outcomes, decisions made, and any consensus reached.
- **Notes**: Allows for additional remarks, observations, or action items that emerged from the meeting, providing a comprehensive account of the proceedings.

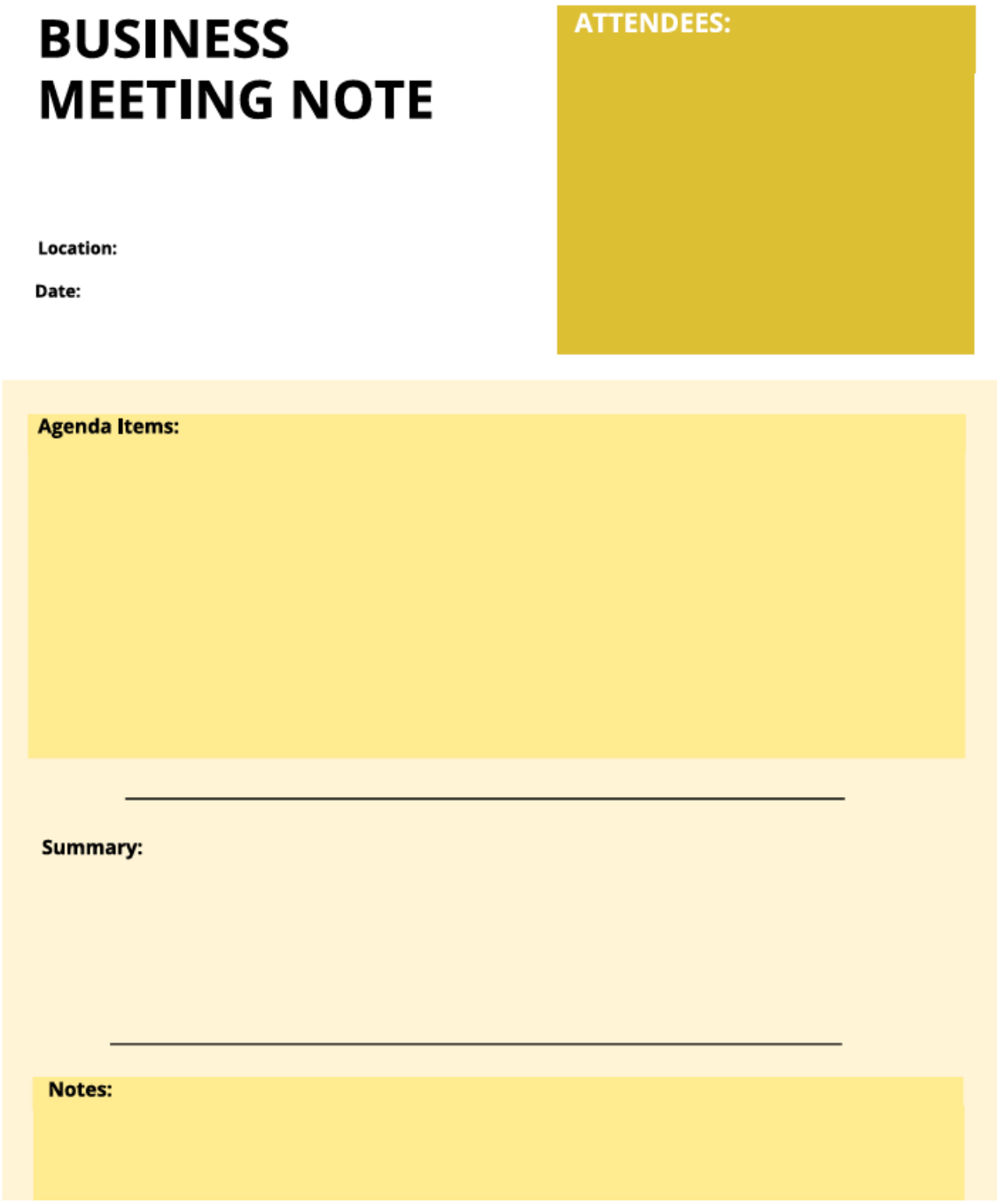



Figure 1: Template created with PDFGeneratorAPI's editor

This template serves as a blueprint, ensuring consistency and coherence across all generated meeting notes documents, regardless of the specific details of each meeting.

