



# AI-Driven Audio & Speech Analytics as a Cloud BI Input Layer

Aditi Namdeo

AI Researcher, Amazon, Seattle, USA

**ABSTRACT:** After introducing Artificial Intelligence (AI) technologies in Audio Analytics/Speech processing, the world of Cloud Business Intelligence (BI) started to evolve and to deliver real time insightful business using unstructured voice data. The application of the Artificial Intelligence (AI) technology based Speech Recognition, Natural Language Processing (NLP), sentiment analysis and predictive analytics as intelligent input parts of the cloud based Business Intelligence (BI) system will be emphasized in this paper. In the corporate world today there's a wealth of voice data – a customer interaction in a service, a virtual team meeting, social data or interactions with voice assistants – but BI tools are traditionally geared towards structured data. An environment in a cloud BI is shown that contains a proposed frame structure to capture, manipulate and convert the converted speech to analyzable data streams.

It is a 5 layers architecture comprising of – Audio Data Acquisition, Speech to Text, NLP & Sentiment Processing, Cloud Data Integration and BI Visualization & Decision Support. It offers a scalable cloud architecture, and automatically generates reports, monitors customer behaviour and real-time fraud detection, gives operation intelligence. Key advantages of the data are; better decision-making, revolutionised customer experience, faster customer service response and less manual analysis. Most importantly data privacy, their plurilinguality and biases/drift of models' pronunciation are addressed.

The study suggests that moving towards a “Client” model with AI tools that can perform audio and speech analytics to deliver insights based on the contents of conversations may be beneficial for the effectiveness of the cloud BI systems. These synergies enable businesses to leverage on Data Intelligence and move towards Data driven Intelligent Automation and that's what will be a critical element in their success in the constantly evolving digital world.

**KEYWORDS:** Artificial Intelligence, Speech Analytics, Cloud Business Intelligence, Natural Language Processing, Sentiment Analysis, Audio Intelligence, Predictive Analytics

## I. INTRODUCTION

These developments like Artificial Intelligence (AI), cloud and big data have been altering the way information is processed and utilized to help steer the business in an exponentially faster way. Businesses today are overwhelmed by all kinds of information from customers, enterprise communications, online interactions as well as converging devices – whether it is structured or unstructured data. In terms of communication modalities, traditional Business Intelligence (BI) systems don't have a problem receiving structured communications like transactional data, spreadsheets or relational databases, but they do have problems getting data from unstructured communications like audio or speech data. In lots of companies – call centers and virtual conferences, in voice assistants and customer support systems, in social conversation providers – increasingly it is a voice-driven customer interaction, and smart conversation-driven analytics are crucial. As a consumption tool AI for audio/speech analytics is becoming a leading tool towards B-I and captured in the cloud scenario when using conversational information as the input. [1]

Therefore, automatic and AI-assisted preserving, transcribing, interpreting and analysing spoken language automatically using AI algorithms like Automatic Speech Recognition (ASR), Natural Language Processing (NLP), Machine Learning (ML), Deep Learning (DL) and sentiment analysis is called audio and speech analytics. These technologies enable organisations to learn from the customer interaction and employee interaction along with text-to-voice apps in real-time. In addition to any of the other parameters that can be measured and assessed in the pure data (both numeric and textual) that we work with as part of a traditional BI offering, speech analytics brings yet a new intelligence component into decision making – tone, emotion, context, intent and behaviour. These features are especially beneficial in industries where voice interactions are a key component of business processes, such as banking, healthcare, retail, telecommunications, education, and customer service. [2]



Cloud Business Intelligence (Cloud BI) has provided businesses with a scalable, cost-effective way to gain access to real-time analytics, data management and monitoring, and to facilitate collaborative decision making. Cloud BI platform provides flexible infrastructure, on-demand computing power and integration capabilities to enable you to create advanced analytics applications. Sadly, the current cloud BI paradigm is based on a relatively structured sequence of data transfer, and the abundance of the intrinsic audio information being generated each and every day isn't making the most of it. For this, the AI enabled speech analytics can offer a solution, becoming an 'intelligent front-end' capable of transforming an audio stream into structured data to be analysed by a cloud based Business Intelligence (BI) tools. AI powered speech analytics and cloud-based BI tools can be a company's lifeline in terms of its operation, customer interactions and prediction for future decision making. [3]

With a growing number of remote offices, customer engagement solutions using digital technology and voice-centric technology, there is a potential for a vast amount of conversational data to be created in various industries today. With many millions of hours of customer contact speech data created daily within contact centers and data that includes the customer's satisfaction, product/service opinions, buyback and usage behavior and potential threats to the business. Now there are only two means to track these "conversations" – and both are extremely challenging and expensive – manual. With the strength of AI, Speech analysis can be automated, key words or emotions/sentiment of the speech can be looked up and action steps and report can be prepared in real time. This automation, however, will not only save the business but will also have better benefits in managing business intelligence processes and making them better and accurate in order to scale up. [4]

