**Solent University**

Faculty of Business, Law and Digital Technologies

**BSc Software Engineering**

**2021**

**Kieron Gillingham**

**“GLAD – A Voice Assistant for People Living With Dementia”**

**Supervisor : Prins Butt**

**Date of submission : May 2021**

# Acknowledgements

With great gratitude my parents for their patience and understanding; to my supervisor Prins Butt for his kind advice; and to my friend Mr Matthew Dear for inviting me to sit at his table that day.

# Abstract

Dementia is among the world’s leading health concerns, affecting approximately 50 million people and rising worldwide. It primarily affects the elderly and causes degeneration of cognitive functioning which inhibits their ability to live independently. As the condition is irreversible, interventions to limit the rate of cognitive loss are the only course of treatment.

This paper describes the implementation of a voice assistant (GLAD) targeted at users living with dementia and related conditions with the aim of assisting them in their daily lives to improve their independence and reduce the rate of cognitive decline. Following a systematic literature review, the architectural design of GLAD is presented and implemented following a software development methodology. An evaluation of its capabilities is then completed and areas of further work suggested.

Contents

[Acknowledgements i](#_Toc70959587)

[Abstract ii](#_Toc70959588)

[Acronyms iii](#_Toc70959589)

[List of Figures 5](#_Toc70959590)

[1 Introduction 6](#_Toc70959591)

[2 Literature Review 9](#_Toc70959592)

[2.1 Method 9](#_Toc70959593)

[2.2 Dementia 9](#_Toc70959594)

[2.3 Assisted Living 10](#_Toc70959595)

[2.4 Digital Divide 13](#_Toc70959596)

[2.5 The State of VAs 15](#_Toc70959597)

[3 Methodology 17](#_Toc70959598)

[3.1 Objectives 17](#_Toc70959599)

[3.2 Development Methodology 17](#_Toc70959600)

[3.3 Management 17](#_Toc70959601)

[3.4 Evaluation 18](#_Toc70959602)

[3.4.1 Software Evaluation 18](#_Toc70959603)

[3.4.2 Process Evaluation 19](#_Toc70959604)

[3.5 Ethics 20](#_Toc70959605)

[3.5.1 Data Collection 20](#_Toc70959606)

[3.5.2 Testing 20](#_Toc70959607)

[3.6 Legal 20](#_Toc70959608)

[4 Planning 21](#_Toc70959609)

[4.1 Requirements 21](#_Toc70959610)

[4.2 Risk Assessment 22](#_Toc70959611)

[4.3 Prototyping 23](#_Toc70959612)

[4.4 Development environment 23](#_Toc70959613)

[5 The Voice Assistant 24](#_Toc70959614)

[5.1 Components of a VA 24](#_Toc70959615)

[5.2 Technologies 24](#_Toc70959616)

[5.2.1 Voice Assistants 24](#_Toc70959617)

[5.2.2 Speech-to-Text 25](#_Toc70959618)

[5.2.3 Intent Detection 25](#_Toc70959619)

[5.2.4 Text to Speech 26](#_Toc70959620)

[5.3 Design 27](#_Toc70959621)

[5.4 Implementation 29](#_Toc70959622)

[5.5 Evaluation 30](#_Toc70959623)

[6 Cognitive Exercises 32](#_Toc70959624)

[6.1 Introduction 32](#_Toc70959625)

[6.2 Exercises 32](#_Toc70959626)

[6.2.1 Serious Games for Dementia 32](#_Toc70959627)

[6.2.2 Music Therapy 33](#_Toc70959628)

[6.2.3 Physical Exercise 33](#_Toc70959629)

[6.3 Implementation 33](#_Toc70959630)

[6.3.1 Number Game 33](#_Toc70959631)

[6.3.2 Radio 34](#_Toc70959632)

[6.3.3 Challenges 35](#_Toc70959633)

[6.4 Evaluation 36](#_Toc70959634)

[7 Considerations for Elderly Users 37](#_Toc70959635)

[7.1 Introduction 37](#_Toc70959636)

[7.2 Research 37](#_Toc70959637)

[7.3 Implementation 38](#_Toc70959638)

[7.4 Evaluation 39](#_Toc70959639)

[8 Companionship 41](#_Toc70959640)

[8.1 Introduction 41](#_Toc70959641)

[8.2 Research 41](#_Toc70959642)

[8.3 Implementation 41](#_Toc70959643)

[8.4 Evaluation 43](#_Toc70959644)

[9 Evaluation 44](#_Toc70959645)

[9.1 Developed System 44](#_Toc70959646)

[9.2 Process 44](#_Toc70959647)

[10 Further Research 46](#_Toc70959648)

[11 Conclusion 47](#_Toc70959649)

[12 References 48](#_Toc70959650)

[13 Bibliography 49](#_Toc70959651)

[Appendix A: Concept Map A-1](#_Toc70959652)

[Appendix B: Ethics Release B-1](#_Toc70959653)

# List of Figures

[Figure 1: The project roadmap. 8](#_Toc70667503)

[Figure 2: MoSCoW prioritisation of requirements 22](#_Toc70667504)

[Figure 3. Open Assistant Architecture 28](#_Toc70667505)

# Introduction

Dementia is one of the leading causes of disability among the global elderly population which causes the deterioration of cognitive functioning. As modern advances in healthcare cause the average age of the world’s population to rise, so too is the number of elderly persons living with dementia that require care and support (World Health Organization 2019). There are estimated to be over 885,000 people in the UK diagnosed with dementia (Wittenberg *et al*. 2019) and over 50 million worldwide (Prince *et al.* 2015). People living with dementia require constant care and, although the nursing care industry is growing, training care-workers is difficult, time-consuming, and expensive. As a result, the number of care-workers is not sufficient to match the growing number of dementia cases.

There is great promise for voice assistants[[1]](#footnote-1) (VAs) such as Apple’s Siri and Amazon’s Alexa to support care-workers and patients by managing routine tasks such as setting medication reminders, carrying out mental stimulation exercises, and alerting human carers when needed. These products are operated through voice commands and can run on existing internet-enabled devices or dedicated hardware. They are capable of monitoring the wellbeing of vulnerable people at a lower cost of time and money than training a care-worker. Such devices could serve as a buffer on the workload placed on care-givers and enable persons living with dementia to have more independence. Unfortunately, current consumer products are not sufficiently reliable enough to provide support in this capacity. Despite being marketed as easy-to-use, these products still require a degree of technical understanding to be used effectively. While younger users (“Digital Natives”) of these products may find it easy to adapt to their use, older generations (“Digital Immigrants”) have more difficulty learning the technology. Furthermore, as these products are intended for general use, they are not suitable for users with special needs.

This project aims to evaluate the potential of VA technology in supporting elderly persons living with dementia and answer the This question will be addressed through the development of a prototype VA called the *General Living Assistance Device* (*GLAD*) that is tailored for users living with dementia. This prototype will include features common to current VAs such as setting reminders, performing search queries, and calling contacts, as well as ease-of-use considerations made for elderly users in general. A shortlist of additional features designed to maintain user independence, monitor wellbeing, and reduce the effects of cognitive decline will be created following a review of best practices and the viability of implementing these features on limited hardware will be assessed.

In this paper, current literature surrounding the challenges caused by dementia for sufferers and carers is reviewed and the current state of VA technology including its shortcomings is summarised. This research is used to list potential features and requirements of a VA to assist people living with dementia. Next, a simple VA that will serve as the foundation of the selected features is designed and presented. This paper then describes the research and implementation of each feature, the difficulties faced, and evaluates their effectiveness. Finally, an evaluation of the feasibility of the entire developed prototype is given and future areas of work are considered.

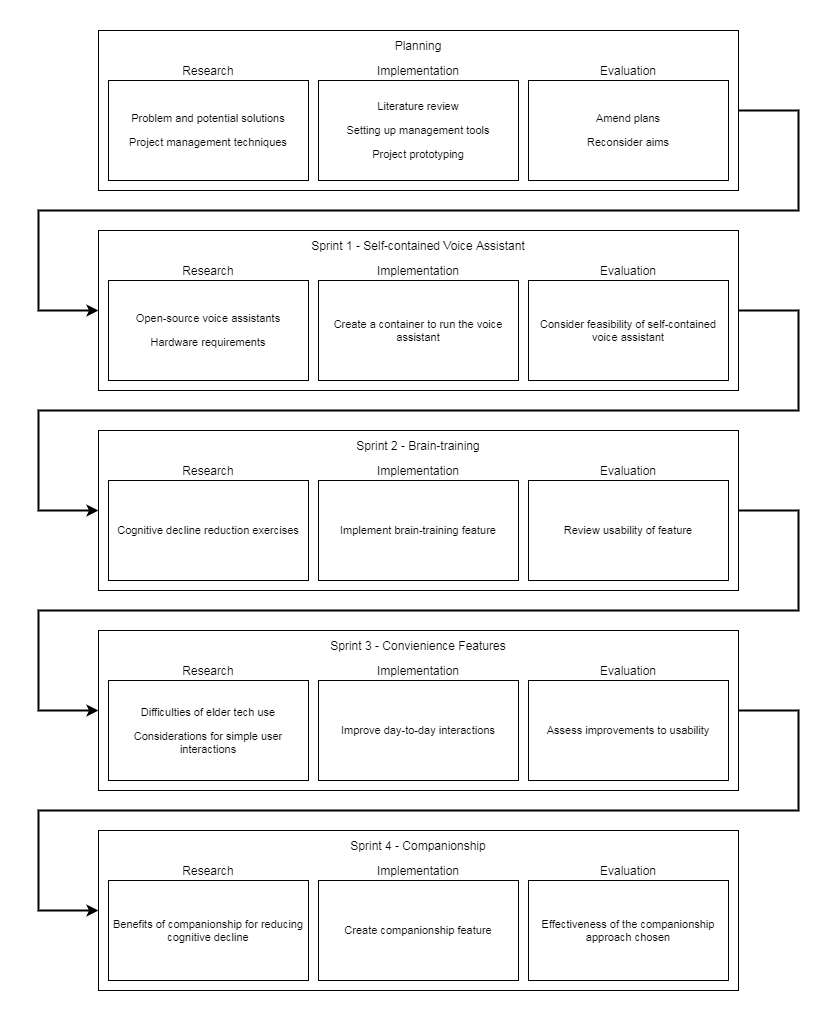


Figure 1: The project roadmap.

