UNIVERSIDADE DE LISBOA
Doctoral Programme in Cognitive Science

SensAI+Expanse

Prediction of Emotional Valence Changes on Humans in Context by an Artificial Agent Towards Empathy

Nuno Andrade da Cruz Henriques
nunoachenriques.net

Advisors

Helder Manuel Ferreira Coelho
Emeritus Professor, Ciências/ULisboa

Leonel Garcia-Marques
Full Professor, Psicologia/ULisboa

Thesis specially developed to obtain the degree of Doctor in Cognitive Science

2020

UNIVERSIDADE DE LISBOA

Doctoral Programme in Cognitive Science

SensAI+Expanse

Prediction of Emotional Valence Changes on Humans in Context by an Artificial Agent Towards Empathy

Nuno Andrade da Cruz Henriques

nunoachenriques.net

Advisors

Helder Manuel Ferreira Coelho
Emeritus Professor, Ciências/ULisboa

Leonel Garcia-Marques
Full Professor, Psicologia/ULisboa

Thesis specially developed to obtain the degree of Doctor in Cognitive Science

Jury

President

Maria Cristina de Castro Maia de Sousa Pimentel
Professora Catedrática e Directora da Área de Literaturas, Artes e Culturas da Faculdade de Letras da Universidade de Lisboa

Members

José Manuel Veiga Ribeiro Cascalho
Professor Auxiliar, Departamento de Ciências da Educação da Universidade dos Açores

Paulo Manuel Trigo Cândido da Silva
Professor Adjunto, ISEL - Instituto Superior de Engenharia de Lisboa

Helder Manuel Ferreira Coelho
Professor Catedrático Jubilado e Emérito, Departamento de Informática da Faculdade de Ciências da Universidade de Lisboa

Ana Luisa Nunes Raposo
Professora Auxiliar, Faculdade de Psicologia da Universidade de Lisboa

2020

Acknowledgements

Humans

Thank you to my advisors Helder Coelho and Leonel Garcia-Marques for the *saltus fidei* with me on this endeavour. Thank you for all the moments we disagreed, the patience while guiding me, and the confidence in my capabilities. The many lessons learned along the way are now part of my long-term memory. An additional acknowledgement goes to Luís Correia for the initial discussions, interest, and effort on steering me and the research proposal towards a successful start.

Thank you to the Cognitive Science scientific committee for the essential scrutiny of each research proposal. The qualification exam is imprinted in my memory: Ana Sebastião asserted on proper experimental procedure; António Branco and Armanda Costa questioned about the human-artificial interaction, to what purpose the agent take some action; Frederico Marques (dearly missed) questioned all about memory, cognition, and personality traits, he taught us how to question and how to be more demanding, his criticism first collides only to be abided next, it was a privilege; João Branquinho challenged us to think beyond our common assumptions, demanded proper arguments, and required the previous definition of any concept in use before moving forward.

I have also benefited greatly from all my class and laboratory mates daily discussions, more fiercely when talking about politics and religion, more enlightening when on science and the unknown. These were contributions difficult to be enumerated yet greatly appreciated and with a feeling. Thanks to you all! An additional thank you to Ana Paiva and all the GAIPS researchers for the warm welcome as a visiting PhD student during November 2016 and April 2017. It was an enriching experience besides the already cool robots ongoing studies.

A special mention is due to some companions along this demanding journey. A heartfelt thank you to Jorge Gomes for being an unofficial advisor, for not letting me slow down by introducing me to the wonders of coffee, for the bike rides, for the precious discussions enlightening the unknown and contributing like no other, and for being a friend. Thanks, Davide Nunes and Fernando Silva for encompassing this quartet, for the support on those rainy days, the criticism keen on scientific rigour which made be rumbled sometimes only to appreciate it later, and for being friends. Thank you, Anders Lyhne Christensen for an altruistic drive towards better scientific research and academic community synergy, for becoming a friend and an advisor always available to discuss the emerging dilemmas.

Although a mostly lonely endeavour yet also accompanied at crucial moments by the well appreciated support of family (father, brother, Diogo, Paulo, Rui, Sónia, Morais, Mok, and all) and dear friends from Portugal and Macao. All the caring was comforting in times of need. I am also privileged to have all that crazy people, that used to gather at Moinhos, interested in supporting and participating in this research study. Thanks to all of you and the incognito participants from this earth that installed SensAI and shared data.

Thank you mother for the crucial support.

Thank you Célia, Rodrigo, and Miguel for being always there with unconditional love and support.

Funding

This work is partially financed by *Universidade de Lisboa* [PhD support grant between May 2016 and April 2019]. Partially supported by *Fundação para a Ciência e Tecnologia* [UID/MULTI/04046/2019 Research Unit grant from FCT, Portugal (to BioISI)].

Logistics

Laboratory of the MAS[1] group of BioISI[2] — Biosystems and Integrative Sciences Institute, Faculty of Sciences, University of Lisbon, Portugal.

This work used the European Grid Infrastructure (EGI)[3] with the support of NCG-INGRID-PT (Portugal). Thank you for all that precious computation cores, memory, storage, and the Internet-connected server as a virtual machine. Without it SensAI+Expanse would be only SensAI restricted to the smartphone local resources.

[1] `http://bioisi.pt/mas/`

[2] Affiliation

[3] `https://www.egi.eu/about/`

Abstract

The field of Cognitive Science is broader enough on the interdisciplinary study of the brain, mind, and intelligence with a scientific research community gaining momentum over the last few decades. Specifically, joining the two fields of psychology and artificial intelligence (AI) one may envision agents, embodied or not, human-like or wearable, with the ability to significantly change the way humans live. This research conceive the artificial agent as a non-anthropomorphic with adaptive empathy for human-agent interaction (HAI) synergy towards better companionship. Therefore, the main objectives of this research are (a) to build a predictive model for each human user on context-based emotional valence changes; and (b) to study the age, gender, and human behaviour neutrality and robustness of the artificial agent regarding the prediction ability. The context include geographically located data from sensors, text sentiment analysis, and human emotional valence self-report, all timestamped events, using a common mobile device such as a smartphone. Also, to analyse and discuss the results on how to leverage such a model to adapt interaction strategies in order to foster higher levels of empathy between a non-anthropomorphic agent and its interacting human. For these goals SensAI+Expanse is developed where SensAI acts as an embodied nearby agent and Expanse encompass the machine learning resources in efficient manner, i.e., a distributed, fault-tolerant, mobile and Cloud-based platform from scratch as a research tool to continuously, online, gather and process data towards automated machine learning (AutoML) and prediction.

The study is designed with a methodology in place to avoid the Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies bias. This goal is accomplished by collecting data in the wild and worldwide by making use of the publicly accessible Google Play repository for the Android™ SensAI smartphone application. Eligible participants are diverse in age, gender, and behaviour on self-reporting emotional valence. In order to balance the gender distribution by age a dichotomy approach using age median ($M = 34$) is used. Regarding participation duration, two thirds (33/49) of the eligible individuals for analysis remain interacting for the required minimum of four weeks. The analysis of the results show evidence of significant behaviour differences between some age and gender combinations regarding self-reported emotional valence. Furthermore, the results from a comparison study between state-of-the-art algorithms revealed Extreme Gradient Boosting on average the best model for prediction ($F1 = 0.91$) with efficient energy use, and explainable using feature importance inspection. Moreover, the artificial agent remained neutral regarding human demographics and, simultaneously, able to reveal individual idiosyncrasies. Therefore, this research contributions include results with evidence, restricted to population and data samples available, of differences in behaviour amongst some combinations of age ranges versus gender. The main contribution is a novel platform for studies regarding human emotional valence changes in context. This system may complement and supersede (eventually) traditional long-list self-appraisal questionnaires. The SensAI+Expanse platform contributes with several parts such as a mobile device application (SensAI) able to adapt and learn in order to predict emotional valence states with high performance, a cloud computing (Cloud) service (SensAI Expanse) with ready-to-action analysis and processing modules towards AutoML. Additionally, smartphone sensing add a contribution for continuous, non-invasive and personalised health check. In the future, developments about human-agent relationships regarding affective interactions are foreseen. Further, the measurement of empathetic reactions and evaluating outcomes may be used to verify and validate health status thus improving care and significantly change the way humans live.

Keywords: emotional valence prediction, spatial and temporal context, cognition, memory, human-agent long-term interaction

Resumo

O campo da ciência cognitiva é suficientemente amplo no estudo interdisciplinar do cérebro, mente, e inteligência com uma comunidade de pesquisa científica em crescimento nas últimas décadas. Especificamente, juntando os dois campos da psicologia e inteligência artificial, é possível antever agentes, incorporados ou não, como humanos ou abstratos numa pulseira, com a capacidade de mudar significativamente a maneira como vivemos. A investigação descrita neste documento concebe o agente não antropomórfico com empatia adaptativa por sinergia da interação humano-artificial no sentido de uma melhor companhia entre agentes. Assim, os principais objetivos desta pesquisa são (a) construir um modelo de previsão adaptada a cada utilizador humano sobre mudanças de valência emocional em contexto; e (b) estudar comparativamente à idade, género, e comportamento humano a neutralidade e robustez do agente artificial na sua capacidade de previsão. O contexto inclui dados geograficamente localizados de sensores, análise de sentimento em texto, e de relatórios de valência emocional pelos humanos, estes eventos incluem informação temporal, usando um dispositivo móvel comum, tal como um telemóvel. Adicionalmente, para analisar e discutir os resultados de como alavancar esse modelo para se adaptarem estratégias de interação, de modo a promover aumento da empatia entre um agente não antropomórfico e o seu utilizador humano em contacto. Para cumprir estes objetivos é desenvolvido o SensAI+Expanse onde SensAI atua como um agente corporizado e de proximidade, e o Expanse abrange os recursos de aprendizagem por computador (*machine learning*) de forma eficiente. Isto é, uma plataforma distribuída, tolerante ao erro, móvel e baseada na nuvem (*Cloud*) informática, criada de raiz como ferramenta de pesquisa para reunir dados continuamente, ligada, e os processar para a aprendizagem automatizada por computador (*automated machine learning*) e a previsão.

O estudo é desenvolvido com uma metodologia específica para evitar o viés das sociedades educadas, industrializadas, ricas e democráticas (conhecido como WEIRD). Este objetivo é alcançado através da recolha de dados em campo e alargado ao mundo inteiro (potencialmente) fazendo uso do repositório Google Play acessível ao público como local de publicação da aplicação SensAI. Os participantes elegíveis são diversos em idade, género, e no comportamento ao informarem a valência emocional. No sentido de equilibrar adequadamente a distribuição de género por idade, uma abordagem dicotómica utiliza a mediana das idades ($M = 34$). Em relação à duração da participação, dois terços (33/49) dos indivíduos elegíveis para análise permaneceram em interação pelo mínimo necessário de quatro semanas. A análise dos resultados mostra evidência de diferenças significativas de comportamento entre algumas combinações de idade e género em relação à valência emocional informada pelos utilizadores. Adicionalmente, os resultados de um estudo comparativo entre os melhores algoritmos atuais revelaram o Extreme Gradient Boosting, em média, o melhor modelo para previsão ($F1 = 0,91$) com uso eficiente de energia e explicável usando inspeção de importância de cada caraterística (e.g., localização específica). Além disso, o agente artificial permaneceu neutro em relação à demografia humana e, simultaneamente, capaz de revelar idiossincrasias individuais. Portanto, as contribuições desta pesquisa incluem resultados com evidência, restritos à população e amostras de dados disponíveis, de diferenças de comportamento entre algumas combinações de intervalos etários e género. A principal contribuição é uma nova plataforma para estudos sobre as mudanças de valência emocional nos humanos e em contexto. Este sistema pode complementar e substituir (eventualmente) os tradicionais questionários com listas longas de questões para autoavaliação. A plataforma SensAI+Expanse contribui com várias partes, tais como (a) uma aplicação de dispositivo móvel (SensAI) com a capacidade de se adaptar e aprender de modo a prever estados de valência emocional com elevado desempenho; e (b) um serviço de computação em nuvem (*Cloud*), o SensAI Expanse, capaz de análise no momento e com módulos de processamento para a aprendizagem automatizada por computador (*automated machine learning*). Além disso, a abordagem de recolha de dados usando os sensores do telemóvel (*smartphone sensing*) adiciona uma contribuição no para análise da saúde ou do bem-estar em contínuo, não invasivo, e personalizado. Num futuro próximo, prevê-se um desenvolvimento interessante sobre as relações humano-agente em relação às interações afetivas. Adicionalmente, a medição das reações empáticas e a avaliação dos resultados das mesmas podem ser usados para verificar e validar o estado de saúde, e assim melhorar os cuidados e mudar significativamente a forma de viver dos seres humanos.

Palavras-Chave: previsão de valência emocional, contexto espacial e temporal, cognição, memória, interação humano-agente a longo prazo

Contents

List of Figures

List of Tables

List of Code Samples

Abbreviations

A-Life artificial life.

AI artificial intelligence.

ANN artificial neural network.

API application programming interface.

AutoML automated machine learning.

BBP Blue Brain Project.

BCI brain-computer interface.

BRAIN BRAIN Initiative.

Cloud cloud computing.

DGG Discrete Global Grid.

ECA embodied conversational agent.

EEG electroencephalography.

EMPA empathetic agent.

EPFL École Polytechnique Fédérale de Lausanne.

ETL extract, transform, and load.

FACS Facial Action Coding System.

FFM Five-Factor Model.

FOSS Free and Open Source Software.

FTS full text search.

GPS Global Positioning System.

HAI human-agent interaction.

HBP Human Brain Project.

HCI human-computer interaction.

HRI human-robot interaction.

HTTP Hypertext Transfer Protocol.

IDE integrated development environment.

IRI Interpersonal Reactivity Index.

JSON JavaScript Object Notation.

LED light-emitting diode.

LRID Likelihood Ratio Imbalance Degree.

LTD Long-Term Depression.

LTP Long-Term Potentiation.

MCC Matthews Correlation Coefficient.

MGRS Military Grid Reference System.

MLP multilayer Perceptron.

PANAS Positive and Negative Affect Schedule.

PL/pgSQL Procedural Language/PostgreSQL.

pp percentage point.

RDBMS relational database management system.

RGB Red, Green, Blue colour model.

SHAP SHapley Additive exPlanations.

SQL Structured Query Language.

TEQ Toronto Empathy Questionnaire.

TLS Transport Layer Security.

UI user interface.

UX user experience.

WEIRD Western, Educated, Industrialized, Rich, and Democratic.

XAI explainable artificial intelligence.

XML Extensible Markup Language.

Glossary

Affective state

A state revealed by an affect of four possible types: emotion; mood; sentiment; personality trait. The discrete distinction of the state, although difficult and still debatable amongst researchers from several backgrounds, may be assessed using three variables: duration, intensity and intention.

Artificial life

Synthetic systems which somehow behave like natural living systems.

Automated machine learning

Machine learning end-to-end process of real world data stream that may change unnoticed to the human. Includes robustness to (a) classifier target classes proportion changes, unfill of one or more; (b) hyperparameters search; and (c) distinct algorithms (linear, non-linear, connectionist) machine learning for best model prediction.

Being (Tipler, 2003, chap. IV)

An entity that codifies information preserved by means of natural selection.

Cloud computing

A computing term based on utility and consumption of shared computer resources over a network to achieve the most efficient way to obtain processing power and storage capacity. It relies on converged, eventually location distributed, infrastructure and shared services dynamically reallocated on demand per user.

Complex dynamic system

A particle set, usually numerous, at any scale of the Universe that constantly interact and in a way that constrains the ability to predict about the future state of any individually and all simultaneously. A non-linear dynamic system with a high and sensible dependency to its initial state is called chaotic.

Emotional valence

A positive-negative emotional dimension including neutral in middle range and able to classify the human intrinsic disposition (mood or sentiment) of current moment, place or experiencing situation.

Empathy

The (affective and cognitive) ability to perceive, understand and act accordingly the affective states of another for a better dyadic and social bonding.

Life (Tipler, 2003, chap. IV)

Information preserved by means of natural selection.

Smartphone

A mobile phone that's able to perform the functions of a computer. Additionally, having sensors such as compass, proximity, ambient light among others besides a sophisticated multi-touch screen interface, Internet access, and an operating system capable of running general-purpose applications.

Smartphone sensing

Mobile phone-based sensing software that uses device's sensors to collect data about the human user and more. Other terms to name this approach include "personal sensing" and "context sensing".

Chapter 1

Introduction

This first chapter introduces the thesis subject presenting the original motivation, the context, and the main goals. The objectives include the research question and the enumeration of the hypothesis. It briefly depicts the work plan and methodology followed. Next, it exposes the main concepts for the adopted technology and, at the end, it explains the document structure.

1.1 Motivation

The search for a thesis subject is more a confluence of interests, explorations, and influences towards the unknown. In this case, the emergency of a desire to create intelligent artificial beings as an ultimate goal. The transdisciplinary Cognitive Science is a broader enough field of studies and with a scientific research community gaining momentum over the last years. Its main fields (cognitive neurology, psychology, philosophy, linguistics, and artificial intelligence) have contributed with exciting results and foreseeing applications able to convince anyone that, embodied or not, human-like or wearable, intelligent artificial agents with the ability to change the way humans live are at reach in our lifetime.

The year of 2013 marks the public announcement of some major players commitment regarding a confluence of interests towards the current global research effort around the human brain, mind, artificial intelligence, and healthcare. The research that emerged already brought[4] value to the society worldwide by means of the impact that new discoveries and solutions constitute. These serious public investments also with strong private partnerships (e.g., Google Brain Team[5]), able to be more or less easily scrutinised, fused fundamental science with applied research and engineering towards knowledge and innovation. It has the following main initiatives in Europe and in the United States of America:

[4] For more about the outcomes: `https://www.epfl.ch/research/domains/bluebrain/blue-brain/news/`; `https://www.humanbrainproject.eu/en/science/highlights-and-achievements/`; `https://www.braininitiative.org/achievements/`

[5] `https://research.google/teams/brain`

Blue Brain Project (BBP) this early project started in 2005 with an agreement between the École Polytechnique Fédérale de Lausanne (EPFL) and IBM which included the installation of a BlueGene supercomputer on the *école* campus. Henry Markram leads the project with results in the following year presenting a model of a cortical column. Next, in 2007, an initial model of the rat cortical column is announced. The research continues to thrive and from 2013 onwards the BBP plays a leading role in the Human Brain Project. The initial goal was "to build biologically detailed digital reconstructions and simulations of the rodent, and ultimately the human brain"[6].

[6] https://www.epfl.ch/research/domains/bluebrain/

Human Brain Project (HBP) is a ten-year (began in 2013) thousand-million euros research endeavour sponsored by the European Commission and funded by the European Union. It directly employs about 500 scientists at more than 100 universities, teaching hospitals, and research centres. The HBP aims to "put in place a cutting-edge research infrastructure that will allow scientific and industrial researchers to advance our knowledge in the fields of neuroscience, computing, and brain-related medicine"[7].

[7] https://www.humanbrainproject.eu

BRAIN Initiative (BRAIN) this investment in Brain Research Through Advancing Innovative Neurotechnologies (BRAIN) is announced in 2013 by president Obama. The mission of this initiative is "to deepen understanding of the inner workings of the human mind and to improve how we treat, prevent, and cure disorders of the brain"[8].

[8] https://www.braininitiative.org

Further articulating all this brain and mind research effort with the field of AI towards artificial life (A-Life), one may argue that all this research deep in the structure that supports our identity poses questions about the meaning of life and the Universe, from the particle chaos of a complex dynamic system to intelligent life. The creation of synthetic beings is one of the several ways to study A-Life, any piece of information may be an individual. One way of doing these studies is by means of agents' interaction in dyadic and social contexts. These interactions may comprise any of the possible combinations amongst artificial and human agents. For this research purposes a combination of human versus artificial is elected though without restricting the platform and software development to further extensions, experiments, and studies comprising other combinations such as artificial versus a mixed environment of several humans and other artificial agents. Moreover, non-invasive brain-computer interface (BCI) such as an electroencephalography (EEG) device may be integrated in order to collect this type of biometric data and process it in future correlations. The multidisciplinary fields of human-robot interaction (HRI) and human-computer interaction (HCI) imply some kind of perceived behaviour by the agents of one another. This research conceives the artificial agent as non-anthropomorphic with adaptive empathy for HAI synergy towards better companionship. Therefore, to study HAI with added affective states perception more than one field of science is required, i.e., research on predictive models of emotional valence state based on human context by an artificial agent requires the joining of two main fields in Cognitive Science: psychology and artificial intelligence.[9]

[9] Stueber (2014), Bavelas et al. (1986), Hegel et al. (2006), Singer (2006), de Waal (2008), Polajnar et al. (2011), Engen and Singer (2013), Potapov and Rodionov (2014), Riedl et al. (2014), and Morgan (2016)

1.2 Objectives

The two research questions as investigative objectives to pursue are (a) how to build a predictive model for emotional valence changes based on user context (geographically located data from sensors, text sentiment analysis, human emotional valence self-report) using a common mobile device such as a smartphone; and (b) how to leverage such a model to adapt interaction strategies in order to foster higher levels of empathy between a non-anthropomorphic agent and its interacting human.

In the context of this work, the operational empathy definition used is: The (affective and cognitive) ability to perceive, understand and act accordingly the affective states of another for a better dyadic and social bonding.

The smartphone has sophisticated sensors including positioning, gyroscope, compass, accelerometer, magnetometer, optical proximity, audible (using the built-in microphone) and camera. This mobile platform has already been used for human's mood detection. MoodScope (Likamwa et al., 2013) is a mood sensor for mobile devices. It relies on user usage patterns and text mining from SMS, email, phone calls, Web browsing, application usage (Twitter and Facebook are well identified though others may be integrated) and location. The affect sensing using a smartphone has several studies underway already with important published results (Rana et al., 2014; Harari et al., 2017). Moreover, in health and wellness relevant research and with applications is being actively promoted and developed (Cornet & Holden, 2018) including exploration on correlates between sensors' data and depressive symptom severity (Saeb et al., 2015). The American College of Medical Informatics (ACMI) has already envisage this path. In the 1998 Scientific Symposium, one of the informatics challenges for the next 10 years was "Monitor the developments in emerging wearable computers and sensors — possibly even implantable ones — for their potential contribution to a personal health record and status monitoring." (Greenes & Lorenzi, 1998).

This work aims to contribute with a useful and sophisticated up-to-date platform to do human emotional valence state studies. Moreover, to complement and supersede (eventually) traditional long-list self-appraisal questionnaires (e.g., Interpersonal Reactivity Index (IRI)[10], Toronto Empathy Questionnaire (TEQ)[11]). In this sense, to contribute with new findings about specific context (e.g., geographically located, moment of the day) emotional valence changes for a human. Moreover, to contribute with findings for the efficient use of this human-agent relationship type in order to improve health care (e.g., to diminish loneliness in humans with more companionship). In the future, also about human-agent relationships regarding affective interaction. Further, the measurement of empathetic reactions and outcomes evaluation may be used to verify and validate health status. The flexibility of any agent may depend on its intelligence, will to interact with the other, and the desire to share common affective states such as emotions, sentiments, and moods.

[10] Davis, 1980; Davis, 1983

[11] Spreng et al., 2009

Therefore, three hypotheses are stated and verified at the end of a research study included in this dissertation:

Hypothesis 1. *The human-agent long-term bonding, using a mobile device and balanced resources consumption, foster enough data in order for the artificial agent to predict human's emotional valence changes in context.*

Hypothesis 2. *The agent's ability on predicting human's emotional valence changes is gender and age neutral.*

Hypothesis 3. *An artificial agent embed in a mobile device, such as a smartphone, is able to leverage sensors and data in order to predict idiosyncratic factors on human's emotional valence changes.*

1.3 Planning

The whole work actual[12] plan (Figure 1.1) spans from April 2016 to January 2020 and has several phases which, some of those, occur simultaneously with the research work and also the SensAI development. There are phases categorised as preliminary such as (a) state of the art detailed analysis; (b) technology review and probe of free software parts ready to use; and (c) development and testing methodology. The state of the art phase is the first one, of course. A detailed investigation about the recent and more relevant research related with this work is paramount. Although already detailed, state of the art also follows along the timeline as periodic, checking for up-to-date published relevant research in the meanwhile. Next, the whole SensAI+Expanse platform is developed with data engineering and science towards this research and

[12] The initial plan duration was three years. Although, too many contingencies such as unavailable out-of-the-shelf software parts required an extraordinary effort and a prolongation to almost four years.

study objectives (Section 1.2, p. 3). Simultaneously, at the final period of yet several months, the thesis report and scientific articles for publication are written.

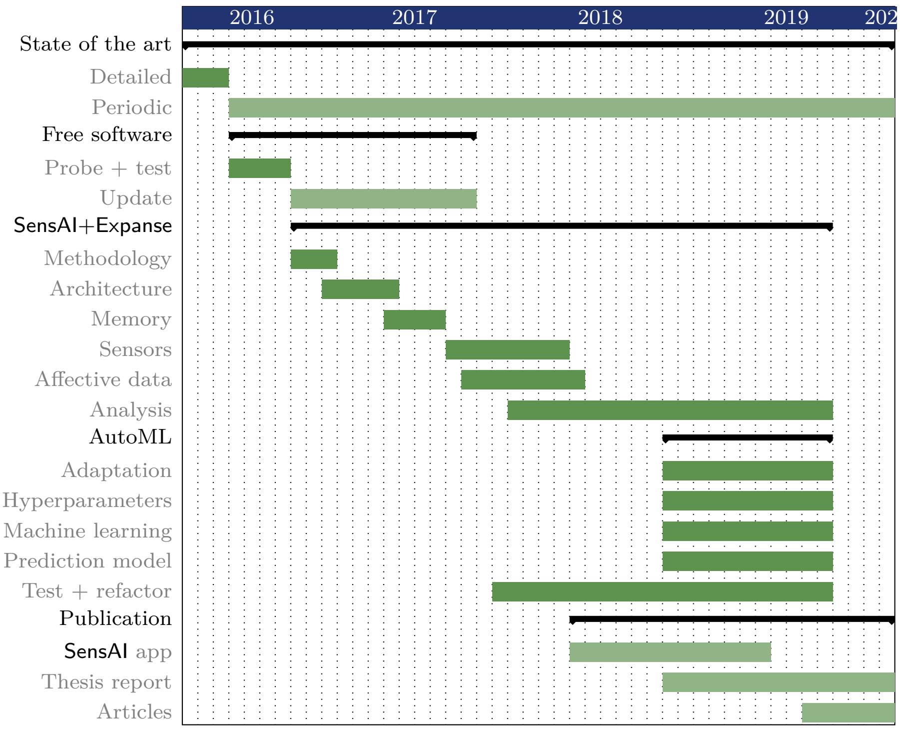

Figure 1.1: Gantt chart for the almost 4-year project plan.

State of the art | **3 + 43 months** | The first period of three months requires a dedication to the investigation of all the related work already published. A detailed analysis is required on relevant matters such as social cognition on interaction, the emotional reasoning in artificial intelligence and its models, architectures, perceptional sensors and human-agent interaction. The next months until the project end will have periodic checks of relevant publications.

Free software | **4 + 12 months** | A period to begin searching for free software parts ready to use in order to save resources (time and effort) on creating each part from scratch. Further, despite having a more or less successful outcome yet the next twelve months remain reserved for updating the search checking on, eventually, new relevant releases.

