Improving Voice Assistant User Experience throughContext Awareness and Personalization

Hetvi Tank[1], Kirit Modi[2], Ankur J. Goswami[3]

MTech. CE(SE), SPCE / Sankalchnad patel University, Visnagar, India [1]
IT Department, SPCE / Sankalchnad patel University, Visnagar, India [2]
CE Department,SPCE/Sankalchnad patel University, Visnagar, India [3]

hetvitank0@gmil.com[1], kjmodi.fet@spu.ac.in[2], ajgoswami.fet@spu.ac.in[3]

Abstract: Context awareness, the system's ability to comprehend and adapt to the user's surroundings, is harnessed through cutting-edge natural language processing, environmental sensing, and machine learning algorithms. To achieve context-awareness, the proposed system employs advanced natural language processing and machine learning algorithms.

The incorporation of context-aware features allows voice assistants to grasp the situational nuances of a conversation. This involves considering the user's prior commands, inquiries, and the broader context of the dialogue. Such awareness enables the voice assistant to provide more relevant and coherent responses,creating a seamless and natural conversation flow.

Personalization plays a crucial role in making voice assistants not only responsive but also adaptive to theunique needs and preferences of each user. Through the analysis of user behavior, preferences, and historical interactions, voice assistants can learn and evolve over time, delivering a more personalized and user-centric experience. This tailored approach not only enhances user satisfaction but also fosters a senseof connection between the user and the voice assistant.

In conclusion, the convergence of context-aware features and personalized responses represents a paradigm shift in voice assistant design. This approach holds the potential to elevate user satisfaction, foster more natural and intuitive conversations, and redefine the future landscape of voice interaction technology.

Keywords: customer satisfaction, expectations confirmation theory, digital assistants, privacy concerns,artificial intelligence.

I. INTRODUCTION

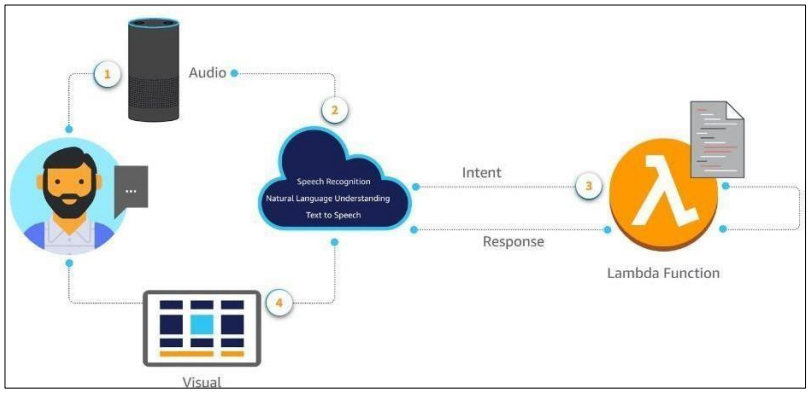

Fig. 1. Alexa Flow Chart

- To enhance the user experience of a voice assistant by leveraging context awareness andpersonalization.
- This project seeks to improve user interaction with a voice assistant by making it context-awareThey are often discussed in terms of the "5W" framework, which stands for Who,
  What, When, Where, and Why.

TABLE I

PROPOSED METHOD

5W of ETL		
No	Condition statisfied	Required
1	1234	0
2	234	1
3	123	0
4	124	3
5	134	0
6	12	3
7	13	0
8	14	3
9	24	13
10	23	1
11	1	3
12	2	13
13	3	1
14	4	123

1	what(Sales,production)
2	who(material)
3	when(date,period)
4	which(arithmatic)
5	where(location)

- The objective of the project is to create an AI, the total elastic application that can capture userspoken/written queries and respond to them accordingly related to sales data.
- This is our proposed 5W structure which fills the gap of fulfilment and satisfies the whole query.
- With this system, we can capture anything in order to create dynamic AI.

II. ALGORITHM

A. Named Entity Recognition

Another significant technique for analyzing natural language space is called entity recognition. It's in charge of classifying and categorizing persons in unstructured text into a set of predetermined groups. This includes individuals, groups, dates, amounts of money, and so on.

B. Text Summarization

As the name implies, NLP approaches can assist in the summarization of big volumes of text. Text summarization is commonly utilized in situations such as news headlines and research studies.