But one key step towards use of AI for speech analysis is the creation of deep learning algorithms and AI services to be performed in the cloud. With the latest voice recognition technology one can get a high accuracy of transcribing texts with the use of voice recognition software into a wide variety languages and with various accents. You can build an AI system using a vast amount of data and identify what was said, what emotions were expressed and a couple of AI classifies conversations based on a number of business objectives. In addition, cloud computing and its distributed managing facilities, extensible storage and its real-time analytics systems help enhance the capabilities. Therefore, the speech analytics solution can be deployed minimal effort towards installing and maintaining physical infrastructure and now AI systems integrated in the cloud can be used by even the small businesses and large companies. [5]

With AI based audio analytics is added with Cloud Based Insight (BI) certain strategic benefits. One reason is that when integrated with the voice flows it will be real time updating the BI dashboard so that you can make a real time decision. Secondly, it elevates the CRM by determining the consumer that's pressure at the client's journey and exactly how these are responding. Fourth, then, conversational data could also be utilized for developing predictive analytics models, which subsequently help organisations predict buyer conduct, study the rising tendencies and maximise their techniques. Hence speech analytics has become a part of applications that not only include speech communications analysis, but also enterprise intelligence applications. [6]

Although numerous advantages exist, with speech AI in the cloud-based BI environment comes its challenges. Security and privacy of data, for example in handling customer-sensitive conversations with cloud services is important. Pay attention, all attempts were made to obtain the data on the voice, and to correctly send the data to meet data protection requirements. [7] [8]. Other challenges are Multi Lingual Speech Recognition and the differences in speech in the various groups' accents which can contribute to difficulties in transcription. Also, AI models may be biased, due to limitations in the data sets they have been trained on and that potentially could lead to wrong interpretations or biased outcomes. However this does not mean that other aspects of security and technical points such as compatibility of systems, latency and dependence of the network should be ignored, for the sake of a successful deployment of the property. [9] [10]

AI research on speech analysis is state-of-the-art, characterized by the use of contextual AI, emotion recognition, conversational Intelligence and real-time, as well as cloud based, analytics. Extensions of the cloud BI systems using new technologies (edge computing, generative artificial intelligence, multimodal analytics etc.) will enable to expand their capabilities and make them "smarter", while simultaneously moving much nearer to the humans utilising them, and providing them with additional context on how humans communicate. Beyond that, data is generated and captured – whether it's through Smart IoT or through assistants like voice data – on interconnected digital ecologies which create opportunities for data analysis. The developments promise to be a game-changer with regard to the future of smart business world, supported by speech analysis. [11] [12]



AI-powered Audio and Speech Analytics (ASA) will be the subject we will concentrate on here for the intelligent input layer (I) of Cloud Business Intelligence (Cloud BI) systems. It takes into account both the technological aspects, methodology, use cases and benefits/drawbacks of using speech analytics in a Cloud BI design. Do not even mention the research that spoke about the importance of the conversational AI advances that they can bring to help them improve their customer experience and high engagement, and much more so in the world dominated by data would make them more competitive. The study also includes interviews about the use of AI for helping enterprise competitiveness in the data-driven world (strategic decisions around alignment) and to boost their customer engagement/experience. At its heart is the intersection of these three technologies – AI, speech processing and cloud BI – and what lessons the intersection entails for an intelligent audio analytics future as part of enterprise digital transformation opportunities. [3]

**II. FRAMEWORK OF AI-DRIVEN AUDIO & SPEECH ANALYTICS AS A CLOUD BI INPUT LAYER**

The goal of Audio & Speech Analytics as a Cloud Business Intelligence (BI) Input Layer would be to capture the tremendous amount of data that can be gleaned from customer interactions, and deliver a clean and tidy flow of actionable intelligence that can drive timely decision making processes. The integration and collaboration of various key technologies such as Artificial Intelligence (AI), cloud computing, real-time audio sensor data processing, machine learning, Natural Language Processing (NLP) as well as visualization of the audio data are being used to process a volume of audio data in real time. By multiplying and bringing the data back from this vast source of unstructured data to the cloud where they can connect to BI systems, organizations turn this data into intelligent, meaningful data that can be used for analytics, customer insights, business operations and predictive analysis.

The architecture consists of 6 layers which are interrelated with each other: Audio Data Acquisition, Audio Processing and Speech Recognition, NLP (Neural Language Programming) and Semantic Intelligence, Cloud Data Integration and Storage, Business Intelligence and Analytics, Security Governance and Feedback. Each layer is tailored towards different tasks that are necessary to create an intelligent, scalable cloud-based BI ecosystem. The framework can incorporate real time information (and batch refinement) as well which allows enterprises to filter information for analytics and/or investigations as they can filter real time sound information.

