Literature Review Report

Paper1: **(Research Papers Summary of Mini Project)**

Artificial Intelligence-based Voice Assistant:

Voice assistant is a major growing feature in the real world which is commonly being used in many smartphones and laptops. Simply AI-based voice assistants are operating systems that recognize the human voice and respond as an integrated voice. The voice assistant has three stages: Text to speech, Text to intention, and intention-to-action. The voice assistant will gather the audio from the human voice and convert it into text, then send it through GTTS (Google text to speech), then GTTS engine will convert this converted text into an audio file in the English Language, then the audio is played by using play sound package of python programming language. Because it is a hands-free application it is very beneficial in the real world. Voice assistants are useful in every field, such as education, basic human needs, and home appliances. Voice assistant is also useful for illiterate people they can get any information just by saying to the assistant.

Paper2:

Exploring the Determinants of Users' Continuance Usage Intention of Smart Voice Assistant:

This summary concludes the continuance usage scenario and approaches for long-term usage or interaction of the person with the assistant and making it more successful and interactive. Through evaluation of the engagement aspect, the current finding contributes to the theory by better understanding how the proposed antecedent determines the continuance intention. It is conceptualized user engagement in three ways: i. Cognitive aspect ii. Affective aspect and iii. Behavioral aspect. The cognitive dimension refers to the user's thoughts, knowledge, concentration, and interest in a specific object. For example, in the case of the voice-assistants cognitive engagement refers to the user's interest in the services provided by these devices. Ex- Music streaming. The second dimension of engagement is the affective/emotional aspect. This dimension refers

to the state of emotional activity i.e., a feeling of positive or negative inspirations related to voice assistants. The third dimension of engagement i.e., the behavioral aspect refers to the user's behavioral manifestations towards an object of interest. Trust and risk factors are highly considerable when we talk about using a certain technology for a very long period.

Paper3:

Security and Privacy on Voice Assistants:

To improve user experience, most operating systems and service providers are gradually shipping smart devices with voice-controlled intelligent personal assistants, reaching a new level of human and technology convergence. While these systems facilitate user interaction, it has been recently shown that there is a potential risk regarding devices, which have such functionality. One can easily control and manipulate the voice commands remotely, issuing arbitrary commands which can greatly expose the users. Having simultaneously several apps using voice input and output has already been shown to expose users' security and privacy there is a definite need from the OS side to consider both microphone and speaker as a unique communication channel and identify apps that try to use both flows when only one is granted by the user. One reason is that voice assistant are backed by artificial intelligence making them more extensible. The security issues become more complicated by the use of intents and the fact that Voice Assistants are constantly activated, which invalidates the assumption of only one application using the communication channel. A rather simplistic, yet powerful defense mechanism against the attacks that target voice assistants would be to utilize biometrics to identify the actual device owners and enable or disable the function of this assistant accordingly.

Paper4:

Goal-Oriented Modelling for Virtual Assistants:

The red dotted arcs indicate the start and end goals of the composite goal, and the black arcs indicate the relations between goals and actions. A goal Net is a mental model of an agent, representing its goals and how it achieves those goals in dynamic environments. Composite goals can be decomposed into atomic and composite goals, while atomic goals are unable to be decomposed any further. Goat net allows for goal and action selection mechanisms to be defined so that an agent can decide what goals to pursue and what tasks to execute based on situational criteria. A simple goal net consisting of a root composite goal, two atomic goals, and an action. A screenshot of the updated interface of the goal net designer in MADE.

Paper5:

Advantages and Constraints of a Hybrid Model K-12 E-Learning Assistant Chatbot:

By understanding the development of E-Learning and the advancement of chatbots mentioned above, we want to explore and evaluate how a chatbot could perform as a learning assistant in an E-Learning environment. We have to see whether using a chatbot could reduce E-Learning issues such as feelings of isolation and detachment. We will compare our chatbot with a teacher counseling service from the E-Learning platform on which our chatbot is based to see how well our chatbot performs. We will compare our chatbot with a teacher counseling service from the E-Learning platform on which our chatbot is based to see how well our chatbot performs. To further improve the user experience of our chatbot, we designed a way for students to switch a conversation from the chatbot to a conversation with a human teacher or counselor inside the same dialogue.