SensAI+Expanse | **35 months** | This phase includes the whole platform development, i.e., the SensAI agent to be used as a tool for the research study and the Cloud infrastructure SensAI Expanse to serve all the agents' data collecting and heavy-duty processing. Several versions are released to fix and improve the detected incidents, interface responsiveness and aesthetics, resources consumption, data format and quality, in order for an adequate user experience for long-term usage — iterative process using a loop cycle of test and refactor. The development includes CogA libraries to integrate offline language detection and translation for text sentiment analysis. Moreover, all the modules of SensAI app are developed from scratch to deal with the specifics of a mobile device embodiment using the Android™ operating system. Specifically, the `SenseiStartStop` and `Homeostasis` modules to cope with the reboot, upgrade, and keep alive cases. To accomplish the goal of learning human emotional valence changes in context, collected data is processed, cleaned and prepared for use in machine learning algorithms. These are the data science and engineering for analysis (statistics) and learn tasks towards prediction.

The continuous interaction, writing, moving by the human whilst using the mobile device contributes with data. The smartphone embodiment is lightweight and with reasonable processing power. It has sophisticated sensors including positioning, gyroscope, compass, accelerometer, light, optical proximity. The entity mood responsiveness is designed using a simple empathy notification bar. Moreover, occasional notifications may happen to improve interaction. The design and implementation of this companion encompass the integration of perceptual and cognitive layers, along with modules for affective state, recognition and anticipation which rely on the memory (short and long-term) modules. The architectural model to implement should be modular to ease endowing a computer with adaptive characteristics towards an autonomous artificial agent with empathetic behaviour in a future version.

This phase is the orchestration of all the research pieces: embodiment, perceptional sensors, affects, memories, cognitive processing, human and artificial entity interaction.

Publication | 21 months | This phase occurs simultaneously with the previous analysis, machine learning, test plus refactor, release cycle tasks. Each SensAI app new version is published in Google Play for both update and new installation by the users. Regarding scientific publications, the thesis dissertation completion (started during previous phases with some notes) includes writing and compiling the research major findings already produced and revealed. There will be, at least, two papers written and submitted to adequate conferences and also open science repositories such as arXiv.

1.4 Technology

Once upon a time there was the Internet and humans start chatting synchronously and mailing asynchronously at instantaneous speeds using wired devices. The Web emerged with its constant changing publications, archives created using permanent storage, file sharing improved, social networks arose, and the Cloud start becoming available to every connected device. And then everybody went mobile and wireless. Devices' location, acceleration, ambient light, camera and other sensors start perceiving the environment, algorithms improved from interfaces interaction to search predictions, and everything started smarting: smartphones, smartwatches, advanced robotics, driverless cars, and intelligent artificial agents towards A-Life.

This work focus on non-anthropomorphic embodied artificial agents. These type of entities may present intelligent behaviour as perceived by humans. Moreover, the agents may reveal emotions, moods and also empathetic behaviour. All these aspects require a friendly interactive environment with sensors and effectors demanding processing power and memory capacity to cope with the day-to-day relation. All these power requirements consume a non-negligible amount of energy. Hence, a current smartphone device and its ability to augment the processing power and memory capacity using Cloud resources seem fit to the task.

In the end, a relevant contribution is achieved by developing, testing, studying, and sharing the results and the research tools able for anyone to build its own platform and configure its own system towards (a) an efficient use of energy such as negligible smartphone battery drain with resource saving diverse techniques[13]; (b) free from the laboratory usual restrictions on samples acquired such as regarding the people from WEIRD[14] societies bias; and (c) with the ability to be extended, modified, and adapted for further studies regarding agents' interaction and affective states.

[13] Detailed in Chapter SensAI+Expanse.

[14] Henrich et al., 2010

Detailed aspects such as software tools used for the whole platform development, study, and published source code repository are described in Appendix A, p. 103, Appendix B, p. 107 and Appendix C, p. 115.

1.5 Structure

This first chapter introduces the path taken for this research work accomplishment and includes, along with this section, Motivation, Objectives, Planning, and related Technology brief descriptions.

The next chapter, "Related Work" describes and briefly discusses relevant research in affective states, empathy, cognition, memory and learning, embodiment, sensors, interaction, and finally smartphone sensing within the context of an artificial agent and its interaction with humans.

The following chapter is "SensAI+Expanse" and presents the whole development of the platform. It describes the conceptual options that guided its implementation, the architecture, and the integration of all the pieces towards data collection, machine learning, and prediction.

The "Study" chapter follows and describes all the tasks related with the HAI between SensAI and its human user. Also, explains the method used including about the participants, design and procedure. The results and analysis are presented as two independent subsections.

Finally, "Discussion" chapter, argues about the limitations perceived, the contributions achieved, and concludes this dissertation document with some future work perspectives, such as the improvements and the current related trends.

Chapter 2

Related Work

This chapter describes the research work related with the thesis subject, i.e., published work closely connected or containing appropriate aspects to the matter in study. This brief and concise review is required in order for sustaining decisions and options along the research development. The main concepts presented and briefly discussed are divided in several sections regarding affective states, empathy, cognition, memory and learning, embodiment, sensors, interaction, and smartphone sensing.

2.1 Affective States

The debate about affect definition of types, processes, and models is still controversial. Regarding recent history starting from late nineteenth century with Wilhelm Max Wundt and his "System der Philosophie"[15], for more than a century several approaches were made in trying to create emotions' models. The challenge was and still is not only the model *per si* but also which emotions to choose — a universal (cross-culture) set of emotions for the human species individual. Moreover, which facial behaviours associated with emotion (Ekman & Friesen, 1971) are universal in the human context? Is still not clear and even arguable[16] which model is the best one in order to sustain research in several fields namely psychology. Nevertheless, some relevant and specific research regarding affective models and states[17] are highlighted and devised to be useful along this research project.

[15] `https://archive.org/details/systemderphiloso00wund`

[16] Shuman and Scherer (2015) present a good review on the four families of emotion theories. It includes the common aspects where all theorists agree and also the divergent issues.

[17] Refer to Affective state in Glossary.

The types of affective states are still under discussion for which ones may prevail unique and independent of the others (N. Frijda, 2008). Although there is disagreement about which theory of emotions, including the duration of emotions that may last from brief moments to ages, yet mood is accepted as a somehow durable affective state. Still, moods are often considered as short as emotions (Barrett et al., 2008; Beedie et al., 2005).

Regarding a duration-based relative sort of affective state types, from shorter to longer duration, it may be hypothesised as: emotion (basic); mood; sentiment; personality trait. This relative order does not imply that emotions are always or even typically short, duration of emotions (affective chronometry) is actually highly variable (N. H. Frijda et al., 1991; Verduyn et al., 2015). Scherer (2009) typology of affective states add two more types between mood and personality trait by replacing sentiment with (a) interpersonal stances (affective stance towards another person in a specific interaction friendly, flirtatious, distant, cold, warm, supportive, contemptuous); and (b) attitudes (enduring, affectively coloured beliefs, dispositions towards objects or persons, liking, loving, hating, valuing, desiring).

Regarding models of affective states, Russell (1980) proposes a circumplex model of affect[18] by disposing eight affect concepts in a circular order. This model[19] is validated by several experiments with human subjects, twenty eight affect words are attributed to the eight categories that were previously arranged around two orthogonal dimensions of positive-negative arousal and pleasure[20]. Russell considers that a cognitive structure for affect consists of an interrelated cognitive set of categories.

[18] Refer also to the Plutchik circumplex model of affect depicted in Figure 3.15, p. 51.

[19] Figure 2.1, p. 10.

[20] Figure 2.2, p. 11.

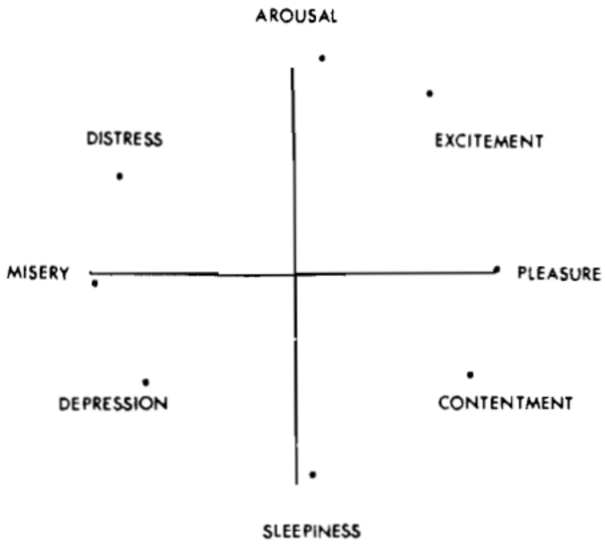

Figure 2.1: Circumplex model of affect.
Source: Russell (1980, Figure 1)

Watson et al. (1988) developed the Positive and Negative Affect Schedule (PANAS) to fill the need for a reliable and valid positive and negative affect scales. It is a 20-item (10-item for each scale) self-report measure. The authors were motivated by the emergence of the positive and negative affect as two dominant and relatively independent dimensions in several cross-culture studies of affect structure. The PANAS is a reliable and valid method for the measurement of these two important mood scales. Although, the original (Watson et al., 1988) hypothesis of independence between the two (positive and negative affect) scales

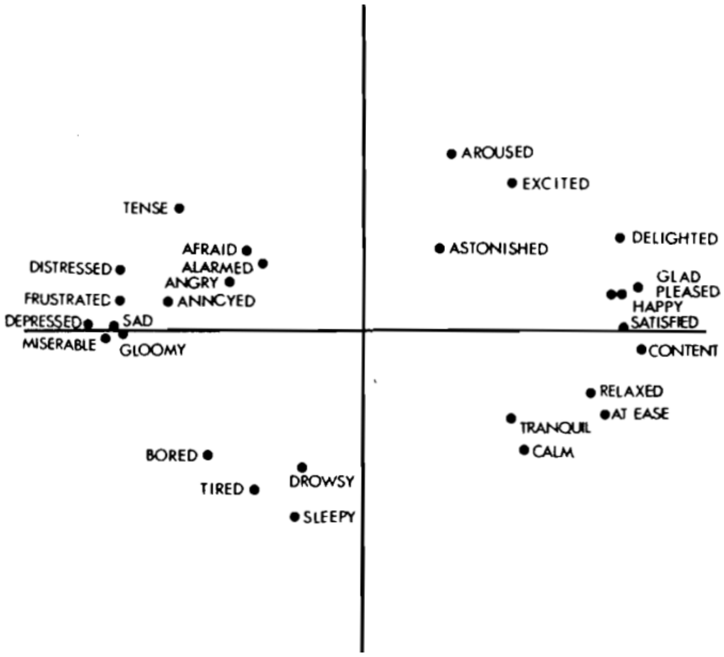

Figure 2.2: Regression weights for 28 affect words as a function of pleasure-displeasure (horizontal axis) and degree of arousal (vertical axis).
Source: Russell (1980, Figure 5)

must be rejected (Crawford & Henry, 2004). The same authors conclude the utility of the PANAS measure to be enhanced by the provision of large-scale normative data.

Paul Ekman[21] (on the study of non-verbal behaviour) has provided strong evidence in support for the hypothesis of a cross-culture association between particular facial muscular patterns and discrete emotions[22]. Ekman and Friesen (Ekman (1968); Ekman and Friesen (1969) *apud* 1971) previously used seven particular emotions (happiness, sadness, anger, fear, surprise, disgust, interest)[23] to hypothesize universality in the relationship between distinctive patterns of the facial muscles and those emotions. Moreover, they suggested that cultural differences may be seen in some of the stimuli, becoming established as elicitors of specific emotions through learning, in the rules for facial behaviour control in some social settings, and in action consequences of emotional arousal. In Ekman and Friesen (1971) the authors use six specific emotions (happiness, sadness, anger, fear, surprise, disgust)[24] to present "evidence of constants in facial behaviour and emotion across cultures [...]" and maintaining "[...] that cultures may not make *all* of the same distinctions among emotions, but does not detract from the main finding that most of the distinctions were made across cultures". Several decades after these findings, Ekman (1999) consolidates his previous work about basic emotions and states that "Each emotion is not a single affective state but a family of related states." as themes with variations where "The themes are the product of evolution, while the variations reflect learning".

[21] `http://www.paulekman.com/paul-ekman/`

[22] Emotion recognition by facial patterns detection.

[23] Seven basic emotions.

[24] Six basic emotions.

Ekman also describes the proposed eleven characteristics (Table 2.1, p. 12) which distinguish basic emotions from one another and from other affective phenomena. This framework proposed by Ekman foresees a future where the pool of the fifteen[25], already appraised, basic emotions may change by adding new ones or removing existent ones in the devised framework context.

[25] Fifteen basic emotions.

Table 2.1: Characteristics which distinguish basic emotions from one another and from other affective phenomena

1. Distinctive universal signals
2. Distinctive physiology
3. Automatic appraisal, tuned to:
4. Distinctive universals in antecedent events
5. Distinctive appearance developmentally
6. Presence in other primates
7. Quick onset
8. Brief duration
9. Unbidden occurrence
10. Distinctive thoughts, memories images
11. Distinctive subjective experience

Source: Ekman (1999, Table 3.1)

Many years later, Ekman (2016) in a pragmatic initiative reports on an emotion survey emailed to 248 scientists researching in the field. The results were a "[...] high agreement about the evidence regarding the nature of emotion, supporting some of both Darwin's and Wundt's 19th century proposals". Moreover, a robust agreement emerged for universals in emotional signals and around five basic[26] emotions: anger; fear; disgust; sadness; happiness. Additionally, the survey reveals a relevant agreement (among the emailed scientists) about the relation of specific moods with particular emotions such as anger to irritability (88%) though the survey did not explore how, the relation of specific personality traits with particular emotions such as fear to shyness (82%), and the relation of specific emotional disorders with emotions such as disgust to anorexia, although more debatable with an agreement of 75%. Ekman (1994) on moods, emotions and traits more than twenty years before the 2014–2016 survey did propose the differentiation of emotions and moods based on time and intensity[27]. Regarding personality traits as one type of affective state, those are considered life-long characteristics.

[26] Five basic emotions.

[27] Emotion versus mood: time and intensity.

McCrae and John (1992)[28] present an extensive introduction to the FFM of personality and its applications. The FFM organises personality traits hierarchically in terms of five basic dimensions[29]: Openness to Experience,

[28] Five-Factor Model (FFM).

[29] Figure 2.3, p. 14.

Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The authors support the comprehensiveness of the model and its applicability across cultures by the use of both natural language adjectives and personality questionnaires. Historically, the lexical approach had a small role when compared with the personality questionnaires which have hundreds of rating scales designed by its researchers to measure more discrete constructs they deemed relevant. Despite the lexical approach small role it gained relevance in several moments of the personality research (past century) history (McCrae & John, 1992):

> "Allport and Odbert (1936) abstracted terms from a dictionary; Cattell (1946) formed them into synonym clusters and then created rating scales contrasting groups of adjectives; Tupes and Christal (1961) obtained observer ratings on these 35 scales and factored them. (Fiske, 1949, had also used a version of Cattell's rating scales in the earliest recovery of the five factors.) Norman used the best 20 rating scales from the Tupes and Christal study in his replications, and that set was subsequently used in many later studies. [...] The most important new studies were the cross-cultural replications by Bond (1979; Bond, Nakazato, & Shiraishi, 1975). Reanalyses of earlier data sets by Digman and Takemoto-Chock (1981) and the meticulous analyses of Goldberg (1981, 1982) revived interest in the lexical approach and reintroduced the FFM to the mainstream of personality psychology. [...] But there is one more compelling reason for studying trait language. Allport and Odbert noted some 4,500 trait terms in English; surely such a wealth of vocabulary testifies to the social importance of personality traits." (p. 181, 184).

> "Most personality assessment has been based on questionnaires with scales designed for specific practical applications or to measure constructs derived from personality theory (Goldberg, 1971). Psychiatric nosology and the theories of Jung (1923/1971), Murray (1938), and Sullivan (1953), among others, have spawned a variety of instruments. [...] Theories of personality have been remarkably diverse, and it might have been anticipated that the questionnaire scales designed to operationalize them would show little resemblance to each other. In fact, however, there is considerable redundancy in what they measure. In particular, many scales measure the chronic negative emotions that are of such great concern to psychiatrists and clinical psychologists, and many others deal with the interpersonal activity so important for social psychol-

> ogists. H. J. Eysenck institutionalized these two dimensions as N and E, and provided useful measures (H. J. Eysenck & S. B. G. Eysenck, 1964, 1975); [...] Both sets of researchers admired H. J. Eysenck's strategy of looking for broad themes by which to organize groups of traits, and sought to extend it to new dimensions. By explaining as much as possible in terms of established factors, and then looking for commonalities in what remained unexplained, researchers could proceed to a systematic mapping of personality traits." (p. 185, 186).

These two personality research paths — the lexical approach and the personality questionnaires — where merged and the contemporary FFM emerged.

Factor		Factor definers		
Name	Number	Adjectives[a]	Q-sort items[b]	Scales[c]
Extraversion (E)	I	Active	Talkative	Warmth
		Assertive	Skilled in play, humor	Gregariousness
		Energetic	Rapid personal tempo	Assertiveness
		Enthusiastic	Facially, gesturally expressive	Activity
		Outgoing	Behaves assertively	Excitement Seeking
		Talkative	Gregarious	Positive Emotions
Agreeableness (A)	II	Appreciative	Not critical, skeptical	Trust
		Forgiving	Behaves in giving way	Straightforwardness
		Generous	Sympathetic, considerate	Altruism
		Kind	Arouses liking	Compliance
		Sympathetic	Warm, compassionate	Modesty
		Trusting	Basically trustful	Tender-Mindedness
Conscientiousness (C)	III	Efficient	Dependable, responsible	Competence
		Organized	Productive	Order
		Planful	Able to delay gratification	Dutifulness
		Reliable	Not self-indulgent	Achievement Striving
		Responsible	Behaves ethically	Self-Discipline
		Thorough	Has high aspiration level	Deliberation
Neuroticism (N)	−IV	Anxious	Thin-skinned	Anxiety
		Self-pitying	Brittle ego defenses	Hostility
		Tense	Self-defeating	Depression
		Touchy	Basically anxious	Self-Consciousness
		Unstable	Concerned with adequacy	Impulsiveness
		Worrying	Fluctuating moods	Vulnerability
Openness (O)	V	Artistic	Wide range of interests	Fantasy
		Curious	Introspective	Aesthetics
		Imaginative	Unusual thought processes	Feelings
		Insightful	Values intellectual matters	Actions
		Original	Judges in unconventional terms	Ideas
		Wide interests	Aesthetically reactive	Values

a. Adjective Check List items defining the factor in a study of 280 men and women rated by 10 psychologists serving as observers during an assessment weekend at the Institute of Personality Assessment and Research (John, 1989a).
b. California Q-Set items from self-sorts by 403 men and women in the Baltimore Longitudinal Study of Aging (McCrae, Costa, & Busch, 1986).
c. Revised NEO Personality Inventory facet scales from self-reports by 1,539 adult men and women (Costa, McCrae, & Dye, 1991).

Figure 2.3: Examples of Adjectives, Q-Sort Items, and Questionnaire Scales Defining the Five Factors.
Source: McCrae and John (1992, Table 1)

It should be noted that the five factors are not exhaustive at an individuals personality description. These factors are a relevant contribution for a representation of a highest hierarchical level of trait description. Further, Ashton et al. (2004)[30] discuss a potential reorganisation of the five-factor model structure for personality assessment and research. A six-factor structure of personality descriptive adjectives is proposed based on psycholexical studies in seven natural languages.

30 Six-factor model. For the HEXACO model and a further review: Ashton and Lee, 2007; Ashton et al., 2014. Also, for a Six-factor comparison with FFM and more: Brocklebank et al., 2015; Barford et al., 2016

All the previous assumptions about (basic or not) emotions by Ekman and others are debatable. Recent theories are taking a constructivist approach in a sense that “Emotions are constructions of the world, not reactions to it.” (Barrett, 2017). This theory of constructed emotion suggests a multi-level view on understanding the basis of emotion in the brain, that is consistent with the emerging computational and evolutionary biological views of the nervous system. “A brain can be thought of as running an internal model that controls central pattern generators in the service of allostasis”. Further, Barrett (2017) concludes the theory of constructed emotion by stating that:

> “It will never be possible to measure an emotion by merely measuring facial muscle movements, changes in autonomic nervous system signals, or neural firing within the periaqueductal gray or the amygdala. To understand the nature of emotion, we must also model the brain systems that are necessary for making meaning of physical changes in the body and in the world. [...] The theory of constructed emotion proposes that emotions should be modeled holistically, as whole brain-body phenomena in context. [...] to solve the age-old mysteries of how a human nervous system creates a human mind.” (p. 16).

Regarding the well-known discussion and claims about emotions detection by facial expressions, Barrett et al. (2019) went further on stating that:

> “[...] facial movements are functionally tied to the immediate context, which includes a person’s internal context (e.g., the person’s metabolic condition, the past experiences that come to mind) and outward context (e.g., whether a person is at work, at school, or at home, who else is present and the broader cultural conditions), both of which vary in dynamic ways over time [...]” (p. 4)

2.2 Empathy

The human entities may be seen as self-consciousness emotion-driven cognitive beings with a bond between the evolutionary way of emotions and their supporting physical structure as proposed by Damásio (2010). Regarding artificial agents, robots that have emotions (Arbib & Fellous, 2004; Parisi & Petrosino, 2010) are not “emotional” robots that only simulate or detect facial emotions. An example of a distinctive trait of emotional robots may be the ability to say “I understand that you’re sad” revealing empathy when interacting with a human. The Wada et al. (2004) approach was to endow artificial entities, baby seal robots, with a high degree of completeness on life-like proactive and reactive behaviour. Specifically, to act accordingly the kind of received care towards a stress reduction goal on elderly people. Non-vision perceptional sensors are used, the focus was on the (therapeutic) impact of empathetic animal-like entities interaction with humans. There were no studies about the artificial entity mood, i.e., entity's affective states and its development over time. This research recognises that the ability of an empathetic interaction doesn't require anthropomorphic embodiment, i.e., entities with human-like body and face.

The enrichment of human-computer interactions with user's affective states capturing and expression of emotions in computers representing and reasoning with affective states have been long discussed (e.g., Paiva et al., 2017). Emotional synthetic characters with feeling and reasoning are proposed by Dias and Paiva (2005) within an interactive virtual environment performing a small case evaluation determining if user positive reactions may arise. The capability known as motor mimicry (Bavelas et al., 1986) is used by Hegel et al. (2006) on an anthropomorphic robot with the size of a four-year-old child to improve human-robot empathy towards better HRI. There were no studies about a non-anthropomorphic empathetic entity. The effects of emotional facial expressions[31] in computer agents and its impact on negotiations with humans (aware of the artificial entity) are explored by Melo et al. (2011) with results indicating the impact similarity as in human-human negotiations. A model for the creation of artificial agents with personality traits is proposed by Doce et al. (2010). The authors reported that the HAI tests resulted on the perception by humans of extroversion, neuroticism, and agreeableness traits but failed for conscientiousness. Castellano et al. (2010) provides an overview and uncover insights about choices to be made when designing an affect recognition system. Artificial agents as human companions and their empathetic behaviour is discussed by Leite et al. (2012) where the risks of a less careful selection of behaviours and its impact on children are pointed out. Inaccurate empathetic behaviours may have a negative effect on children emotional state (anxious or nervous instead of happy)[32]. To keep the human in a positive affective state empathy must be adaptive. There were no studies concerning physical interaction as in tactile and its impact on the adaptive empathy.

[31] Although, humans' facial muscles, gestures and body movements, as signs of specific emotions, can vary on an individual and cultural basis in context (Schwark, 2015; Barrett et al., 2019).

[32] Findings (adults case study) also supported by Cramer et al. (2010).

Regarding recent research studies making use of the empathy concept, there are many in a healthy interest in better interactions between humans and artificial entities. Vernon et al. (2016) include empathy as one aspect inside the social cognition ability of a robot. This feature is a requirement for "[...] robots that can interact effectively with people requires a special focus on building systems that can perceive and comprehend intentions in other agents". The authors recognise the "[...] challenge of creating cognitive robots that can read intentions". The use of empathy in virtual and embodied conversational agents resulted in some advancement and even products being developed and used. Siddique et al. (2017) developed Zara the Supergirl for interaction with humans and the results confirm the authors' assumption that people have different preferences regarding personality traits of the other. Richards et al. (2018) use the human's perception ability of empathetic dialogue cues to tailor empathy individually. Sakurai et al. (2019) developed VICA, a Visual Counselling Agent directed to reduce stress in humans by means of a first-person interaction and building trust in those relations. Modelling and evaluating empathy in ECA poses some challenges being addressed by several authors (e.g., Yalçin, 2018; Yalçin, 2019; Perusquía-Hernández et al., 2019; Rashkin et al., 2018). Some authors claim that empathy is not in our genes and is simultaneously (a) agile by enhancing it with novel experience; and (b) fragile such as being broken by social commitment to a change in an opposite direction (Heyes, 2018).

The empathy (Stueber, 2014) concept is usually a starting point for social glue bringing better interaction, communication, and mutual helping which range from dyadic to many-to-many interactions. de Waal (2008) reports on the evolution of empathy and the perception of another's emotional state activating shared representations causing a matching emotional state in the observer. This reporting is based on extensive study and review mainly about primates (humans and apes) observed behaviour. Morgan (2016) argues that instead of the compassion used in healthcare practice, which lacks conceptual richness and clarity, one should focus more on a broader concept of empathy that includes both affective and cognitive components. Vaes et al. (2016) findings suggest that empathy can be triggered for non-human entities as long as they are seen as minimally human, i.e., a human agent may be empathetic with several distinct types of non-human agents (e.g., dog, ant, mobile device) as long as some human-attributed feature such as making conversation is recognisable in it. Conversely, Ranjbartabar et al. (2019) claims that further studies are required to understand better when to emphasise empathetic dialogue and when to interact neutral. Further, the results presented evidence of human diversity on the appreciation of a more or less emotional-aware dialogue. Regarding the controversial results presented, this research considers and develops the inclusion of empathy as a score to be perceived by the human as the HAI empathy level. This metric is sensible to the frequency of human reporting. It decays over time (e.g., 24-hour cycle) and the decay rate may change with other actions such as pausing the data collection. Empathy level notifications, are kept silent, only notifying when the human is interacting.

2.3 Cognition, Memory, Learning

> "Memory primes cognitive function and constrains learning providing a structure for the acquisition of new information." (Wood et al., 2012, p. 94)

The integration of memory and cognition is paramount towards learning and prediction. Regarding humans, a diversity of memory types have been devised and still being investigated for further understanding of inner and integrative mechanisms. The complexity on clearly comprehending these workings resides on memory not being located in a specific place of the human brain but widely distributed and integrated in a network of cognitive processes.