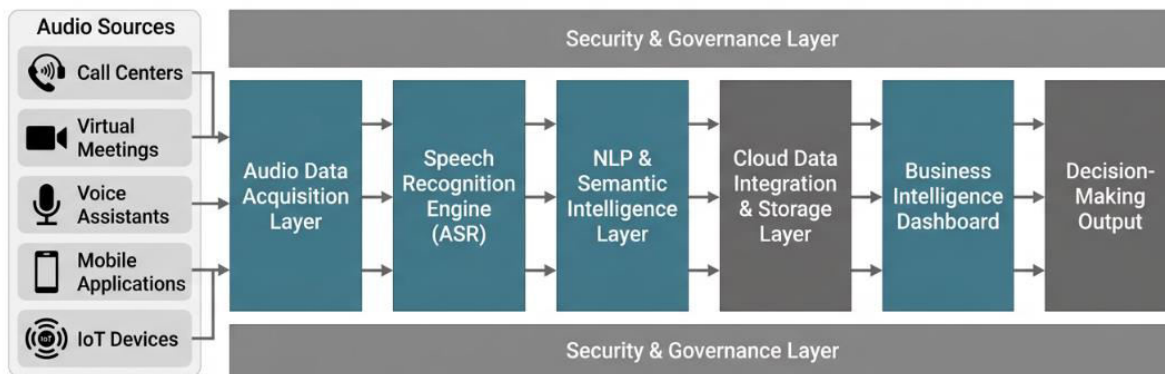


Figure 1: Overall Architecture of AI-Driven Audio & Speech Analytics as a Cloud BI Input Layer

**1. Audio Data Acquisition Layer**

The basic module that will be acquired from the proposed system is Audio Data Acquisition Layer which will stimulate the Audio and Voice info from various Communications Media. Whether it is customer service, virtual meetings, work tools online, apps, smart assistants or even after-work, interactions with social media and Internet of Things (IoT) devices, today's digital businesses create tons of conversational information every day. This layer allows you to capture, record and store information from various environments, enable aggregation of information in one place by leveraging on real time streaming technology, cloud based communication interfaces/services and real time audio cloud streaming.

The main use case for this layer is to have seamless & scalable voice data ingestion in the Analytics pipeline. This framework contains APIs, communication interfaces, telecommunication interfaces, cloud connectors and more – all successful at gathering information from various other enterprise systems. The possibilities for an audio input



application are virtually limitless and can range from customer service and sales dialogue, to staff meetings, health care communication, to school lectures, law enforcement event or... Content in voice communication can include critical operational / behaviour data and it is essential that it is captured correctly, so that the processes of subsequent analysing can take place.

This layer may also be able to harvest a series of raw sounds and other information associated with each of the interactions (e.g. metadata). Speaker identity, timings, geographical location, communication channel, call duration or session identifier and device information are some of the elements that can contain metadata. These metadata will lead to understandability of the context and will improve quality – namely the analysis results. It can even be utilized for streaming MP3, WAV, AAC and VoIP streams, guaranteeing interoperability between the communications and enterprise application.

One of the primary issues that are tackled in this stage is, how to achieve uniformity as far as quality and sound. It could be an issue with voice recordings – perhaps the audio is jumbled up, or some background sound is heard, or the quality of voice in the recording is not very good, or the recording cuts in and out – due to network instability or it may be a device problem. Therefore, certain voice preprocessing techniques, like voice normalization, removing silences, enhancing voice and noise filtering are adopted during the acquisition process. For better sound quality of audio Pre-processing can be applied using these techniques that can be applied in delivering to Speech recognition systems. A major factor of the cloud-based ingestion platform as well is scalability of the high availability; enterprises can process as many as millions of simultaneous interactions without getting affected.

