Initial Plan: Human-IA interaction for behaviour change

Author: Conor Kemp

Supervisor: Parisa Eslambolchilar

Module Number: CM3203

Module Title: One Semester Individual Project

Credits: 40

Project Description

Artificial Intelligence (AI) is a vast field consisting of multiple sub-fields that intersect with different subjects. The vastness of the subject results in various definitions being used to describe it. Within this project, AI will refer to “the study of the design of intelligent agents” (Poole et al. 1998). An intelligent agent (IA) refers to “a system that acts intelligently” (Poole et al. 1998). Defining traits of IAs are that they are flexible to changing goals and environments, they will learn from experience, can make appropriate choices given perceptual limitations and finite computation (Poole et al. 1998). An agent is “something that acts in an environment” (Poole et al. 1998).

AI is becoming more prevalent in our society and thus is given more media coverage, which highlights both the positive and negative aspects of AI. Due to this and the way in which some individuals consume media content, unbiased information regarding AI can be difficult to obtain. This results in individuals possessing varying viewpoints and a limited understanding of AI. Some individuals will readily adopt new AI technologies whereas some will reject the use of it.

Human agents are capable of giving advice to other humans through the use of their natural intelligence. Some ‘AI assistants’, such as Alexa, are examples of IA as they use sensors to receive requests and then accomplish their goal by returning results. These ‘AI Assistants’ are also capable of providing advice to a human, which results in the question: Are humans more likely to accept or refuse an IA’s advice in comparison to a human agent? The aim of this project is to identify what physical and interactive attributes are crucial for an IA to possess in order to be considered trustworthy for providing behaviour change advice. Identifying and cataloguing the impact of these attributes will allow developers to implement IA systems that are deemed trustworthy by a user.

Project Aims and Objectives

Overall Summary

The aim of this project is to identify what physical and interactive attributes are crucial for an IA to possess in order to be considered trustworthy for providing behaviour change advice. In order to accomplish this, the project will be separated into sections whose aims and objectives will be discussed below.

The sections will be:

- Application Platform
- Modular IA
- Testing Period
- Study Period
- Final Report

Application Platform

- Objectives
    - Create a modular Android application that will collect and store data
    - Application must be battery efficient
    - User data must always be kept on the device
    - Application must support a wide variety of devices
    - Application must have a consistent intuitive user interface

In order for the IAs to offer advice they will need data to analyse to determine what advice to give. This will be done by creating an Android based application that will serve as a platform to compare attributes against each other. The application will collect sensor data, system data, user inputted data and environmental data. Data collected will be stored locally on the device and will only be used by the IA to provide advice to the user and so it will never leave the users device. This ensures that there is no ethical dilemma when regarding a user's personal data. Data stored will be assigned a 'tag', which will be used to determine how the data should be used by the IA.

'User Generated' data will be the primary stored data tag. It is data that is created by the user through their everyday routine. This is the data the IA will be offering advice about improving, such as if the user has been inside for multiple days in a row then the IA would advise the user to go for a walk.

The 'Environmental' tag consists of data created by the user's environment which is beyond the user's control, such as the weather. 'Environmental' data will be checked by the IA to see if a trend it has identified has a reason, such as if the user has been inside for multiple days in a row due to rain, and if so, the agent will evaluate if the advice should still be given.

'User Input' is data that the user has inputted into the application directly. This will be in the form of daily questions. This will be used to determine why data may correlate, such as if the user is sad because they have spent multiple days inside due to rain.

Users will have control over the data collected in the 'User Generated' and 'User Input' data tags. If a user disables or enables a specific data type it will be logged. This aims to identify what data users consider to be sensitive and do not trust allowing an IA access to.

The following is a list of data types that the application will aim to store:

- Sensor **(User Generated)**
    - Steps
    - Distance Walked Per Day
    - Location
    - Device recorded temperature

- System **(User Generated)**
    - Screentime (Time spent using applications)
    - Sleep Tracking (Estimation)

- External **(Environmental)**
    - Daily Weather
    - Outside Temperature
    - Sunrise Time

    - Sunset Time

- User Inputted **(User Input)**
    - Socialness (Interaction with friends)
    - Mood (User rating of mood)

Modular IA

- Objectives
    - Create a modular IA advice template whose attributes can be toggled
    - Each variation of the IA can be enabled via a passcode
    - IA should consume a reasonable amount of power
    - IA should provide reasonable advice

To measure impact of each attribute, the participants of the study will be randomly divided into groups and then each group will be randomly given a different variation of the IA. Each member of the group will be given the same IA. This aims to eliminate any anomalous data produced and produce a fair assessment of the IA's attributes.

Each attribute will be implemented as a standalone feature by separating its functionality into different packages within the application. The table below displays an example of how gender will be assessed.

Name	Attribute 1: Analysis	Attribute 2: Advice	Attribute 3: Gender	Attribute 4: Interaction	Attribute 5: Output
Agent 1	Machine Learning	Advice: With Justification	Female	High	Speech output
Agent 2	Machine Learning	Advice: With Justification	Male	High	Speech output

The 'Analysis' attribute will either use machine learning or traditional programming (Android Developers 2019) to determine what advice a user should receive. Machine learning will be the primary way in which the IA is 'intelligent' and having a comparison against a traditional programming solution will determine which implementation of data analysis results in a user following the advice given. However, there is an inherent problem with this testing method in that if one or both of the machine learning or traditional programming implementations are flawed then the advice given to a user may be bad which would corrupt the result of the study. Potential flaws should be identified within the testing period.