Concerning the inner mechanisms at the cellular and molecular level, both Long-Term Potentiation (LTP) and Long-Term Depression (LTD) are two well-studied and characterised forms of synaptic plasticity towards neural correlates of memory (e.g., Bliss and Cooke, 2011) and consequent support for cognition and learning processes. LTP and LTD induce specific patterns of neural activity in order to persist changes in the synaptic strength of affected neurons in a set of brain regions. LTP is the mechanism facilitating and enhancing the chemical transmission on a synaptic connection, i.e., individual nerves modifying themselves in order to sustain a synaptic response to a repeated stimulation — long-term memories. Conversely, LTD is responsible for the weakening of affected synaptic connection, this may be useful, for instance, at the hippocampus region for the removal of unnecessary memory traces. Figure 2.4, p. 19, depicts evidence of these structural changes associated with LTP and LTD where the authors describe the mechanism as:

> "[...] (A) Synaptic strength correlates with spine volume and the area of the postsynaptic density (orange). Note that the PSD[33] in potentiated synapses is often perforated. (B) LTP can also lead to the appearance of new spines. Within 30 min of triggering LTP (30 stimuli applied to the presynaptic axon at 10-sec intervals paired with depolarizing current injection into the postsynaptic neuron; black bar) of a synaptic connection in the hippocampus, new spines appear." (Lüscher & Malenka, 2012, Figure 4)

[33] postsynaptic density

Regarding memory persistency types and integration in the central nervous system, it is still debatable the acceptance of a clear separation between the short-term and the long-term high-level main categorisations. Nevertheless, there is some evidence suggesting that these two systems

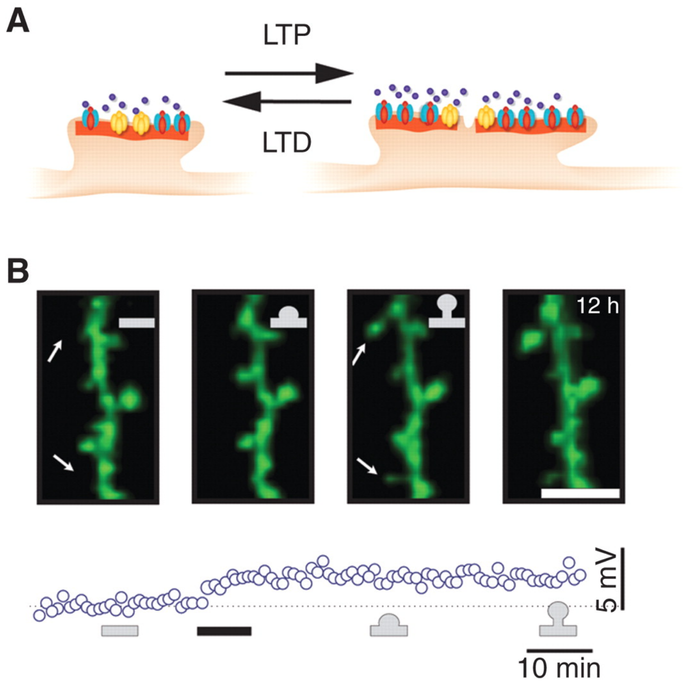

Figure 2.4: Structural changes associated with LTP and LTD.
Source: Lüscher and Malenka (2012, Figure 4)

are indeed distinct in the whole memory structure. Regarding short-term, an entity usually depend on this type of memory for current reasoning processes yet long-term memory may be recruited to assist on giving more information for a better decision-making process. Regarding the long-term persistence, it is far more complex including explicit (e.g., the knowledge accumulated and required to write an essay) and implicit (e.g., a composition of actions required to tie a shoelace) main categorisations. The explicit type can be further subdivided into semantic and episodic categories. The semantic deals with the concepts, facts, meanings, and knowledge about everything independently of the spatial and temporal context. Conversely, the episodic memory stores the representations of contextual knowledge, experiences and specific events including associated emotions.

Regarding identity, social bonding, representations and simulations of past and future scenarios, all these require a healthy memory ability of persistently remember but also to forget when required (or to let some memories decay and loose importance). Further, time and interference are essential for the "forgetting to remember" (Altmann & Gray, 2002) process mainly in the short-term memory. Concerning artificial companions, there is a review about memory models for intelligent social agents by Lim (2012). These memory-related mechanisms able to improve cognitive processes and, simultaneously, save limited brain resources on processing collected data are of interest for this research. Nevertheless, the goal is to keep both cognition and memory-related mechanisms very simple and efficient towards learning and predictive abilities.

The biologically-inspired or brain-inspired cognitive architectures are being developed and have been tried without much success in the past (Goertzel et al., 2010). Mainly the symbolic (which is more psychology-inspired than brain-inspired) architectures which seemed in theory very promising by means of universal representational and computational power, although tend to underperform in learning and memory (episodic and associative). Thus the current focus on hybrid and specific challenge-related architectures. This kind of approach promise to achieve faster results by means of the solution specificity, i.e., simply and straightforwardly towards the goal by leaving optimisations to the next stage. There are the obvious advantages of *ad hoc* architectures, specialised and eager for best results yet much more challenging for replication and scrutiny. The focus on brain-inspired instead of mind-inspired seems to be thriving with better promises (Ugazio & Ruff, 2017; Disney et al., 2016; Duch, 2017). The main components set of these kind of architectures must include memory and cognition towards learning.

Regarding artificial agents, machine learning encompasses data mining and predictive modelling by means of algorithms and statistical models, making use of patterns and inference towards learning, i.e., a machine able to learn about a specific task without explicit instructions being given. The way humans learn and how better replicate in a machine is still debatable. However, considerable efforts are being made in order to understand the involved mechanisms, even proposals for a unifying theory of generalisation by means of a Gaussian process function learning model, and making use of Bayesian optimisation processes (Schulz, 2017; Lorenz et al., 2017). These approaches applied to machine learning as predictive analytics and also as a requirement to reduce the search space efficiently for a solution are of interest for this research. An example of an estimator hyperparameter tuning by means of a Gaussian process optimisation (limited to fifty iterations) is depicted in Figure 2.5, p. 21. This convergence plot shows the successful minimisations avoiding an exhaustive exploration or a subset of the space comprised by all hyperparameters configurations. Thus with less time-consuming evaluations attain equivalent (or even better) solutions.

The memory mechanisms make use of duration, decay, and interference regarding different goals on (short and long-term) persistence as previously referenced. Regarding humans, there is also evidence of the influence of perceptual fluency (the ease in which a stimulus is processed) on judging about past and also future happenings (Palma et al., 2018). Regarding artificial agents, this fluency as a valid cue for past and future judgements may inspire algorithms that weights differently the data

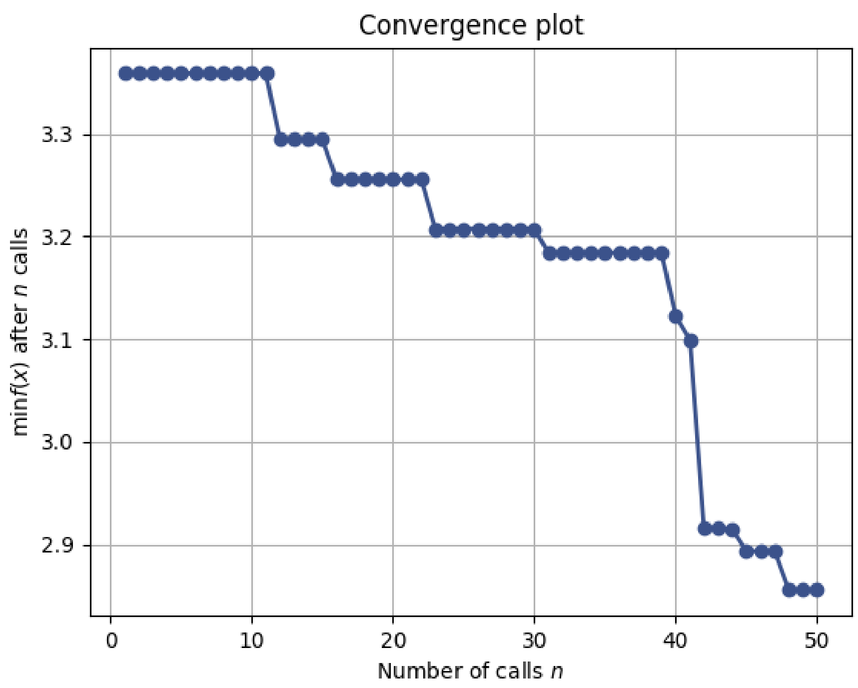

Figure 2.5: Gaussian process optimisation convergence plot on a function $f(x)$ (limited to 50 calls).
Source: `https://scikit-optimize.github.io/auto_examples/hyperparameter-optimization.html#convergence-plot`

easier to percept and process. This may include proper and adaptive rhythms of data acquisition, and sensors with efficient energy use better than others. These approaches regarding rhythm and efficiency are of interest to this research, and are further discussed in SensAI+Expanse, Chapter 3, p. 29.

2.4 Embodiment, Sensors, Interaction

The concept of interaction requires a reciprocal action or influence. In order to achieve this reciprocity, an understanding of the other must emerge from the two or more agents interacting. Regarding humans and anthropomorphic embodiment, Bavelas and Gerwing (2007) report on aspects of the human brain and cognition. For instance, the facial displays and hand gestures in face-to-face dialogue serve as empathetic glue for a better social integration and improved learning on parent-child interaction. Despite the existence of an efficient channel of communication using speech (interaction by human spoken language) other channels such as these of hand gestures and facial displays are in place for its usefulness. Thus natural communication evolved making use of multimodal interaction.

Regarding the human-artificial interaction, there have been a myriad of products, projects, and research ranging from simple pragmatic question-answer communication to affective-aware and even empathetic transmission. Life-like artificial agents such as the ambitious BabyX (Figure 2.6, p. 22) from Sagar et al. (2016) are being developed in an effort to give the illusion of life in anthropomorphic virtual agents for realistic HAI towards affective connections. The authors use (and contribute to its development) the Auckland Face Simulator platform[34] in order to produce this and others highly realistic avatars (synthetic embodied representations of humans) with the ability of real-time interaction.

[34] `https://www.auckland.ac.nz/en/abi/our-research/research-groups-themes/laboratory-animate-technologies.html`

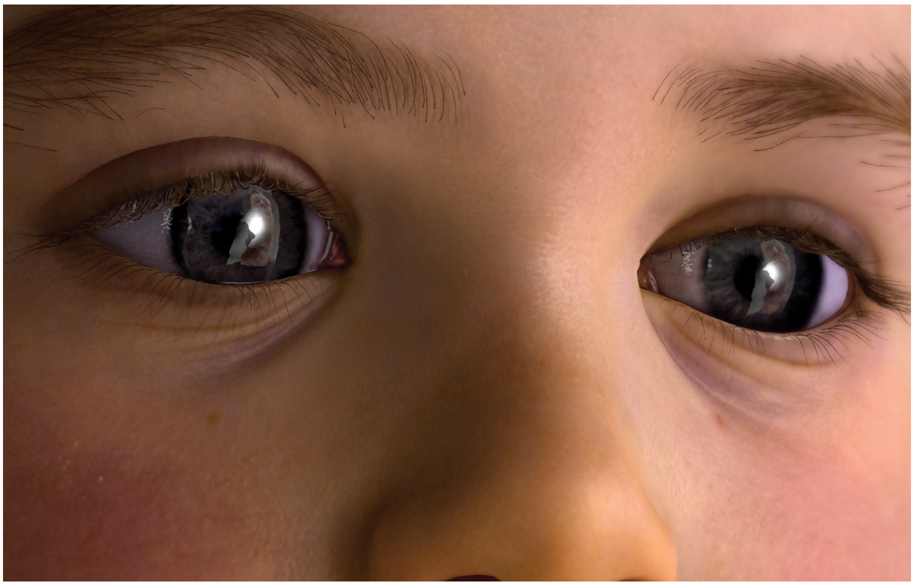

Figure 2.6: BabyX from real-time interactive psychobiological virtual infant simulation.
Source: Sagar et al. (2016, Figure 4)

Furthermore, these avatars are beyond simple imitation of human behaviour. The long-term goal towards autonomous agents, able to face-to-face interaction as humans, is driving Sagar et al. (2016) research holistic approach to explore paths such as the development of several biologically inspired cognitive architectures. These models support the dynamic behaviour of the face which emerges from a myriad of systems interacting on multiple levels, from low-level biological mechanisms to high-level social interaction. The integration of several pieces such as a neural system and subsystem models towards a synthetic nervous system is supported by a developed modular simulation framework called Brain Language (Figure 2.7, p. 23)[35]. This framework aims to integrate neural networks with real-time computer graphics and sensing. It supports a wide range of neuroscience models from simple spiking neurons to self-organising maps and able to interconnect in order to form larger neural networks. All these aim at online learning to happen in a live context gathering spatial and temporal data.

[35] Description from the source: “Screenshot of interactive BL viewer: (left) BL scopes viewing activity of a single neuron (top) or an array of retinotopic neurons (bottom) during a live interaction; (right) partial view of BabyX’s virtual connectome, which can be explored interactively; connections light up (green or red) when activated”

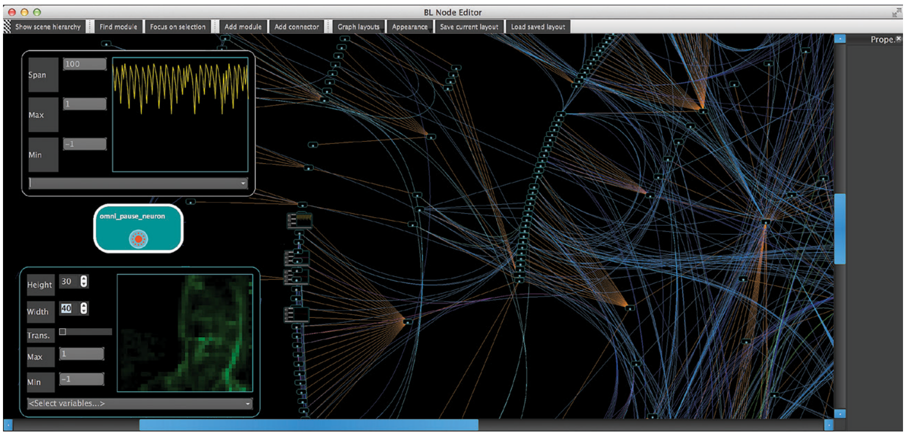

Figure 2.7: An interactive Brain Language viewer example.
Source: Sagar et al. (2016, Figure 2)

Although modelling human behaviour and dyadic interaction is a colossal task, a fusion of multimodal data sources for emotion detection and recognition may be used to improve accuracy (Castellano et al., 2016; Castellano et al., 2013; Castellano et al., 2008). Multichannel interaction may be also useful to cope with cases where not all channels are in use (e.g., the human is not able to speak hence no audio channel available). The anthropomorphic near-realistic faces of artificial agents also poses some challenges such as the *uncanny valley*[36] of perception. The term coined by Masahiro Mori in 1970 is for human-like aesthetics and behaviour on robots and does not make any reference to non-anthropomorphic devices like computer tablets. Meaning that this effect may be avoided on non-anthropomorphic agents.

[36] Piwek et al. (2014) for a review about the *uncanny valley*. Moreover, the authors claim that achieved results show no evidence to support the further "deepening of the valley with motion" (originally predicted by Masahiro Mori in 1970).

Regarding this type of artificial agents, a software architecture is proposed and developed for a rather original social robot named YOLO by Alves-Oliveira, Gomes, et al. (2019). The authors aim on a creativity-stimulating agent through storytelling activities for children. To the purpose, artificial intelligence software is developed in an attempt to enable nurturing capabilities of the social robot towards fostering human creativity. In this sense "[...] this software allows the creation of Social Behaviors that enable the robot to behave as a believable character". In order to accomplish this believability, the different behaviour capabilities "[...] were based on psychological theories of personality and developed using children's input during co-design studies". Furthermore, still on the believability, one may argue that trust and reputation (Granatyr et al., 2017) are of interest to implement in HAI. This characteristic reveal the need for an ethics-related approach to the development of artificial agents in which the trust-based approach, mixing emotion with reasoning to create expectation on the outcome in the HRI, will be an advantage (Koyama, 2016). Concerning this complexity of fusing affects with moral aspects towards decision-making, Coelho et al. (2010) propose a moral agent core architecture where a context-sensitive approach has strong influence regarding the decision ability. This proposal includes a mix of both affective and morality managers, and several modules such as about reactive ability, perceptional sensors, and utility function towards a deliberation component feeding the decision-making. Further, Da Rocha Costa and Coelho (2019) show that in order to compose the reputation-based moral regulation mechanisms for the negotiation processes in agents' societies by means of social exchanges, the same agents need to be sensible to moral sanctions. The authors have considered both self-interest and emotions to ground the agent's sensibility to moral sanctions, but only the self-interest was taken into account regarding the proposed moral systems, reputation systems, and evolutionary games articulation. Moreover, the authors conclude by clearing that the possibility of negotiation of moral sanctions and creation of moral norms were not explored.

Asada (2019) even further proposes, as a working hypothesis, artificial pain as a required feature in order to shape conscious minds of robots through the developmental process of empathy, morality, and ethics regarding the emergence of the self concept. Still on conscience in artificial agents, now with feeling, Man and Damasio (2019) propose a new class of sentient machines with physical constructions maintained by homeostasis and thus sharing some core traits with all natural living systems. The authors claim that the introduction of an aware risk to self (vulnerability) will give rise to true agency "[...] when it acts with a preference for existence over dissolution".

Regarding HAI advancements, research in the fields of HRI and HCI has been contributing with knowledge exchange by means of research findings in order to improve dyadic interactions. Sophisticated new ways of enhancing touch interaction on humans, screens, and everyday objects were proposed by Sato et al. (2012) and, more recently, by Liu and London (2016) researching on a tangible AI interface. These latter authors claim that the proposed morphing touchable add-on for a mobile device, providing the perception of physical presence, improved the users' empathetic connection with the virtual intelligent system, leading to a better human-agent communication experience. This study used a text-based conversational agent, i.e., a mobile device (Android™) chat application. The participants were separated into two groups, a control group where the subjects used only the text-based interaction, and a second group where the physical interactions are in place during the text-based conversation. Although an interesting research yet the results are only preliminary due to under-representation of female gender and age ranges in both groups, and also some bias regarding the introduced pauses in tangible AI versus the text-based only. Conversely, without shape-changing physical add-ons, wearable computing for HCI is also using state-of-the-art touch devices such as the Google Glass depicted in Figure 2.8, p. 25, which seamless integrate a computer with a touch interface, and several sensors including a video camera, in a human glass frame. These not yet mainstream solutions include capabilities of geographical location, natural language processing, and machine learning, providing augmented reality with numerous applications ranging from production lines to entertainment services.

Figure 2.8: Google Glass Enterprise Edition 2 wearable device.
Source: `https://www.google.com/glass/tech-specs/`

An avant-guard research by Zhao et al. (2016) on emotion recognition by means of wireless signals reflected off the body shows how to measure physiological signals without invasive, nor wearable, sensors. The authors claim affect recognition on a two-axis emotion model, similar to the circumplex model of affect by Russell (Figure 2.1, p. 10), with four discrete emotions: sadness (negative valence and negative arousal), anger (negative valence and positive arousal), pleasure (positive valence and negative arousal), and joy (positive valence and positive arousal). Moreover, the results presented are accurate enough for comparison with the on-body electrocardiogram collecting the ground truth values, i.e., on a par with state-of-the-art emotion recognition systems. Further, the authors envision future research about heartbeat morphology in the context of non-invasive health monitoring and diagnosis.

Still on the use of multimodal sensing, Subramanian et al. (2018) research and develop a database connecting personality traits with emotional states. The result is the ASCERTAIN database encompassing FFM personality scales and emotional self-ratings of fifty eight humans. It includes the recorded electroencephalogram, electrocardiogram, galvanic skin response, and facial activity data collected while viewing affective film clips. This database of emotional attributes and personality traits relation was used to study the influence of personality differences on humans' affective behaviour. The authors conclude envisioning "The fact that personality differences are observable from user responses to emotion-wise similar stimuli paves the way for simultaneous emotion and personality profiling". Still on biometric signals, Bota et al. (2019) present an interesting review including challenges and future opportunities on emotion recognition using machine learning and physiological signals. The authors caution for the lack of evidence regarding "[...] which feature combinations of which physiological signals are the most relevant". Moreover, the unavailability of public datasets regarding emotion recognition and the possible modalities in unconstrained daily-life scenarios is also an open issue.

The mobile devices, smartphones, tablets, and wearables, have been elected as complementary (soon to be principal) hardware support tools for non-invasive use in healthcare and well-being. This trend is being driven by a growing interest in cognitive science, specifically psychiatric and psychology fields combining with affective computing and making use of machine learning from AI. Conversational agents, embodied and emotion-aware or even with affective capabilities, are being actively researched and developed for several years now. Empathetic ability is included in cases where affective interaction with the humans is required or of interest (e.g., Sorbello et al., 2016; Lin et al., 2019). Although full of challenges yet developments and proposals of cognitive and affective architectures are thriving (e.g., Pérez et al., 2016). Inkster et al. (2018) developed a study using an empathy-driven ECA named Wysa regarding mental health. The authors claim to have identified "[...] a significantly higher average improvement in symptoms of major depression and a higher proportion of positive in-app experiences among high Wysa users compared with low Wysa users". Vaidyam et al. (2019) in a recent review about conversational agents in mental health conclude that "[...] evidence shows that with the proper approach and research, the mental health field could use conversational agents in psychiatric treatment". However, the same authors also caution that "[...] further research with standardized outcomes reporting is required to more thoroughly examine the effectiveness of conversational agents". On a field study regarding an empathetic robot for group learning, Alves-Oliveira, Sequeira, et al. (2019) explore HRI with empathy to address if these conditions achieve more collaborative learning and even better positive educational outcomes in a long-term interaction with groups of students in a learning context. The authors state that there were no evidences of relevant learning gains provided by the long-term educational interaction study. Concluding that "This result reflects the need to perform more long-term research in the field of educational robots for group learning".

Regarding the affect sensing using a smartphone, there are several studies underway already with important published results (Rana et al., 2014). These are studies focused on (affect) sensing instead of a simple interaction between a human and a non-anthropomorphic agent (e.g., smartphone embodied). Nevertheless, some interaction still occurs when the human is required to assess (by means of a questionnaire or similar) the current accuracy of the smartphone's agent results. These two concepts of affect sensing and agent's accuracy assessment by the human are of interest for this research regarding a cognitive architecture. Moreover, Guo and Hoe-Lian Goh (2016) results suggest that the perceived usefulness is higher when humans interact with an affective embodied agent, and lower when interacting with a non-affective or neutral agent. Though the learning outcome (digital game-based learning) was the same with either types of agents.

The mobile device non-invasive sensors required for affect recognition may vary ranging from camera (visual perception) and audio to an accelerometer for movement patterns. Regarding visual perception, although facial display and movements are being actively used on emotion recognition yet Barrett et al. (2019) advise for further and proper research in order to support claims about generalisation of the findings. Affect recognition in speech has already some interesting advancements regarding emotions and personality traits in solo speech and also in a multimodal approach. The current research includes end-to-end deep learning, support vector machines models, and a fusion of heuristics, algorithms, and feature extraction techniques in order to achieve the best comparable accuracy (Siddique et al., 2017; Fayek et al., 2015; Solera-Ureña et al., 2016; Carolis et al., 2016; Trigeorgis et al., 2016). However, Cevher et al. (2019) caution regarding off-the-shelf tools for emotion recognition in speech and face patterns versus neural transfer learning approach for the same recognition from text sources. The authors find that the tested off-the-shelf tools are not ready yet for in-car context speech interactions emotion recognition without further adjustments. Thus affective state assessment by means of text sentiment analysis emerge as of interest to be included in a multimodal approach to affective recognition. There are several tools and extensive research on this type of affective analysis ranging from sophisticated natural language processing using machine learning (Calefato et al., 2017) to simple rule-based models using lexical features (Hutto & Gilbert, 2014). Moreover, text sentiment analysis has already been included in recent multimodal approaches to affective recognition (Soleymani et al., 2017; Provoost et al., 2019). The simple rule-based model by Hutto and Gilbert (2014) for sentiment analysis on short texts is of interest to this research because of (a) its efficient energy use and instantaneous analysis; and (b) the state-of-the-art accuracy and generalisation across contexts better than any other tested benchmark.

The integration of smartphone sensing, AI, affective and cognitive research, and well-being towards a better life experience for humans is already happening. The affect sensing using wearable or mobile devices such as a smartphone was envisage back in the 1998 Scientific Symposium by the American College of Medical Informatics (ACMI) — one of the informatics challenges for the next 10 years (Greenes & Lorenzi, 1998). More than twenty years have passed since this awakening and smartphone sensing for behavioural research is thriving with active discussions (Harari et al., 2017; Cornet & Holden, 2018; Denecke et al., 2019). A common finding within the several available reviews is the agreement on a myriad of new possibilities that this interdisciplinary collaboration between experts from informatics and health fields foster. Simultaneously, there is a concern relating to methodological, clinical integration, and privacy issues. More research with proper methodology and transdisciplinarity is needed for sustained achievements (Place et al., 2017). Nevertheless, there is a continuous effort to develop tools regarding smartphone sensing for mental health and well-being such as the EARS tool from Lind et al. (2018), and the component-based mobile and Cloud-integrated system for rapid prototyping towards more efficient research focused on science instead of technology issues. The smartphone sensing in therapy context has already some products and developed applications such as HappyHour from Carmona et al. (2015). The authors argue that fusing several sources of data may lead to a more accurate detection of emotions, specifically data from smartphone sensors, an electrocardiogram, and weather information. Moreover, the goal of HappyHour is to positively impact its users' mood through moderate walking exercise. In order to accomplish the proposed goal it uses machine learning to infer emotional states and recommend actions such as a walk in the park (using points of interest). Although an interesting effort towards the users' well-being yet no proper scientific studies and results were presented regarding the claim of mood improvement. Conversely, Lathia et al. (2017) did a proper study using longitudinal data from ten thousand individuals who downloaded a mood-tracking smartphone application and self-reported about happiness and physical activity. Simultaneously, movement pattern data from the accelerometer sensor was collected. The results reveal a strong correlation between physical activity and happiness, i.e., the more physically active the happier. A perceived limitation is the possibility of missing some activity from the data collected due to the case where the smartphone is not being carried during the exercise. Nevertheless, presented findings demonstrate the usefulness of smartphone sensing regarding "[...] health-related phenomena as they naturally occur in everyday life".

Chapter 3

SensAI+Expanse

This chapter describes the platform comprised of the SensAI agent and its SensAI Expanse resources — SensAI+Expanse. A distributed, fault-tolerant, mobile and Cloud-based platform from scratch as a research tool to continuously, online, gather and process data. The general HAI is initially restricted by its parameter values which drives SensAI. This behaviour may be influenced by the agent's context along the interaction timeline and changes may emerge. Integration[37] of the application and the platform services architecture is described along with some development details about its cognition, memory, sensing, embodiment, and interface.

[37] Refer to Development Tools, Appendix A, p. 103 for specific information about the repositories and application (a) as a product; and (b) how is distributed and accessible in order to develop new interactions for further research.

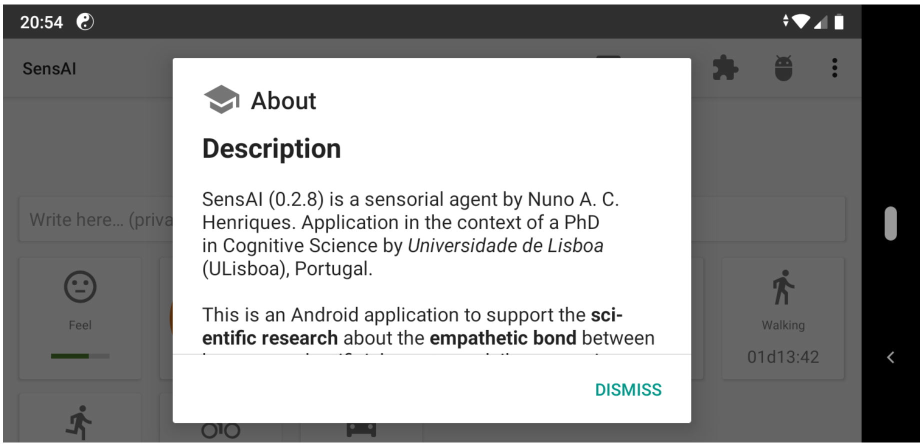

Figure 3.1: SensAI application live in an Android™ device.