# Literature Review

## Method

To assess the difficulties caused by cognitive decline and potential technologies that could mitigate them, a review of existing literature was conducted. The literature was gathered through multiple search queries on ScienceDirect using combinations of the following keywords: dementia, “caring for”, “home care”, palliative, helping, challenges, difficulties, needs, prevention, independence, enabling, preferences, “assistive technology”, “smart home”, capability, voice, assistant, “consumer product”, adoption, elderly, “digital immigrants”, design, “early onset”. Studies mentioning “social media”, student, or diagnosis were excluded. Each search query was refined until it produced 30 or fewer results and exclusion criteria were applied.

## Dementia

Dementia is the impairment of higher brain functions such as memory and cognitive processing and is commonly caused by Alzheimer’s disease (AD). The World Health Organization (2019) estimated 50 million people worldwide living with dementia burdening individuals, caregivers, and healthcare services with an approximate global cost of US$ 818 billion in 2015. Dementia cases are expected to reach 82 million by 2030 and 152 million by 2050. This inflation is the result of a global aging population caused by increased life expectancy worldwide. Galende et al. (2021) assessed the social impact of dementia to advise healthcare policies in Spain, concluding that robust healthcare programs were essential. Their review found between 4% and 9% of people in Spain over 65 affected by dementia with rates increasing proportional to age. These proportions are typical of other countries.

The decline of cognitive functioning causes difficulties in safety, autonomy, and quality-of-life (QoL) affecting both patients and carers; and various focus groups, interviews, and workshops have attempted to formalise the design requirements of care services based on these concerns (Morgan et al. 2002; Thoma-Lürken et al. 2018; Lockerbie and Maiden 2020). Focus groups with formal and informal caregivers conducted by Thoma-Lürken et al. (2018) revealed 6 recurring domains of problems preventing aging-in-place for people living with dementia: Self-reliance, safety, social, behavioural, formal Services, and cognition; however they did not address the causal relationships between these domains. The most common issues raised were patients suffering a loss of independence and inability to perform activities of daily living (ADL). In a review by Alexopoulos et al. (2002), mid- and late-life depressions were found to exacerbate cognitive decline, and the increased loss of independence produced further depression. One study that focused on support for informal caregivers found that many felt they could continue to provide sufficient care for their dependant for longer if they were given better education and relief as they lacked appropriate training and were hindered by time constraints (Chi et al. 2020).

## Assisted Living

As technologies such as robotics, the internet-of-things (IoT), and machine learning are developed worldwide, these innovations have been applied to improve the quality-of-life (QoL) for both the fully cognitive elderly and those suffering from cognitive decline (Li, Lu, and McDonald-Maier 2015). Traditionally, caring for people living with dementia would be performed at home by family members until they were unable and the dependent was moved into formal care (Kemp, Ball, and Perkins 2013); however financial pressure placed on both individuals and healthcare services has motivated research into assisted living (AL) technologies to promote “aging-in-place” and allowing dependents to maintain their autonomy. These advances aim to relieve pressure on healthcare services and informal carers.

Telecare refers to technologies that provide remote healthcare directly to patients in their own homes such as monitoring sensors (Barlow, Bayer, and Curry 2006; Roberts and Mort 2009). These approaches are also referred to as telemedicine and telehealth inconsistently across different authors. Some of the earliest telecare solutions include the EU-ACTION project that began in 1997 – a system intended to introduce ICT into home environments to educate home carers and dependants in correct care techniques (Magnusson et al. 2002).

The ubiquity of IoT devices has led to the notion of intra-connected smart homes in which multiple sensors and devices can communicate and be controlled through a unified interface such as a voice assistant. Cooper et al. (2008) describe “intelligent environments” similar to the newer idea of smart homes, noting how they could assist individuals with cognitive impairments using reminders, directional guidance, or monitoring. They recognise the importance of technology understanding the context of a situation (such as a user’s location or task), an issue that has been addressed more recently with machine learning. Other early work on smart homes also foresaw their use in telecare for elderly and disabled people including those with cognitive impairment (Chan et al. 2008; Chan et al. 2009). Belley et al. (2015) present a practical algorithm for detecting erratic behaviour in people with cognitive decline by analysing power usage of smart devices. Liu et al. (2016) also focused on the benefits of health monitoring of elderly in smart homes, however they conclude that smart homes were not capable of completely supporting the elderly. Rumeau et al. (2020) investigated co-living spaces for elders to reduce isolation with a tangential experiment related to smart home technology.

Shishehgar, Kerr, and Blake (2018) outline a variety of robotics projects for supporting elders including companion robots, mounted mechanical arms, electronic wheelchairs and walking assistants, domestic cleaning robots, and health and time management robots. Wilson et al. (2019) discuss how these robotics projects can be integrated with smart homes.

Despite the number of care projects that have been created, few have seen widespread adoption or made it past early pilot stages due to poor evidence of cost-effectiveness (Obi, Ishmatova, and Iwasaki 2013; Clarkson et al. 2017). The effectiveness of these solutions is difficult to measure because the majority of studies use qualitative means of assessment or failed to apply their findings to a formal framework (Siegel and Dorner 2017). Dodd et al. (2020) were also unable to find an existing measure for assessing the effectiveness of care solutions with respect to the key desired outcomes of stakeholders. Limited study sizes also question the validity of any positive findings in these studies. Among those projects that were deployed, adoption is likely hindered by the deep-rooted social stigma related to dementia and AL technologies; particularly in rural areas (Morgan et al. 2002).

End-of-life care should preserve an individual’s dignity and QoL. Östlund, Brown, and Johnston (2012) reviewed palliative care studies to assess how well recipient’s dignity was addressed. Palliative care includes solutions to ease the pain of conditions without addressing the cause of the problem. Dementia care is considered palliative as it improves the comfort of the patient and may reduce deterioration, but cannot reverse any existing damage. None of their reviewed studies addressed patients concerns regarding the impact of their own death on their surviving friends and family. Rich relationships, autonomy and control, knowledge, and improved mental health were identified as the desired outcomes of care solutions (Dodd et al. 2020). Lockerbie and Maiden (2020) created a model for defining the QoL goals for people with dementia through workshops with four experienced UK care workers, also concluding that improving independence and social connectivity were desirable outcomes of support. Gómez (2015) discusses the nature of autonomy for elders and the sustainability of solutions that support their independence, arguing that autonomy should not be accepted as a guaranteed improvement to QoL. Hersh (2014) developed a framework for assessing the outcomes of ICT support.

## Digital Divide

The concept of a digital divide between elderly (digital immigrants) and younger (digital natives) users of technology is well documented. Digital immigrants are characterised by their struggle or resistance to adopt technology because of decreased learning capabilities, a rapidly changing industry, limited or poor experiences, or lack of confidence; instead using technology only when necessary.

Mobile phones have the capability of improving QoL for elders (Plaza et al. 2011). Many applications are available for encouraging personal health and wellbeing.

Despite the existence of tech-savvy elders and the inevitable generational shift as digital natives continue to age with technology, it is crucial to consider the difficulties caused by natural aging and late-life disabilities that are barriers to assistive technologies (Fischer et al. 2014). The results of a survey into motivations behind elder’s technology adoption by Sintonen and Immonen (2013) found that prior experience with technology their own physical limitations are key deciding factors for frail elderly. Their population had an approximately 1:2 split of frail and well-coping elderly. Hawley-Hague et al. (2014) found that concern for their own safety was a another crucial factor in elders adoption of AL technology. A review by Song and van der Cammen (2019) was concerned with how AL technology affects elders living alone.

The needs and preferences of elderly users should be considered before designing any AL technology. Jacelon and Hanson (2013) discuss the benefits of involving elders in the design process for smart homes to ensure they meet the practical needs of this specialist group. Gkouskos and Burgos (2017) also highlight the importance of involving elders in the design process of any AL technology.

Detweiler and Hindriks (2016) formalised a taxonomy for value sensitive design of AL technology and raise the issue of limited coverage of research into all permutations of their identified values, technologies, and contexts.

After interviewing elders who consider themselves technologically savvy, Kania-Lundholm and Torres (2015) question the importance of age as a factor in the Digital Divide; instead finding socio-economic explanations. The elders interviewed were generally highly educated and had used computers as early as the 1970s.

Castilla et al. (2013) created a software tool for the elderly that streamlines common computing features such as email and telecommunication. They realised that simply enlarging icons and text was insufficient for making software accessible. To better aid user’s synchronous learning of what capabilities were offered and how to perform them, they concluded that no more than three options should be available at any time. Similarly, Iancu and Iancu (2020) suggest principals to be considered when designing mobile phones for Elders and found that multiple paths of completing the same action were confusing to users. They recommend familiarity and consistency in the design of tasks. Alternative human computer interaction (HCI) technologies have been considered to facilitate digital immigrant’s engagement with modern devices.