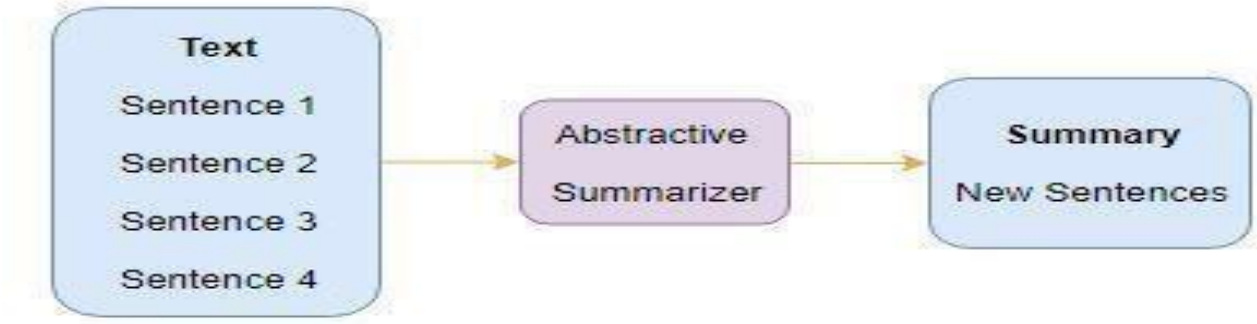

Fig 2. Text summarization

C. Bag of Words

This paradigm represents a text as a bag (multiset) of words, neglecting syntax and even word order whilekeeping multiplicity. In essence, the bag of words paradigm generates a matrix of incidence. These word frequencies or instances are then employed as features in the training of a classifier.

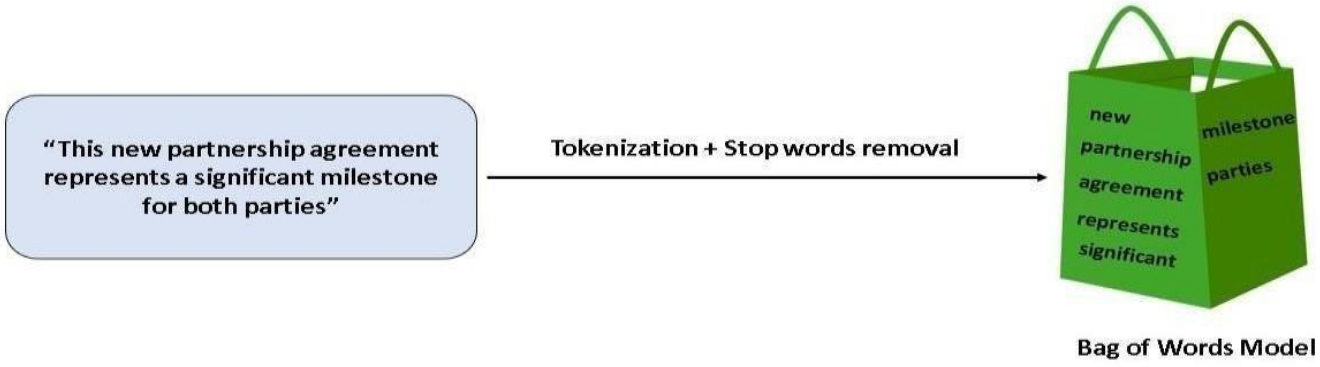

Fig 3. Bag of words

D. Keyword Extraction

Keyword extraction is one of the most important tasks in Natural Language Processing, and it is responsible for determining various methods for extracting a significant number of words and phrases from a collection of texts. All of this is done to summarize and assist in the relevant and well-organized organization, storage, search, and retrieval of content.

III. EXISTING FRAMEWORK ARCHITECTURE

A. Alexa Skills Kit (ASK):

- Framework: ASK is Amazon's official framework for building Alexa skills. It serves as the foundation for developing voice interactions and applications for Alexa-enabled devices.

B. AWS Lambda

- Serverless Backend: AWS Lambda is a key component of Alexa skill development. It allows you to run code in response to voice requests from Alexa. Most Alexa skills use AWS Lambda as theirbackend service.