The technological choices for a functional prototype development should be the most suited for the project specifics. The requirements, commitments, and compromises to be made on choosing the software development platform were based on the criteria of (a) performance as in efficient; (b) free licence as in Free and Open Source Software (FOSS); and (c) community momentum, support, and compatibility as in pervasive.

3.1 Introduction

An artificial agent may have interest in adjusting empathetically towards the human's current affective state (Picard, 2003; Castellano et al., 2010). The interaction is based on the way each entity perceives contact, together with the perception of human's affective states using a multimodal approach. Regarding sentiment, social network posts (Twitter status) and in-application diary written texts will be subjected to sentiment analysis in order to collect emotional valence from this modality source. The ground truth is obtained from the user on self-reporting about current mood, i.e., positive, neutral, and negative emotional valence. The agent will be subjected to a simple adaptive process by means of interaction with humans. SensAI presents an empathy score value during this interaction. The score decays over time, it also changes with some factors such as the frequency of human reporting. This visual adaptive metric should be perceived by the human as current human-agent empathetic score (Figure 3.2, p. 30). Empathy (Stueber, 2014; de Waal, 2008) is used as a starting point for social glue bringing better interaction, communication, and mutual helping ranging from dyadic to many-to-many interactions.

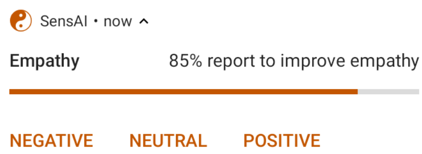

Figure 3.2: SensAI empathy notification including valence report buttons.

A novel approach is taken by this work where (distinct and complementary) algorithms will be used to endow an artificial agent with the prediction ability of the human emotional valence state in spatial and temporal context. The developed and also off-the-shelf algorithms include rule-based (Hutto & Gilbert, 2014) lexical processing of human written text in order to obtain a valence value, i.e., sentiment analysis on demand avoiding overload of the mobile device limited resources such as battery life by using efficient machine learning techniques. Moreover, fine-tuning of some off-the-shelf algorithms such as (a) offline language translation in order to support Portuguese additionally to English; and (b) emoticon fail-safe cascading from ineligible text to be used as primary valence value.

The SensAI agent is non-anthropomorphic, non animal-like, faceless, and will not have the ability of locomotion as in robots. It is a mobile device (smartphone) embodied thus a near-body daily companion. The agent has distributed memory using Cloud resources for full data persistence besides local device data cache. Moreover, its expanded cognition comprise processing and learning capabilities beyond the ones available in the physical structure. The agent's embodiment encompass the human-agent interaction interface and perceptional sensors which are configured (developed algorithm) for a balanced data acquisition rhythm (e.g., $active = 2s$, $inactive = 8s$, $f = 1/5Hz$, $D = 20\%$). This setup avoid too much pressure on resources (power, memory, processor availability) simultaneously assuring relevant data. The artificial agent empathy score will depend on the human-agent multimodal interaction. The continuous reporting and moving by the human whilst carrying and interacting with the agent (smartphone) will contribute with data. The entity responsiveness is designed using a simple notification with the above referred empathy score. Additionally, a dashboard with physical activity information is available with real-time and also aggregated data including a circular histogram of the self-reported emotional valence (Figure 3.3, p. 31).

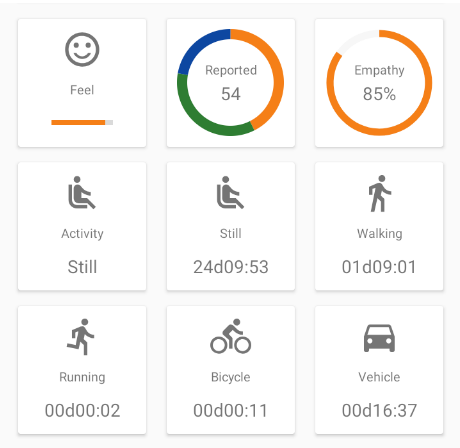

Figure 3.3: SensAI dashboard with real-time and aggregated data.

Regarding the possible affective models for perceiving the human emotional state changes, a simple valence-based analysis will be used. This is similar to text-mining polarity models, a continuous scale ranging from -1 to 1 will be used to score human's current emotional valence such as the implemented solution by Hutto and Gilbert (2014). Despite being a continuous scale, three classification ranges are devised where a score of $[-1.0, -0.5]$ is negative, $(-0.5, 0.5)$ is neutral, and $[0.5, 1.0]$ is positive. These classification ranges are helpful on correlating with the ground truth data from the human reporting, i.e., values from three buttons aligned in a polarity scale emoticon-style such as [☹] [😐] [☺]. Moreover, the three classification ranges are in the domain of continuous values able to support computations of diverse (weighted) valence sources and, simultaneously, capable of computing a discrete result amongst negative, neutral, and positive ranges extremes -1.0, 0.0, and 1.0 respectively.

The design and implementation of this companion will encompass the integration of perceptual and cognitive layers, along with modules for affective states recognition which rely on the memory and learning modules. The architectural conception is modular enough to ease endowing a computer with adaptive characteristics towards an autonomous artificial agent with empathetic behaviour. A continuous time-model with iterated evaluation for testing and improvement is devised and implemented. The resulting system (a) supports each agent cognition and memory distributed resources; (b) learns each human-specific valence changes model; and (c) has the ability of valence state prediction in context for each human. A simplified view of the SensAI+Expanse system is depicted as a conceptual data flow in Figure 3.4, p. 32.

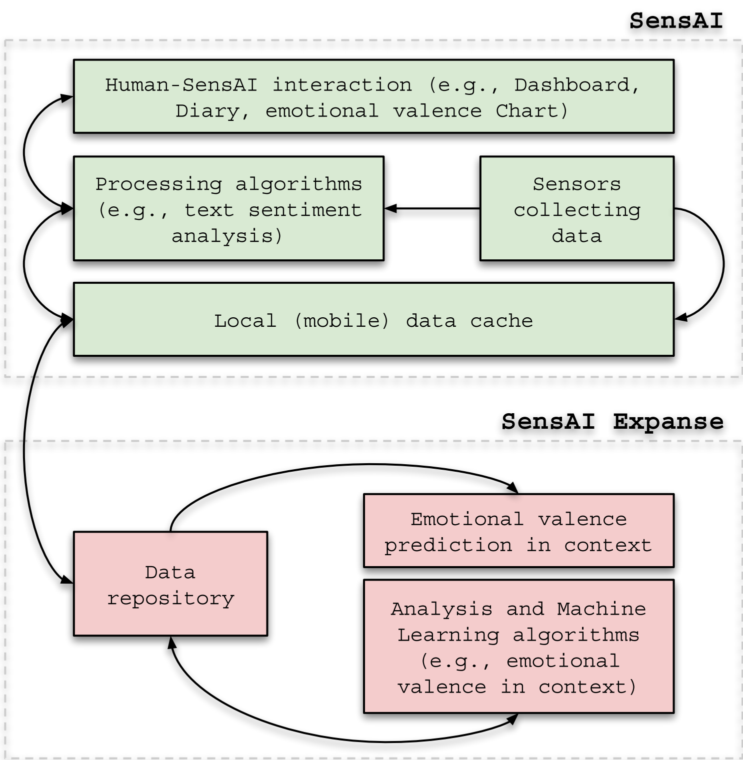

Figure 3.4: SensAI+Expanse conceptual data flow.

The Expanse assures data storage and machine learning towards prediction for all SensAI accompanied individuals. Moreover, resource consumption tasks such as learning, may be extended, connected to the database and exported as a Web service to be consumed by a mobile embodied agent. The Expanse is versatile enough to serve as an offline system enabling data analysis and also as an online service offering functionalities to SensAI. For instance, a Jupyter[38] Notebook script for statistical analysis over collected data and a machine learning algorithm may be used for the SensAI capabilities and functionalities expansion. (Figure 3.5, p. 33)

[38] https://jupyter.org/

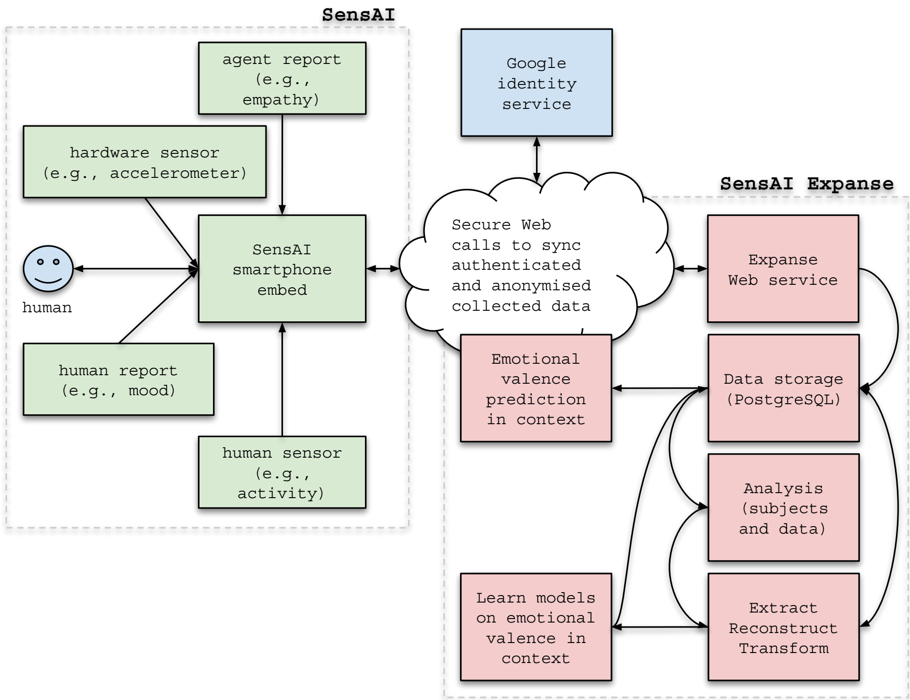

Figure 3.5: SensAI+Expanse data flow.

3.2 SensAI

SensAI is an embodied agent for interaction with humans towards an artificial empathetic companion. It is a sensorial interacting agent developed as an application for the Android™-based mobile devices. It enables smartphone sensing[39] in a multimodal approach including in-application diary written texts and social network posts subjected to sentiment analysis. The human individuals delivery ground truth emotional valence by means of touch screen buttons. The agent gives back some statistics and charts, physical activity data, and a few messages stating current (e.g., last three hours) emotional valence state.

[39] Cornet and Holden (2018)

3.2.1 Approach

General guidelines and restrictions are devised for an almost non-invasive interaction, i.e., a quiet agent that will sense and adapt based on human data towards empathy. SensAI is (a) non-anthropomorphic; (b) non-animal-like; (c) faceless; (d) a near-body daily companion using mobile device sensors; (e) accepting written texts as diary notes and thoughts; (f) accepting human one-click emotional valence in a three-value scale (negative, neutral, positive); (g) sensing physical activity and showing on a dashboard; and (h) collecting geolocation data periodically and event-triggered amongst other sensors. All the sensing and event data collecting must have a balanced rhythm to save power, processing, and memory resources whilst assuring relevant data. Also, SensAI has to cope with memories, short-term for sensing, long-term to support and consolidate context about some other agent behaviour and dyadic relation. Also, working memory to process and make decisions about actions and persistence reinforcement or cleansing.

The human supervision on interacting with the artificial agent must be available. The dyadic bonding and social conditioning (e.g., like the parental supervision in humans and most mammals) cases are used towards group inclusion by abiding to the same rules and baseline behaviour of every other group member (family and more embracing). Thus including the possibility to suppress the interaction by disconnecting the agent sensing, i.e., the human may turn off all data feed with one-click setting. This feature is a suspension option called "pause". Using this mode, SensAI will be disengaged and will not be collecting data (all sensors off) autonomously although it will remain available and waiting for some interaction such as engaging once again. Also, Expanse will remain active in a continuous learning process regardless. The agent will be always ready for the human's decision to re-engage.

An agent towards empathy[40] should nurture the relation between itself and the other agents (human and artificial). There is more than one way for this goal achievement. For instance, an agent detecting a lowering emotional valence may tend to present more positive (valence) aspects, engaging in a congruent interaction that pleases the other towards the improvement of the dyadic bonding (Figure 3.6, p. 34).

[40] Section 2.2, p. 15

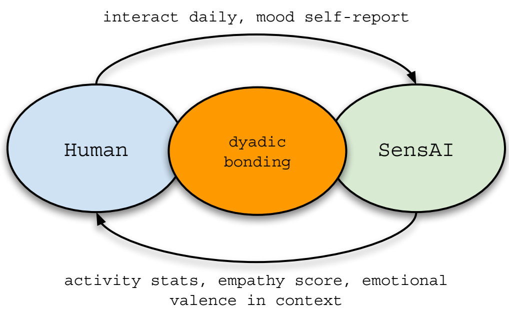

Figure 3.6: SensAI mechanism towards empathy.

Moreover, the engagement with varying affective behaviour is not without risks. Cramer et al. (2010) findings indicate the importance of some restraint when implementing social behaviour. An agent's mistake in assessing the affective state of the user and, consequently, an inaccurate empathetic behaviour may have a detrimental effect on trust by the other agent such as a human. When the likelihood of wrong inferences on user affect is high then the empathetic behaviour should be restrained. Rukavina et al. (2016) highlights the "[...] need for multimodal emotion classification since facial expressions as the sole basis are not always indicative of the truth that is important for the task at hand". Also, physiological data acquisition is referred as a good complementary mode. Another complimentary solution is the use of sentiment analysis by means of text mining from chats and messages written by the subjects when interacting with other humans, agents, or simply by posting on social networks. Additionally, the gender matters when analysing the resulting facial expressions as a reaction to the positive feedback. Results from Rukavina et al. (2016) show that more female subjects reacted negatively to the positive feedback than the male ones.[41]

[41] Even though the sample size is small (30 subjects) and considered a limitation by the authors.

Furthermore, humour is used for social bonding between humans. HAI (robot and software) may also improve from joke telling by the artificial agent (Tay et al., 2016) specially the non-disparaging ones[42]. The SensAI includes already prepared mechanisms such as notifications and in-diary messages able to present jokes (written) as a dyadic bond reinforcement.

[42] Jokes that avoid derogatory remarks about human condition and sexuality.

The multimodal approach currently implemented purposely exclude (a) microphone use to capture audio; and (b) camera use for still and moving pictures able to be used on facial display feature extraction towards emotional status appraisal. These exclusions are supported by the almost non-invasive approach where privacy must be preserved. Moreover, facial expressions and movements are tied to the immediate context such as (a) internal metabolic state or past experiences coming to mind; and (b) external context being at work or school and who else may be present at that moment. Thus stereotypes must be avoided. Despite the human face being an efficient tool for social communication, one should be aware of the strong context-dependency of facial movements when inferring or judging from perceived face pattern. Barrett et al. (2019) further recommend:

> "Instead, the facial configurations [...] are best thought of as Western gestures, symbols or stereotypes that fail to capture the rich variety with which people spontaneously move their faces to express emotions in everyday life.
>
> [...] a prototype is the most frequent or typical instance of a category (Murphy, 2002), whereas a stereotype is an oversimplified belief that is taken as generally more applicable than it actually is. [...] emotional expressions are more variable and context-dependent than commonly assumed [...]" (p. 46)

An in-context supervision feature is available as feedback given by the human interacting with three buttons in the main UI, also present in the notification message. These represents a discrete polarity scale of negative, neutral and positive. Hitting one of these buttons will provide on-demand ground truth polarity registered at a given moment and geographically referenced.

3.2.2 Application

The SensAI agent encompass activities and dialogues to interact with the human. As a mobile device application, it is comprised of several functional modules and layers. Additionally, there are active services to keep all the work running detached from the interaction activities. Most of the work is done asynchronously with scheduling and also on demand (e.g., collecting data from a new sensor event). The CogA[43] libraries for text sentiment analysis, language identification and translation are embed in the SensAI services. A local (device) storage is used for data cache (kept for several days) while waiting for the synchronisation with Expanse. Java is the main programming language used for the software development. Activity, service, and special keep alive modules are no more than developed classes and instantiated objects. Some run on demand, others periodically, as services and activities on dedicated threads and also on the main UI thread. The software architecture modules are depicted in Figure 3.7, p. 36. All the relevant activities and services developed and included in the SensAI embodiment are described below:

[43] Source Code, Appendix C, p. 115.

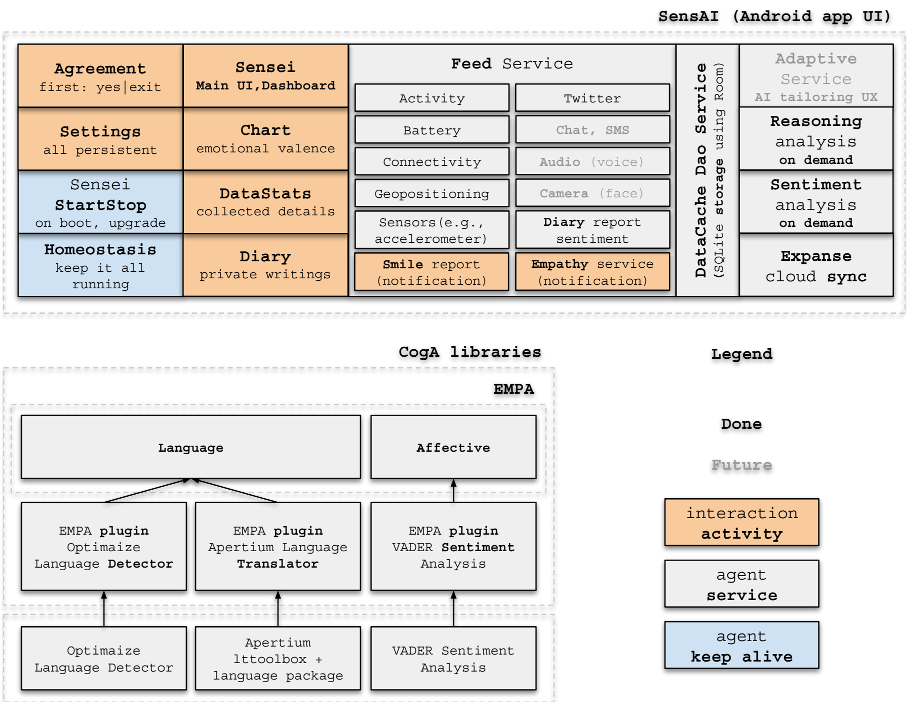

Figure 3.7: SensAI software architecture modules.

Sensei is the main activity, acts as a maestro which conducts all the other pieces at start up and presents the main interaction interface with the human. This display includes a dashboard of activity information (e.g., walking accumulated time) besides shortcuts to the diary writings, and emotional valence self-report buttons additionally to the ones present in the notification mechanism. Moreover, it manages all the required permissions to be given by the human (e.g., location data using Global Positioning System (GPS) sensor) and also the first interaction workflow presenting the agreement among other requirements. An active, in the wild, SensAI main UI is depicted in Figure 3.8, p. 37.

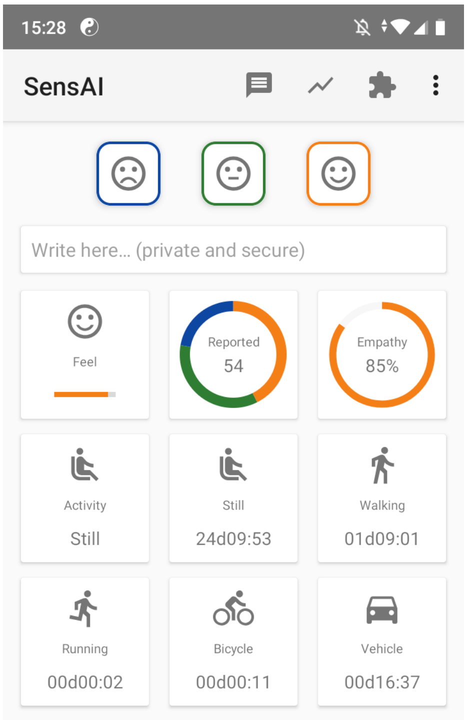

Figure 3.8: SensAI main UI and system notification bar.

`Agreement` is a one-time written text stating the application purpose and requesting consent for collecting data. The human may agree or simply exit. By design, no data whatsoever will be collected before the agreement consent.

`Settings` are for the human to define some behaviour (e.g., pausing data collection) and to provide some (required for the scientific study) demographic data such as birthdate and gender (Female, Intersex, Male, Neutral). These definitions may be changed and extended in the future with few to none code adaptations.

`Homeostasis` in order to guarantee some fault tolerance and keep the agent in a good health, is a periodic, scheduled, service. For each run, it checks for critical aspects such as database health and data feed. On detecting an issue it will resource to proper action in order to solve it, some common failures are already identified such as `Feed` service not running. Moreover, it adapts itself to the interaction state, i.e., if at rest then database optimisations and repair actions may run, conversely, updating notifications such as empathy score level only happen when the human user is paying attention. This mechanism prevents potentially disturbing events such as too frequent device's screen awaking just for an empathy level adjustment. The homeostasis-like solution for the SensAI application is complemented with `SenseiStartStop` required to protect and guarantee `Homeostasis` service to run as expected.

`SenseiStartStop` is a system event receiver to assure persistence and robustness against the mobile device failures and reboots. It is responsible for doing a system registration at SensAI first start in order to be called on device boot and on application upgrade to deal with those special states. This registration also signals Android™ operating system to revive SensAI in case of unexpected crash and removal from running state.

`Chart` presents icons of emotional valence value and associated data over time. The chart currently includes (a) human self-report (ground truth) emoticon-style concerning emotional valence value, i.e., ☹, 😐, and ☺; (b) social network posts subjected to sentiment analysis (e.g., Twitter logo for tweets); and (c) a text message symbol for the diary writings also subjected to sentiment analysis. The timeline displayed (horizontal axis) defaults to the time range of the presented information yet with a superior limit restricted to the local memory setting (e.g., twenty eight days). The vertical axis displays a scale set by the emotional valence range segmented in negative, neutral, and positive intervals from bottom to top. Also, the emotional valence respective colour are used to emphasise each one of the three segments. Moreover, screen touch and gestures may zoom, pan, and select additional information over the icons. An active SensAI sentiment chart is depicted in Figure 3.9, p. 38.

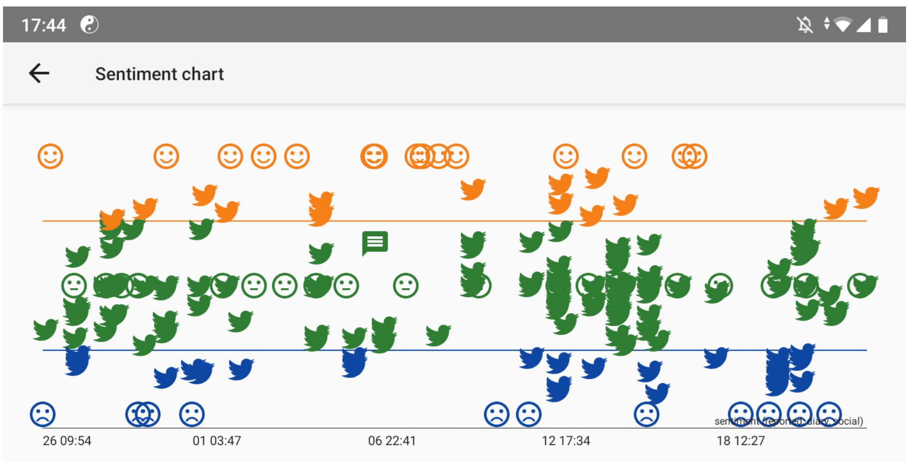

Figure 3.9: SensAI sentiment chart.

DataStats was first designed as a debugger screen during development, next was adapted for full disclosure of which, when, and how the sensor events are collected. Moreover, it displays agent's active percentage which is typically near 100% yet may be below that value when the pause option is activated (no data being collected) or the application is forced to shutdown and Feed service stopped. Additionally, information about Expanse syncing status is also presented.

Diary is a private place for writing thoughts, only the text sentiment analysis value is synced with the Expanse. The written texts never leave the mobile device thus may be lost if the application is uninstalled. This module is already designed and prepared to be a chat such as in a text-based ECA towards an improved HAI. Moreover, SensAI already posts periodic messages providing information about human emotional valence (e.g., average value over the last three hours) and physical activity such as walking (Figure 3.10, p. 39).

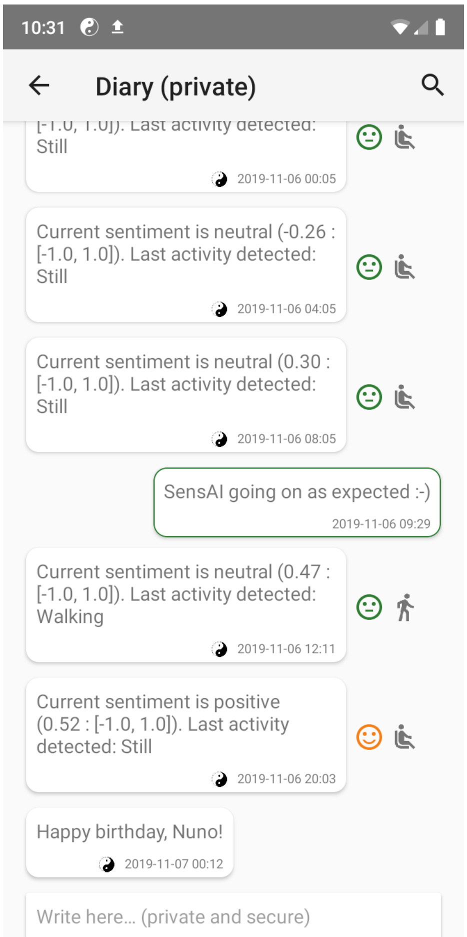

Figure 3.10: Human diary with some SensAI messages regarding human data.

Feed is a started service running autonomously in the background. SensAI includes several other services that only run on demand by scheduling or explicitly calling. This module encompasses and manages all sensors data collecting such as the Android™-device hardware included accelerometer. Moreover, it seldom interacts using notifications triggered by the occasional emotional valence human self-reports (using the emoticon buttons). The reports and other events trigger an empathy level adjustment which is displayed in a progress bar inserted in the application persistent notification message. It also manages the empathy score decay. This is the core service providing all the passive data available from the mobile device sensors.

`Sentiment` analysis utility which includes integration with language detection and translation facilities provided by the CogA libraries and services. A custom heuristic[44] is developed[45] in order to adapt the analysis to human idiosyncratic aspects such as mixed languages and emoticons when writing short messages. All in the best effort to get the sentiment value from text plus emoticons along with the language identifier.

[44] Minsky, 1961; Romanycia and Pelletier, 1985 for more about heuristics in AI.
[45] Detailed description in "Cognition and Memory", p. 45.

`Reasoning` contains common algorithms as utilities to help SensAI on reasoning about the human. It contains a specific algorithm to assess the current emotional valence value from all the collected data relevant for the computation (e.g., human self-reports, diary messages sentiment analysis). More reasoning helpers may be added on improving SensAI capabilities.

`DataCacheDaoService` the mobile device local data cache service for the collected data. It includes the functionalities able to set (insert) and reset (update) database data. The intent received must include the action request along with the new data which is processed upon verification of the intent. This service provides a staging (limited-term) memory-like database before sync with a permanent (long-term) storage. This is a one way data service providing a channel for the other SensAI services and activities to store the collected data. Moreover, the `DataCache` data access object is the one providing all the database available actions including the ones called by this service and also the others related with data retrieval (query) and deletion. Furthermore, `DataCache` includes (re)engineering of the full text search (FTS) feature in some database tables. Additionally, it deals with the migration between database versions to avoid data loss on application upgrades.