## 3.4 Evaluation Metrics

To assess the effectiveness and quality of the AI-generated business meeting notes, I employed a set of established evaluation metrics from the field of **Natural Language Processing** (*NLP*). These metrics, namely **BLEU**, **ROUGE**, **METEOR**, and **Cosine Similarity**, offer quantitative insights into various aspects of the generated text, from its similarity to reference texts to its linguistic quality and relevance.

- **BLEU** (*Bilingual Evaluation Understudy*) is a precision-based metric that compares n-grams of the generated text with n-grams of a reference text, calculating the percentage of matches. It's commonly used in machine translation but is also applicable to any task involving text generation.
  BLEU focuses more on precision than recall, which can sometimes undervalue the importance of capturing all relevant information. It may not fully capture the semantic accuracy and fluency of the generated text.

- **METEOR** (*Metric for Evaluation of Translation with Explicit Ordering*) initially developed for machine translation evaluation, compares the generated text with reference texts at both the precision and recall levels, incorporating synonymy and stemming. It also adjusts for sentence structure and fluency.

- **ROUGE** (Recall-Oriented Understudy for Gisting Evaluation) is a set of metrics designed to evaluate the quality of summaries by comparing them to one or more reference summaries. ROUGE-N measures the overlap of n-grams between the generated text and the reference, while ROUGE-L considers the longest common subsequences, emphasizing the importance of longer, meaningful phrases.

- **Cosine Similarity** measures the cosine of the angle between two non-zero vectors in a multi-dimensional space, in this case, the term frequency vectors of the generated text and the reference text. It quantifies the similarity between two documents irrespective of their size, making it useful for comparing the semantic similarity of the meeting notes to the actual discussions.

While these metrics provide valuable quantitative assessments, they also have their limitations, particularly in capturing the nuanced and context-dependent nature of business meetings.
The quality of meeting notes is not solely defined by their similarity to a reference but also by their ability to concisely capture the essence, decisions, and action items of a meeting in a manner that's accessible and useful to stakeholders.
Moreover, the absence of standardized reference summaries for business meetings poses a challenge, as these metrics rely on comparisons to a "gold standard" that may not exist or may vary significantly across different meetings.
This variability underscores the necessity of **human evaluation** in the assessment process.

# 4 Workflow

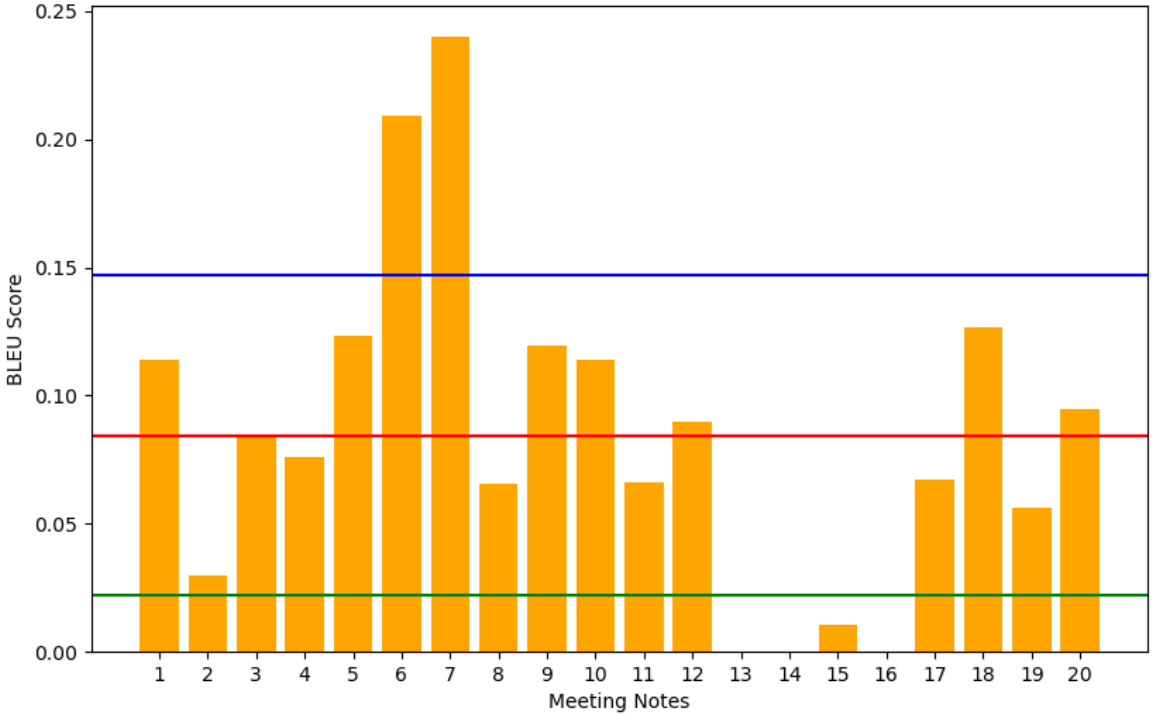
The dataset required minimal preprocessing, containing only *mp3* data, it was ready for the transcription.
The initial phase of transcription was adeptly handled by **Speechmatics**, whose advanced capabilities in speech recognition and speaker diarization were pivotal. This technology not only transcribed the spoken word with high accuracy but also distinguished between the various speakers, a critical feature given the collaborative nature of meetings.
Then comes the text generation phase using **OpenAI**'s GPT models, the challenge of dealing with the considerable length of these meetings came to the forefront.
Meetings, often extensive and multifaceted, needed to be broken down into digestible sections to ensure the coherence and manageability of the generated notes.
For the same reason GPT-4 was preffered insthead GPT-3.5, having larger input and output size capabilities.
One of the key challenges using these models is **prompt engineering**:
The prompts were carefully crafted to encapsulate the essence of the discussions, ensuring that the generated notes contained all the fields created in the PDF Template, in a easy manipulation format.
**PDFGenerator-API** allowed to design a versatile template that accommodated the segmented nature of the meeting notes, with fields for Title, Location, Date, Attendees, Agenda Items Discussed, Summary, and Notes to fill with the content of the same fields generated by the AI.
Evaluating the quality of the generated meeting notes necessitated a blend of automated metrics like **BLEU**, **ROUGE**, **METEOR**, and **Cosine Similarity**, alongside **human judgment**.