C. Alexa Developer Console

- Development Environment: The Alexa Developer Console is an online platform where you design, build, test, and manage your Alexa skills. It provides a graphical interface for configuring your skill and testing it with simulated voice interactions.

D. Interaction Model

- Architecture: The interaction model defines how Alexa understands and responds to user input. It includes intents, slots, and sample utterances. You design and configure this model in the Alexa Developer Console.

- Process: After developing and testing your skill, you can submit it for certification through the Alexa Developer

Console. Amazon's certification process ensures that your skill meets its guidelines and quality standards.

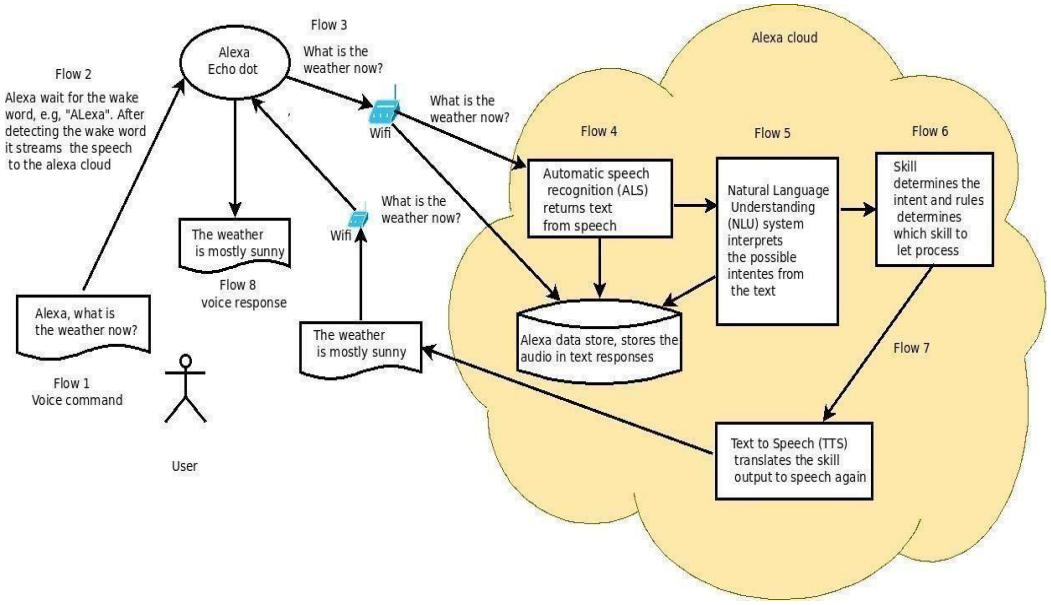

Fig. 4. Architecture of Alexa

E. Motivation

The underlying drive of this research is to offer valuable insights and actionable solutions to real-world challenges. By tackling genuine problems, we seek to make meaningful contributions that address pressingissues and generate practical outcomes. This motivation propels our efforts to drive positive change and provide tangible benefits to individuals, communities, and society at large.

IV. LITERATURE REVIEW

Sr. No	Authors and references	Paper Title/Journal	Main ideas	Simulation	Advantages	Limitations
1	Tom Brill Laura Muno z	Siri, Alexa,and Other Digital assistant: A Study of customer satisfaction with artificial intelligent applications	Siri, Alexa, andother digital assistants are rapidly being embraced by consumers and businesses. Digital assistants arespeech-enabled integrated artificial intelligence (AI) technologies (generally referenced as conversation-enabled applications).		This study affirmsthe role of the expectations confirmation processin customer satisfaction evaluation. Further, it provides insights that allow managers to understand the drivers and the degree of customer satisfaction with digital assistants. This study also	Customer satisfaction is not the sole goal inthe evolving landscape of digital assistants. Future research should explore diverse user dimensions, preferences, and generational differences to understand the full scope of Opportunities and challengesin