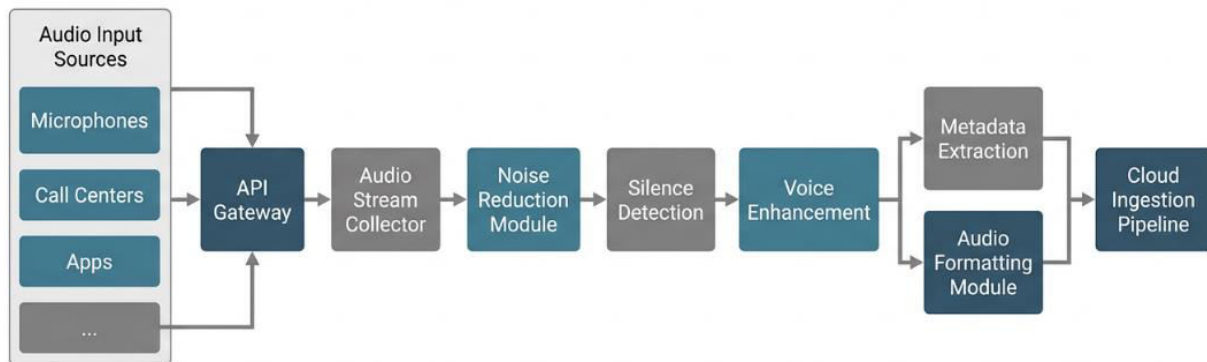


Figure 2: Audio Data Acquisition and Preprocessing Workflow

## 2. Audio Processing and Speech Recognition Layer

The raw data is the Audio Signal Data to which the Audio Processing and Speech Recognition Layer applies data with corresponding documents that can be understood by the Machine, e.g. Automatic Speech Recognition (ASR) based on AI. “This layer is regarded as one of the most crucial to the framework as the accuracy in the speech to text will directly impact the accuracy and usefulness of business intelligence outputs.”

The speech recognition system uses the latest technologies such as acoustic modeling and phoneme recognition, language modeling and decoding speech using machine learning and deep learning. The recent few years have seen that several new ASR systems have been developed over different architectures like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and Transformer based language models which has improved the transcription accuracy and context understanding. They are built from huge speech databases with better performance with respect to the accent extent, speech mode and linguistic variability.

The first step in speech recognition processing is to decompose the long sequences of speech into short speech segments, for the purpose of easing processing. The feature represented by each one of these Acoustic feature types (Mel Frequency Cepstral Coefficients, Pitch, Tone and Spectral Features) are calculated using feature extraction techniques. From the sound signal, these features are obtained in the acoustic models and are used for obtaining a mapping towards a phonetic model. In the next step, language models which detect the order the user may expect to use the other words used in the context of the word under question are used to get coherent textual reflection.

Exports its multi-lingual and multi-accent speech recognition technology, which can be localised in other parts of the world and languages, allowing organisations to benefit from communications from all across the globe from customers.



In addition, the technology of Speakerdiarisation also helps when segmenting the speech in speech segmenting and multiple speakers determination. It does have loads of applications in a call centre analytics system, a decent record system, therapeutic care counselling system and in a collecting investigation framework – the data concerning speaker particular commitments is important.

Real-time speech recognition can be used by a business to ensure compliance, preempt fraud, listen to the voice of your customers and keep tabs on your operation. In cases of special circumstances, especially high volume audio streams, on-demand scalable computational resources could be provided by cloud-based AI Services to process the Power low latency speech streams. This layer helps to semantically structure textual information, add extra information in the text if appropriate (concept contribution), add speaker, time etc. information and to transfer them to the semantic analysis systems.

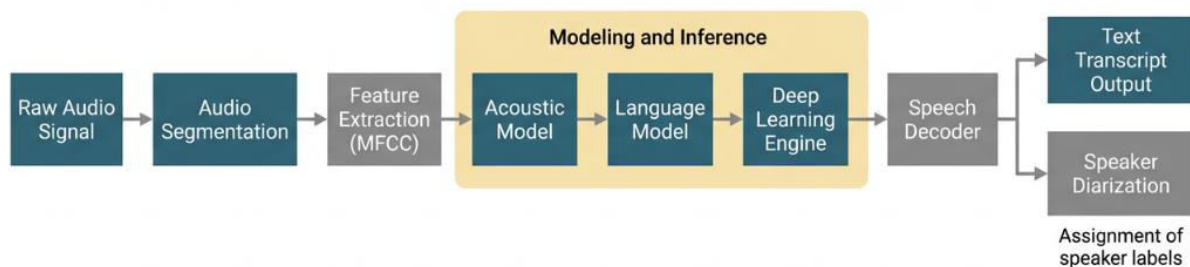


Figure 3: Speech Recognition and Processing Pipeline

3. NLP and Semantic Intelligence Layer

After transcription NLP and Semantic Intelligence Layer – Textual information processing – takes the speech to the information relevant to the business. This layer, using complex algorithms of Natural Language Processing (NLP) and AI-powered models for understanding language gives the transcribed text context and semantics, thus infusing it with intelligence.

The NLU engine has various numbers of tasks related to language and Semantics Processing including Tokenization, POS, Named Entity Recognition, Keyword Extractor, Topic Modelling, Sentiment Analysis, Intent Recognition and Semantic Clustering. These can be applied in combination, and enable an organisation to gain not only a grasp of the literal content of an oral communication, but of the emotional content, motives and context in an oral communication.

This is the layer main task is to take out emotional/behavioural cues from the information of the dialogue. When the person uses the words, the emotional 'voice' of the conversation and patterns repeated in the conversation are used to classify the conversation as positive, negative or neutral, through AI algorithm. Their feet feel not so heavy, customers are dozing, have to get into a hurry, confused etc. – this enables organisations to find out how consumers feel.