# 5 Results obtained

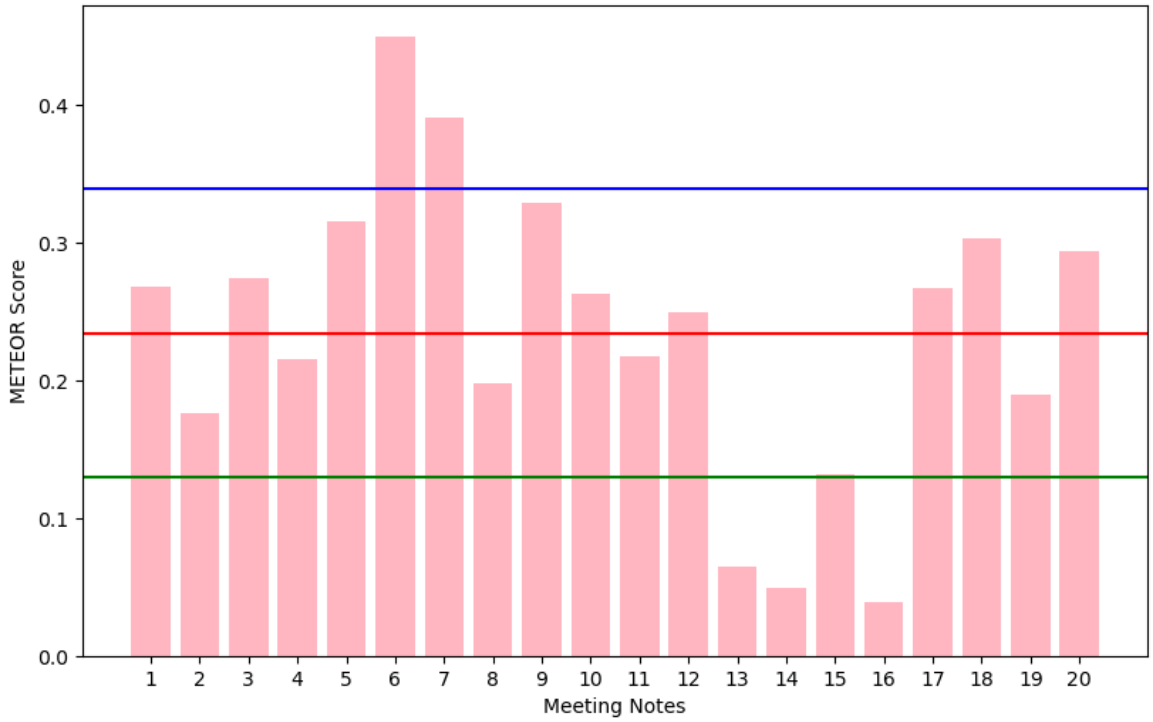
Following the method explained above, I generated Meeting Notes' PDFs for 20 meetings of the dataset.
Subsequently, using the metrics mentioned, I conducted an evaluation of their quality.