					provides recommendations as to where management should focus its priorities in order to assist users in gaining greater value with digital assistants.	this Rapidly advancing field.
2	Langzhou Chen,Volker Leutnant	AcousticModel Bootstrapping Using Semi-Supervised Learning	Speech recognition, semi-supervised training	Alexa Skill Set	It uses semi-supervised speech recognition.	Limitedlabeled datain semi-supervised acoustic model bootstrappingg can lead tosuboptimal performance, impacting accuracy and robustness. time- sensitive.
3	Che-Wei Huan g, Rol and Maas, Sri Harish Malli di, Björn Hoff meist er	A Study for Improving Device-Directed Speech Detection toward Frictionless Human-Machine Interaction	This paper is an update of prior work on detecting device-directed speech, or identifying utterances intended for Alexa.		The task can be phrased as a binary utterance-level classification problem that we approach with a DNN-LSTM model using acoustic features and features from the automatic speech recognition (ASR) decoder as input	Effectively integrating diverse acoustic and ASR decoder features for binary utterance-level classification using a DNN-LSTM model, which may pose challenges and impact overall classification accuracy.
4	AbdalGhani Abujabal Judith Gaspers	Neural Named Entity Recognition from Sub word Units	Named Entity Recognition (NER) is an important task in spoken language technology applications, such as voice-controlled smart assistants like the Amazon Echo or Google Home		To evaluate our models, we use the CoNLL script [3] to compute precision, recall, and F1 scores on a per-token basis. We report the average F1 score	Limitation in neural named entity recognition from sub-word units couldbe a reduced ability to capture nuanced semantic relationships due to the model's reliance on sub word represent at ions, potentially leading to less accurate identificationn of named

						entities in complex contexts.
5	Jaime Lorenzo Trueba, Thomas	Towards achieving robust universal neural vocoding	Statistical parametric speechsynthesis (SPSS) has seen a paradigm change recently, mainly thanks to the introduction of several Autoregressive models [1, 2, 3, 4,5, 6], turning into what can be termed statistical speech waveform synthesis (SSWS)[5].		This evaluation considered 2 female and 1 male speaker (the ones used to trainthe 3Spk vocoder).	Resultssuggest thatthe proposed vocoder, trained on varied materials (74 speakers and 17 languages, all recorded in studio conditions) can significantly outperform speaker dependent vocoders in clean unseen scenarios (relative MUSHRA score of 98%).
6	ChiehChi Kao,Ming Sun, Yixin Gao	Sub-band Convolutional Neural Networks for Small-footprint Spoken Term Classific action	With the rapid development of publicly available datasets (e.g. spoken term classification [1],speaker identification [2,3], acoustic event classification/detectio n [4, 5], etc.),state-of-the-art models for various acoustic applications can be trained with alarge amount of annotated data. CNN-based architectures have achieved state-of-the-art performance in keyword spotting [6], speech recognition [7, 8] speaker identification [2, 3], acoustic event classification [9]	Alexa SkillSet	They proposed a sub-band CNN architecture and explored it for spoken term classification.	effectively capturing nuanced acoustic features for small-footprint spokenterm classification using sub-band Convolutional Neural Networks, potentially impacting accuracy in diverse acoustic environments.

V. PROPOSED FRAMEWORK

A. Problem Statement

Alexa's personalization and context-awareness could involve addressing the need for improved adaptive capabilities. This may include developing mechanisms to enhance Alexa's understanding of individual user preferences and refining its ability to dynamically adapt responses based on real-time context. Challenges may include optimizing the balance between personalization and privacy, as well as ensuring seamless integration of context-aware features for a more natural and effective voice-assistant interaction.

B. Technology

- Alexa Developer Consol
- AWS Lambda

C. Alexa Developer Consol and AWS Lambda

The Alexa Developer Console and AWS Lambda form a crucial duo in the development and deployment of Alexa Skills. The Alexa Developer Console, as a web-based platform, serves as the central hub for skill creation, management, and testing. Developers utilize its features to design voice user interfaces, define interaction models, and simulate user interactions for testing. Moreover, the console facilitates the certification process, ensuring that developed skills meet the necessary standards before publication. On the other hand, AWS Lambda, a serverless computing service, plays a pivotal role in executing the code associated with Alexa Skills. Developers commonly use Lambda as the endpoint for their skills, where code for handling Alexa requests and generating responses is hosted. The serverless nature of Lambda allows for automatic scaling based on usage, providing scalability and cost-effectiveness. Additionally, AWS Lambda enables seamless integration with other AWS services, offering a robust infrastructure for Alexa Skill execution. Together, the Alexa Developer Console and AWS Lambda empower developers to create, refine, and deploy sophisticated voice-enabled applications with ease.