`Expanse` is a periodic, scheduled, service for data sync between the mobile device local cache and a persistent (long-term) memory aggregator in the Cloud — Expanse. It provides a fault-tolerant service by means of a mechanism similar to a (relational database) transaction, i.e., only successfully transferred data is marked as such (able to be deleted after local cache persistence time limit). Moreover, on lack of a suitable data connection available it will increase verification frequency for later try to sync. This mechanism of local cache syncing with the Cloud is paramount in order to restrict memory resources consumption and, simultaneously, guarantee proper data collection persistence.

Regarding information security, all SensAI+Expanse communications are secured using end-to-end encryption by means of Hypertext Transfer Protocol (HTTP) over Transport Layer Security (TLS). The Google Identity Platform[46] is used as a trusted third-party between SensAI and Expanse providing user authentication previous to data anonymisation and storage sync. This mechanism guarantees that only the owner of the data may be able to securely store it and, eventually, retrieve it. (Figure 3.11, p. 41)

[46] https://developers.google.com/identity/

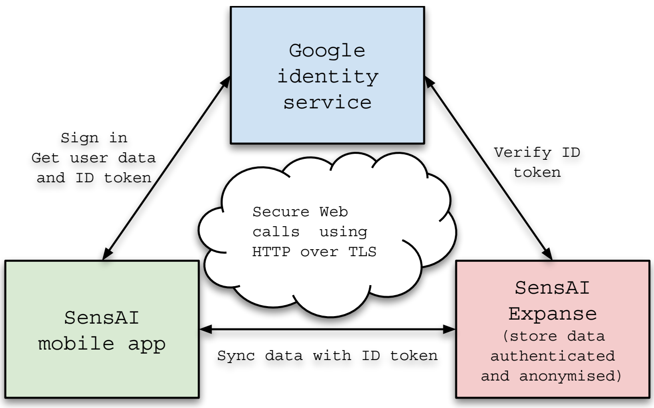

Figure 3.11: SensAI+Expanse secure data flow.

3.3 Expanse

The Expanse is the augmentation spread of the SensAI limited smartphone resources (e.g., memory, processing, and power). It stores data from all SensAI agents anonymously to guarantee that the human's privacy is kept when the data flows to analysis. It provides a repository with historical data, processing algorithms, and services of machine learning towards prediction of emotional valence in context, i.e., augmented memory and cognition towards predictive analytics. Moreover, in a best effort to process all eligible data through available learning algorithms and heuristics towards AutoML (a) it adapts to the diverse human behaviours reflected in the data set; and (b) it uses a Bayesian efficient auto discovery on parameters. Software architecture modules and services are depicted in Figure 3.12, p. 42. Each piece relevant for this system is described[47]

[47] A `test` package for Python modules unit testing is also developed in order to comply with the software engineering best practices: https://gitlab.com/nunoachenriques/sen

in the following sections. Third-party used packages and frameworks are described in appendix C, p. 115.

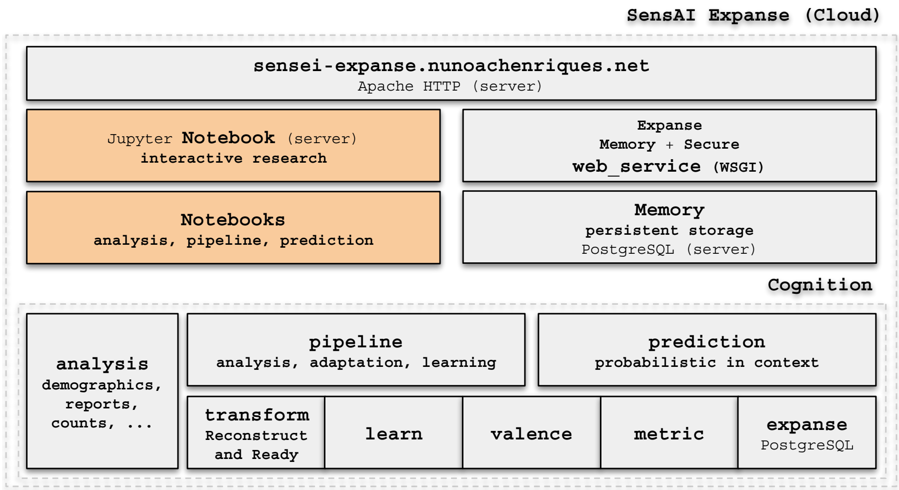

Figure 3.12: The Expanse software architecture modules and services.

3.3.1 `cognition` Package

This package concerns all about processing, analysing, and learning from data towards prediction of emotional valence in context. The extract, transform, and load (ETL) process from relational database data management inspired the `cognition` design and conception.

`analysis` provides a utility class for statistical analysis. It is able to gather some data aggregations from the database, find eligible entities based on some criteria, load the data into some algorithms, obtain insights, and produce charts with information obtained from the processed data.

`pipeline` streamlines the whole learning process towards prediction models. It encompasses heuristics, input, output, and processing decisions. It has the ability to analyse and adapt pipeline processing regarding different human behaviours on moving and sharing emotional valence. This adaptation as in AutoML is restricted by some boundaries in order to obtain viable data for the machine learning algorithms.

`transform` is paramount on data preparation (clean, fix, reconstruct) and transformations such as long to wide, reconstruction of discarded data to save memory, and more on all sensors of interest. Moreover, transforms over reconstructed data to be ready for classification and prediction algorithms. This transformation may require upsampling data within proper boundaries related to collecting parameters. Furthermore, includes a data class (negative, neutral, positive) imbalance (reports count) degree from Zhu et al., 2018. Also, a custom valence class count check and restrict in order to adapt the learning process in cases such as emotional valence reported for only two (or even one) classes. The final eligible entities' data is achieved after these valence class count and imbalance degree processing.

`learn` it is called after all the `transform` actions, it provides the machine learning services towards contextualised emotion valence state models. Includes unsupervised algorithms such as HDBSCAN[48] for clustering location coordinates, accurately dropping outliers (the estimated less relevant places for each entity) when at the `pipeline`, and, finally, mapping the relevant ones in a Discrete Global Grid (DGG) such as MGRS. An additional automatic step is in place for running HDBSCAN clustering algorithm on the different `min_samples` provided in order to find the best `min_cluster_size` parameter. This module also includes some supervised algorithms for multi-class classification such as an Extreme Gradient Boosting by `xgboost` and, additionally, a custom multilayer Perceptron (artificial neural network (ANN)) by `tensorflow` using Keras application programming interface (API). Moreover, model estimation includes hyperparameter auto tuning with cross validation for specific models. It learns with the best model parameters using Bayesian optimisation from the `scikit-optimize` package as a replacement[49] for the less efficient `scikit-learn` `GridSearchCV`.

[48] `https://hdbscan.readthedocs.io/`

[49] `https://scikit-optimize.github.io/notebooks/sklearn-gridsearchcv-replacement.html`

`valence` comprise utility functions such as the one to convert an emotional valence value from a continuous domain ($[-1.0, 1.0]$) to a discrete class classification amongst negative, neutral, and positive ranges extremes -1.0, 0.0, and 1.0 respectively.

`metric` has utilities for measurement related functions such as a score normalisation regarding the $[0.0, 1.0]$ codomain.

`expanse` for database connection, it is the layer to deal with all the input and output including data bulk extraction with and without filtering.

`prediction` provides the required service for the prediction functionality of the emotional valence in context of each human. It requires the availability of the resulting models from the `pipeline` process completion. Includes a custom cache system in order to accelerate consecutive calls to each individual. It encompasses four main modes available to obtain a prediction given an entity identifier and the required features (geographic coordinates and a timestamp). These modes are comprehended by (a) all estimators (with or without probabilities for each valence class); (b) estimator with best score model; (c) specific estimator model; and (d) all estimators average score weighted prediction.

3.3.2 `memory` Package

The package with all the requirements for a long-term data persisting memory and also the Web service layer for input and output. It encompasses (a) a relational database schema to implement the data source model for the long-term memory; and (b) a secure Web service implementation to assure the repository input and output.

`secure` is the provider of authentication, session, anonymisation layers, and the Web service. It establishes an end-to-end encrypted communication channel in order to protect the access and guarantee, simultaneously, data origin, privacy, and anonymity. It first secures a (authenticated and encrypted) connection with SensAI, and then generates an abstract and unique identifier (hash) in order to anonymise data. In a sense, the Web service may be seen as an augmented memory and cognition provider by means of this connection to Expanse.

`source` is the memory structure supporting all the system persistence. It comprise the required `*.sql` files for a relational database management system (RDBMS) schema implementation. This implemented structure provides the long-term memory several components relation with proper attributes. Moreover, the SQL dialect follows the ISO-92 norm with a few necessary PostgreSQL extensions programmed using a procedural language, i.e., PL/pgSQL. The actual RDBMS is a PostgreSQL version 11+ although it has been also successfully tested during several weeks in a 10+ version.

3.4 Integration

The system integration has some compromises and adaptations made along the development process which are relevant to be briefly described. These decisions are organised in the following two subsections: one for memory and learning, and another for the embodiment and interface.

3.4.1 Cognition and Memory

The brains (natural, artificial, and biologically inspired) to some extent may be seen as mostly about cognitive processing and memory use on learning in order to adapt to the embodied environment to put it simplistically. Regarding humans, emotions are paramount on regulating these functions by means of helping with the decision-making and episodic memory's persistence amongst other things[50]. In an apparent paradox, "Forgetting to remember"[51] by means of decay and interference is also necessary to sustain proper processing and storage in order to cope with the information stream that flows through perception sensors. Therefore, the concept of cognitive economy is included encompassing some data exclusions from memory in order to foster savings by means of (a) sustaining the last collected data value without change from each sensor during a well-defined time interval (e.g., 15 minutes); and (b) discarding data below relevant thresholds (e.g., location displacement less than 10 meters from the previous values). The resulting information gaps in (a) are filled *a posteriori* in the Expanse cognition process. This reconstruction is done strictly with the same time interval values used by SensAI thus avoiding any data bias.

[50] Goertzel et al. (2010), Damásio (1994), and Wood et al. (2012)

[51] Altmann and Gray (2002)

The SensAI agent includes a custom heuristic developed in order to adapt the sentiment analysis to the idiosyncrasies of short text messages written by humans such as the social networks' posts. These short messages may vary from plain English and Portuguese[52] to a mixture of abbreviations and emoticons. In order to deal with this rich and sometimes creative written content a best effort approach is described[53] in a workflow depicted in Figure 3.13, p. 46.

[52] The two supported languages.

[53] Appendix C, p. 115 for the source code of `emotion/Sentiment.java`

The Expanse developed custom pipeline to support the SensAI learning process encompasses a myriad of heuristics and other algorithms. Amongst this variety it is included a data class (negative, neutral, positive) imbalance (reports count) degree from Zhu et al. (2018). Also, a custom[54] class count checking and restricting in order to adapt the learning process in cases such as emotional valence reported for only two or (even one) classes. The final eligible entities list is obtained after these valence class count and imbalance degree processing step. Further, the `do_learn_model` of the pipeline process flow (Figure 3.14, p. 48) uses four[55] algorithms elected to be representative of sufficiently diverse and relevant paradigms of the state-of-the-art machine learning:

[54] `https://gitlab.com/nunoachenriques/sensei-expanse/blob/1ae8fc2667d89255d582e81fefc6d3909f45069c/cognition/pipeline.py#L849`

[55] Several other classification algorithms were tested from different code implementations such as Random Forest and Support Vector Machines from `scikit-learn`.

Random provides a baseline, i.e., the other algorithms should perform better than this one. Making use of the `scikit-learn DummyClassifier` (Dummy).

Linear it is a fast and simple estimator with better performance than baseline. Making use of the `scikit-learn LogisticRegression` (Logistic Regression) classifier.

Non-linear paradigm is supported by a decision tree ensemble model by

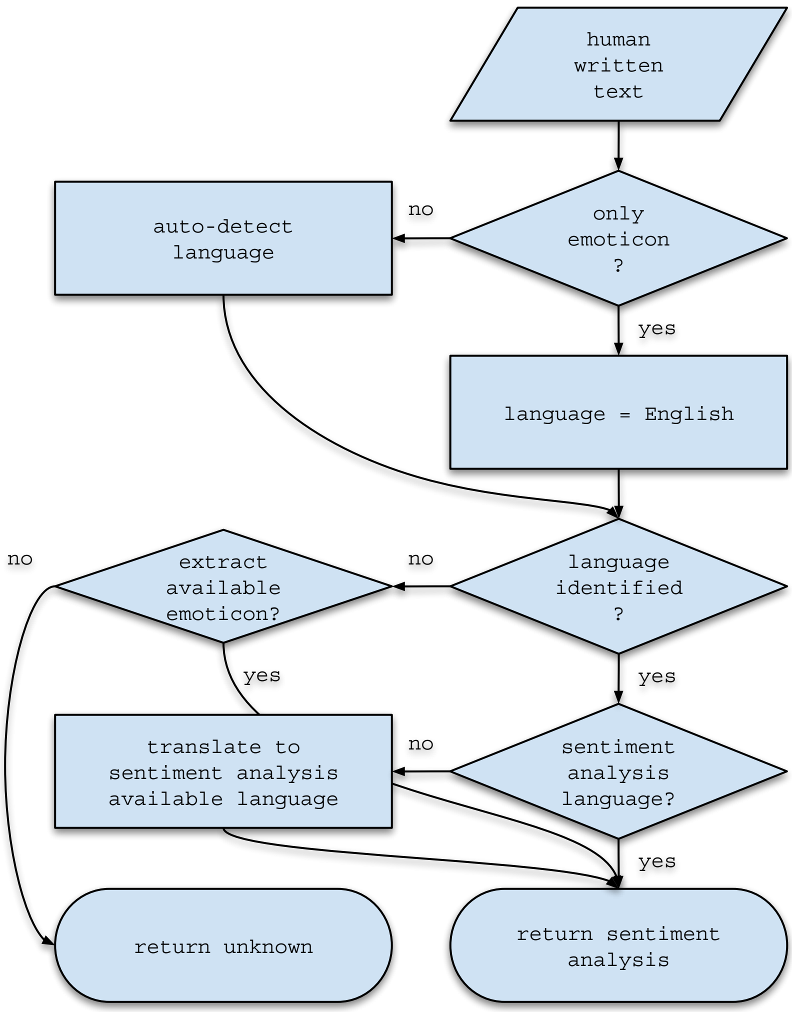

Figure 3.13: Sentiment analysis heuristic regarding short text messages.

making use of a `xgboost XGBClassifier` (Extreme Gradient Boosting). It is a state-of-the-art estimator for machine learning with very good performance and also strongly explainable[56] regarding the predicted result.

[56] For any prediction result each feature importance (e.g., average gain of splits which use the feature) is available.

Connectionist paradigm is required as a deep learning completion to this set. Making use a custom multilayer Perceptron (MLP) as an ANN model supported by `tensorflow KerasClassifier` (TensorFlow Keras MLP). It is a state-of-the-art algorithm for machine learning with very good performance but weakly explainable and with resources high consumption.

Moreover, the MLP sequence is defined as

```
input data
    |
Dense(ReLU) > Dropout >  Dense(ReLU) > Dropout > Dense(Softmax)
                                                      |
                                             output classification
```

where ReLU (rectifier) is used as activation, Dropout as fast regularisation, and Softmax to assure an output of a probability distribution over the predicted classes. It is compiled with the Adam optimiser and the `SparseCategoricalCrossentropy` loss function. A code snippet is shown in Code sample 3.1, p. 47.

Code Sample 3.1: Neural network model

```
model = tf.keras.models.Sequential(
    layers=[
        tf.keras.layers.Dense(units, activation="relu",
                              input_shape=input_shape, name="first"),
        tf.keras.layers.Dropout(dropout_rate, name="first_dropout"),
        tf.keras.layers.Dense(units, activation="relu", name="second"),
        tf.keras.layers.Dropout(dropout_rate, name="second_dropout"),
        tf.keras.layers.Dense(3, activation="softmax", name="output")
    ],
    name="MLPClassifier"
)
model.compile(
    optimizer=tf.keras.optimizers.Adam(amsgrad=True),
    loss=tf.keras.losses.SparseCategoricalCrossentropy(),
    metrics=["accuracy"]
)
return model
```

After all the memory reconstruction from previous savings, and data transformations including upsampling and downsampling, several adaptations to the current data set of each individual are in place, as depicted in the `do_learn_model` box of Figure 3.14, p. 48. The pipeline proceeds, as shown in the item 4 of the same box, with machine learning hyperparameters autotuning by making use of Bayesian optimisation by means of an included (default) Gaussian Process in the cross-validated search provided by `BayesSearchCV`[57], and the steps required to complete

[57] https://github.com/scikit-optimize/scikit-optimize/blob/master/skopt/searchcv.py

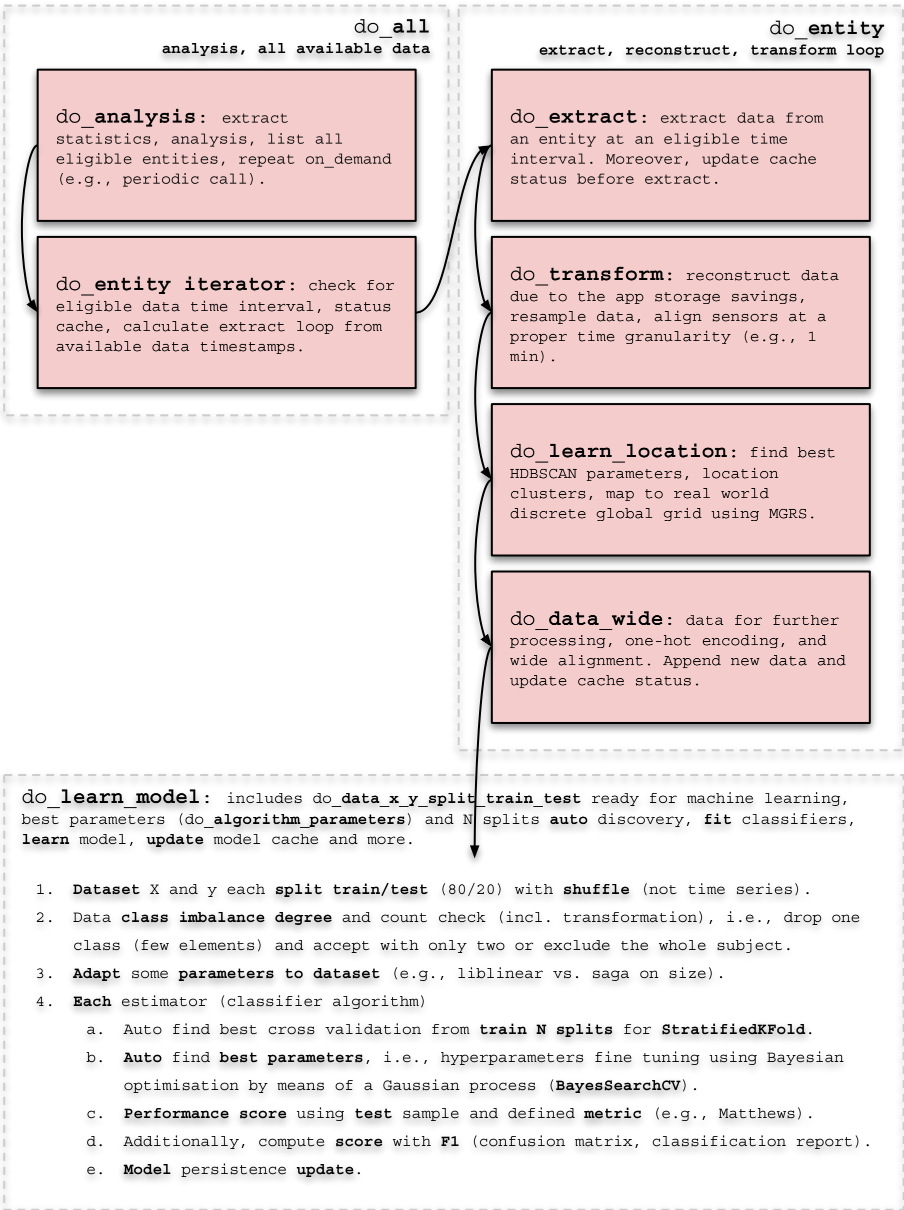

Figure 3.14: Expanse machine learning pipeline data flow.

this automated machine learning along the pipeline, and only then a prediction ability is attained. This capacity of predicting emotional valence in spatial and temporal context is persisted by the memory structure and also able to be updated on demand. The SensAI+Expanse platform is ready to serve prediction[58] requests in a secure and efficient way via Web service after the resulting models for each individual are properly computed and available. The whole system is conceived and designed to ease expanding the set of functionalities and services required to be integrated.

[58] Refer to Subsection 4.2.4, p. 68, for more about the prediction results, models, and Web service API.

3.4.2 Embodiment and Interface

The current smartphones do excel regarding interaction by means of several sensors such as audio input and output (microphone and speakers), multi-colour light-emitting diode (LED), and multi-touch gesture-ready display. Moreover, these mobile devices provide and make use of displacement and positional sensors such as accelerometer and GPS, optical sensor for light intensity measurement, and others. This rich set of available sensors in each mobile device poses and opportunity for the SensAI agent embodiment regarding the perceptional mechanisms required for a proper interaction with the human towards an empathetic companion. A common mobile device of daily use such as an Android™-based smartphone comply with this requirements. Thus leveraging the agent's ability to perceive emotional valence in context supported by the data gathering from the available sensors — smartphone sensing.

The accomplishment of a quiet, almost non-invasive, agent approach[59] requires some restriction on the humans perceived and also the actual data collecting actions. Accordingly, the camera and the microphone[60] sensors are not used despite some interest on this two data rich modalities, i.e., no images are taken and no video or speech is recorded. Moreover, the use of imaging and audio sensors poses accessibility challenges for humans with disabilities such as facial (partial) paralysis, or hearing impaired if spoken natural language is required to be used in the interface. Therefore, there is a focus on display touching, text[61] messages, and notifications, whilst still quietly collecting from the available regular hardware sensors (e.g., accelerometer).

[59] Subsection 3.2.1, p. 33

[60] Audio collection to recognise emotions has already some interesting research on going such as Fayek et al. (2015) and Trigeorgis et al. (2016).

[61] Using VADER (Hutto & Gilbert, 2014) for an efficient as instantaneous and low resources consumption sentiment analysis.

The ground truth of the emotional valence for each human, required to assess the resulting model performance of the machine learning, is of each individual's responsibility. A self-report mechanism is provided by means of three emoticons as described in Subsection 3.2.2, p. 36. Additionally, an always present shortcut for this self-evaluation is available in a persistent system notification. Moreover, this mechanism is robust to interaction bias such as high-frequency repeated button clicks and mistaken valence when promptly corrected by an additional hit on a different emoticon. This includes a simple yet effective heuristic of accounting only for the last hit during a time interval. All these events always occur in context, i.e., the location and a timestamp are always collected along with the event.

The SensAI+Expanse communication on visual interaction must be consistent in all display activities in order to facilitate an easy understanding such as a clear distinction between emotional valence negative, neutral, and positive ranges. This approach requires the interface to always make use of the same respective valence colours on different graphical artefacts regarding distinct negative, neutral, and positive ranges. A self-explanatory example of this consistency was already depicted in Figure 3.9, p. 38, regarding the SensAI application sentiment chart. In order to support this whole system chromatic communication, and foster association between UI artefacts and the discrete (class) values representing the three emotional valence ranges, some colour definitions are persistently stored and shown in Table 3.1, p. 50. Each Red, Green, Blue colour model (RGB) component uses normal range $[0.0, 1.0]$ values.

Table 3.1: Emotional valence colours for consistent chromatic communication

Valence	Value	Colour RGB
Negative	-1.0	`(0.050980392, 0.278431373, 0.631372549)`
Neutral	0.0	`(0.180392157, 0.490196078, 0.196078431)`
Positive	1.0	`(0.960784314, 0.498039216, 0.090196078)`

The mapping between colours and emotions is not trivial and even controversial[62] due to relevant variance across cultures and even from persons with shared social identities. Therefore, the chosen colours for the emotional valence three ranges are loosely based on the Plutchik circumplex model[63] depicted in Figure 3.15, p. 51. The Positive orange colour is a compromise between the yellow from `joy` and the design requirement of proper contrast regarding the interface theme background. Moreover, colour-blindness visual disability may impair some humans from distinguishing amongst the three colours used for emotional valence reference values[64]. In order to solve this issue some basic accessibility principles[65] were used such as always having a text alternative describing the value in each artefact.

This colour-valence mapping is also consistent across SensAI on communicating an emotional valence such as in the previously referenced SensAI application sentiment chart (Figure 3.9, p. 38) and the `Diary`

[62] Kaya and Epps (2004) and Madden et al. (2000) both works reveal the need for further research on colour-emotion perception across cultures.

[63] Plutchik (2001, Figure 6) describes the relations amongst emotion concepts using a wheel of colours.

[64] Table 3.1, p. 50

[65] `https://www.w3.org/WAI/fundamentals/accessibility-principles/#alternatives`

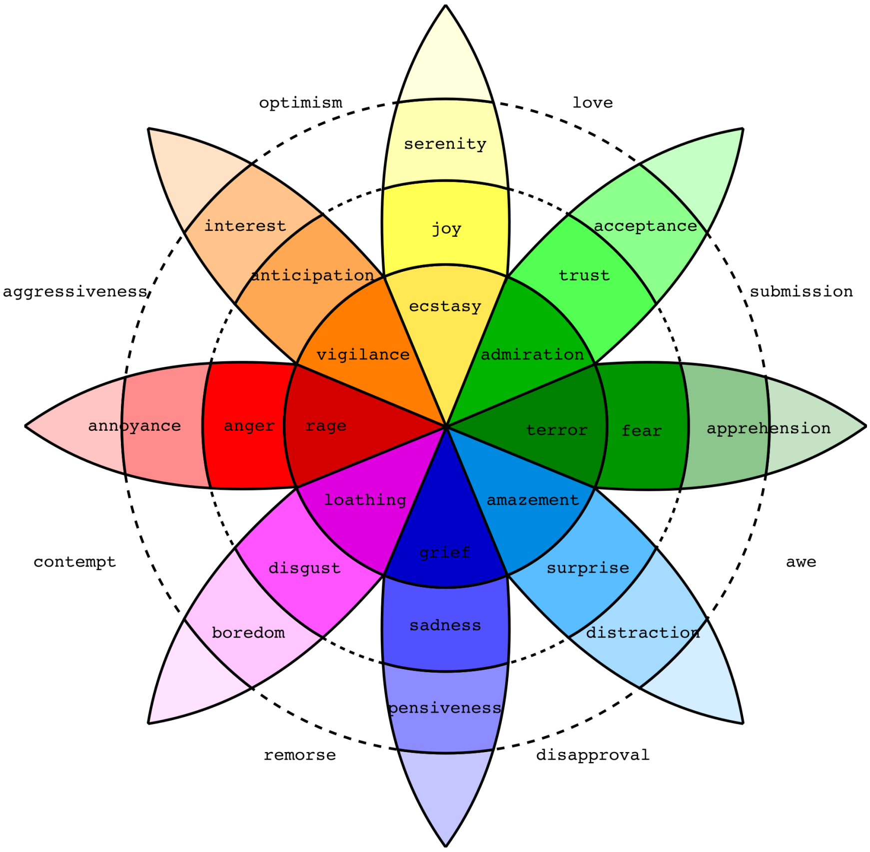

Figure 3.15: Plutchik wheel of emotions colour coded.
Source: `https://commons.wikimedia.org/wiki/File:Plutchik-wheel.svg`

messages sentiment analysis as depicted in Figure 3.16, p. 52. The same colours regarding the three emotional valence ranges are applied in the text message box borders in order to reveal the text sentiment analysis range of negative, neutral, or positive. Moreover, the emoticon artefacts regarding the current average emotional valence are also coloured accordingly.