## 5.1 NLP Metrics

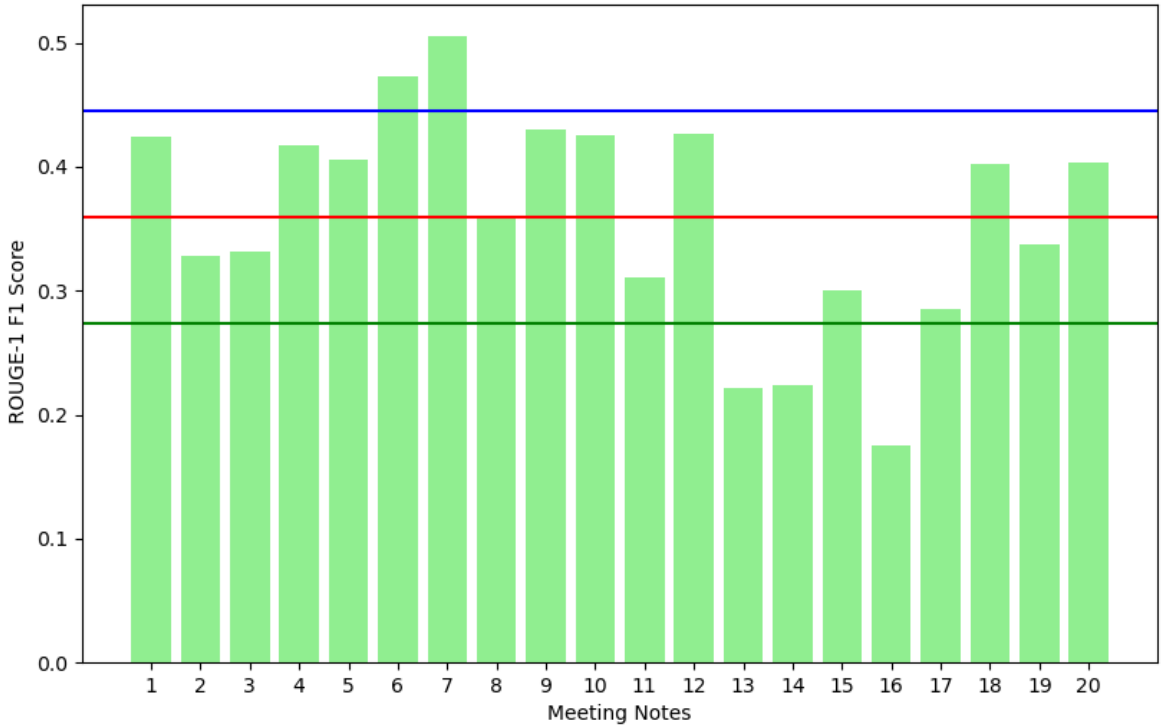
### BLEAU



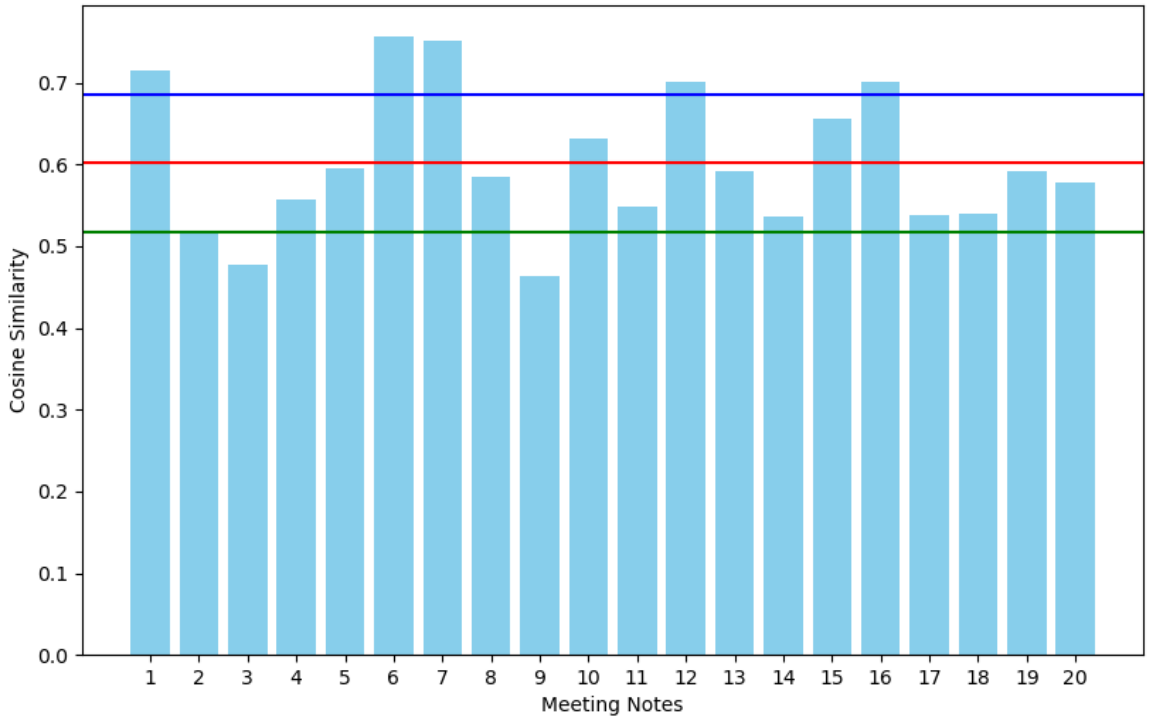

### METEOR



## ROGUE




## COSINE SIMILARITY



The **METEOR** scores vary, with some meeting notes achieving scores closer to 0.4, which is reasonably high for this metric, indicating a good level of correlation with reference texts. However, several notes scored below 0.2, suggesting room for improvement in capturing the semantic meaning and syntactical structure of the reference texts.

The **BLEU** scores show significant variation and are generally lower, with many notes falling under the 0.1 threshold. Since BLEU primarily measures precision in the overlap of n-grams, the lower scores may indicate that the exact phrasing and terms from the reference texts are not being replicated in the generated notes. This isn't necessarily negative in summarization tasks, as the goal is to convey the same meaning rather than replicate the exact wording.

The **ROUGE-L** scores are more consistent, with most meeting notes hovering around the 0.4 mark and some even reaching just above 0.5. This demonstrates a decent overlap in terms of the longest common subsequences between the generated text and the references, pointing to a moderate level of summary quality.

The **Cosine Similarity** results are quite promising, with many notes scoring above 0.5 and some nearing 0.7. These scores suggest a high degree of semantic similarity between the generated notes and the reference texts, which is encouraging as it implies that the generated notes are contextually similar to the reference material.

The results present a complex picture.
The generally high Cosine Similarity scores indicate that the notes are semantically on track, while the ROUGE-L scores suggest that the structure and flow of the original meetings are somewhat preserved.
However, the METEOR and BLEU scores hint at discrepancies in linguistic and semantic precision.

Given these mixed results, **human evaluation** becomes even more critical. Quantitative metrics can guide and inform the summarization process, but they may not fully capture user satisfaction, readability, or actionability. The generated notes must not only be technically accurate but also practically useful for the stakeholders. Therefore, combining these metrics with thorough human review would provide a more comprehensive assessment of the notes' quality and effectiveness.

## 5.2 Human Evaluation

For the human evaluation were established clear criteria for the evaluation based on the goals of the meeting notes:

- **Relevance**: Do the notes capture the key points and decisions made during the meeting?
- **Clarity**: Are the notes understandable and free of ambiguity?
- **Conciseness**: Do the notes convey the information succinctly without unnecessary detail?
- **Flow and Organization**: Is the information presented in a logical order that reflects the flow of the meeting?
- **Speaker Attribution**: Are the speakers correctly identified, and are their contributions accurately represented?