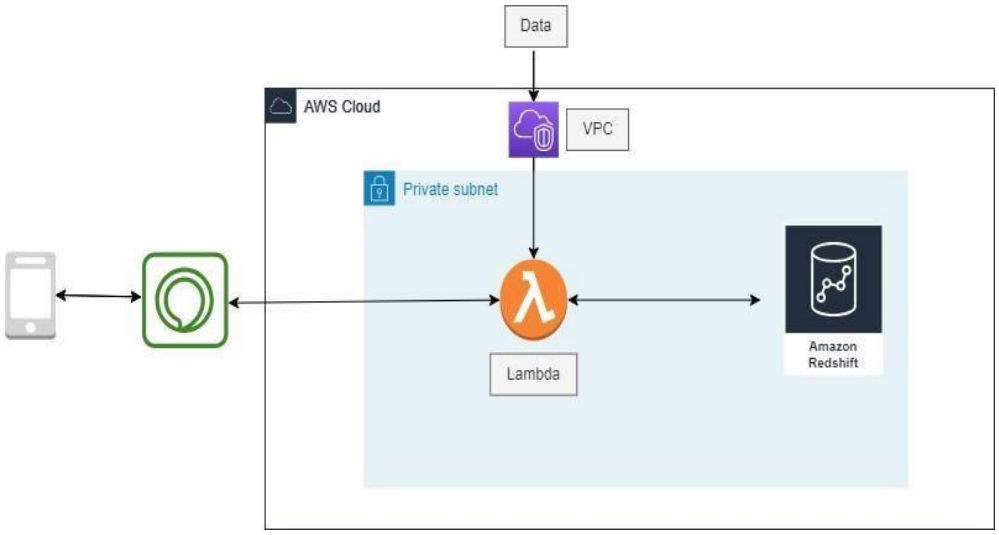

Fig 5. Architecture of Voice Assistant User Experience through Context-Awareness and Personalization

D. Project Planning and Design

- Define the objectives and scope of your Alexa skill. Determine the key functionalities and userinteractions.

E. *Skill Configuration*

- Access the Alexa Developer Console and create a new skill project.
  Configure the basic settings of your skill, including the skill's name, language, and region.

F. ***Lambda Function Integration***

- Create an AWS Lambda function to serve as the backend for your skill.
- Configure the Alexa Developer Console to link your skill to the Lambda function.

G. ***Code Development***

- Write the backend code for your skill's logic in the programming language supported by AWSLambda (e.g., Node.js, Python, Java).

H. Certification and Publishing

- Submit your skill for certification through the Alexa Developer Console.