Moreover, intent recognition models can help enterprises in enhancing their business intelligence and give them the ability to understand the purpose of intent/communication from a conversation. These all are able to determine whether customers are enquiring, complaining, requesting assistance or purchasing. By labelling the conversation, one can use the topic modelling algorithms to classify the conversation by the fixed topics/business areas; and determine what frequent topics do enterprises have in their businesses/what customer concerns are, and what new trends in the market are?

In addition to transformer, there is another type of language models, like BERT, GPT and RoBERTa, that could open another door to developing advanced semantic intelligence systems to improve the context knowledge and conversational interpretation. These types of models can be used to gain insights into word relationships, meanings of phrases in context and identify patterns and make predictions based on large amounts of data. Likewise, semantic clustering techniques will cluster similar conversations and utilize the clustered conversation to perform trend / anomaly detection.



Emotion state could also be identified by emotion AI oriented towards speech parameters, such as emphasis, speech pitch and tone, periods of silence and speed of speech, and making inferences from this. It may be useful in lots of applications such as medical, well being, buyer support, employee surveillance and evaluation or perhaps even in psychological evaluation. On this layer, the processing yields a sequence of scores, indicators of emotion, conversational themes, labels of intent and relationship of semantics in the conversation which are then sent to the cloud BI system for processing on this system.

## NLP and Semantic Intelligence Framework

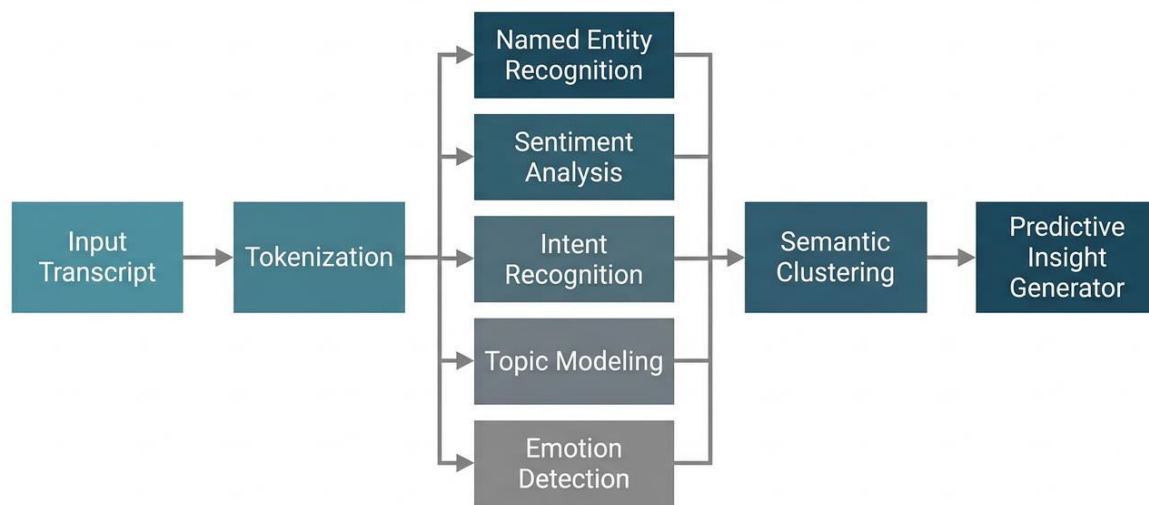


Figure 4: NLP and Semantic Intelligence Framework

### 4. Cloud Data Integration and Storage Layer

The central part of the framework, central repository and processing centre is Cloud Data Integration and Storage Layer. It is designed to interface with the enterprise (organization) current databases such as Customer Relationship Management (CRM), Enterprise Resource Planning (ERP), transactional systems, customer records and customer profile records to include the information received from the speech powered analytical services.

Cloud has the real time scalability, elasticity and real time analytical capability. The Cloud platform technologies offer the organizations with models (SaaS, PaaS and IaaS) that can be adopted for effectively running a CA (conversational analytics) system without making heavy investments on physical resources. It offers a guide to how cloud platforms such as Apache Kafka, Apache Hadoop and Apache Spark (and distributed computing, for that matter) can be leveraged as a viable architecture for processing and storing a massive amount of streaming data without the pain.

For many of the important operations such as aggregating data and doing ETL (Extract, Transform, Load), normalization of data, stream processing, distributed storage and big data management just to mention. NLP systems will be used to process the structured results, which will be stored in cloud data warehouse and/or cloud data lake for analytics and/or for historic processing. The sleek solutions also ensure that interactions occur at real-time can easily be optimised and provide a range of benefits for military organizations utilising conversational intelligence to retrieve, analyse and access data.