Hsiao et al. (2017) present natural hand motion controls for desktop applications with limited success. Other means of natural HCI that have been used include voice control and eye tracking (Li, Lu, McDonald-Maier 2015).

## The State of VAs

Recently, consumer VAs such as Google Assistant and Amazon’s Alexa have become familiar presences in households and on mobile devices (McLean and Osei-Frimpong 2019). These consumer VAs are frequently tasked with performing web queries and making online purchases. Current VAs are still somewhat limited, despite considerable advances in natural language processing, however, privacy concerns are a common factor restricting their usage (Ho-Sam-Sooi, Pieters, and Kroesen 2021).

VAs are able to integrate with Smart Home technology and provide a conversational means of controlling the devices within. Conversational controls are accessible even for elders with cognitive decline. Chatbots and natural language interaction computers have been used for various healthcare services (Adamopoulou and Moussiades 2020).

Trust in the VA is an essential requirement as carers and family members will be unwilling to place their dependent’s well-being in jeopardy. Poushneh (2021) explored the perception factors of artificial personalities in mobile voice assistants. Hu, Lu, and Gong (2021) investigate how user interactions with and trust of AI are affected by human-like qualities, determining that the humanness of voice output does not impact competence-related trust. However, Hu et al. (2021) found evidence that improving the perceived intelligence of VAs results in more frequent use. Chattaraman et al. (2019) conducted a usability experiment with elders and found that for users with cognitive impairment, an informal personality in a VA was less effective and caused difficulty.

Intent detection involves translating a natural language command into a digital instruction and is a crucial component of a VA. The varying nature of natural language in its terminology, intonation, speed, and context makes this a difficult task that is accomplished only through machine learning such as the multi-layered neural network used by Firdaus et al. (2019) or the deep neural network used by Lin and Xu (2019) to learn new intents.

Mulfari et al. (2021) approached the task of designing VA system for users with speech disorders by using keyword spotting algorithms for intent detection. Kumar, Deepak, and Santhanavijayan (2020) propose an efficient emotion detection algorithm, although limited. As dementia significantly impacts speech, an appropriate method of understanding a user’s request or inferring their need is crucial.

# Methodology

## Objectives

In this project a VA intended to reduce the rate of cognitive decline in people diagnosed with dementia was created with the aim of answering the research question: Can a VA help reduce the rate of loss of cognitive functioning?

## Development Methodology

The project was structured using elements of Agile development practice. The project was broken down into stages (Appendix A) each focusing on a particular feature in line with the Agile principal of continuous delivery. Because of the limited timeframe of the project, each stage is aligned with a single sprint. Each stage of the project would also feature its own research into the specifics of that feature to determine how best to implement it.

Development took place between February and May of 2021.

Unlike typical Agile development, which is designed to facilitate good collaboration between team-members and stakeholders in projects with changing requirements, this project was an independent effort with fixed requirements.

## Management

The development of the project was managed using GitHub tools to track required tasks and milestones. For each stage of the project, the tasks within that stage were outlined as GitHub Issues. These Issues were assigned to Milestones for each sprint and tracked in a GitHub Projects Kanban board.

The key metrics of success for software development projects are velocity and effort (Kupiainen, Mäntylä and Itkonen 2015). Effort accounts for the time required and difficulty of a task. High velocity projects deliver small improvements to the product frequently.

Time management was handled by a burndown chart (Appendix B). Although this approach assumes equal effort is required for each stage, it ensured that time was distributed evenly throughout the project, and no stage would reduce the quality of another.

## Evaluation

### Software Evaluation

Ideally, the VA would be tested by users with cognitive impairment over an extended period to ascertain a change in their rate of cognitive decline, however there are several issues with this approach. Elders with cognitive decline are vulnerable persons and obtaining ethical clearance to work with these individuals is beyond the scope of this project; additionally, people living with dementia may not be able to provide informed consent of participation. The results of this evaluation would also be questionable as the benefits of dementia care technologies are often only perceivable with longer follow-up times and the timeframe of this project would not allow such an extended study. As cognition loss is inevitable with or without intervention, identifying whether a change in cognition is attributable to the VA would also require a large sample size.

Instead, the suitability of the developed artefact was measured by comparing it to existing frameworks for dementia care technologies. Additionally, each new feature is informally evaluated throughout development. This recurring evaluation falls under the iterative development ideology of Agile and highlights inadequacies of design, poor usability, and poor performance as quickly as possible so they can be rectified.

It is difficult to compare VAs because of varying functions and capabilities (Jiang et al. 2015).

A list of voice commands covering the use of the implemented features was created after the development stages to measure the speed and accuracy of the VAs responses. These commands were spoken to the VA for evaluation at the end of the project. Each command was repeated five times to collect a range of results and reduce the impact of outliers. These commands were decided following development as the number and nature of the included features was unknown beforehand.

The time-to-process (TTP) of each command was measured from the output logs of the VA beginning with the timestamp where the VA detected the user had begun speaking to the timestamp when the VA finished processing the command (including the time to compose a response, but not including the time to speak it).

The accuracy of transcription was measured using Levenshtein distance to compare the difference between the phrase the user knowingly spoke and the automatic transcription of the phrase that was heard.

Additionally, the expected outcome of each command was compared to the actual outcome.

* :Robustness
  + Count frequency of errors during example use (test plan)
  + Accuracy of TTS
  + Correctness of intent detection
  + Speed of transcription

The basic interaction will be measured by the accuracy and response time of its intent analysis component against a library of sample commands.

An automated test-suite for chatbots and VAs, Botium, was considered to evaluate the VA. Botium is able to simulate human input errors which could be adapted to represent unclear speech inputs from people with cognitive decline. Ultimately, Botium was not used for testing as it was not feasible to develop a custom adaptor for the VA in addition to developing the VA itself within the project timeframe.

### Process Evaluation

The project management was evaluated using the burndown chart to confirm that the assigned hours were completed at a consistent rate. Additionally, by comparing the developed software to the requirements initially set out, it can be determined that the time was spent appropriately by covering all initial determined if to determine if all requirements were met.

## Ethics

### Data Collection

The developed artefact constantly collects auditory samples of users’ voices and coincidental background sounds while active. Although the software does not request sensitive information, it may inadvertently collect such data if spoken aloud in proximity to the attached microphone. Any information collected is only parsed for keywords relating to the VA and is not transmitted outside of the software or persisted in memory after processing.

### Testing

There are no ethical requirements related to testing the GLAD VA as it will not involve user testing. An ethics approval was obtained for the project.

## Legal

The project has been created under a GNU GPLv3 license allowing open modification and redistribution of the software, provided any derivative is not closed-source. Any third-party technologies or other software used within this project are provided under a license that allows these same permissions.

# Planning

## Requirements

Following the initial literature review, a concept map based on common keywords found in the literature was created (Appendix). This concept map supported the definition of the key needs and possible features for the proposed VA.

The needs and features were initial divided into functional and non-functional (Table 1), and then further sorted using MoSCoW prioritisation (Figure 2). This helped to plan the development stages.

|  |  |
| --- | --- |
| Functional | |
| FR1 | Speech to text for staggered speech |
| FR2 | Natural language processing |
| FR3 | Text to speech output |
| FR4 | Offline functionality |
| FR5 | Minimal user input needed (i.e. yes or no questions) |
| Non-functional | |
| NFR1 | Set Reminders |
| NFR2 | Cognitive training exercises |
| NFR3 | Music therapy |
| NFR4 | Consistent personality |
| NFR5 | Independent Action |
| NFR6 | Extendable |

Table 1: Functional and non-functional requirements

|  |  |
| --- | --- |
|  |  |

Figure 2: MoSCoW prioritisation of requirements

The must-have requirements for GLAD include the defining elements of a VA: Speech-to-text (STT), natural language processing (NLP), and text-to-speech (TTS). The requirements that should be met cover the features that distinguish GLAD from other VAs, specifically its ability to function offline and minimal interaction. Could-have requirements cover the main features that are intended to be included: Reminders, cognitive training, and extendibility. The remaining requirements – music therapy, consistent personality, and the ability to take independent actions – are sorted under would-have, meaning they will only be included as supplementary features if feasible but are of the lowest priority.

## Risk Assessment

Risks to the project completion were also evaluated at this stage (Appendix), and mitigation strategies were considered and implemented for the most severe risks.

## Development environment

An Ubuntu 20.04.2 virtual machine (VM) was set-up for development and testing through Oracle VirtualBox 6.1.16. This machine had 4GB of RAM and 10GB of storage. Development of the python program was done using PyCharm Community 2021.1. During development, as a result of memory issues that arose, the machine was recreated with 20GB of storage.

A second VM with 2GB of RAM and 8GB of storage running Debian GNU/Linux 10 to simulate a Raspberry Pi for testing the VA on limited hardware. Problems installing the required python libraries prevented this VM from being used.

## Prototyping

An initial prototype for the VA (Appendix) was created to evaluate the use of a self-hosted back-end for Mycroft (Mycroft AI Inc. n.d.) – an existing open-source VA – with the intention of developing further features as skills. The back-end services required too much disk space and processing power to run on low-performance hardware and would have done so at a considerable disadvantage. As this could not be completed in a timely manner and would have negatively impacted performance, the Mycroft-based prototype was not used for the full development.