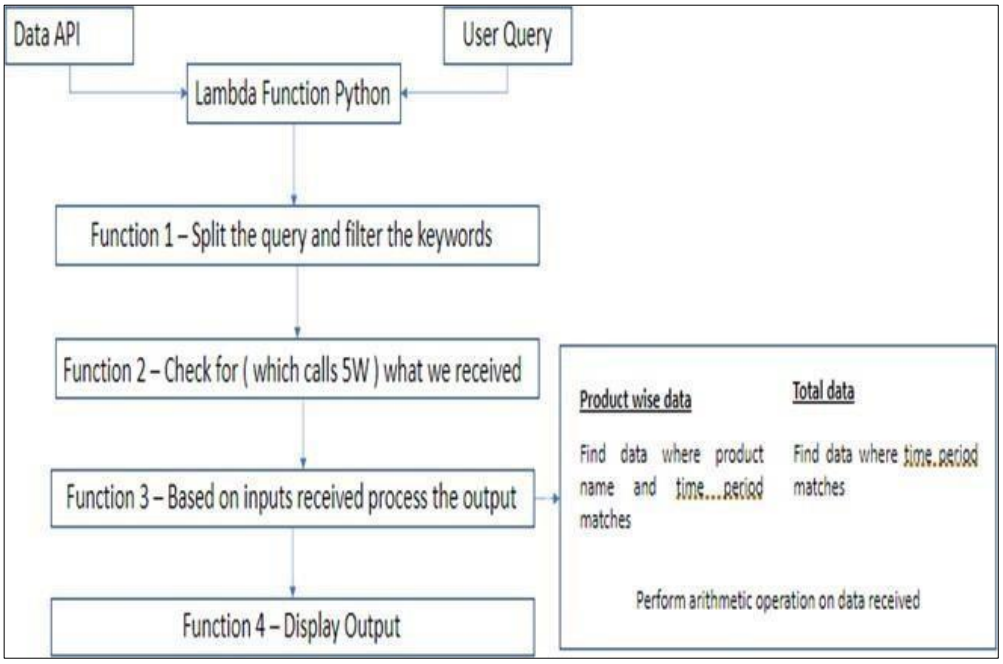

Fig. 6. Flow of Improving Voice Assistant User Experience through Context-Awareness and Personalization

Privacy and Security: Using a Virtual Private Cloud (VPC) is a crucial aspect of enhancing security in a cloud environment. VPC provides a private and isolated network space within the cloud, allowing you to control and secure your resources effectively. Here's how VPC contributes to security:

Network Isolation*:* VPC enables you to create isolated network environments, ensuring thatyour resources are not directly accessible from the internet. This isolation adds an extra layer of security by preventing unauthorized access.

Controlled Access: With VPC, you can define and configure security groups and network access control lists (ACLs) to control inbound and outbound traffic. This allows you to specify which IP addresses or ranges can access your resources, reducing the attack surface.

Subnet Segmentation: You can further enhance security by dividing your VPC into subnets.This segmentation enables you to organize and control traffic flow, creating zones with different security requirements for various components of your infrastructure.

VPN and Direct Connect*:* VPC allows you to establish secure connections between your on- premises infrastructure and your cloud resources using Virtual Private Network (VPN) or AWS Direct Connect. This

ensures secure communication and data transfer.

Encryption: Implementing encryption for data in transit and at rest is crucial for security. VPC provides options for encrypting communication between instances within the VPC and offers integration with other AWS services that support encryption.