The evaluation was made by 3 people, analyzing 20 generated meeting notes and giving giving a score from 1 to 10 to every criteria.
The mean results were the followings:

| **Relevance** | **Clarity** | **Conciseness** | **Flow and Organization** | **Speaker Attribution** |
|---|---|---|---|---|
| 8.13 | 8.56 | 7.42 | 8.71 | 7.11 |

These scores provide valuable feedback on the system's current capabilities and highlight areas for improvement. Specifically, focusing on enhancing the conciseness of the notes and improving speaker attribution would likely yield a better product.

Only 3 people evaluating is clearly a simplified version of how this should be done, but it's enough to have a initial human feedback of all the process.
The evaluators, while external to the actual meetings, provided an outsider's perspective on the clarity and organization of the content. Their impartiality can be particularly valuable in assessing the general quality of the notes, on the other side meeting participants can offer a unique perspective on the accuracy of the content, especially in terms of relevance and speaker attribution,their familiarity with the subject matter

and the context of the conversation could lead to a more nuanced evaluation of the notes.
However, it's also important to consider that meeting participants might have subjective biases based on their personal experiences during the meeting.
In future evaluations, it would be beneficial to include both meeting participants and non-participants, as this would combine the depth of insight from those with direct experience of the meetings with the objectivity of external reviewers. This approach would likely yield the most comprehensive understanding of the AI-generated notes' quality and utility.

# 6 Conclusions

This investigative journey began with a pivotal question: *Can generative AI create comprehensive and informative Business Meeting Notes using only the recording of the meeting?* It becomes clear that the answer to this question is not binary but rather a spectrum of potential and progress.
The project's results, while promising, reveal the complexities inherent in the automation of meeting note generation.
The use of *Speechmatics* for speech-to-text conversion laid a strong foundation, capturing the diverse voices and discussions within the meetings with remarkable accuracy.
The subsequent employment of OpenAI's GPT-3.5 and GPT-4 models for text generation has showcased the impressive capabilities of AI in creating lengthy conversations into concise summaries and with the assistance of PDFGeneratorAPI, these summaries were successfully transformed into structured, professional-looking documents.

However, the limitations of current technology have also come to light.
The **token limits** in AI models, for instance, pose a challenge when dealing with extensive meeting recordings, necessitating the division of transcripts into segments that can fit within the models' processing constraints. This segmentation, while a necessary adaptation, has implications for the continuity and contextual understanding of the AI.
The evaluations conducted, albeit with a small sample of human reviewers, have indicated that the AI-generated notes are relevant, clear, and well-organized.

Nevertheless, the variability in scores, particularly in conciseness and speaker attribution, suggests that there is still room for improvement.

Looking ahead, the future is bright for the application of generative AI in creating business meeting notes.
As AI technology continues to evolve, with advancements in token limits and enhanced contextual understanding, we can anticipate more sophisticated systems capable of handling entire meetings in a single pass. The trajectory of this project indicates that, with further development and more robust datasets, AI could soon become an invaluable tool for businesses, aiding in the efficient and accurate documentation of meetings.

In answering the initial question, I find myself with a qualified: *yes, generative AI can indeed create comprehensive and informative business meeting notes, yet the technology is still maturing.* The path forward involves not only technological refinement but also a continued partnership between AI and human oversight, ensuring that the final product serves its intended purpose to the fullest.
The journey thus far has been one of discovery and learning, and it paves the way for a future where AI can fully realize its potential in service of business communication.

# 7 Appendix

## 7.1 GitHub Repository

The code used is available in the GitHub repository:
*https://github.com/MarcoBolo001/Meeting-Notes-Generation*
To replicate the code it's necessary to get API Key from OpenAI, SpeecMatics and PDFGeneratorAPI.

## 7.2 Meeting Notes Examples

The result of the process described in this document is shown in the following examples of meeting notes, other examples can be found in the GitHub directory, at the address: */output/pdf/*

# BUSINESS MEETING NOTE

**budget and appropriations council meeting**

**Location: long beach convention center**

**Date: 08/09/2022**

## ATTENDEES:

- speaker 0 (councilwoman sanchez, councilwoman allen, councilman spooner)
- speaker 1 (councilman alston)
- speaker 2 (councilwoman price, councilwoman sarah)
- speaker 3 (vice mayor richardson)
- speaker 4
- speaker 5
- councilman super now
- councilman austin
- councilman allen
- vice mayor richardson and council member muranga

## Agenda Items:

- item 11: increase in general fund for the city manager department to support the friends of the long beach public library.