Furthermore regarding the `Diary`, the SensAI current text sentiment analysis facility is able to process text in both English and Portuguese independently of the device language setting. Concerning the example depicted in Figure 3.16, p. 52, this chat-style interface presents a diary with sentiment-analysed human-written text messages in English yet SensAI reports in Portuguese as per the device language option set by the user.

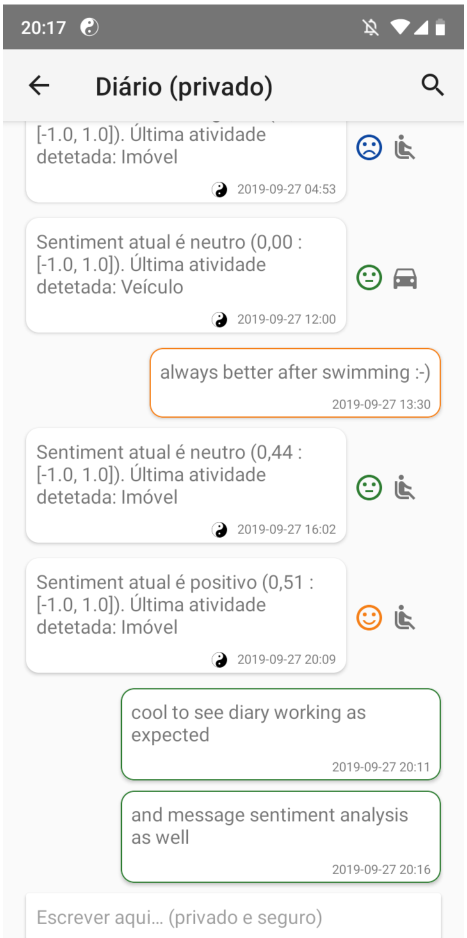

Figure 3.16: Diary with sentiment-analysed human-written text messages in English.

Chapter 4

Study

This chapter describes the main study, i.e., how the results were obtained, the scientific method, and a brief analysis. The first section describes the method regarding the participants, study design, and the procedure using the SensAI+Expanse developed system. It briefly describes participants eligibility, retention, and some gender statistical evidence. Although the population available for the study is small yet an effort was in place to avoid the well-known WEIRD[66] societies bias by using SensAI-based design and supported procedure. The second section shows and describes some relevant details about the results and findings including population demographics, agent learning aspects and, finally, the prediction capability and performance. The final section explain and briefly discuss attained results including some idiosyncratic aspect's analysis.

66 Henrich et al., 2010

4.1 Method

This first section is about the study operational design and procedure. It includes descriptions of how the participants were selected, how diversity was fostered, how the procedure was designed in order to achieve a proper scientific study, attaining results and findings able to be studied and scrutinised.

4.1.1 Participants

The primary objective is to indirectly gather participants of all kinds and creeds, i.e., to avoid the laboratory usual limitations such as sampling only from WEIRD societies known as a frequent bias. This goal is accomplished by the choice[67] to collect data making use of smartphone sensing[68]. This technique poses several advantages over the alternative proprietary devices and mechanisms restricted to a specific laboratory.

67 Appendix B, p. 107, for details about this option.

68 Cornet and Holden, 2018

For instance, by developing a free software application for Android™-based devices, and by publishing that application in the secure and well-know Google Play repository, anyone of the several millions of potential subjects[69] may access and install SensAI from that trusted source. Moreover, preliminary studies assessing the agent capabilities regarding interaction smoothness and proper data collection were in place. This procedure was supported by an efficient and incremental application improvement cycle due to the Google Play developer facilities, including the Store and the Console, of easy deployment of application upgrades. This central repository for applications' distribution additionally manages the technical requirements concerning the device eligibility for SensAI installation. In order for the application to be widely available, the required resources and device features were kept at a minimum: Android™ version 6 or greater[70], a low-end memory and processing capabilities device, featuring a common GPS sensor, and basic access to the Internet. Furthermore, the same platform for publishing SensAI also gives back information about installation country of origin, brand and model of the device, and application serious faults (e.g., stopped responding). All this information concerning each application installation idiosyncrasies and running behaviour is paramount on fixing the failures and deploying upgrades.

[69] Appendix B, p. 107, for more about the potential subjects.

[70] The major version at the time was already 8. The current major version of the Android™ operating system is 10.

Regarding the population sample relevant statistics, this study eligible versus ineligible human entities data is depicted in Figure 4.1, p. 54, showing forty nine eligible participants out of fifty seven. These fifty seven (total) are the ones which downloaded, installed, and tried SensAI application during the study time interval considered for this research[71]. The eligible forty nine are obtained after discarding seven without demographic data (gender and birthdate) and one that never reported emotional valence.

[71] Restricted to the Doctoral Programme in Cognitive Science context.

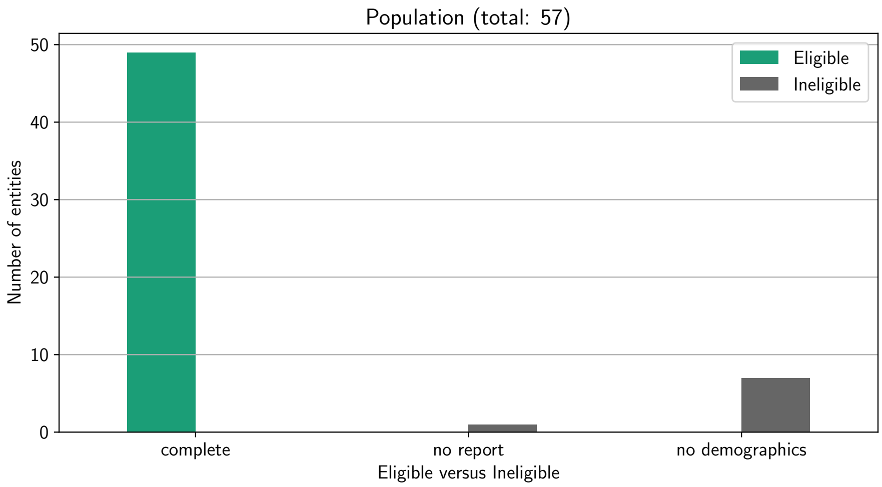

Figure 4.1: Population of participants: eligible versus ineligible.

The retention of entities[72] is depicted in Figure 4.2, p. 55, concerning the ineligible ones with an expected short-term interaction, i.e., individuals that did not participate for long. These participants did not share demographic or emotional valence self-report data. Conversely, Figure 4.3, p. 55, depicts the retention for the eligible entities by gender and during the same ranges: less than a day, less than a week, less than four weeks, and equal or more than four weeks. The majority of eligible participants that did share demographic data also opted for long-term (four weeks or more) participation and data sharing.

[72] Each person is an entity.

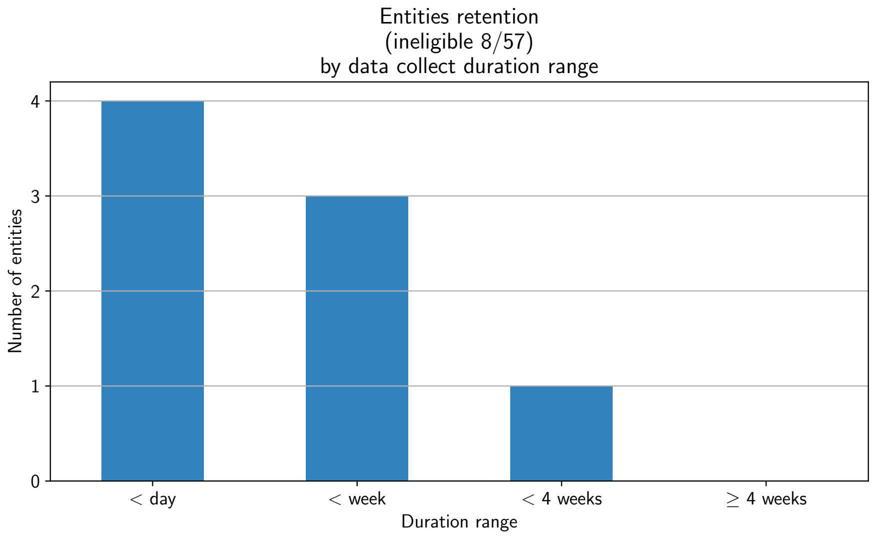

Figure 4.2: Retention of ineligible participants.

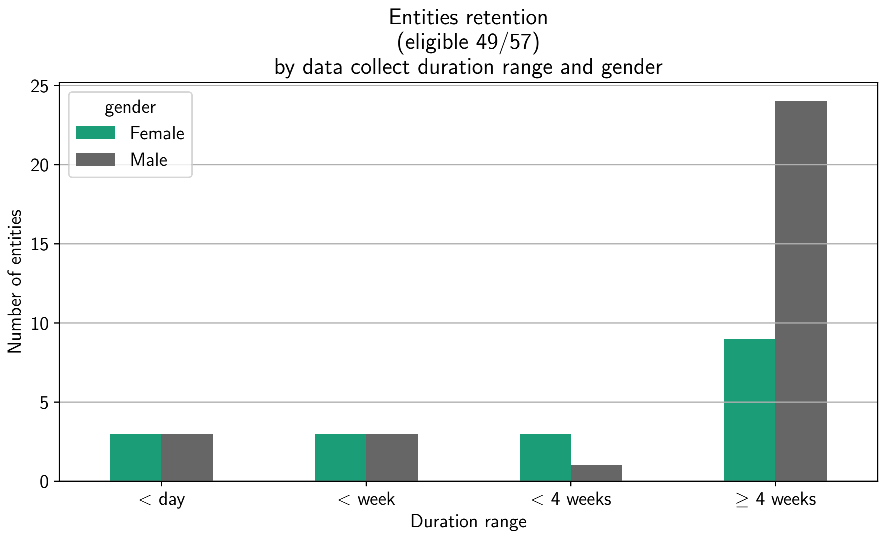

Figure 4.3: Retention of eligible participants.

Furthermore, closely observing the gender distribution along the duration ranges, there is evidence of more (eligible) participants committed to long-term versus short-term interaction. Additionally, male gender seems more polarised than female, i.e., less male participants than female in the medium-term [1 week, 4 weeks) ranges and more than double in the long-term as depicted in Figure 4.3, p. 55. This evidence is restricted to a limited and gender unbalanced population. It should not be taken as statistical significant for wider generalisation.

4.1.2 Design and Procedure

The experimental design and procedure is supported by the developed SensAI+Expanse[73] platform. A mobile device (smartphone) is used for interaction and data acquisition. The mobile device embodied agent SensAI has available two ways of collecting data which are (a) passively using several sensors (smartphone sensing)[74] such as an accelerometer and a GPS; and (b) actively by interacting with the human using system notifications and display buttons providing facilities such as the three emoticons available for emotional valence self-reporting. Moreover, SensAI displays information about the human physical activity aggregated time by type (e.g., running)[75] and also the current emotional valence status, self-reporting statistics, and the current empathy score. Additionally, useful information is communicated making use of: (a) a chronological sentiment chart with self-reports, diary messages, and social network posts emotional valence value provided by instantaneous text sentiment analysis; (b) a private diary[76] for writing messages to self where text is subjected to sentiment analysis, SensAI will report current averaged emotional valence status along with physical activity every three or so hours; and (c) a textual dashboard encompassing several statistics regarding sensors' event count and more.

[73] Chapter 3, p. 29, for a description.

[74] Cornet and Holden, 2018

[75] Figure 3.8, p. 37

[76] Figure 3.10, p. 39

The concept guiding SensAI is one of a very passive almost non-invasive agent. Nevertheless, some mechanisms are in place for seldom communication such as a message asking for the human current emotional status due to a full day without any kind of reporting. This notification occurs depending on a few parameters: time of the last occurrence, empathy below a defined threshold, human agent actively interacting. Additionally, SensAI is robust to human behaviour idiosyncratic aspects while interacting including age and gender neutral on messaging. Some special care was taken regarding these issues to avoid any kind of prejudice or undesired segregation.

> "If some wrong ideas about gender, race, and culture created substantial barriers between humans and even eroded the universality of social cooperation [...] then we should not implement them into AI [...]" (Vallverdù et al., 2019, p. 89)

Furthermore, the interface design followed best practices such as making use of ready-made icons by professional designers, recommended and provided by Google. Even though there is implemented a, already described[77], chromatic strategy for sustaining an easy and coherent communication concerning the emotional valence (negative, neutral, positive) range values, two additional options are available for theming the interface (a) a lighter display with white background as default; and (b) a darker one as an alternative option to the lighter default theme.

[77] Subsection 3.4.2, p. 49

4.2 Results

This section briefly describes the foremost findings and results regarding the study in question. There are several distinct aspects encompassing the results collection which are (a) the specific demographic details regarding gender and age ranges representation; (b) the human emotional valence self-reported behaviour; (c) a machine learning algorithms comparison study; and (d) the attained prediction capability for each participant.

4.2.1 Demographics

The population demographic[78] data is collected by requesting the birthdate and gender at first application use. The total number of entities (humans) participating in this study is small (less than sixty) compared to tens of thousands and even millions which compose contemporary societies. Even smaller (almost fifty) after excluding ineligible[79] entities. For that reason, there are not enough subjects for two-level (age range, gender) segmentation with the devised ten-year age range bins, i.e., some age range classes are under-represented as depicted in Figure 4.4, p. 57.

[78] Demographic data may be further used for more sophisticated research such as to contextualise inferences and learning as shown by Garten et al. (2018).

[79] First eligible assessment requires demographic and emotional valence self-report data.

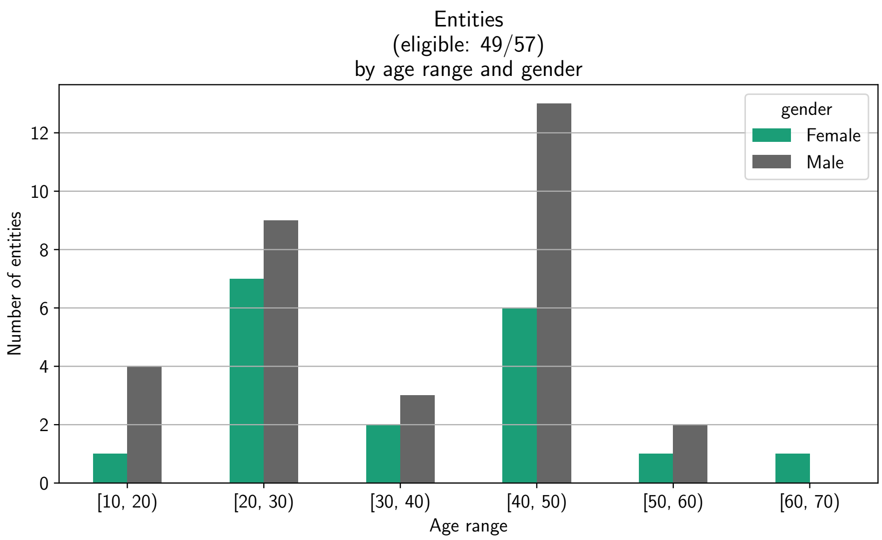

Figure 4.4: Participants demography showing some ten-year age ranges under-representation.

Therefore, an age range dichotomy approach using age median is devised as a proper solution for a viable study. Accordingly, a gender reasonably balanced distribution by age is achieved with this age range dichotomy approach. The result is depicted in Figure 4.5, p. 58, where despite some gender disproportion yet the age range (left to right) entity count relative difference attained is only 4.2% (equivalent to one individual) as further presented in Table 4.1, p. 58.

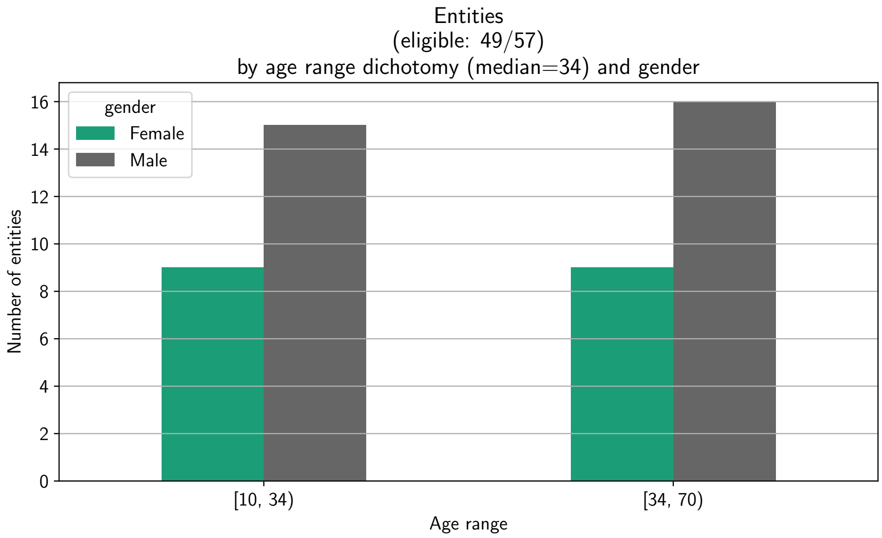

Figure 4.5: Participants demography age range dichotomy solution.

Table 4.1: Eligible population after dichotomy solution

	Age range	
	$[10, 34)$	$[34, 70)$
Female	9	9
Male	15	16
Total	24	25

4.2.2 Behaviour

The studied population general behaviour on self-reporting emotional valence may be observed by means of evidence depicted in both the total count and also the relative proportion of each negative, neutral, and positive valence classes. The generated analysis has a focus on small-time interval cycles such as the weekday and hour of the day. These short-term choices are supported by the statistical requirement of a sufficient number of cycle repetitions regarding the majority of the subjects, i.e., most eligible participants[80] are comprised in the long-term range ($\geq$ 4 weeks) versus the medium and short-term ($<$ 4 weeks). A detailed observation of Figure 4.6, p. 59, (total count) and Figure 4.7, p. 59, (proportion) reveals evidence of almost equal reporting across the weekend including both

[80] Figure 4.3, p. 55

total count and emotional valence discrete class proportion. Moreover, this similarity regarding both Saturday and Sunday is slight different from the other five weekdays whereas the relative proportion of positive class is smaller.

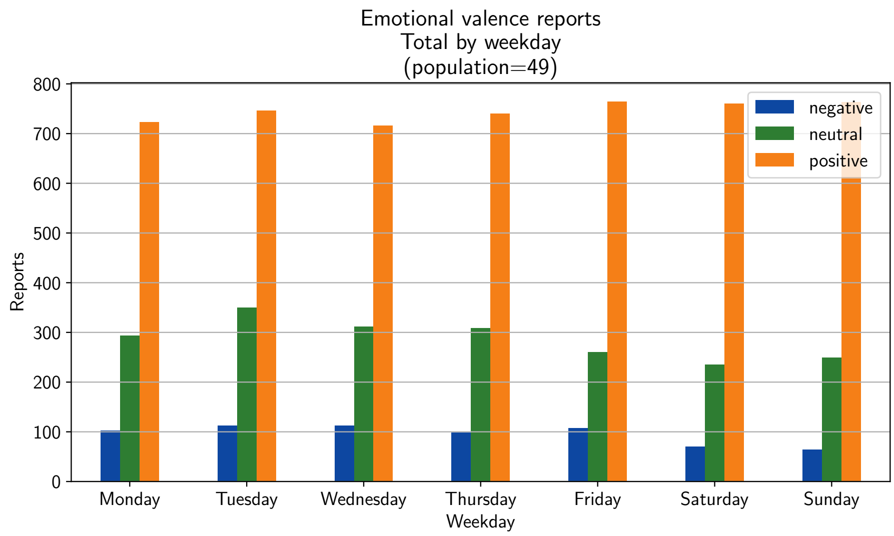

Figure 4.6: Emotional valence reports distributed by weekday.

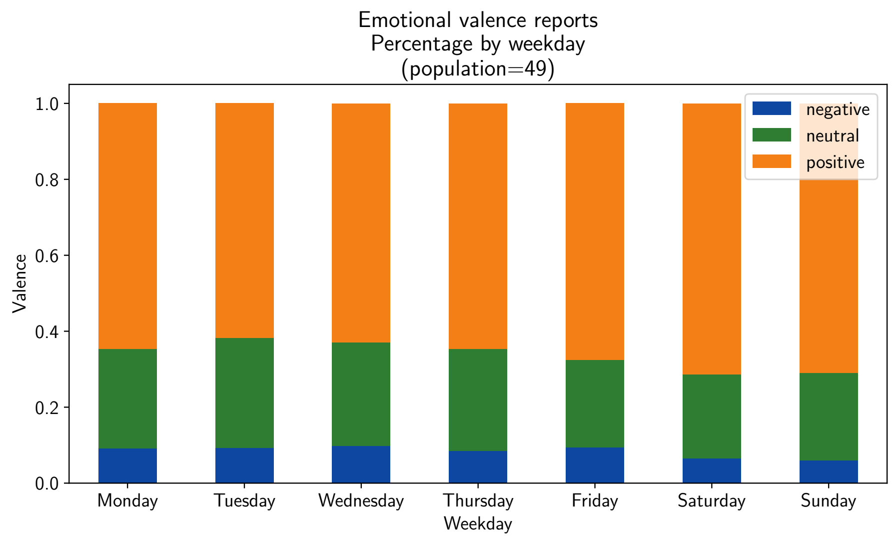

Figure 4.7: Emotional valence reports percentage distributed by weekday.

Regarding the hourly distribution, the emotional valence depicted in Figure 4.8, p. 60, (total count) and Figure 4.9, p. 60, (proportion) shows some interesting differences across the twenty four hours of the day. The self-reports total count at specific day hours (between 3 a.m. and 5 a.m. slots) is too low to be taken into account for a proper study with statistical significance. Accordingly, more data is required and of interest for further research regarding this behaviour peculiarities happening in each individual context and not only on average for the whole population sample. Moreover, the depicted charts despite showing aggregated data by time, yet the source of that data is from each eligible individual and includes even more associated context able to be explored such as location, physical activity, and weather making use of any reliable provider.

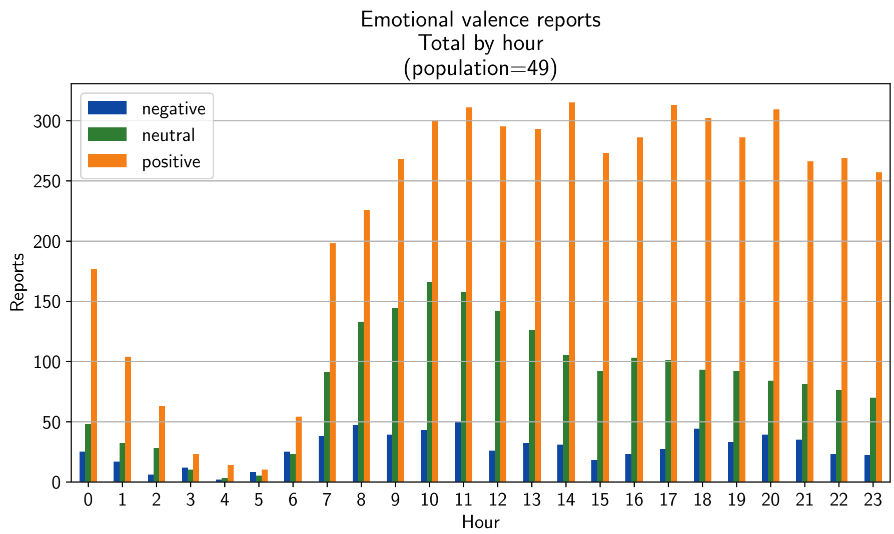

Figure 4.8: Emotional valence reports distributed by hour.

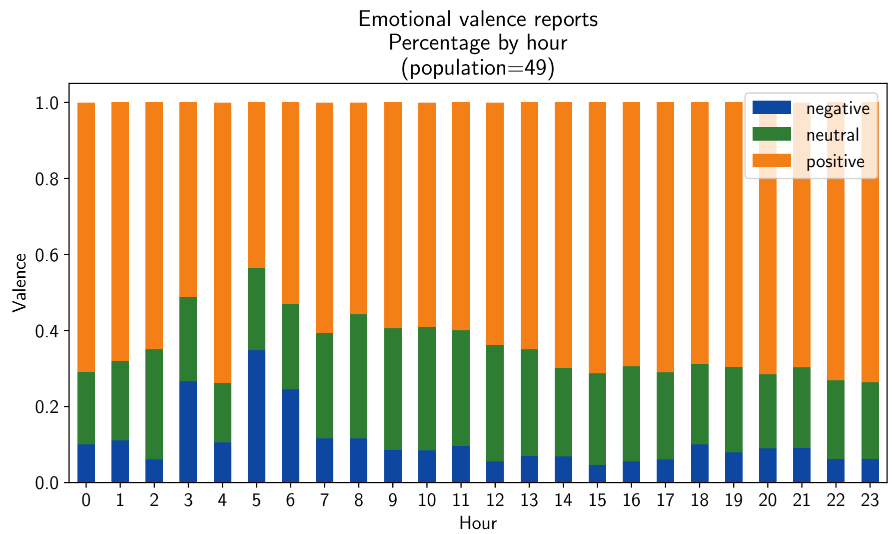

Figure 4.9: Emotional valence reports percentage distributed by hour.

Furthermore, this course of action may be able to reveal identifying idiosyncrasies of each participant. This future findings may contribute to more research regarding each person set of identities. This collection of identifying aspects of each individual span widely from medical and physiological interest (e.g., sleeping period and duration) to cultural behaviours such as a comparison of age ranges and gender concerning emotional valence at specific hours, identifiable locations (e.g., night club), different countries and regions. The science field of social behaviour may be also interested in using SensAI+Expanse platform (or simply the same type of approach) in order to explore comparative studies worldwide in a seamless procedure.

Regarding the specific behaviour by age range and gender, a close observation of the data depicted in Figure 4.10, p. 61, and Figure 4.11, p. 61, reveals evidence of valence proportion differences between some groups of age and gender combination.

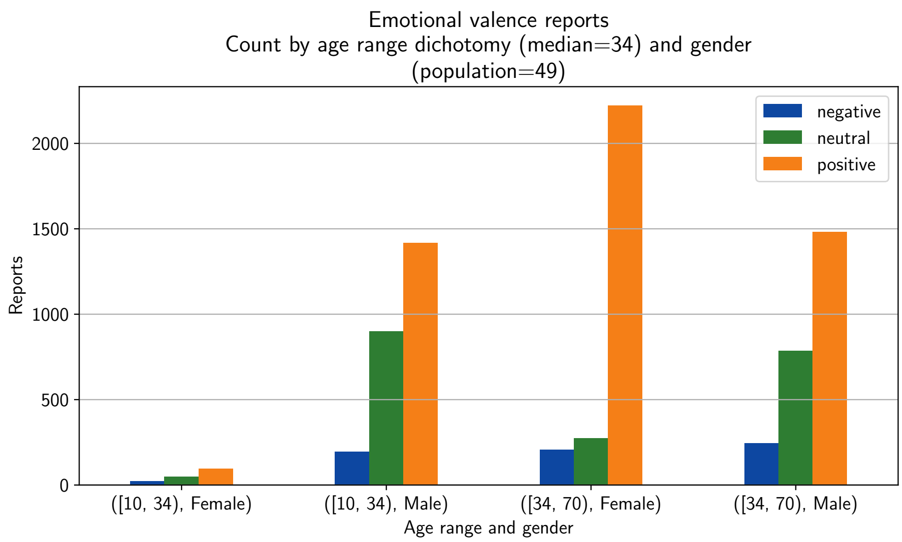

Figure 4.10: Emotional valence reports by age range and gender.