# The Voice Assistant

## Components of a Voice Assistant

The principal components of a voice interaction system are STT, NLP, and TTS. Together, these technologies are able to register a spoken command, identify and act upon it, and vocalise a response; allowing for conversational software interfaces. These components may operate programmatically or utilise machine learning techniques. Open-source implementations of these components were evaluated and selected to create the basis for GLAD.

## Technologies

### Voice Assistants

Although there are many familiar consumer VAs such as Amazon’s Alexa, Google’s Google Home, Apple’s Siri, and Microsoft’s Cortana, digital privacy concerns have prompted the development of a number of open-source alternatives.

Mycroft (Mycroft AI Inc. n.d.) is an open-source VA comparable to most popular consumer products and designed to run on a range of devices including desktops, Raspberry Pis, or on specialised hardware – the Mycroft Mark I and Mark II. It features an existing library of installable skills and its back-end consists of a suite of cloud-hosted services for STT, NLP, and TTS, and supporting infrastructure such as account management.

Open Assistant (Open Assistant n.d.) is designed for controlling computer operations using voice commands. It is written in Python and is capable of running entirely self-contained and without requiring internet access. Unlike other VAs, it does not respond to a wake word and instead is “always on”, listening constantly for any of its pre-defined commands. Its key-phrase-based intent detection and strict pattern matching limits its effectiveness.

The Jasper Project (Jasper Project n.d.) is an open-source voice interface for Raspberry Pis. As with Open Assistant, Jasper is always-listening but has few native capabilities. It was considered for the initial GLAD prototype but was rejected as it is written in the now unsupported Python 2.

### Speech-to-Text

Modern STT technology takes advantage of machine learning techniques to effectively identify spoken words. DeepSpeech is deep learning STT engine developed by Mozilla Corporation (2020) that uses Google TensorFlow. There are pre-trained models available for TensorFlow covering many languages and these can be easily integrated into DeepSpeech clients.

Most STT technologies are designed around users with an average level of diction; however this assumption cannot be made when designing for the cognitively impaired. Mulfari et al. (2021) demonstrated the use of AI-driven speech detection for users with staggered speech. A keyword spotting model was trained on audio samples of ten users with speaking impediments caused by cerebral palsy and similar conditions. The principal difficultly encountered was the limited sample audio available on which to train the model, resulting in a model only capable of identifying specific keywords.

### Natural Language Processing

Automated intent detection through NLP is generally used in marketing feedback to ascertain customer opinions on products and services. For VAs, NLP is the component that determines how to interpret a spoken sentence and can include identifying the phrase as a command, identifying parameters of the command (such as the time to set an alarm for), or determining whether a spoken utterance is directed at the VA.

Consumer VAs typically employ ML techniques for NLP as in Firdaus et al. (2019) and Lin and Xu (2019). In other cases such as Mulfari et al. (2021), keyword matching can be an effective low processing approach.

### Text to Speech

Voice synthesis or TTS involves generating audio clips from input text. Historically this was accomplished by deconstructing the text into syllables and concatenating phenome samples from a given library into a sentence. While these results were comprehensible, they were distinctly non-human. ML and improved algorithms have led to modern projects such as Resemble.AI (Resemble.AI, 2020) being able to produce lifelike voices in real-time.

For VAs, the chosen TTS implementation should produce a clear voice in a quick manner. In addition to the clarity of the chosen voice, the personality of a VA can impact its success (Poushneh 2021). The personality is affected by the sound of the voice and the tone of the language it uses. From their experiment with cognitively impaired and unimpaired individuals, Chattaraman *et al.* (2019) found that users with cognitive impairment found an informal tone in VAs more difficult to understand because of its indirect nature, instead preferring instructions to be clear and brief.

## Design

As mentioned previously, an initial prototype was created based on running an offline version of the Mycroft back-end services, although the concept was not successful. This would have allowed a Mycroft instance to serve as the interface to the VA by connecting to a local back-end and the other features would then be created as skills for Mycroft. This idea was discarded because the required back-end services were too memory intensive for a self-contained offline version and the prototype was not completed.

GLAD was instead developed based on Open Assistant, which has a modular architecture comprised of the following components: Ear – refers to the module responsible for receiving auditory data; Speech Recognition – refers to the STT module that processes the data; Minds – simple intent detection modules that compare the text with known commands; Abilities – modules that are triggered by commands to perform a task and generate a response; Voice – the STT module; and Sound – the audio playback module. Each component operates on a separate thread, allowing concurrent processing of audio, intents, and any feature the user has activated. Each thread listens to its own input queue on a loop and can output messages to the input queue of another thread. A central “hub” is responsible managing the messages between components (Figure 3).

Ear uses a peripheral microphone to listen for audio. When the audio reaches a volume threshold, it begins recording until the volume drops below the threshold for an extended duration. A buffer is also used to capture leading audio before the recording. The volume threshold dynamically adjusts by monitoring background noise. When a complete recording is assembled, the data is forwarded to the input queue of Speech Recognition.

Speech Recognition takes the recorded audio and processes it to detect spoken words using PocketSphinx (cmusphinx n.d.). This implementation is limited to detecting words from a pre-defined list.

A mind is a program running within the VA. Each mind is a python script with functions that are run when a tagged keyword is detected by Speech Recognition. Only one mind can be active at a time, and speech input is only checked against the active mind’s keywords. The active mind can be swapped during runtime by functions that may be triggered by other minds. Functions may send system events, run external programs (such as a media player), shutdown the VA, or activate the audio out modules: Voice and Sound.

Voice uses pyttsx3 (Bhat 2017) to generate a synthesised voice from an input string. pyttsx3 is a wrapper library for TTS engines that works offline.

The Sound module can be used to play local MP3 sound files. It is used to provide audio cues in addition to the vocal responses given by Voice.

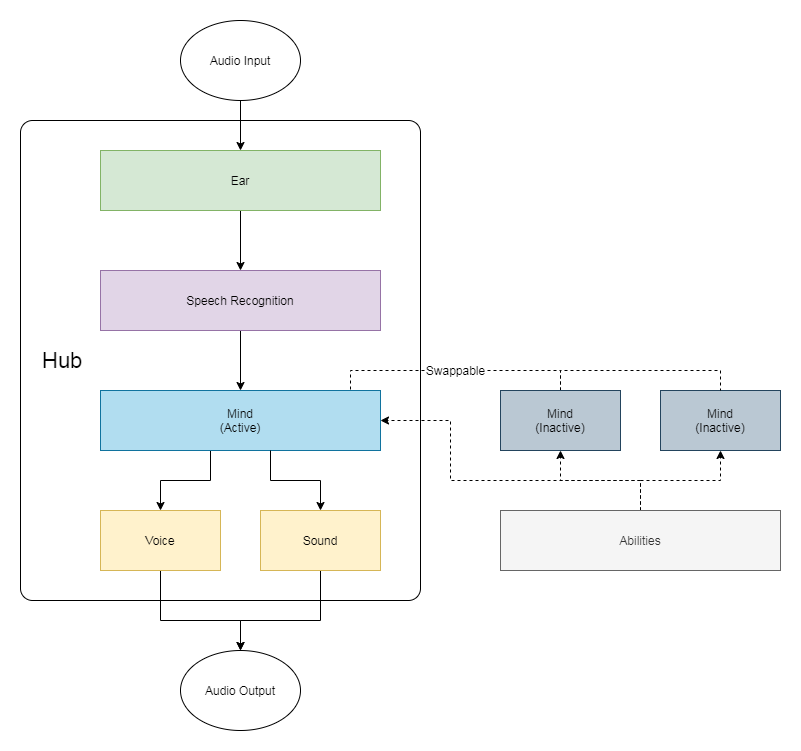


Figure 3. Open Assistant Architecture

## Implementation

GLAD, being based on the Open Assistant architecture, is written in Python 3. The modular design allowed for some components to be replaced as needed.

The included speech recognition module was insufficiently robust for GLAD’s needs as it was only capable of understanding words which were specified as keywords in the loaded minds. As a consequence, all spoken words were interpreted as the closest matching keyword. While this was suitable for recognising pre-defined phrases, it gave inaccurate results with single words; resulting in frequent misunderstandings. Additionally, the intent detection algorithm required that spoken commands must exactly match the pre-defined phrases, severely limiting the VAs adaptability and suitability for cognitively impaired users.

To address the robustness concern, the Speech Recognition module was modified to instead implement Mozilla *DeepSpeech* (Mozilla Corporation 2020). A pre-trained model was included in the VA that DeepSpeech uses to recognise a larger vocabulary than the PocketSphinx implementation. Additionally, the intent detection algorithm was improved to be more lenient when matching phrases by searching the strings for keywords rather than complete matches.

## Evaluation

The basis for GLAD was completed within the allotted timeframe of the first sprint.

The default voice synthetisation for Open Assistant is less sophisticated than other available options; however it was sufficiently clear for the needs of GLAD and was not replaced.

The pre-trained model for DeepSpeech is intended for general use and clear voices. Although it may not be suitable for the intended users, it was not feasible to obtain the data necessary and train a new model within the timeframe of this project. The design allows the model to be easily replaced at a later point with a more suitable one.

Being fully self-contained, the VA fulfils the requirement of being offline and Open Assistant’s minds and abilities – comparable to skills in Amazon’s Alexa – fulfils the extendibility requirement.

Open Assistant is licensed under GNU GPLv3, and Mozilla DeepSpeech is licensed under the Mozilla Permissive License. Both technologies are open-source and their licenses provide the required level of restriction.