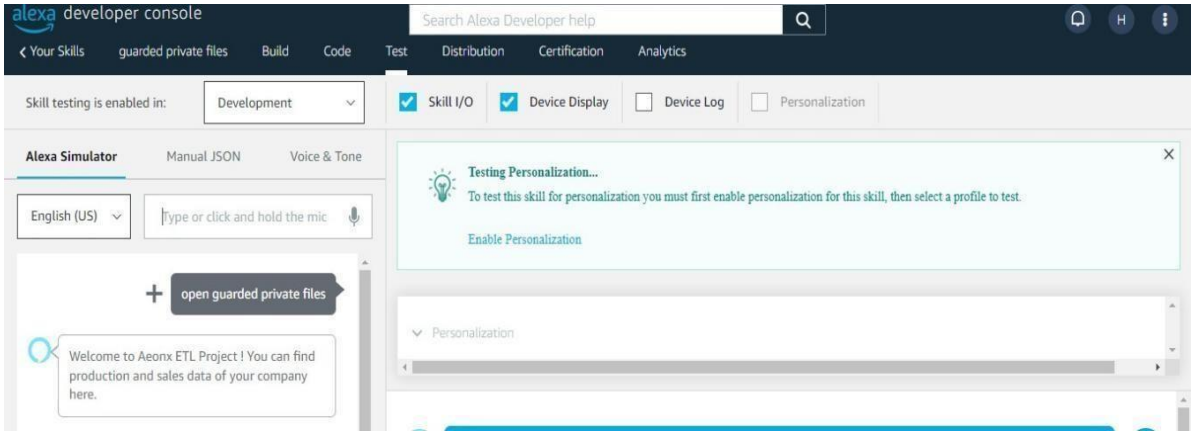

Fig. 7. Wake word

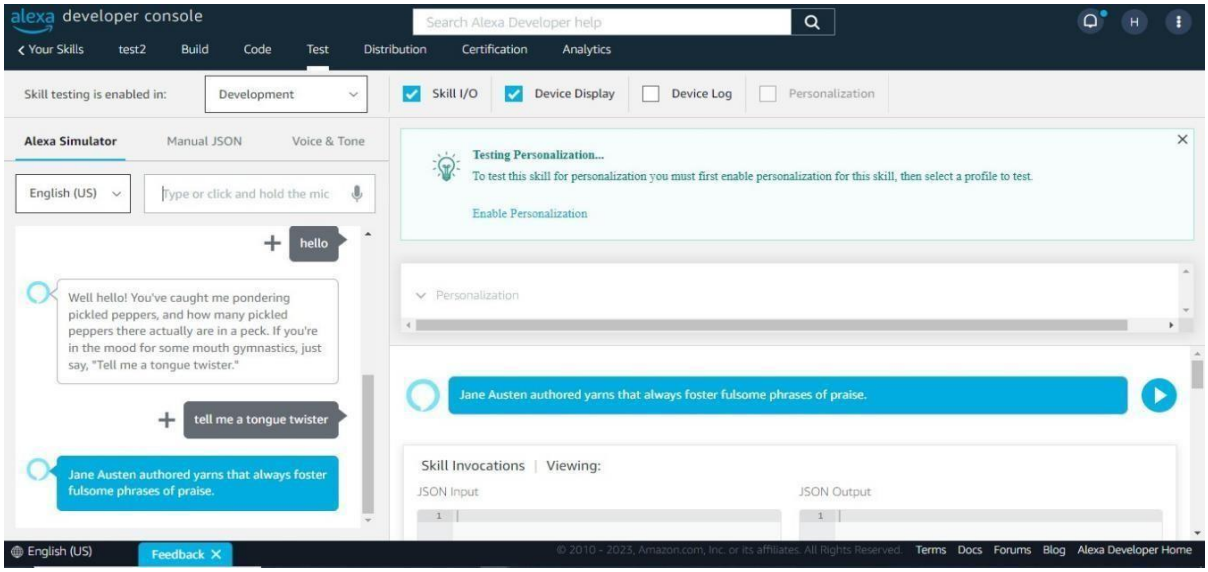

Fig. 8. Custom questioning

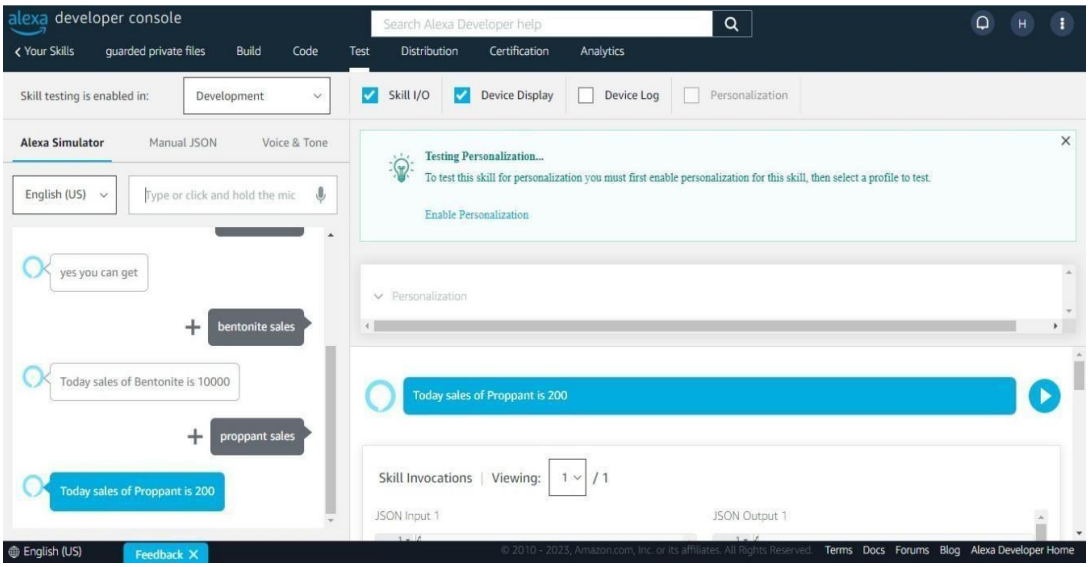

Fig. 9. Custom questioning

Controlled Access: With VPC, you can define and configure security groups and network access control lists (ACLs) to control inbound and outbound traffic. This allows you to specify which IP addresses or ranges can access your resources, reducing the attack surface.