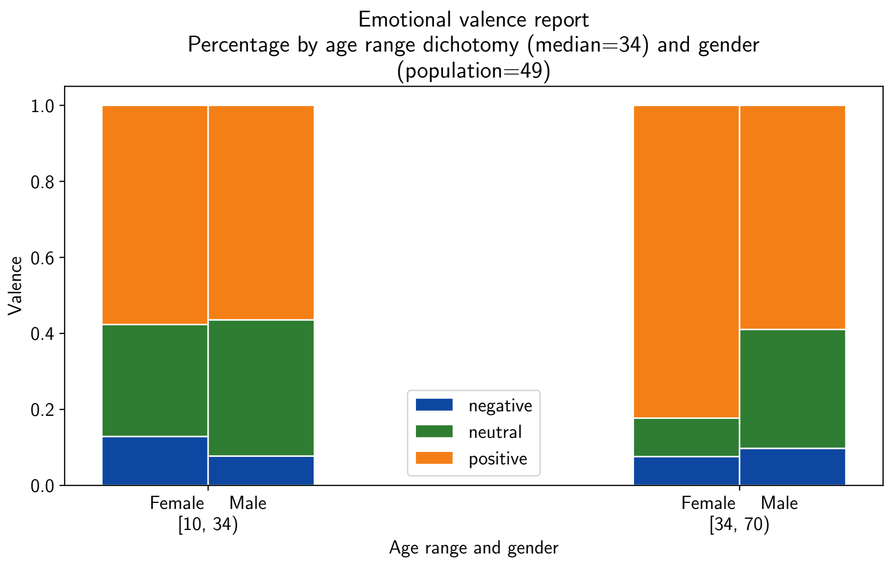

Figure 4.11: Emotional valence reports percentage by age range and gender.

Regarding those apparent differences, previous to the procedure of testing for statistical significance[81] it may be useful to first find if the several groups data is normally distributed[82] in order to become aware of the available options regarding proper statistical significance testing. A test based on the one of D'Agostino and Pearson[83] is used where the null hypothesis (H_0) is: the sample is drawn from a normal distribution. The test results presented in Table 4.2, p. 62, show that H_0 can be rejected for all the two age ranges and the four gender combinations, i.e., none of the groups seem to have data drawn from a normal distribution.

[81] A result has statistical significance when it is very unlikely to have occurred given the null hypothesis.
[82] `https://en.wikipedia.org/wiki/Normal_distribution`
[83] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html`

Table 4.2: Emotional valence reports by age and gender (Figure 4.10, p. 61 and Figure 4.11, p. 61): normal distribution test results

Age and gender	p value	Meaning ($\alpha = 0.05$)
$[10, 34)$	2.163×10^{-53}	H_0 can be rejected ($p < \alpha$)
$[10, 34)$ female	3.138×10^{-5}	H_0 can be rejected ($p < \alpha$)
$[10, 34)$ male	3.877×10^{-49}	H_0 can be rejected ($p < \alpha$)
$[34, 70)$	2.012×10^{-250}	H_0 can be rejected ($p < \alpha$)
$[34, 70)$ female	3.578×10^{-240}	H_0 can be rejected ($p < \alpha$)
$[34, 70)$ male	3.785×10^{-62}	H_0 can be rejected ($p < \alpha$)

Therefore, the Mann-Whitney U (non-parametric) test[84] is a robust option to assess statistical significance between two not normally distributed samples. The null hypothesis (H_0) enunciates that: two sets of measurements are drawn from the same distribution. The results presented in Table 4.3, p. 62, show that H_0 can be rejected for all but two tests, i.e., there is evidence of significant differences on three comparisons of two groups each as described next:

[84] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html`

$[10, 34)$ **vs.** $[34, 70)$ show differences between two age ranges mainly driven by the female gender.

$[10, 34)$ female **vs.** $[34, 70)$ female evidence a difference in behaviour where the older age group reported significantly less negative and neutral emotional valence by contrast to an order of magnitude more positive reports.

$[34, 70)$ female **vs.** $[34, 70)$ male evidence the overwhelming positive reports by the female versus male gender in this age range group.

Table 4.3: Age and gender groups comparison (Figure 4.10, p. 61 and Figure 4.11, p. 61): Mann-Whitney U test results

Age and gender	p value	Meaning ($\alpha = 0.05$)
$[10, 34)$ vs. $[34, 70)$	1.161×10^{-30}	H_0 can be rejected ($p < \alpha$)
$[10, 34)$ female vs. $[34, 70)$ female	5.539×10^{-14}	H_0 can be rejected ($p < \alpha$)
$[10, 34)$ male vs. $[34, 70)$ male	1.561×10^{-1}	H_0 cannot be rejected ($p > \alpha$)
$[10, 34)$ female vs. $[10, 34)$ male	3.938×10^{-1}	H_0 cannot be rejected ($p > \alpha$)
$[34, 70)$ female vs. $[34, 70)$ male	7.027×10^{-67}	H_0 can be rejected ($p < \alpha$)

After this brief and general behaviour analysis of the emotional valence self-reports by the population sample eligible for this study, further research is envisioned to be drawn from this approach. Specifically, enriching the demographics-based comparison with the already collected data (e.g., location) or even adding more parameters of interest.

4.2.3 Learning

The machine learning pipeline data flow is already described in Chapter 3, Section 3.3, p. 41, and illustrated in Figure 3.14, p. 48. The four estimators (algorithms) applied in the learning process towards prediction were selected as state-of-the-art representatives of the four paradigms deemed more relevant. These encompass the random, linear, non-linear, and connectionist distinct approaches to machine learning. The four estimators also provide the opportunity for a performance comparison study, and are already explained in Subsection 3.4.1, p. 44.

The performance measurement mechanisms (available metrics for classification) are part of the machine learning process and required for model evaluation, i.e., to be able to quantify the quality of the attained predictions. Accordingly, several classification metrics[85] were tested for the estimators performance score assessment. After some preliminary testing and repeated running with several data sets (from live SensAI installs) through the pipeline, two were selected for comparison: F1 score[86] and MCC[87]. The selection criteria has three requirements, supported by both metrics, which are (a) multi-class ($n = 3$) capability; (b) appropriate for the task and data context; and (c) stable with very good software support[88] for machine learning integration.

Regarding the findings, although the MCC is considered to have considerable advantages[89] over F1 score yet such claim was not evidenced in this research with the current eligible population[90] and data sets. On the contrary, MCC scores seemed too high (an apparent overfitting) for all four estimators used to run SensAI learning process as depicted in Figure 4.12, p. 63 and Figure 4.13, p. 64.

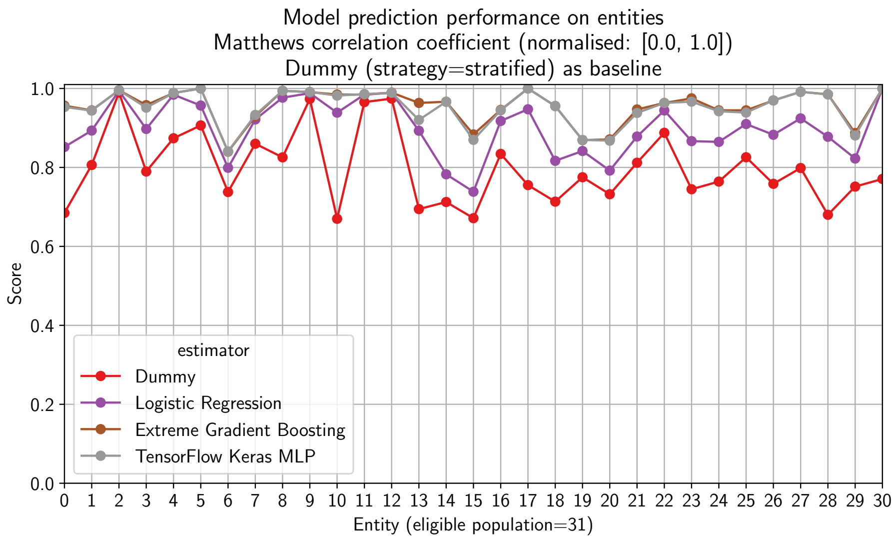

Figure 4.12: Model prediction performance by entity using MCC.

Moreover, this apparent overfitting on all four estimators prediction

85 `https://scikit-learn.org/stable/modules/model_evaluation.html#classification-metrics`

86 `https://en.wikipedia.org/wiki/F1_score`

87 `https://en.wikipedia.org/wiki/Matthews_correlation_coefficient`

88 `https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html` and `https://scikit-learn.org/stable/modules/generated/sklearn.metrics.matthews_corrcoef.html`

89 Chicco, 2017

90 Population is further restricted and reduced due to a class count and balance strict requirements for the machine learning process.

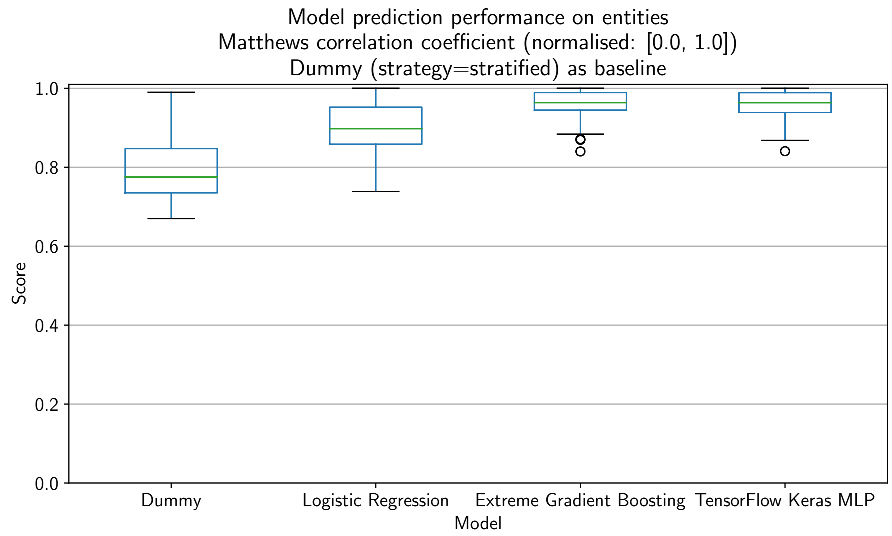

Figure 4.13: Model prediction performance statistics using MCC.

performance (making use of the MCC metric) raises doubts because (a) the two state-of-the-art estimators have several values with a perfect 1.0 score; (b) the linear logistic regression has also several 1.0 prediction scores; and (c) the baseline (Dummy) with a median value near 0.8.

Conversely, the F1 score measurement for all four estimators present a more reasonable and expected result as depicted in Figure 4.14, p. 64, by entity, and Figure 4.15, p. 65, for each estimator statistics.

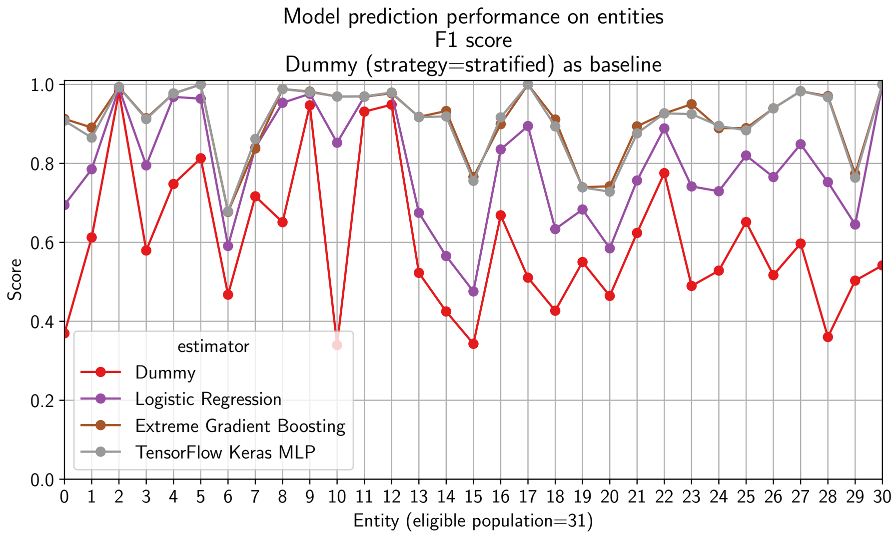

Figure 4.14: Model prediction performance by entity using F1 score.

One relevant evidence is the baseline (Dummy) estimator median value near the expected 0.5 mark. Another one is the median value of the linear logistic regression just below the 0.8 and the performance results of the population are more sparse than the condensed values in the MCC case[91]. Moreover, the two metrics were applied with similar experimental conditions, i.e., the same pipeline parameters and algorithms, equal

[91] Figure 4.13, p. 64

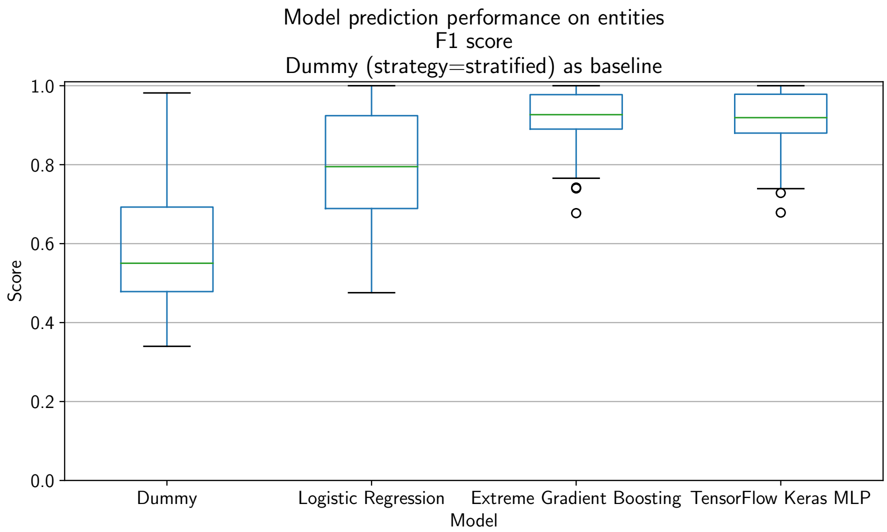

Figure 4.15: Model prediction performance statistics using F1 score.

population and data set.

In order to be fully aware of the available options regarding proper statistical significance testing, it may be useful to first find if both data sets are normally distributed. The Dummy estimator (amongst the four selected for the study) is elected for the measurement results comparison: F1 versus MCC. As before (regarding emotional valence self-reported behaviour in Subsection 4.2.2, p. 58) a test based on the one of D'Agostino and Pearson[92] is used where the null hypothesis (H_0) is enunciated as: the sample is drawn from a normal distribution. The test results presented in Table 4.4, p. 65, show that H_0 cannot be rejected for the two metrics, i.e., both F1 and MCC seem to have data drawn from a normal distribution.

[92] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html`

Table 4.4: Model prediction performance on entities (Figure 4.16, p. 66): normal distribution test results

Metric	p value	Meaning ($\alpha = 0.05$)
F1	2.738×10^{-1}	H_0 cannot be rejected ($p > \alpha$)
MCC	2.745×10^{-1}	H_0 cannot be rejected ($p > \alpha$)

Therefore, the Mann-Whitney U (non-parametric) test[93] is a robust option to assess statistical significance between these two samples. The null hypothesis (H_0) is enunciated as: two sets of measurements are drawn from the same distribution. The test results presented in Table 4.5, p. 66, show that H_0 can be rejected for F1 versus MCC test, i.e., there is evidence of significant differences.

[93] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html`

Furthermore, although the Mann-Whitney U test results in a clear "differ significantly" with a p value of $4.317e-06$ which is four orders of magni-

Table 4.5: F1 vs. MCC (Figure 4.16, p. 66): Mann-Whitney U test results

Metrics	p value	Meaning ($\alpha = 0.05$)
F1 vs. MCC	4.317×10^{-6}	H_0 can be rejected ($p < \alpha$)

tude lesser than the α of 0.05 yet further inspection is taken. Accordingly, relevant cases were selected for a confusion matrix[94] inspection on F1 versus MCC comparison. This procedure revealed that F1 score seems more accurate in all cases thus should be the selected metric for further computation and prediction. Additionally, making use of the classification report[95] no pattern was identified correlating the inspected cases, i.e., all seem to have distinct data shapes. Moreover, there is a majority of 26 out of 31 cases (83%) where F1 versus MCC relative difference is greater or equal to $10pp$[96]. This evidence is depicted in Figure 4.16, p. 66, and Table 4.6, p. 67.

[94] `https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html`

[95] `https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html`

[96] percentage point (pp)

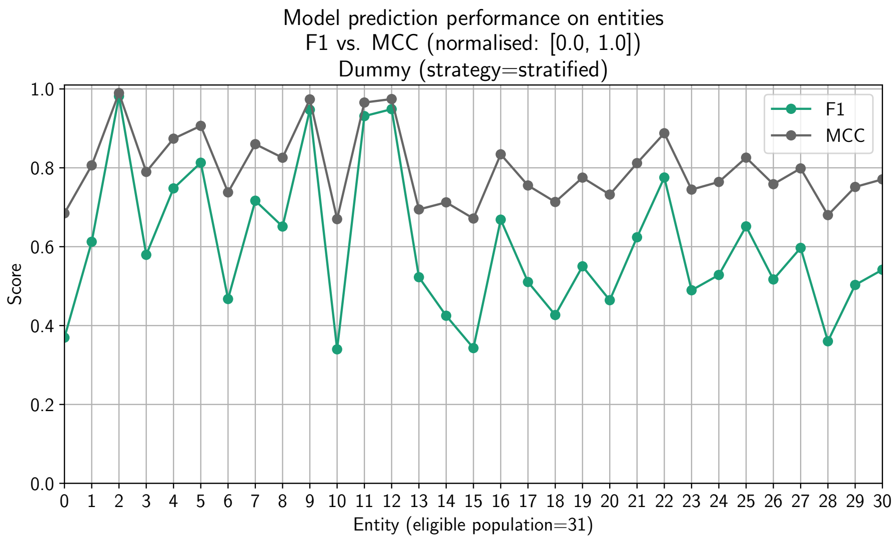

Figure 4.16: F1 vs. MCC model prediction performance comparison by entity using Dummy estimator.

Further, two examples are selected with the criteria of comprising the largest difference in percentage points, i.e., two entities with $33pp$ on F1 versus MCC score result differences. These examples (entities 10 and 15) are depicted in tables 4.7, p. 67, and 4.8, p. 68, of inspected confusion matrices and further support the argument that F1 score is a more appropriate metric than MCC for this research study. These findings also illustrate, as expected, the low prediction capability of a model using the Dummy estimator (as a baseline) which resembles by chance performance. As presented in both tables, the diagonal of correct predictions (normalised: $[0.0, 1.0]$) ranges between 0.31 and 0.39 for each one of the three classes. Accordingly, the prediction performance (Dummy estimator) for both entities is (rounded) $F1 = 0.34$ (Table 4.6, p. 67).

Table 4.6: Dummy F1 vs. MCC cases with relevant (>= 10*pp*) score differences

Entity	F1	MCC	F1 vs. MCC (rounded *pp*)
0	0.37	0.68	32
1	0.61	0.81	19
3	0.58	0.79	21
4	0.75	0.87	13
6	0.47	0.74	27
7	0.72	0.86	14
8	0.65	0.83	17
10	0.34	0.67	33
13	0.52	0.69	17
14	0.42	0.71	29
15	0.34	0.67	33
16	0.67	0.83	17
17	0.51	0.76	24
18	0.43	0.71	29
19	0.55	0.78	22
20	0.46	0.73	27
21	0.62	0.81	19
22	0.78	0.89	11
23	0.49	0.74	26
24	0.53	0.76	24
25	0.65	0.83	17
26	0.52	0.76	24
27	0.60	0.80	20
28	0.36	0.68	32
29	0.50	0.75	25
30	0.54	0.77	23

Table 4.7: Confusion matrix for entity 10: Dummy $F1 = 0.34$ (Table 4.6, p. 67)

		Predicted			
		Negative	Neutral	Positive	Support
True	Negative	0.33	0.31	0.36	75
	Neutral	0.27	0.33	0.39	157
	Positive	0.25	0.39	0.35	127

Table 4.8: Confusion matrix for entity 15: Dummy $F1 = 0.34$ (Table 4.6, p. 67)

True \ Predicted	Negative	Neutral	Positive	Support
Negative	0.31	0.4	0.29	746
Neutral	0.31	0.39	0.3	899
Positive	0.3	0.38	0.32	676

4.2.4 Prediction

The prediction performance outcome after the machine learning process is summarised in Figure 4.17, p. 68. The model resulting from making use of the Extreme Gradient Boosting estimator and the F1 score metric is selected as the best option amongst a total of four models available[97] (including Dummy[98] as baseline). The option for this model is also supported by an interest towards (a) explainable artificial intelligence (XAI)[99] predictions; and (b) efficient energy use[100] besides the overall score attained.

[97] The four distinct estimators are already explained in Subsection 3.4.1, p. 44.

[98] `https://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyClassifier.html`

[99] `https://en.wikipedia.org/w/index.php?title=Explainable_artificial_intelligence&oldid=941615766`

[100] `https://en.wikipedia.org/w/index.php?title=Efficient_energy_use&oldid=937914269`

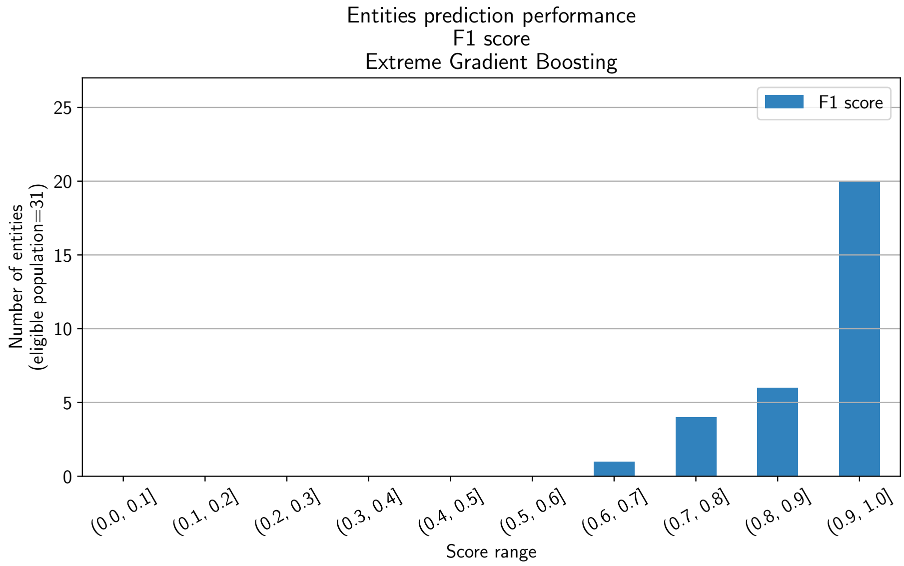

Figure 4.17: Model prediction performance aggregated by F1 score range using Extreme Gradient Boosting.

Regarding the first interest about XAI, the Extreme Gradient Boosting estimator already includes feature importance[101] scores for each entity model thus proper contextualised XAI (e.g., a specific location feature with the highest score amongst the features list). Concerning an efficient energy use, evidence is presented in Table 4.9, p. 69, where Extreme Gradient Boosting emerged as the best performer although marginally to the second in rank yet for less than a tenth of the processing duration using the same computing resources. Therefore, the solution making use

[101] The calculation of the feature importance may be computed by several distinct methods. For instance, `gain` — average gain of splits which use the feature: `https://xgboost.readthedocs.io/en/latest/python/python_api.html#xgboost.plot_importance`

of the Extreme Gradient Boosting estimator is simultaneously, on average, the best achieved model for prediction with efficient energy use, and also easy explainable by feature importance inspection.

Table 4.9: All four estimators average (population) F1 score and total duration

Model	F1 score	Duration
Dummy	0.600	00:00:46
Logistic Regression	0.795	01:46:22
TensorFlow Keras MLP	0.907	24:35:11
Extreme Gradient Boosting	**0.910**	**01:54:34**

The Expanse Cloud API comprise several modes for accessing its prediction capability. Some of those modes may be further combined with individual classification class probability, i.e., instead of receiving only the most probable classification one may request each negative, neutral, and positive class probability for a specific context. The currently available modes are:

all delivers full emotional valence predictions in context given an entity identifier and required features (location and moment). It comprises a list of all available estimator's prediction (`null` if unavailable) with probabilities (optional) for each class. Also, feature importance scores for the cases where the estimators support it. Moreover, several distinct requests may be sorted in a batch request and the answer will deliver results by the same order.

best delivers the emotional valence predictions in context as **all** but restricted to the result returned by the estimator (model) with the best score (e.g., Extreme Gradient Boosting in entity 23 case as depicted in Figure 4.14, p. 64).

estimator delivers the emotional valence predictions in context as **all** but restricted to a given (requested and available) estimator.

voting delivers the emotional valence estimators weighted average score prediction in context given an entity identifier and the required features (location and moment). The result will not include probabilities but only a single class or a list of values if different contexts were comprised in a batch request.

Regarding the prediction service requests, the Expanse Web service is a secure, end-to-end encrypted, restricted to the identified SensAI agents, and only serves content from and for the authenticated requesting entity. This prediction service (request name: `get_prediction`) accepts (besides the whole identification and authentication process) one of the four available modes, a context list with one (or more) location and moment, (optionally) an estimator name, and an indication if class probability is to be also delivered. A request example is presented below in Code description 4.1, p. 70.

Code Sample 4.1: Expanse prediction request example in JSON format

```
{
    "oauth_id_token": "abcdef0123456789..."
    "expanse_request": {
        "get_prediction": {
            "mode": "all",
            "context": [
                {
                    "moment": "2020-02-02T00:00:42.000+00:00",
                    "latitude": 38.725267,
                    "longitude": -9.150019
                }
            ],
            "estimator": "xgboost.sklearn.XGBClassifier",
            "probabilities": true
        }
    }
}
```

The geographic coordinates given in the context requests are translated to the real world DGG in use: MGRS[102]. The precision option is 1000 m square side due to (a) reasonable cell size for the same place sentiment; and (b) a feasible number of cells able to be used as features in the learning process. Therefore, the prediction results are valid for this coarse location where the geographic coordinates are bounded, i.e., emotional valence value is predicted for the whole cell[103] at a given moment. The Table 4.10, p. 70, presents evidence of how complex may be to give an exact prediction in the context of an entity.

[102] `https://en.wikipedia.org/w/index.php?title=Military_Grid_Reference_System&oldid=921197599`

[103] A SensAI+Expanse custom cell reference such as `mgrs_1000_33LUL0624` where `mgrs_1000` is a prefix indicating DGG system (MGRS) and precision (1000 m), `33LUL0624` is the MGRS grid cell reference.

Table 4.10: Extreme Gradient Boosting (F1) prediction for entity 24: seven locations (highest overall influence) by MGRS reference, equal moment 2049-10-05 08:00 (local time)

	Class probability		
Location	Negative	Neutral	Positive
33LUL0624	0.730	0.205	0.064
33LUL0422	0.117	0.569	0.314
33LUL0423	0.001	0.002	0.997
33LUL0514	0.143	0.200	0.657
33LUL2512	0.606	0.026	0.369
33LUL0522	0.421	0.207	0.371
33LUL2509	0.396	0.005	0.599

The same table presents some locations (e.g., 33LUL0423) with a probability polarised to one emotional valence class where a clear prediction is devised, i.e., positive emotional valence with a probability of 0.997. Conversely, other locations such as 33LUL0522 have a divided probability between negative, neutral, and positive, thus is more complex to devise and select a proper emotional valence state in this discrete scale (negative, neutral, positive). Moreover, the context[104] for which SensAI+Expanse is unable to give an adequate prediction it will return `NaN` for all emotional valence classes. Furthermore, an *ad hoc* visual tool (Figure 4.18, p. 71) was developed[105] in order to easily explore prediction results in context of geographical location and moment.