# Cognitive Exercises

## Introduction

Because dementia and related conditions are largely irreversible, the only course of treatment is to slow the rate of cognitive decline. To this end, efforts have been made to incorporate cognitive exercises in the form of serious games into the GLAD VA. Additionally, music therapy has been shown to have positive effects on people living with dementia, and a radio feature has been developed for the VA.

## Exercises

### Serious Games for Dementia

Serious games can be an effective tool for improving cognition as they promote engagement and exercise perception and memory functions through an enjoyable medium. Although typically used in educational environments, serious games have seen successful applications in many fields including dementia care. These games have been developed specifically for individuals with cognitive impairment with the aim of exercising their short-term memory, planning skills, spatial awareness, and reactions.

Chi, Agama, and Prodanoff (2017) created a Smart Thinker, a suite of four games to improve player attention and memory including High-Low, which has players compare two numbers; Color Game in which players must select squares matching a given colour; Rock-Paper-Scissors, a popular hand game; and Find Me, a text game of keeping track of objects.

Serious games have also found use as a tool for screening people for cognitive impairment through emulating ADLs and scoring players on their success (Zucchella *et al.* 2014; Tong, Chan, and Chignell 2017).

Imbeault, Bochard, Bouzouane (2011) highlight the importance of serious games being adaptive and supportive for users with cognitive impairment. They apply AI techniques to evaluate user performance during a cognitive skill test and adjust difficulty accordingly.

### Music Therapy

Music therapy is suggested as another means of reinforcing memory skills by triggering old memories. Music therapy is effective and enjoyable, with benefits to the recipients’ memory. It has also been successful in improving patient relationships with care-givers.

Music therapy carries a risk of triggering unpleasant memories; however this cannot be practically anticipated.

### Physical Exercise

In addition to brain-training, there is evidence of the effects of physical exercise on cognitive functioning (Heyn, Abreu, Ottenbacher 2004); however the results of such studies have found only a moderate correlation between physical exercise and improved cognition. There are other psychological benefits to physical exercise (Thuné-Boyle *et al*. 2012) such as improved confidence and mood, and a VA could potentially provide guidance for a physical exercise routine to encourage active behaviour in users.

## Implementation

### Number Game

A simple number recognition game inspired the *High-Low* game by Chi, Agama, and Prodanoff (2017) was created for the VA. This game tests short term memory and basic numeracy by tasking players with recalling a number spoken by the VA and identifying whether a second number is larger than, smaller than, or equal to the first.

This game uses the VAs voice detection to receive commands and responses from the player, eliminating the need for graphical interfaces or physical controllers which can be challenging for older adults.

There a five levels of gameplay in the number game with increasing ranges of possible numbers and lengthening delays between the numbers given. To maintain simplicity of design for elder users, players are automatically progressed to the next level of difficulty when they are performing well, and regressed to an easier level if they appear to be struggling. User profiling would allow tailoring the difficulty for specific individuals; however this would require storage of personal information that could be considered sensitive.

### Radio

The VA was given a digital radio player that is programmed to the URLs of m4d Radio (m4d Radio, n.d.), a collection of five digital radio stations designed for people living with dementia. These stations are available from streaming URLs and feature popular music from different decades of the 20th century with the intention of stimulating positive memories from listener’s youth and young adult years.

The radio can be triggered by the keywords music or radio. To reduce the user input required, the possible stations are hard-coded and one is selected at random whenever music is requested. The user is however able to request a different station, prompting another station to be randomly selected.

The radio stations are streamed using a python library for *VLC media player.*

### Challenges

To facilitate these features, the VA required more robust interaction capabilities. A mechanism was integrated into the intent detection module to allow users to provide an answer to a prompt, and adjustments to the keyword search algorithm allowed keywords with wild card characters.

The prompt-answer interaction works by keeping a global dictionary of user options relating to a previous response. Minds are able to set the content of this dictionary to include options the user may select and functions to run for each answer. For example: when asking the question “Do you want to play the number game?”, the mind would set the dictionary to include two answers – “yes” and “no” – a provide functions to run when each answer is detected. If these options are set, then any detected command is first checked against the dictionary if there is an outstanding question. If no match is found, or there are no options set, the normal flow of data (from speech detection to the active mind) continues. When an answer is given, the dictionary is cleared until needed again. This mechanism of question and response provides a more conversational interface for the VA.

With an increasing number of features, the limitations of using keywords for intent detection became more apparent. A function for parsing wild card characters in keywords was included to enable more versatile pattern matching.

This new interpretation of spoken commands had the consequence of a spoken command matching the pattern of multiple functions. In these cases the most common function is used.

## Evaluation

The benefits of each of these features on user cognition have not been tested. It could be compared to a framework.

The number game very simple and unengaging. The entertainment factor of the game has not been evaluated. The difficulty of the game has not been evaluated.

Use of the radio requires an internet connection to receive broadcasts. This is in violation of the offline-usage requirement. To rectify this, a catalogue of songs could be included.

The radio is usable and its operation has been kept simple.

# Considerations for Elderly Users

## Introduction

As dementia and related conditions are most prevalent among the elderly, it should be assumed that the primary users of the VA will include digital immigrants with limited technological expertise and early-stages of cognition loss. The VA should therefore make efforts to support this unique subset of users by following guidance on designing technology for both the elderly in general and those with limited cognition.

## Research

Avoid complex UI

Not relevant to VA

Single path for each task

Iancu and Iancu (2020)

* + Mobile design considerations for elders
  + Single path for tasks
  + Consider speech speed

There is also a population of early adopters and innovators of IT who have now matured into elder age. The interviews Kania-Lundholm and Torres (2015) did with this population found that many of them still maintain a competence with technology. This suggests that current younger tech users may also preserve a measure of technological competence as they mature, and a growing number of elderly people adopting technology has been observed (Fischer et al. 2014).

It should not be overlooked that caregivers have needs of their own. A VA is in a unique position among AL technologies in that is can additionally provide support to the care-giver themselves. These care-givers are likely to be more competent with technology than their care-recipients, although they may lack formal training.

Thoma-LÃ¼rken,Theresa;Bleijlevens,Michel H. C.;Lexis,Monique A. S.;de Witte,Luc P.;Hamers,Jan P. H. 2018

When designing software for users unfamiliar with IT, it is important to make clear the options available and consider appropriately identifiable images, auditory cues, and terminology (Castilla et al. 2013). For a VA, this applies to the choice of language used, avoiding technical words in favour of metaphors that may be familiar to elderly users, and sound cues to strengthen mental connections between tasks and a goal.

* UI
* Navigating a world was tricky

Limited hand-eye co-ordination makes traditional computer interfaces unsuitable for elderly users, making alternative methods such as touch screens, eye-tracking, gesture control (Hsiao et al. 2017), and voice control desirable.

* Plaza,Inmaculada;MartÃ¬n,Lourdes;Martin,Sergio;Medrano,Carlos 2011
* Chattaraman,Veena;Kwon,Wi-Suk;Gilbert,Juan E.;Ross,Kassandra 2019

## Implementation

The audio detection module was configured with longer pause times to support users with staggered or slow speech.

The speed of GLADs spoken speech was lowered to 100 words-per-minute. Commands were included to adjust this speed during use. Changes to this option are persisted in a configuration file.

Yes-no questions were possible with question-answer mechanism implemented previously, however their prevalence required a more uniform way of creating them. A function was created that standardised the options of yes-no questions. As part of this it included extra affirmations (“sure”, “okay”), denials (“nah”, “don’t”), and homonyms (“shore”, “know”).

To encourage better use of the VA, several soundbites were created and included. These include jingles for when the VA starts and stops, and confirmation and denial sounds for use in the number game when a player gives an answer. The soundbites were created in a digital audio editor using soft piano and bell synthesisers to evoke a comforting atmosphere for users. The sound module is used to play the effects when called.

A feature that was not completed was the independent monitoring. The VA does not attempt to engage with the user without prompting, requiring the user to initiate any interaction. The VA could be improved with a feature in which it suggests playing music if it detects a prolonged silence. This unprompted response could be intrusive and would need to be configurable.

## Evaluation

Needs of elderly users were considered during the early stages of development and played a part in selecting the technologies that would form the basis of the VA. Only a few changes to the existing architecture were required at this stage.

The bulk of this stage of development saw improvements and modifications to the existing architecture including the removal of excess features from Open Assistant that were not required for GLAD, and stability improvements.

# Companionship

## Introduction

Companionship and social relationships have been consistently shown to be very important for maintaining both physical and mental health, however for the elderly, social relationships are hampered by difficulty communicating (A. Palmer et al. 2016). The proposed VA could help overcome this obstacle and assist early-stage dementia sufferers with maintaining their social relationships and provide a degree of companionship itself.

## Research

Real and popular TTS voices were evaluated by Cambre *et al.* (2020) to identify those that were considered preferable for reading long content. While more “humanlike” voices were rated highly by listeners, they note that no generalizable metric for determining the quality of a TTS voice could be created.

Lord et al. (2020) express the need for dementia support technologies and treatments to be personal, respectful, and consistent for both patients and carers. They also highlight two successful interventions in line with their model that enabled extended autonomy and home-care of persons living with dementia: Maximising Independence at Home (MIND) and the New York University Spouse Caregiver Intervention (NYUCI). These interventions focus on group education and therapy for patients and family, supporting the idea of social activity prolonging patient cognisance.