Subnet Segmentation: You can further enhance security by dividing your VPC into subnets. This segmentation enables you to organize and control traffic flow, creating zones with different security requirements for various components of your infrastructure.

VPN and Direct Connect: VPC allows you to establish secure connections between your on-premises infrastructure and your cloud resources using Virtual Private Network (VPN) or AWS Direct Connect. This ensures secure communication and data transfer.

Encryption: Implementing encryption for data in transit and at rest is crucial for security. VPC provides options for encrypting communication between instances within the VPC and offers integration with other AWS services that support encryption.

VI. CONCLUSION

Personalization takes context awareness a step further by tailoring responses and recommendations based on the user's individual preferences, historical data, and behavior patterns. By learning from past interactions, the voice assistant can adapt its responses to align with the user's unique preferences, creating a more personalized and human-like experience.

Customer satisfaction has long been a focal point of extant marketing and information technology literature. This study advances our understanding of the theoretical foundations for customer satisfaction as related to a new AI technology platform involving digital assistants.

Given the relative infancy of current digital assistant adoption and utilization, there is limited empirical work directly related to the consumer experience and customer satisfaction. This study affirmed the role of the expectations confirmation process in the customer satisfaction evaluation. Further, it provides insights that allow managers to understand the drivers and the degree of customer satisfaction with digital assistants. These elements can influence customer satisfaction evaluations.

VII. FUTURE SCOPE

In the future, improving the Voice Assistant User Experience through context awareness and Personalization holds promising prospects. Advanced context-aware features could encompass real-time environmental data and emotional cues, contributing to a more adaptive user experience. The integration of multimodal interactions, combining voice with gestures or facial expressions,

could offer a comprehensive and intuitive interface. Future systems might focus on continuous learning algorithms to dynamically adapt to evolving user preferences, ensuring a personalized and evolving interaction over time. Privacy-preserving personalization will be crucial, addressing concerns about data security as personalization becomes more intricate. Tailoring voice assistants to specific domains, fostering collaboration with third-party services, and prioritizing accessibility for diverse user groups are avenues for development. Additionally, recognizing and adapting to global cultural nuances and integrating with emerging technologies like augmented reality could further enhance the inclusivity and capabilities of voice assistants. In summary, the future holds potential for more sophisticated, adaptive, and culturally aware voice assistant interactions, transforming the landscape of human-computer interaction.

REFERENCES

[1] Rospocher, M., van Erp, M., Vossen, P., Fokkens, A., Aldabe,I., Rigau, G., Soroa, A., Ploeger, T., and Bogaard, T.(2016). "*Building event-centric knowledge graphs from the news.Web Semantics: Science, Services, and Agents on the World Wide Web*" In Press.

[2] Knight, K., & Langkilde, I. " *Preserving ambiguities in generation via automata intersection*" In AAAI/IAAI (pp. 697-702).

[3] Liddy, E. D. (2001). Natural language processing.

[4] Che-Wei Huang, Roland Maas, Sri Harish Mallidi, Björn Hoffmeister., "*A Study for Improving Device-Directed Speech Detection toward Frictionless Human-Machine Interaction*", 2019.

[5] Thomas M. Brill, Laura Munoz & Richard J. Miller Siri, Alexa, and other digital assistants: "*A study of customer satisfaction with artificiaIntelligence applications*", 2019.

[6] D. Reich, F. Putze, D. Heger, J. Ijsselmuiden, R. Stiefelhagen, and T. Schultz, (2011) "*A real-time speech command detector for a smart control room*," in Proceedings of Interspeech

[7] E. Shriberg, A. Stolcke, D. Hakkani-Tr, and L. Heck, "*Learning when to listen: Detecting system-addressed speech in human-human-computer dialog*," in Proceedings ofInterspeech, 2012.

[8] T. Yamagata, T. Takiguchi, and Ariki, "*System request detectionin human conversation based on multi-resolution Gabor wavelet features,*" in Proceedings ofInterspeech, 2009.

[9] H. Lee, A. Stolcke, and E. Shriberg,(2013) "*Using out-of-domain data for lexical addresseedetection in human human-computer dialog*," in Proceedings of HLT-NAACL,

[10] Wang, D. Hakkani-Tur, and G. Tur "*Understanding computer directed utterances in multi-user dialog systems,*" in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013.