Current commercial chatbots typically use two types of models to create responses: a retrieval-based model and a generative-based model. On comparing with the teacher counseling services provided by the E-Learning platform on which our chatbot is based, our chatbot does have advantages in chatting with students.

Paper 6:

Using AI to attack VA: A Stealthy Spyware Against Voice Assistances in Smart Phones

As Virtual Assistance plays an important role in our day-to-day life which conduct various tasks, these types of tools usually come with high privileges which can access our smartphones or another system. A comprised VA is a stepping stone for attackers to hack into users' phones. With the new stack, the attacker uses voice commands to activate the VA and launch several tasks which include leaking private information and sending inappropriate messages or emails. In this paper, A novel attack approach is proposed namely VASPY which includes An attacking environment sensing module to choose an optimal attacking time and voice volume making the attack unnoticeable by the users, and found that VASPY can launch attacks unnoticeable by the users. The spyware can't be detected by anti-malware tools from both industry and academia.

Paper 7:

A Google glass based real-time scene analysis for the visually impaired:

In this paper, Blind and Visually Impaired People (BVIP) persons are considered who face difficulties with different tasks that involve scene recognition. Paper represents a system that is simply based on the design of google glass which is further being used as a Visual Assistant. Because it is a visual assistant, A camera is there which capture the image of the different scene (in general Surroundings) which is being analyzed using the custom API from Azure services provided by Microsoft. After capturing the images, The vision API is converted to speech, which is heard by the BVIP users wearing google glass. A dataset is created using different types of images to improve the efficiency of the system for that purpose, The vision API is trained first and tested on this dataset, which increases the mean average precision (Map) score from 63% to 84%. The overall response time was measured to be less than 1 second. It can be concluded that the VISION API help the BVIP better understand their surroundings in real-time.

Paper 8:

The Feasibility of injecting inaudible voice commands to voice assistants:

Voice Assistants like Apple's Siri, Microsoft's Cortana, and Google Assistant now become very popular human-machine interaction methods and have made various systems voice controllable. In this work, a completely inaudible attack named as Dolphin_Attack modulates voice commands on ultrasonic carriers to achieve inaudibility. By leveraging the nonlinearity of the microphone circuits, the modulated low-frequency audio voice commands can be successfully demodulated, and more importantly interpreted by the voice assistants.

By injecting a sequence of inaudible voice commands, which include activating Siri to initiate a face time call on an iPhone, and activating Google Now to turn on airplane mode. This work validates that it is feasible to detect Dolphin Attacks by classifying the audios using a supported vector machine (SVM), and suggests to re-design VA to be resilient to inaudible Voice Commands Attacks.

Paper 9:

Personal VA Security and Privacy – A Survey:

Personal VAs is used as an interface between a human and a machine in digital environments. In these, voice commands are used to interact with the phone, smart homes, cars, or offices. A survey found that, In the United States alone, no. of smart speakers like Amazon's Echo and Google Home, has increased by 78% to 118.5 million, and 21% of the US population has at least one device. Due to the increasing dependency on the VAs, Security and privacy have become a major concern of users, developers, and policymakers. In this, attacks and countermeasures are discussed. This survey describes research areas where the countermeasure is lacking.

PVAs are now commonplace and are significantly changing the way users interact with computer systems. Users increasingly depend on PVAs as the main or even single interface to computer systems and smart environments. Consequently, the security of these devices has become the focus of public attention and research efforts. Likewise, privacy is a concern for most users as PVAs record and observe speech, the most fundamental form of human interaction.

Paper 10:

Short Research on Voice Control Systems Based on AI Assistant:

This paper proposes a voice control system that is entirely based on AI assistants. The AI assistant system using Google assistant, a representative service of open API artificial intelligence, and the conditional auto-run system, IFTTTF (If This, Then That) was designed. It const-effectively implemented the system using Raspberry Pi, a voice recognition module, and open software. The proposed system is expected to be applied to various control systems based on voice recognition.

In this study, the design of a voice control system using embedded systems and open API AIs is detailed. The proposed system is a voice command system that operates modules, such as a relay module, through a user's voice commands.