## Implementation

At this stage of the project, a name and personality were selected for the developed VA. These aspects of the device were crucial to developing an atmosphere of trust and reliability for the device. Additionally, further improvements to the intent recognition were made, reducing misunderstandings and errors that could jeopardise user’s opinion.

In creating a personality for the VA, the name GLAD was decided to evoke a helpful tone.

As even advanced voice synthesisers struggle to believably introduce emotion to generated speech, it was considered preferable to not fake a human voice and personality for GLAD. Instead, GLAD speaks politely and clearly in a formal manner akin to a butler. GLADs simple voice synthesis was also preferred to keep it portable and suitable for installation on an independent device such as a Raspberry Pi. Raspberry Pi OS includes eSpeak, the voice engine used for GLAD, as standard.

To further improve the interaction between the user and the VA, the VA was programmed to respond to the user thanking it. This was accomplished by tracking the time since GLAD last spoke to the user and listening for the word “thank” within ten seconds. After this time, it is assumed that the user was not addressing GLAD, and the keyword “thank” is ignored.

An early concept for the VA included the ability to contact friends, family, and carers and have spoken conversations. This feature was not able to be implemented in the final design. Furthermore, it would have required an internet connection and storage of contact details to use. It is also possible that users may have preferred to communicate with through more familiar means such as a telephone.

At this stage efforts were made to include a timer/alarm function for the VA allowing it to provide assistance in other ADLs such as cooking or medication scheduling. Two obstacles were encountered during this implementation: The existing architecture was not able to extract parameters from given commands; and the grammatically complex task of translating a spoken time (such as “quarter-to-eight in the evening”) into a computer-readable time (19:45). The necessary changes were made to the architecture to support custom parsing of a spoken command during the function that it called; however there was insufficient time to implement the parsing function for time selection.

Time date functions

## Evaluation

The VA is limited in its ability to offer companionship without user profiling.

# Evaluation

## Developed System

### Requirements

At the completion of development, the overall system was considered against the requirements set-out initially. The following requirements were met: FR1, FR2, FR3, FR4, FR5, NFR2, NFR3, NFR4, and NFR6. NFR1 and NFR5 were not met.

As discussed previously, the alarm feature that would have covered NFR1 was not implemented within the timeframe because of unforeseen complexity in parsing spoken times. NFR1 was prioritised as a “could have” requirement and was not critical to the operation of the VA.

Autonomous action was not able to be implemented in the VA because of delays caused by issues arising in earlier sprints, missing requirement NFR5. As NFR5 was a lowest-priority requirement, its omission is not a considerable loss.

* Artefact not ready for real-world use.
* Challenges to development.
* Poor extendibility.
* Difficulty evaluating.

While persons suffering with late-stage dementia will likely still require constant support, this system should allow persons with early stage dementia and pre-dementia to continue living independently for a longer period before requiring more consistent care.

Comparing the capabilities of the developed VA to existing commercial products would be an unproductive evaluation as they are designed to operate on an extensive infrastructure and can deliver superior performance.

The final artefact was not completed to a satisfactory level despite the majority of requirements being addressed.

Most dementia cases are present in countries with low income. Sufferers in these countries would not directly benefit from the development of the proposed voice assistant; however reducing the dependency on human care-workers in developed countries will open the possibility of aid for lower income countries. Additionally, studies on dementia are often done in high income countries, how effective the techniques developed are for low income countries is uncertain.

### Speed and Accuracy

33 voice commands covering the developed features of the system (including greetings, the number game, the radio, and affirmations) were decided upon. In some cases, multiple phrasings of the same command were tested. These commands were then spoken to the VA; with each repeated 5 times.

117 of the 165 total commands were correctly transcribed without error. 48 commands had some error (<100% accuracy) with 18 having considerable errors (<50% accuracy).

144 of the 165 commands were accurately processed, regardless of transcription accuracy. Of the 21 that did not give the expected response, no command was misinterpreted as another command.

Overall, the VA had an 87% accuracy of transcription and 87% accuracy of intent detection. It should be noted that 27 commands were incorrectly transcribed but produced the expected response.

The average time to respond to a command was 3.574 seconds, with the fastest response being 2.204 seconds and the slowest being 8.931 seconds.

Variations in processing speeds can be attributed to . The processing

It was found during development that the STT component of GLAD can take noticeably longer to process longer spoken phrases. Although no excessively lengthy commands were used during this test, there is a correlation between the length of the transcribed phrase and the processing time.

Figure 4. Chart of processing time by length of phrase

Processing speed includes cases where a sound-effect plays before a response is given, which causes some delays.

## Process

Many dementia care technology projects strongly recommend developing the system alongside potential users including carers and people living with dementia. This co-development approach ensures that the desires of end-users are incorporated more effectively, and improves the relationship between developers and users making the ultimate adoption of the technology more likely. Although this would have been ideal, it was not possible with this project because of the aforementioned restrictions working with vulnerable persons in this scope.

Ultimately, the day-to-day breakdown of hours was not strictly followed because of other commitments.

Poor assessment of task complexity. Alarm complexity unexpected.

Poor planning of development unknown complexity of features.

Initial literature review method failed and a newer approach was needed.

# Further Research

A completely auditory assistant requires the user to have a good short-term memory to track the state of the VA and tasks. As this is a notable weakness for people living with dementia, a visual display may be more effective.

Integrating the VA into an existing smart-home environment would allow for greater amount of functions and features to be included. Alternatively, the VA should be installed on a physical device or dedicated platform.

Efforts could also be made to provide localisation or language options by moving command declarations and responses to external files to be selected at runtime.

Additionally, more features for the VA could be developed including further games (possibly utilising A.I. for more dynamic experiences).

# Conclusion

This project involved the design and creation of a VA to assist people living with dementia with the aim of slowing their rate of cognitive decay.

A systematic literature review was conducted to assess existing dementia care technologies. From this review, possible features of the VA were considered and the requirements for the system were outlined. The design and development process of the project was then explained. Each stage of development is then described, highlighting the features created, supporting technologies, and challenges faced. The finished artefact was then compared to the requirements initially set-out and evaluated. Finally, suggestions for further work to prepare the VA were given.

This project has shown that it is feasible to create a lightweight VA for people living with dementia that can operate without an internet connection.

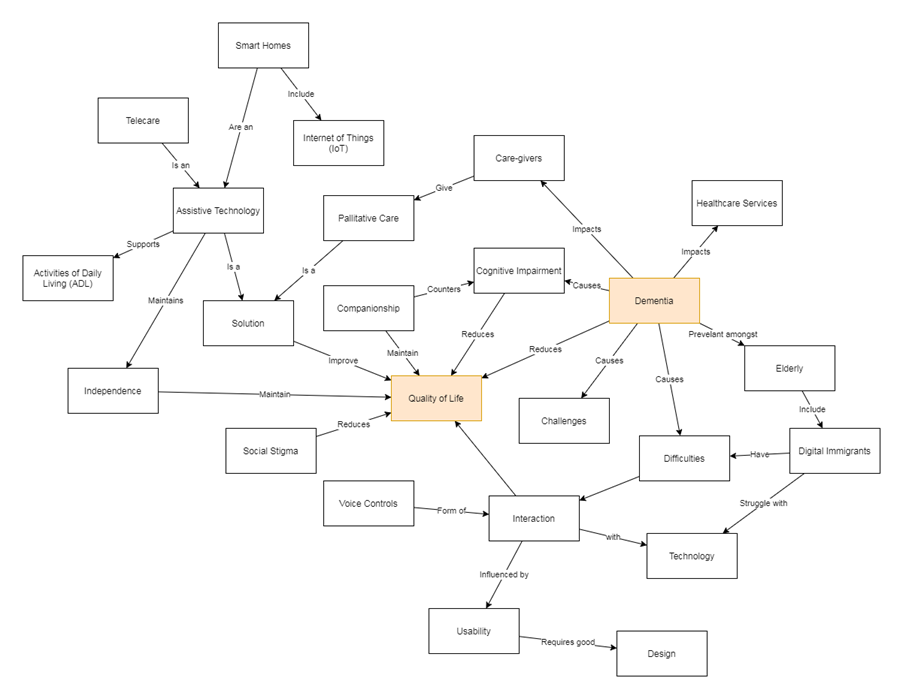
The developed VA has not been tested by potential users and it cannot be confirmed that it is suitable for the intended usage.

The VA lacks many convenience features, and provides only limited functionality compared to existing products.