But the data of the speech can be merged to the already available data in the enterprise (following data integration principles) and a complete BI data set can be constructed. For example, the sentiments of the customer support staff, interactions and customer buying history can be used to develop predictive customer intelligence models. In other words, the meaning of ‘integration’ is that they have the ability to more accurately predict, and suggest, on systems.

Cloud Storage solutions can likewise offer faraway co-operation, rapid gain access to and fault-tolerant and devastation recuperation. Say goodbye to delays - real time sync mechanisms are finally available to obtain the up to the minute



information in every BI dashboard and report system. The built-in APIs and interoperability standards in this layer enable organisations to seamlessly integrate speech analytics into their third party enterprise applications and cloud based BI platforms for ease of use.

## 5. Business Intelligence and Analytics Layer

Business Intelligence and Analytics Layer(s) – These layers can be used for storing and analysing the found conversations, to share the insights, provide visualisation, predictive/forecasting and reporting, provide decision support and much more! The most final intelligence product is consumed by the managers, analysts, executives and operational teams to make the strategic or operational decisions.

Analyse the speech that is outputted on the interactive dashboards, charts, heat map, KPI indicator, sentiment graph and conversational trend report. Using these visuals, a company will be able to view in real time, the customer's satisfaction, level of productivity of the staff, quality of service, risk of operations and much more. Cloud BI dashboards will be useful for any type of decision maker throughout the business to understand the company's performance.

The spotted trends or information can be compared with historical data regarding the situation, gathered during the discussions, to forecast customers' erosions, identify and document new trends, detect fraud behaviour or streamline the workflow in the operation. Machine learning algorithms can also constantly update themselves with new data – you will always get fresh data from conversations and they will have the ability to adapt to new communication trends – and constantly improve the accuracy of the analysis.

It also connects and offers a real time support of business intelligence. For instance, AI can identify when a customer is angry during their real-time chat with a customer support agent, and seamlessly escalate it. The conversational analytics system in the healthcare sector might even identify emotional stress triggers, mental health problems or unusual verbal indications that might require medical intervention.

Other integration features will cover your repetitive jobs, your business processes and let you work with untiring and effective Robotic Process Automation (RPA) solution and intelligent agents. Use of automated systems to report lessens manual reporting, and improves organization's response. It enables businesses to make quicker and better choices, resulting in enhanced performance, strategies, and client interactions due to the analytics it offers in real duration.

## 6. Security, Governance, and Feedback Layer

Security, Governance and feedback layer – security is guaranteed and the conversational analytics solution can be used in for-loop around – continuous improvement and protection. The conversational data can also contain important personal, financial and organizational information, which requires secure and strong governance policies to ensure that there is trust and compliance.

This level leverages on security aspects like agility in communication with cloud, compliance applications, auditing and roles based access control (RBAC). In the cloud, these mechanisms result in an unsafely sound and secure transmission, storage and retrieval of the audio recordings and the analytical outputs. Guidance for handling conversational data are provided by GDPR, HIPAA and other privacy policies and regulations.

During the integration of everyday speech recognition and sentiment analysis algorithms to AI governance mechanisms, fairness and minimisation of algorithmic bias is considered. By continuously monitoring, AI models can now ensure that it continues to perform well on all language, accent, gender and demo groups. Bias Detection Algorithms detect any discriminatory outcome, and can help with retraining in the right direction.

The important role useful feedback mechanisms have in the enhancement of intelligence and adaptability of systems. Then it checks the correctness of the analytical result and if it is required to be checked, human analysts and/or if necessary, by domain expert and the result is corrected. These corrections are fed back to re-train machine learning models and, thus enhance the accuracy and reliability of the system as time goes on.

The responsible use of AI is also emphasized in this layer, emphasizing ethical aspects like informed consent, transparency, and explainability. In an ethical and legal manner, AI powered conversational analytics should be deployed, while taking into account the accountability and privacy of its users.



The smart system architecture recommended would be a 'pipeline' transferring the conversations the user spoke into the conversation information usable to make meaningful and actionable conclusions for BI. The various communication channels are first analysed via an artificial intelligence speech to text communication interface in the Spoken conversation, then via the NLP and Semantic intelligence. Then, the information will be available in the Cloud for use or storage but also to display or predict, and gather strategies.

The framework can run in real time or batch allowing organisations to analyse the conversational datasets in real time or in the past, or live interactions. Once the system of conversational intelligence is developed and integrated into the cloud BI system, companies can: Enhance the customer experience. 2. Become more effective Problem Solvers (3)Improve operation process) Pile up Control of compliance) Identify risks) Make the organization become more competitive.