The 'Advice' attribute will give the user advice with or without justification. Justification being the reason why the advice is being given, such as graphs displaying a trendline of the user's data. This attribute will help to determine how much trust is put in the IA. If a user blindly trusts and follows the IA's advice without justification can it be linked to another attribute?

The 'Gender' attribute will involve changing the name of the IA and the colour scheme of the application. The name and colour scheme of each of the gender options will be kept the same to ensure a fair comparison of all attributes i.e., all male IAs will have the same name and colour scheme. Male gendered IAs will be assigned a masculine name and female agents a feminine name. These names will be chosen from the Office for National Statistics (Office for National Statistics 2020). Similarly, the colour scheme will represent colours traditionally associated with the gender such as blue and grey, which are traditionally masculine colours (Michelle Vatral 2018). This attribute aims to assess the impact of gender on advice given by an IA. This might show bias to a specific gender being trusted more or less.

The ‘Interaction’ attribute will have 2 settings: High and Low. At high interaction, advice will be given twice per day and for low interaction advice will be given once per day. This attribute aims to identify what the ideal interaction amount is between the IA and user.

The ‘Output’ attribute will involve either speech output or text output. Speech output will only involve verbal output, therefore better mimicking the method in which a human agent would give advice. Text output is the traditional method in which a computer would display information to the user. This attribute aims to identify what the ideal interaction method with which a user will receive advice.

Testing Period

- Aims
    - Test the application and IA to ensure there are no inherent flaws

This testing period aims to identify bugs and any flaws with the IA and application.

Study Period

- Aims
    - Give each participant in the study an IA

Each participant in the study will be given a link to download the application. They will also be provided with a unique passcode to enter on the initial start of the application. This will activate the IA that has been randomly assigned to that participant and start their study period. Once the user has downloaded the application, they will be prompted with a questionnaire regarding their preferences and knowledge about the attributes being tested (preferred advice, preferred interaction amount, etc). At the end of the study the questionnaire will be provided again. The answers of these questionnaires will be compared against the results of the IAs to identify critical attributes. The duration of the study period will be 1 week. This may be extended to 2 weeks if the application and IA are implemented ahead of schedule.

Final report

- Aims
    - Assess what attributes had the greatest impact on advice being followed

After the study period both the usage data generated and questionaries will be assessed to identify the effect of the attributes on the advice being followed. The results of the initial questionnaire will be compared against the results of the final questionnaire to see if participants opinions have changed. The final questionnaire’s results will then be used to identify what the participants think would be the most and least important attributes for an IA to possess. This will then be compared against the usage data to see if it correlates.

The usage data for each of the different IAs will be compared to identify which attributes had a positive effect on advice being followed, which attributes had a negative effect on advice being followed and which attributes had no effect. A positive effect would entail more advice being followed, negative would entail advice being ignored and neutral effect where there is no discernible positive or negative effect.

Work Plan

Date	Week	Description	Notes
08/02/2021	1	Initial User Interface and Database implemented Data in the Sensor category are automatically stored in the database	
15/02/2021	2	Data from the External and User Inputted categories implemented and stored in the database	Review Meeting, Ethics approval form (https://www.cs.cf.ac.uk/ethics/COMSC_Ethics_form.pdf) submitted to: comsc-ethics@cardiff.ac.uk
22/02/2021	3	Implementation of System category, Implementation of Initial Questionnaire along with an automatic countdown to final questionnaire	
01/03/2021	4	Real-world testing of data storage, research into implementing Analysis attribute	Review Meeting
08/03/2021	5	Analysis and Advice attributes implemented	
15/03/2021	6	Gender, Interaction, Output attributes implemented	Review Meeting, Participant recruitment inquiry emails sent to 1st & 2nd Years and Student Union.
22/03/2021	7	Testing Period	
29/03/2021	8	Testing Period	Review Meeting
05/04/2021	9	Study Period	
12/04/2021	10	Study Period	Review Meeting
19/04/2021	11	Final Report	
26/04/2021	12	Final Report	Review Meeting
03/05/2021	13	Final Report	
10/05/2021	14	Final Report	

(Review meetings are subject to change)

References

Android Developers. 8 May 2019. What's new in Android machine learning (Google I/O'19). Available at: https://www.youtube.com/watch?v=wpKJpeOy-68 [Accessed: 23/02/2021]

David Poole; Alan Mackworth; Randy Goebel. 1998. Computational Intelligence: A Logical Approach. 1st Ed. New York: Oxford University Press

Michelle Vatral. 2018. The Current Role Of Color Psychology In The Practice Of Gender Marketing. BSc Hons. University of North Georgia.

Office for National Statistics. 2020. Baby names in England and Wales: 2019. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/babynamesenglandandwales/2019 [Accessed: 05/02/2021]