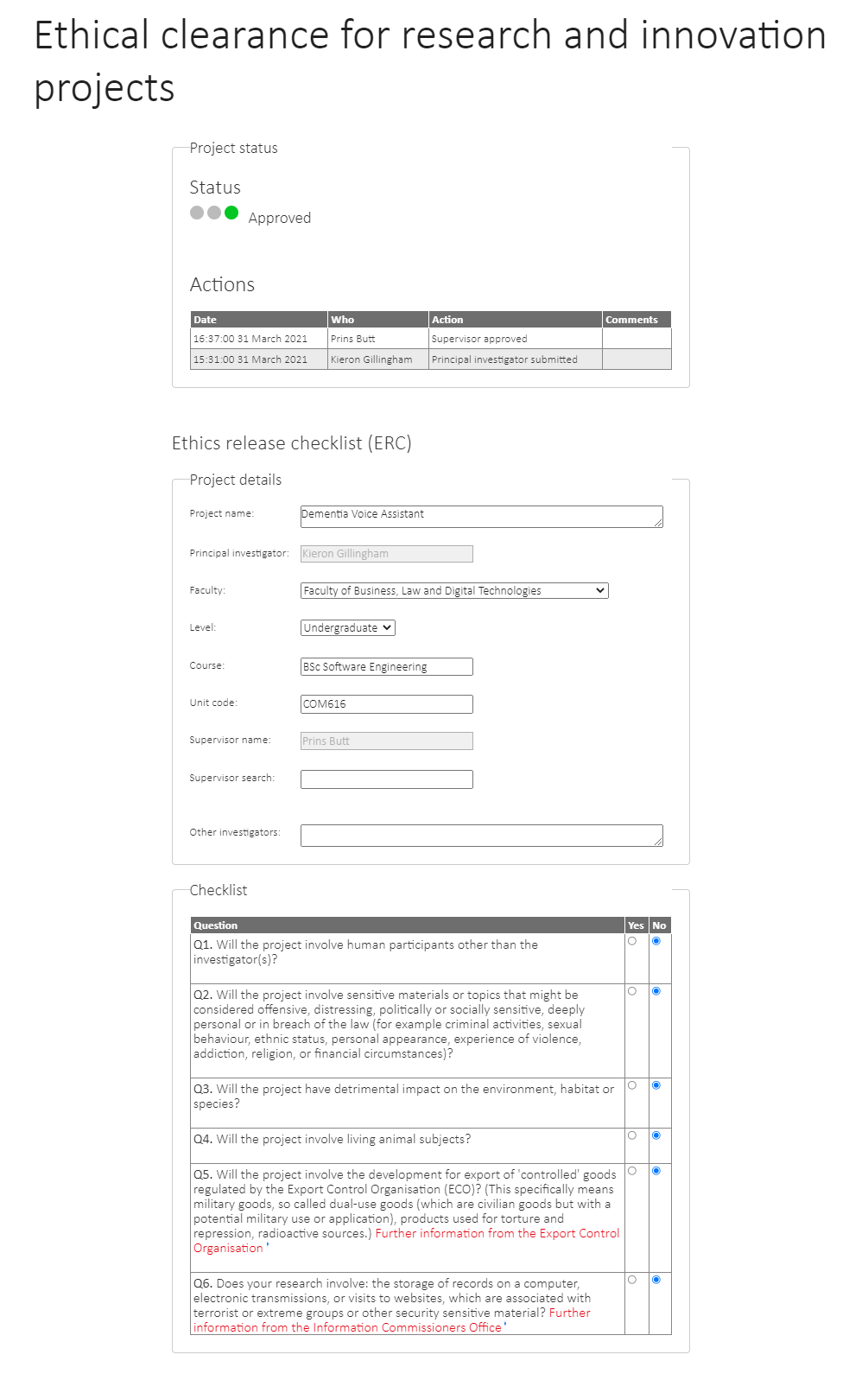
# References

# Bibliography

# Appendix A: Concept Map



# Appendix B: Ethics Release



# Appendix C: Results

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID | Prerequisites | Spoken Phrase | Heard Phrase | Accuracy of Transcription | Average Accuracy | Time to process (TTP) [seconds] | Average TTP [seconds] | Expected outcome | Outcome success | Outcome Accuracy |
| **1.1** | VA is active | Hello | HELLO | 100% | 100% | 03.850 | 03.087 | Respond "hello" | Yes | 100% |
| **1.2** | HELLO | 100% | 02.973 | Yes |
| **1.3** | HELLO | 100% | 02.907 | Yes |
| **1.4** | HELLO | 100% | 02.961 | Yes |
| **1.5** | HELLO | 100% | 02.742 | Yes |
| **2.1** | VA is active | What time is it | WHAT TIME IS IT | 100% | 100% | 08.931 | 04.982 | Respond with current system time | Yes | 100% |
| **2.2** | WHAT TIME IS IT | 100% | 04.025 | Yes |
| **2.3** | WHAT TIME IS IT | 100% | 03.651 | Yes |
| **2.4** | WHAT TIME IS IT | 100% | 03.939 | Yes |
| **2.5** | WHAT TIME IS IT | 100% | 04.362 | Yes |
| **3.1** | VA is active | What is the time | BOYS THE TIME | 62% | 92% | 04.717 | 04.462 | Respond with current system time | Yes | 100% |
| **3.2** | WHAT IS THE TIME | 100% | 05.331 | Yes |
| **3.3** | WHAT IS THE TIME | 100% | 03.527 | Yes |
| **3.4** | WHAT IS THE TIME | 100% | 04.981 | Yes |
| **3.5** | WHAT IS THE TIME | 100% | 03.756 | Yes |
| **4.1** | VA is active | What day is it | WHAT DAY IS IT | 100% | 100% | 04.373 | 04.029 | Speak the day of the week | Yes | 100% |
| **4.2** | WHAT DAY IS IT | 100% | 04.415 | Yes |
| **4.3** | WHAT DAY IS IT | 100% | 04.472 | Yes |
| **4.4** | WHAT DAY IS IT | 100% | 03.350 | Yes |
| **4.5** | WHAT DAY IS IT | 100% | 03.533 | Yes |
| **5.1** | VA is active | Who are you | WHO ARE YOU | 100% | 91% | 03.602 | 02.956 | Introduce self | Yes | 60% |
| **5.2** | HOW ARE YOU | 82% | 02.746 | No |
| **5.3** | WHO ARE YOU | 100% | 02.887 | Yes |
| **5.4** | WHO ARE YOU | 100% | 02.860 | Yes |
| **5.5** | WHERE ARE YOU | 73% | 02.684 | No |
| **6.1** | VA is active | What is your name | WHAT IS YOUR NAME | 100% | 100% | 03.937 | 03.328 | Introduce self | Yes | 100% |
| **6.2** | WHAT IS YOUR NAME | 100% | 03.246 | Yes |
| **6.3** | WHAT IS YOUR NAME | 100% | 03.215 | Yes |
| **6.4** | WHAT IS YOUR NAME | 100% | 03.179 | Yes |
| **6.5** | WHAT IS YOUR NAME | 100% | 03.065 | Yes |
| **7.1** | VA is active | What can you do | WHAT CAN YOU DO | 100% | 96% | 04.172 | 03.692 | List functionality | Yes | 100% |
| **7.2** | WHAT CAN YOU DO | 100% | 03.927 | Yes |
| **7.3** | WHAT CAN YOU DO | 100% | 03.502 | Yes |
| **7.4** | WHAT CAN YOU DO | 100% | 03.187 | Yes |
| **7.5** | BUT CAN YOU DO | 80% | 03.674 | Yes |
| **8.1** | VA is active | Put some music on | PUT SOME MUSIC ON | 100% | 90% | 04.519 | 04.182 | VA receives command to play the radio | Yes | 100% |
| **8.2** | BUT SOME MUSIC ON | 94% | 03.528 | Yes |
| **8.3** | BUT SOME MUSIC ON | 94% | 03.609 | Yes |
| **8.4** | THAT SOME MUSIC ON | 82% | 03.513 | Yes |
| **8.5** | THAT SOME MUSIC ON | 82% | 05.743 | Yes |
| **9.1** | VA is active | Play the radio | PLAY THE RADIO | 100% | 99% | 03.488 | 04.095 | VA receives command to play the radio | Yes | 100% |
| **9.2** | PLAY THE RADIO | 100% | 03.214 | Yes |
| **9.3** | PLAY THE RADIO | 100% | 03.805 | Yes |
| **9.4** | LAY THE RADIO | 93% | 03.828 | Yes |
| **9.5** | PLAY THE RADIO | 100% | 06.140 | Yes |
| **10.1** | VA is active | I want the radio on | I WANT THE RADIO ON | 100% | 100% | 04.642 | 04.096 | VA receives command to play the radio | Yes | 100% |
| **10.2** | I WANT THE RADIO ON | 100% | 04.049 | Yes |
| **10.3** | I WANT THE RADIO ON | 100% | 04.375 | Yes |
| **10.4** | I WANT THE RADIO ON | 100% | 03.739 | Yes |
| **10.5** | I WANT THE RADIO ON | 100% | 03.674 | Yes |
| **11.1** | Radio is playing | Change the station | JANE THE STATION | 83% | 77% | 03.642 | 03.488 | Change the radio station | Yes | 100% |
| **11.2** | CHANGE THE STATION | 100% | 03.822 | Yes |
| **11.3** | CHANGE THE STATION | 100% | 03.448 | Yes |
| **11.4** | CHANGE THE STATION | 100% | 03.348 | Yes |
| **11.5** | CAN TESTATION | 0% | 03.178 | Yes |
| **12.1** | Radio is playing | Put a different station on | BUT A DIFFERENT STATION ON | 96% | 65% | 04.397 | 04.716 | Change the radio station | Yes | 100% |
| **12.2** | CARTERET STATION ON | 0% | 04.502 | Yes |
| **12.3** | TO A DIFFERENT STATION ON | 88% | 04.757 | Yes |
| **12.4** | PORTENDUERE STATION ON | 62% | 04.905 | Yes |
| **12.5** | THERE DIFFERENT STATION ON | 81% | 05.018 | Yes |
| **13.1** | Radio is playing | Put the volume up | BUT THE VOLUME AT | 82% | 70% | 03.699 | 03.693 | Turn the radio volume up | No | 40% |
| **13.2** | BUT THE VOLUME UP | 94% | 03.281 | Yes |
| **13.3** | BUT THE VOLUME UP | 94% | 03.327 | Yes |
| **13.4** | BUT THE VALET | 0% | 03.650 | No |
| **13.5** | BUT THE VOLUME AT | 82% | 04.507 | No |
| **14.1** | Radio is playing | Lower the volume | HERE THE VALLEY | 50% | 85% | 03.285 | 03.730 | Turn the radio volume down | No | 80% |
| **14.2** | THE LOWER THE VOLUME | 75% | 04.265 | Yes |
| **14.3** | LOWER THE VOLUME | 100% | 03.949 | Yes |
| **14.4** | LOWER THE VOLUME | 100% | 03.361 | Yes |
| **14.5** | LOWER THE VOLUME | 100% | 03.788 | Yes |
| **15.1** | Radio is playing | Stop the music | STOP THE MUSIC | 100% | 96% | 03.581 | 03.877 | Stop the radio | Yes | 100% |
| **15.2** | STOP THE MUSIC | 100% | 03.673 | Yes |
| **15.3** | STOP THE MUSIC | 100% | 03.576 | Yes |
| **15.4** | STOP THE MUSIC | 100% | 04.316 | Yes |
| **15.5** | STOPPED THE MUSIC | 79% | 04.241 | Yes |
| **16.1** | Radio is playing | Turn off the radio | TURN OFF THE RADIO | 100% | 93% | 03.881 | 04.044 | Stop the radio | Yes | 80% |
| **16.2** | TURN OFF THE RADIO | 100% | 04.521 | Yes |
| **16.3** | TURN OFF THE RADIO | 100% | 04.377 | Yes |