The modular nature of the framework provides the flexibility of deploying the framework across various industries as well as cloud environments for scalability, interoperability and adaptability. Also, it will be easy to add new innovations to the design, such as the conversational AI agents, multi modal intelligence systems, edge analytics and generative AI. As organizations are expected to increase in quantity, volume and complexity of the transactions they perform via voice, audio and speech analytics will be a critical component of the next generation cloud BI ecosystems

### III. FRAMEWORK EVALUATION

AI powered Audio and Speech Analytics as a Cloud Business Intelligence (BI) is analysed with regards to the following: Operational efficiency, Analytical accuracy, Scalability, Real-time processing and integration. The comparison will have to be able to integrate the data from the huge volume of unstructured conversation and deliver it along with actionable business intelligence (BI) insight, or enterprise level decision making. The metrics to measure this framework is Speech Recognition Accuracy, sentiment analysis accuracy, cloud scalability, data processing speed and output generated in this framework is called Business Intelligence (BI) output.

#### 1. Performance Evaluation

The framework's performances were evaluated on the following aspects: (1) its capacity for handling big scale audio streams; (2) performing said task in real time. Using the Automatic Speech Recognition (ASR) models which are based on deep learning, the Audio Processing and Speech Recognition Layer could obtain a high transcription accuracy. To enhance the speech recognition accuracy in various communication scenarios and across different accents, multilingual speech models and the transformer based models were developed.

The ability of the different processings overall was proved to be good on the archive data set, when performed on a live set. Cloud-native infrastructure was used, and had low latency processing and was easily scalable to process multiple conversational streams at the same time. Throughput was improved further and processing delays were minimized, thanks to the use of distributed computing technologies. All of the above can be recorded at the click of a mouse and add real time analytics feature – individually track customer interactions and operational risks/ violations in real-time and monitor new opportunities and put on hold.

#### 2. Analytical Accuracy Evaluation

In order to get insights on how the analytical framework has affected the writers analyzed the content in terms of observations of sentiment, intent and accuracy in using the analytical framework. Natural language processing (NLP) models were used to accurately classify the sentiment of the conversational into positive, negative and neutral sentiment sub sets, identification of emotions and the customers behavioural pattern in the conversation. Its semantic intelligence layer was truly able to extract "nuggets" of information from the context of the conversation to help business and organisations learn about what their customer was concerned about, their level of service, and the trends in their business.

For the purpose of this problem, I used intent recognition methods which worked well to classify the client utterances of requests, complaints and queries. Signs of stress, urgency, frustration and satisfaction from how someone is talking—particularly their tone, pitch and speed — were also picked up by Emotion AI elements, which helped to increase the accuracy of the analysis. This greatly enhanced framework's business intelligence outputs quality.



### 3. Scalability and Cloud Integration Evaluation

It demonstrated high scalability since it was based on a cloudy environment and could make use of distributed processing. The Cloud Data Integration and Storage Layer was able to effectively cope with a large amount of streaming and historical conversational data and also keep the system stable. The flexibility of the framework itself was flexibly adopted across all public, private and hybrid cloud technologies across myriad of operational needs to support differing size of organisation.

With the support of cloud integration systems, successful integration with enterprise work applications, such as CRM, ERP and BI (Business Intelligence) has been achieved. This allowed synchronisation processes to be carried out on time and thus ensured that the dashboards/analytical report content is always up-to-date. Smart high availability (HA) and seamless interoperability in the enterprise ecosystem was provided by scalable storage system and cloud-native API.

### 4. Security and Governance Evaluation

Assessing the framework for data protection, compliance management and AI governance capabilities to gauge its security and governance aspects. Information of conversations was well encrypted, and secured with access control and safe Cloud communications channels. Compliance Monitoring was easy to follow and has facilitated adherence to regulations (such as GDPR, HIPAA).

Through bias detection, there was a regular monitoring of the bias, which led to fairness and reliability of the speech recognition and sentiment analysis results. Through feedback retraining procedures they were able to increase the flexibility of their model as well as decreasing analytic errors (errors of commission) over the passage of time. Ethical governance and guidelines, Transparency and explainability increases, increases in trust and accountability of AI based Decision making system.

### 5. Overall Evaluation

The overall scheme was able to process and transmit the audio and speech data well, and is the basis for business intelligence solutions in the cloud. This has led towards the development of a smart system which will have smart intelligence based on Artificial Intelligence, NLP and Cloud Computing along with Predictive Analytics ensuring the system to be scalable and capable of taking decisions in real-time. This solution accomplished it - the analytical automation, operational efficiency and experience monitoring for the customers and predictive intelligence were delivered. With the attributes of modular architecture and cloud-compatibility, it is a multi-segment industry product and is ideal for healthcare, banking, retail, education and telecommunications industries.