- item 12: increase in special advertising and promotion fund for the end of summer celebration support.

- item 13: general fund increase for a donation to the jazz angels.

- item 14: general fund increase for a donation to the little lion foundation.

- item 16: general fund increase for contributions to various organizations including casa korero and friends of long beach public library.

- item 28: general fund increase for a donation to ron palmer summit.

---

## Summary:

the budget and appropriations council meeting held at the long beach convention center on august 9th, 2022, addressed several key items related to appropriations and donations from the city's budget. the council discussed increases in appropriations for various community projects and organizations, indicating a strong commitment to supporting local initiatives that benefit the community. notable items included substantial contributions to public library friends, the end of summer celebration, and various nonprofits like the jazz angels and the little lion foundation. furthermore, the meeting progressed to consider reports related to the city's financial contributions to local control programs, sales tax agreements, and several important city projects including enhancements to the local aquarium and the enactment of a healthcare worker ordinance. the meeting emphasized the council's role in ensuring that budget allocations reflect the community's needs and priorities, aiming for a comprehensive and swift approach to deliberations. a unanimous vote on the discussed appropriations showcased a cohesive and collaborative effort among council members to support community welfare.

---

## Notes:

the meeting started with speaker 4 acknowledging the agenda and proposing discussions on specific items related to fund appropriations for various content functions. significant contributions towards community services and projects were highlighted, reflecting the council's continual support towards enhancing local amenities and services. the discussion also included procedural aspects such as public comments and the voting process, ensuring transparency and community involvement in the decision-making. the unanimous 9-0 vote on the motions underlines the council's unified approach towards budget allocations and community support. the session concluded with reminders of the packed agenda focusing on budget hearings, emphasizing that the council intended to expedite the process while covering all necessary discussions regarding city budgeting priorities, especially focusing on critical areas like fire, police, and parks, through simultaneous presentations and comprehensive question sessions to follow.

# BUSINESS MEETING NOTE

**discussion on council bill 118764: technology matching fund program**

**Location: seattle city council**

**Date: 09/06/2016**

## ATTENDEES:

- speaker 1
- speaker 2
- speaker 3
- speaker 0: herbold
- johnson
- o'brien
- sergeant bagshaw
- burgess
- president harrell

## Agenda Items:

- overview of council bill 118764 and its objectives
- the allocation of $320,000 from seattle information technology to ten community providers
- the goal to increase technology access, literacy, and digital equity opportunities
- details on the maximum disbursement allowed per project
- the impact of these grants in supporting community technological needs
- specific examples of projects funded by the grants
- voting process and outcomes

---

## Summary:

the meeting focused on discussing council bill 118764, an ordinance related to the technology matching fund program, which aims at enhancing technological access and literacy among seattle residents. the bill proposed allocating $320,000 from seattle information technology to ten community providers. these allocations are intended to foster digital equity by increasing technology access, literacy, and by providing significant investments in community enterprises. it was highlighted that each project could receive a maximum of $50,000, with the grants being a crucial part of the city's community investment strategy. the grants aim to empower individuals across all age groups and backgrounds with essential skills for navigating the 21st-century job market. the year's projects are anticipated to offer over $470,000 in community matching resources, reaching over 2,500 residents. special mentions included initiatives by el centro de la raza, native girls code, and the urban league of metropolitan seattle, illustrating the diversity and impact of the funded projects. the discussion also covered the voting process on the bill, which saw a favorable outcome, leading to its passage. the meeting concluded with anticipation of positive impacts from the funded projects on communities, especially on immigrants, refugees, seniors, and individuals with disabilities.

---

## Notes:

- the bill's passage marks a significant investment in digital equity, aiming to address the technology gap among seattle's diverse population.
- over 2,500 residents, including 580 immigrants and refugees, 1,240 seniors, and 1,100 people with disabilities, are expected to benefit from these initiatives.
- the committee's emphasis on community matching funds exemplifies a collaborative approach to enhancing the city's technological infrastructure and accessibility.
- the projects funded through this initiative serve as a model of how technology grants can be leveraged to support educational and professional development in underserved communities.
- the successful voting outcome (8 in favor, 0 opposed) reflects a unanimous council support for strengthening the city's commitment to digital inclusion and equity.