| **16.4** | TURN OFF THE RADIO | 100% | 03.614 | Yes |
| **16.5** | THEN OTHER RADIO | 67% | 03.829 | No |
| **17.1** | VA is active | Play the number game | FLAY THE NUMBER GAME | 95% | 95% | 03.843 | 04.077 | VA receives command to play the number game | Yes | 100% |
| **17.2** | PLAY THE NUMBER GAME | 100% | 04.037 | Yes |
| **17.3** | PLAY THE NUMBER GAME | 100% | 04.265 | Yes |
| **17.4** | PLAYS A NUMBER GAME | 80% | 03.995 | Yes |
| **17.5** | PLAY THE NUMBER GAME | 100% | 04.247 | Yes |
| **18.1** | VA is active | Let's play a game | LET'S PLAY A GAME | 100% | 100% | 03.439 | 04.120 | VA receives command to play the number game | Yes | 100% |
| **18.2** | LET'S PLAY A GAME | 100% | 03.766 | Yes |
| **18.3** | LET'S PLAY A GAME | 100% | 03.949 | Yes |
| **18.4** | LET'S PLAY A GAME | 100% | 03.846 | Yes |
| **18.5** | LET'S PLAY A GAME | 100% | 05.601 | Yes |
| **19.1** | Game has started | What are the rules | WHAT ARE THE RULES | 100% | 100% | 03.764 | 03.940 | Explain the number game rules | Yes | 100% |
| **19.2** | WHAT ARE THE RULES | 100% | 03.879 | Yes |
| **19.3** | WHAT ARE THE RULES | 100% | 03.560 | Yes |
| **19.4** | WHAT ARE THE RULES | 100% | 04.377 | Yes |
| **19.5** | WHAT ARE THE RULES | 100% | 04.122 | Yes |
| **20.1** | Game has started | Higher | HIGHER | 100% | 100% | 03.443 | 03.593 | Accept "higher" as answer for question | Yes | 100% |
| **20.2** | HIGHER | 100% | 03.574 | Yes |
| **20.3** | HIGHER | 100% | 03.582 | Yes |
| **20.4** | HIGHER | 100% | 03.700 | Yes |
| **20.5** | HIGHER | 100% | 03.664 | Yes |
| **21.1** | Game has started | Lower | LOWER | 100% | 80% | 03.530 | 03.403 | Accept "lower" as answer for question | Yes | 80% |
| **21.2** | NO | 0% | 02.459 | No |
| **21.3** | LOWER | 100% | 03.736 | Yes |
| **21.4** | LOWER | 100% | 03.732 | Yes |
| **21.5** | LOWER | 100% | 03.560 | Yes |
| **22.1** | Game has started | The same | THE SAME | 100% | 100% | 03.624 | 04.246 | Accept "the same" as answer for question | Yes | 100% |
| **22.2** | THE SAME | 100% | 04.065 | Yes |
| **22.3** | THE SAME | 100% | 04.594 | Yes |
| **22.4** | THE SAME | 100% | 04.609 | Yes |
| **22.5** | THE SAME | 100% | 04.338 | Yes |
| **23.1** | Game has started | Stop playing | STOP PLAYING | 100% | 100% | 03.048 | 03.587 | Stop the game | Yes | 100% |
| **23.2** | STOP PLAYING | 100% | 03.638 | Yes |
| **23.3** | STOP PLAYING | 100% | 03.886 | Yes |
| **23.4** | STOP PLAYING | 100% | 03.738 | Yes |
| **23.5** | STOP PLAYING | 100% | 03.627 | Yes |
| **24.1** | VA is active | Talk faster | TALK FASTER | 100% | 100% | 03.170 | 03.426 | Increase VA talking speed | Yes | 100% |
| **24.2** | TALK FASTER | 100% | 03.238 | Yes |
| **24.3** | TALK FASTER | 100% | 03.164 | Yes |
| **24.4** | TALK FASTER | 100% | 03.296 | Yes |
| **24.5** | TALK FASTER | 100% | 04.261 | Yes |
| **25.1** | VA is active | Talk slower | TAKE ALL OUR | 36% | 44% | 03.371 | 03.438 | Decrease VA talking speed | No | 40% |
| **25.2** | TAKE OUR | 0% | 03.398 | No |
| **25.3** | TO LOWER | 0% | 03.458 | No |
| **25.4** | TALK LOWER | 91% | 03.502 | Yes |
| **25.5** | TALK LOWER | 91% | 03.462 | Yes |
| **26.1** | User has been asked to confirm an action | Yes | YES | 100% | 80% | 02.310 | 02.458 | Action is confirmed | Yes | 80% |
| **26.2** | YES | 100% | 02.342 | Yes |
| **26.3** | YES | 100% | 02.513 | Yes |
| **26.4** | THE | 0% | 02.648 | No |
| **26.5** | YES | 100% | 02.478 | Yes |
| **27.1** | User has been asked to confirm an action | Okay | OKAY | 100% | 100% | 02.494 | 02.780 | Action is confirmed | Yes | 100% |
| **27.2** | OKAY | 100% | 02.612 | Yes |
| **27.3** | OKAY | 100% | 02.567 | Yes |
| **27.4** | OKAY | 100% | 03.592 | Yes |
| **27.5** | OKAY | 100% | 02.634 | Yes |
| **28.1** | User has been asked to confirm an action | Sure | SHE | 0% | 25% | 02.743 | 02.529 | Action is confirmed | No | 20% |
| **28.2** | SHOW | 25% | 02.425 | No |
| **28.3** | SHE | 0% | 02.398 | No |
| **28.4** | SO | 0% | 02.471 | No |
| **28.5** | SURE | 100% | 02.610 | Yes |
| **29.1** | User has been asked to confirm an action | No | NO | 100% | 80% | 02.267 | 02.348 | Action is denied | Yes | 100% |
| **29.2** | NO | 100% | 02.204 | Yes |
| **29.3** | NOW | 0% | 02.352 | Yes |
| **29.4** | NO | 100% | 02.572 | Yes |
| **29.5** | NO | 100% | 02.343 | Yes |
| **30.1** | User has been asked to confirm an action | Nah | NO | 0% | 20% | 02.227 | 02.478 | Action is denied | Yes | 80% |
| **30.2** | NOT | 33% | 02.373 | Yes |
| **30.3** | NOT | 33% | 02.574 | Yes |
| **30.4** | THAT | 0% | 02.549 | No |
| **30.5** | NOT | 33% | 02.665 | Yes |
| **31.1** | User has been asked to confirm an action | Don't do that | DON'T DO THAT | 100% | 100% | 02.931 | 03.133 | Action is denied | Yes | 100% |
| **31.2** | DON'T DO THAT | 100% | 03.354 | Yes |
| **31.3** | DON'T DO THAT | 100% | 03.098 | Yes |
| **31.4** | DON'T DO THAT | 100% | 03.080 | Yes |
| **31.5** | DON'T DO THAT | 100% | 03.202 | Yes |
| **32.1** | VA is active | Turn off | TURN OFF | 100% | 100% | 02.907 | 02.782 | VA receives command to shut down | Yes | 100% |
| **32.2** | TURN OFF | 100% | 02.724 | Yes |
| **32.3** | TURN OFF | 100% | 02.746 | Yes |
| **32.4** | TURN OFF | 100% | 02.767 | Yes |
| **32.5** | TURN OFF | 100% | 02.767 | Yes |
| **33.1** | VA is active | Shut down | SHUT DOWN | 100% | 87% | 02.963 | 03.140 | VA receives command to shut down | Yes | 20% |
| **33.2** | SO DOWN | 67% | 03.314 | No |
| **33.3** | SHOT DOWN | 89% | 03.090 | No |
| **33.4** | SHOT DOWN | 89% | 03.086 | No |
| **33.5** | SHOT DOWN | 89% | 03.248 | No |

1. There is no consensus on a general term for this class of products. Alternative terms include intelligent virtual assistant (IVA), intelligent personal assistant (IPA), and smart speaker. For this document the term voice assistant (VA) will be used. [↑](#footnote-ref-1)