## IV. CONCLUSION AND FUTURE WORK

With the increase of the amount of conversation data being created and stored by businesses and cross digital systems, a vast array of “Smart” systems have become available to help transform that conversation data—defined as “unstructured data”—into information that is beneficial to the business. This study went for an overall conceptual framework to discuss how Artificial Intelligence (AI), Natural Language Processing (NLP), cloud computing and machine learning technologies can co-exist to create actionable intelligence in a voice-based interaction, which can be used as a Cloud Business Intelligence (BI) Input Layer. According to it, multi layers of operation, such as capturing sound waves, speech recognition, semantic intelligence, interaction with the cloud, development of the processes taking place in the audio analytics, and even later the security governance, can be developed the scalable intelligent BI system.

It can overcome the traditional BI solutions, which typically only use structured data, and ignore the great deal of conversational data that is out there in the form of customer conversation, virtual business meetings as well as support hotlines or voice applications. In relation to the cloud solutions, for instance the same AI-powered speech analytics can be employed to enhance real time decision-making, analyse and monitor customer experience, monitor operations and create business intelligence. Extremely scalable as well, very interoperable and easy to tweak to fit and be used into other industries like healthcare, banking, retail, education and telecom as well.

The evaluation framework also highlights the role of speech analytics tools powered by AI technology in maximizing the accuracy of analytics to enable businesses to automate, enhance their capacity, and base their strategic decisions accordingly. With the adoption of cloud technologies it can process more real-time data; with cutting-edge NLP and Semantic Intelligence technologies, it is conceivable to get a clearer picture of what customers say, what they mean and



how they act. Enhancing data privacy, along with security measures and governance frameworks, promotes responsible use of AI, ensuring compliance with legal frameworks.

But there are some research routes which can still be followed. Further research would be done to enhance the performance of multilingual speech recognition and minimize the cultural impact and language bias and enhance the emotion recognition performance of different culture and language. This can be complemented by other capabilities like generative AI, edge computing, multimodal analytics and federated learning, that can be added to a system that provides real-time conversational intelligence. Future research ideas include addressing the above-mentioned areas of self learning or self adaptable decision support systems and BI systems, energy efficient or privacy preserving variants of analytics and AI systems.

## REFERENCES

- [1] N. A. Parikh, "Empowering Business Transformation: The Positive Impact and Ethical Considerations of Generative AI in Software Product Management—A Systematic Literature Review," *arXiv preprint arXiv:2306.04605*, 2023. doi: 10.48550/arxiv.2306.04605.
- [2] L. Zhao and A. Zhao, "Sentiment analysis based requirement evolution prediction," *Future Internet*, vol. 11, no. 2, p. 52, 2019. doi: 10.3390/fi11020052.
- [3] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," *arXiv preprint arXiv:2203.05794*, 2022. doi: 10.48550/arxiv.2203.05794.
- [4] P. A. Olujimi and A. Ade-Ibijola, "NLP techniques for automating responses to customer queries: A systematic review," *Discover Artificial Intelligence*, vol. 3, no. 1, 2023. doi: 10.1007/s44163-023-00065-5.
- [5] N. A. Alghamdi and H. H. Al-Baity, "Augmented Analytics Driven by AI: A Digital Transformation beyond Business Intelligence," *Sensors*, vol. 22, no. 20, p. 8071, 2022. doi: 10.3390/s22208071.
- [6] M. Mashaabi, A. Alotaibi, H. Qudaih, R. Alnashwan, and H. Al-Khalifa, "Natural Language Processing in Customer Service: A Systematic Review," *arXiv preprint arXiv:2212.09523*, 2022. doi: 10.48550/arxiv.2212.09523.
- [7] M. M. Kabir, M. F. Mridha, J. Shin, I. Jahan, and A. Q. Ohi, "A survey of speaker recognition: Fundamental theories, recognition methods and opportunities," *IEEE Access*, vol. 9, pp. 79236–79263, 2021. doi: 10.1109/ACCESS.2021.3084299.
- [8] K. Wei, B. Li, H. Lv, Q. Lu, N. Jiang, and L. Xie, "Conversational speech recognition by learning audio-textual cross-modal contextual representation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 2432–2444, 2024. doi: 10.1109/TASLP.2024.3381234.
- [9] D. Snyder, D. G. Romero, G. Sell, D. Povey, and S. Khudanpur, "X-vectors: Robust DNN embeddings for speaker recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, Canada, 2018.
- [10] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "LibriSpeech: An ASR corpus based on public domain audio books," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, QLD, Australia, 2015, pp. 5206–5210.
- [11] S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, *et al.*, "CNN architectures for large-scale audio classification," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, USA, 2017.
- [12] D. Salvati, C. G. Drioli, and L. Foresti, "A late fusion deep neural network for robust speaker identification using raw waveforms and gammatone cepstral coefficients," *Expert Systems with Applications*, vol. 222, p. 119750, 2023. doi: 10.1016/j.eswa.2023.119750.