[104] The context currently encompass the location and the moment. It may be further improved to include other features such as human activity (e.g., running).

[105] Using the SensAI Expanse facilities and the `bokeh` Python package in a Jupyter notebook.

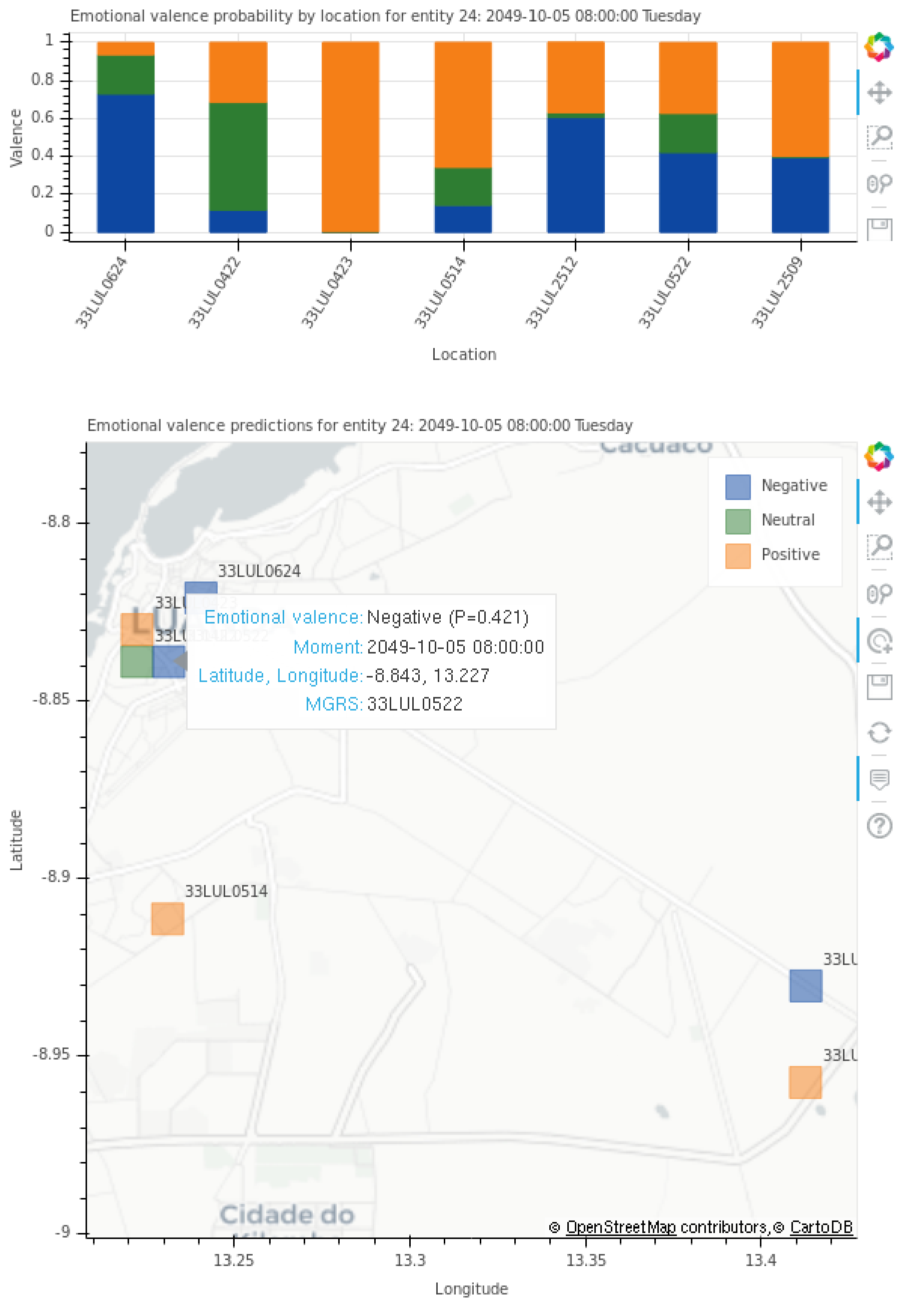

Figure 4.18: Tool for predictions exploration (snapshot of a live run).

Moreover, this tool serve as a functional prototype of a future feature to be included in the SensAI agent application. The visualisation of the emotional valence value in context is accompanied by additional information such as full spatial and temporal context, prediction class, and also the probability of the result. This graphical interactive tool includes the options to change the zoom level and select a geographical region of interest, to pan and navigate smoothly to wanted areas, and always presenting the emotional valence layer cells integrated with tiles from OpenStreetMap. All this information may be further integrated and used in advanced application features such as future SensAI proactive events. For instance, SensAI may take the initiative to interact more empathetically at a given moment, and smooth a context where a negative emotional valence is predicted with a high probability.

4.3 Analysis

The results presented (Section 4.2, p. 57) show evidence, restricted to the population and the data samples in this research, of differences in behaviour amongst some combinations of age ranges versus gender. Nevertheless, SensAI was able to adapt, and learn to predict emotional valence states with high performance (Figure 4.15, p. 65 and Figure 4.17, p. 68). Therefore, the findings and achieved results confirm both hypotheses[106] 1 and 2. Moreover, there is evidence amongst the population individuals of idiosyncratic aspects such as distinct features overall influence.

[106] Refer to page 4 for hypotheses description.

Almost all the humans subjected to the presented study in this research revealed sensitivity to the moment (temporal) dimension. Specifically, the weekday[107] is the most influential, 64.5% of the cases, for emotional valence predictions, followed by hour[108] for 25.8%, and location[109] concerning only 9.7%. These percentages ranking is obtained using SHapley Additive exPlanations (SHAP) on all entities prediction model regarding the feature ranked first on each entity. Therefore, for almost all the participants the moment, specifically the weekday and hour, is decisive to obtain a prediction whereas for a few entities some locations strongly compete for emotional valence state prediction. Thus Hypothesis 3, p. 4, is also confirmed, i.e., SensAI is able to reveal idiosyncratic factors on human's emotional valence changes. Moreover, adding features to the learning process may reveal other distinct factors not yet discovered such as the influence of physical activity (e.g., riding a bike).

[107] Feature: `moment_dow`

[108] Feature: `moment_hour`

[109] Feature: `mgrs_...`

The following figures (4.19, 4.20, 4.21) show evidence of different configurations on location versus time factors for three entities emotional valence prediction model. For instance, Figure 4.19, p. 73, regarding entity 5 feature overall influence, shows a location (`mgrs_1000_29SMC8689`) ranking first but strongly challenged (both mean(|SHAP value|) close to the same impact value) by weekday (`moment_dow`) ranking second in overall influence but first of the time features by a relevant difference (observable by comparison with the remaining features). A similar case of competing features, regarding both spatial and temporal dimensions, may be observed in 4.20, entity 12. The differences in this case are that the ranking positions swapped, weekday ranking first and location ranking second, and there is a third feature, day quarter[110], ranking third just below and very close to the location value.

[110] The 24-hour day segmented in four parts each comprising six one-hour slots: 0–5; 6–11; 12–17; 18–23.

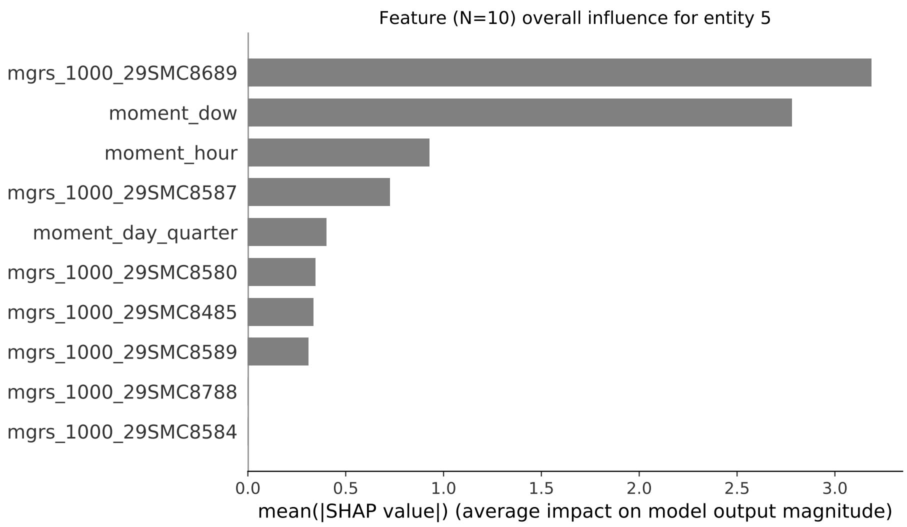

Figure 4.19: Entity 5: feature overall influence.

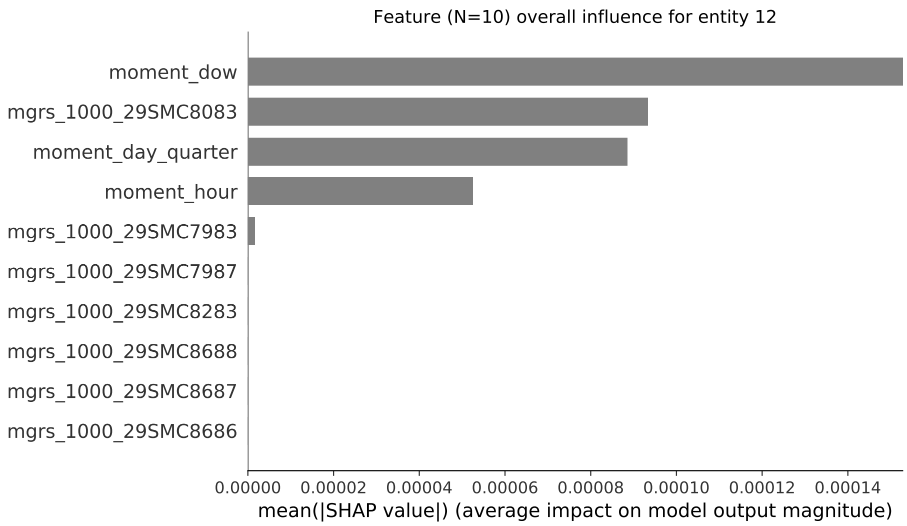

Figure 4.20: Entity 12: feature overall influence.

Regarding the third figure (4.21) depicting entity 24 feature overall influence, a close inspection reveals evidence of three temporal features strongly influencing the emotional valence prediction. The features ranking first and second are weekday and hour respectively, the third is the day quarter. Moreover, in this case it should be noticed that the three time features have relevant differences regarding the average impact on model output magnitude. Further, the location feature ranking fourth overall is strongly competing with a value close to the one of day quarter.

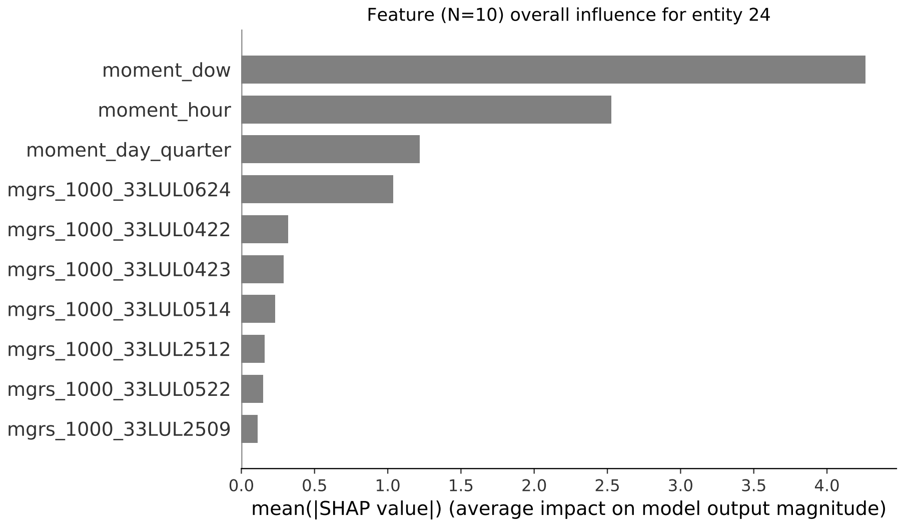

Figure 4.21: Entity 24: feature overall influence.

Furthermore, making use of the tool presented in the previous Subsection 4.2.4, p. 68, to explore deeper in the emotional valence states relation with the context regarding entity 24 (Figure 4.21, p. 74) there is evidence of relevant emotional valence change between two weekdays, Sunday and Tuesday, concerning both the same location cells and hour of the day, as depicted in figures 4.22 and 4.23 (probability charts), and also 4.24 and 4.25 (emotional valence cells over maps). All these charts and maps are generated and exported on interaction with the tool[111].

[111] Subsection 4.2.4, p. 68

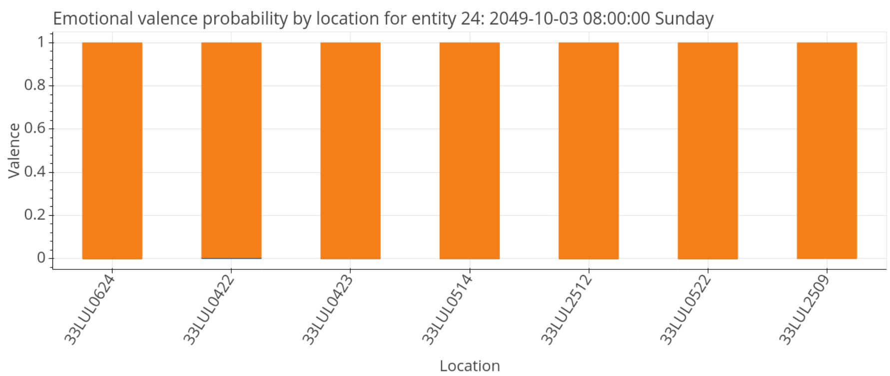

Figure 4.22: Emotional valence probabilities by location: entity 24, 2049-10-03 08:00 Sunday.

Additional evidence of competing influence between time and location dimensions is presented in Figure 4.26, p. 76, where the strong impact of a specific location causes a similar prediction for some weekdays despite the weekday factor higher influence (Figure 4.21, p. 74). Monday and Tuesday emotional valence predictions seem strongly influenced by the weekday factor whereas remaining weekdays are apparently impacted by the location itself.

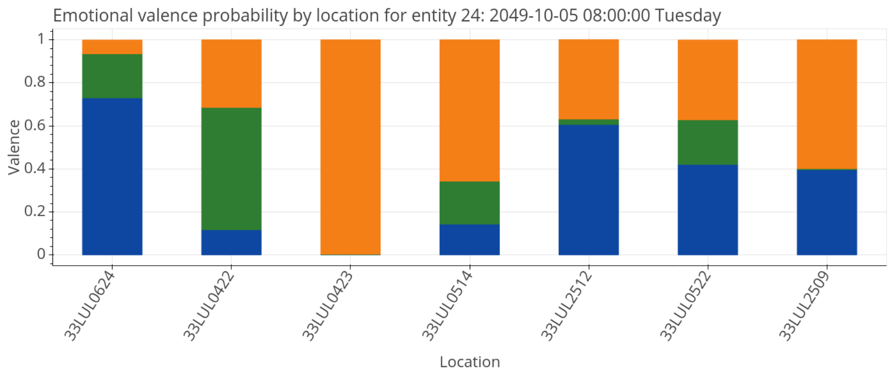

Figure 4.23: Emotional valence probabilities by location: entity 24, 2049-10-05 08:00 Tuesday.

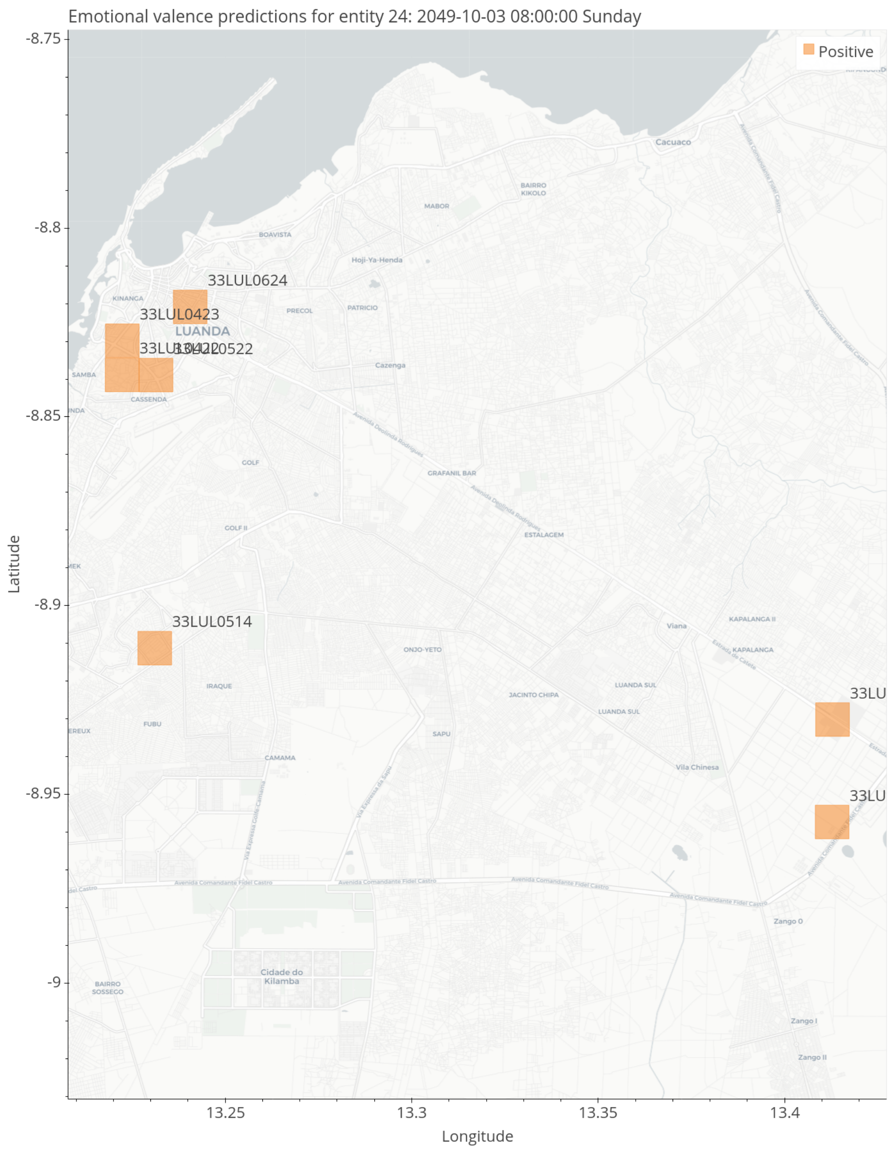

Figure 4.24: Emotional valence probabilities by map grid cell: entity 24, 2049-10-03 08:00 Sunday.

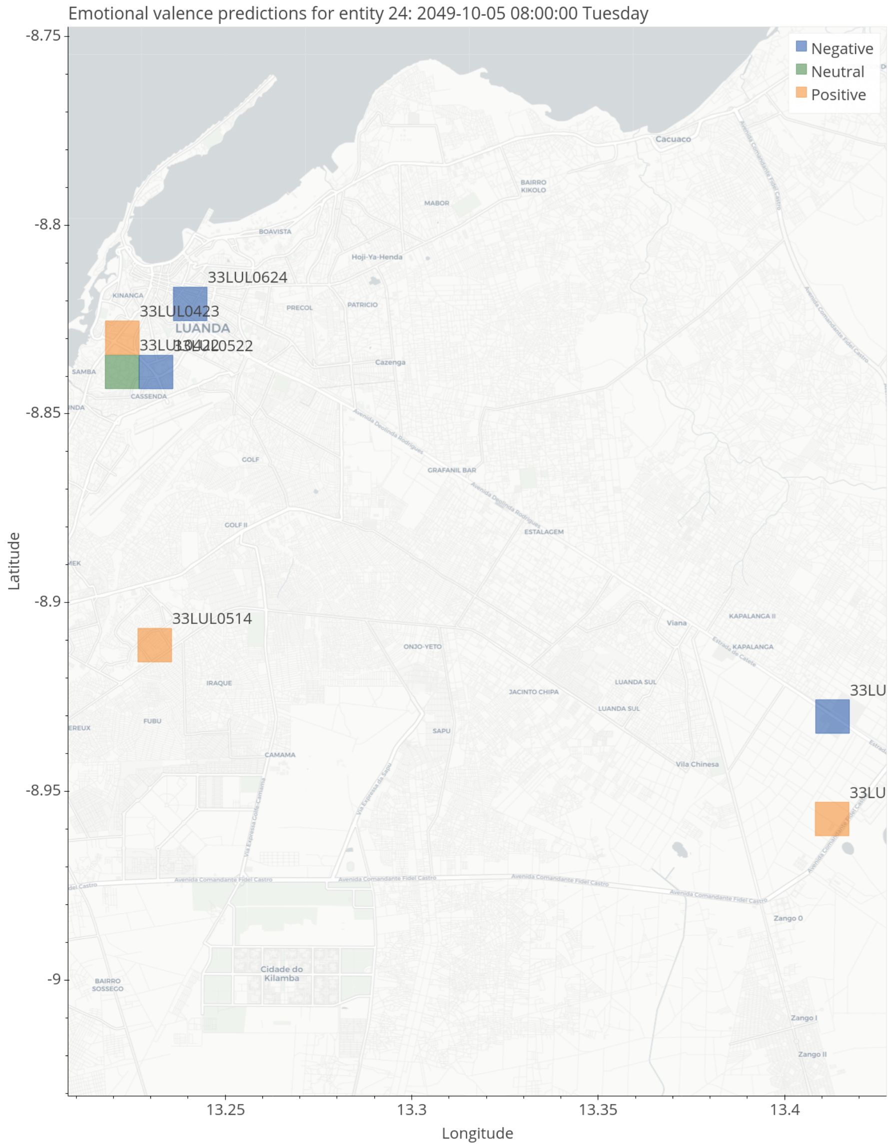

Figure 4.25: Emotional valence probabilities by map grid cell: entity 24, 2049-10-05 08:00 Tuesday.

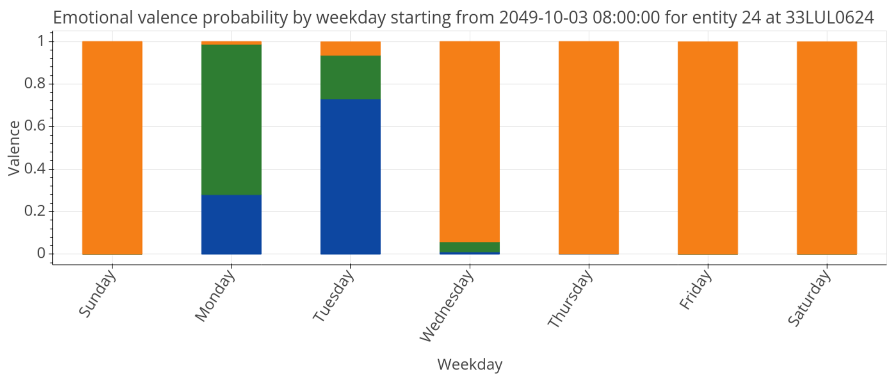

Figure 4.26: Emotional valence probabilities by weekday at 33LUL0624: entity 24, 2049-10-03–09 08:00.

All these findings by exploring the emotional valence predictions in context regarding a specific entity, raises more questions about other factors influence such as the already mentioned physical activity taking place before, during, and after the moment (e.g., hour) of the prediction. Future studies may answer this and other queries.

Chapter 5

Discussion

This is the closing chapter which discloses perceived limitations, some restrictions, and also relevant contributions of the presented research work and results described in the previous chapters. Furthermore, future trends and already devised improvements are also presented.

5.1 Limitations

The specific context, methodology followed, participants number and diversity, and data samples restrict the scope of every scientific study regarding HAI. This work is no exception. An obvious limitation is the absence of laboratory-based control over all the possible variables that may interfere with the study. This is common to the type of research and methodology where smartphone sensing is used as data collector. In order to clarify under which conditions the results should be interpreted, the identified limitations of this work are enumerated:

1. There is no prior health information about the users, particularly their past or ongoing clinical history, diagnosis or treatment, and presence of co-morbidities that may impact the engagement effect including prediction results bias.
2. An easy user-based feedback regarding the interaction experience is missing. This lack of information on user experience (UX) using SensAI limits the data available to gain insights towards UX improvement. Nevertheless, some application usage data (e.g., user entering and leaving each SensAI activity) is already being collected in order to diminish this limitation in future upgrades.
3. Regarding statistical comparisons and analysis, the current inability to account for socio-economic status may impede further analysis of

interest such as available resources and income which may correlate with stress and impact on emotional valence state.

4. People matter. Interacting with a non-anthropomorphic versus human-like agent may impact emotional valence state. Results about an increase of brain activity, including emotion-related areas, when interaction is human-human versus human-computer are evidenced in Schindler and Kissler, 2016.

5. Moral agency (e.g., Coelho et al., 2010; Da Rocha Costa and Coelho, 2019) is not included, i.e., a set of moral values to avoid undesirable situations such as the positive reinforcement of a human bullying another. SensAI is kept very simple and passive. A sophisticated conversational agent with moral sense and upgraded empathetic behaviour is envisioned as future work.

6. The human user is not questioned about privacy perception and expectations on HAI as described in Leite and Lehman, 2016. Nevertheless, information about privacy is presented in the initial user agreement. Moreover, data is anonymised and secured end-to-end on Cloud syncing. Further, privacy is enforced by keeping diary messages and other directly identifying information exclusively in SensAI and never disclosed, i.e., this private data is lost if the application is uninstalled from the mobile device.

7. Affective reactivity and regulation gender differences on emotional response to context are not considered as proposed by McRae et al., 2008. Although, gender and age neutrality is achieved by SensAI+Expanse results on predicting emotional valence states. Moreover, there is no evidence of any bias in the prediction scores achieved regarding the individual gender.

8. A smartphone may be considered a wearable device only to some extent. Although most of the time it is near the human body yet distinct from a true wearable device such as a wrist band. An example of this distinction it is the case of doing some aqua sports (e.g., swimming) and the smartphone left in a locker and unaware of the actual ongoing activity.

Regarding available resources, it may be also considered a limitation the finite computation facilities which impose compromises on options and forces decisions concerning the viability of results in a timely manner. In order to comply with the doctoral programme timeline, some grounding relating number of studies and explorations versus available computation and time is required.

5.2 Contributions

This research work contributions include a study results with evidence, restricted to the population and data samples available, of differences in behaviour amongst some combinations of age range versus gender. The main contribution is a novel system for studies regarding human emotional valence changes in context. This proposed method may complement and supersede (eventually) traditional long-list self-appraisal questionnaires[112]. The SensAI+Expanse developed system contributes with several parts[113] such as a mobile device application (SensAI) able to adapt and learn in order to predict emotional valence states with high performance, and a Cloud service (SensAI Expanse) with ready-to-action analysis and processing modules towards AutoML. It is free software thus able to be deployed in a private or public Cloud and modified as fit for each case. Moreover, publicly publishing the application for one or more mobile device platforms (e.g., Android™ on Google Play) available worldwide[114] avoids the laboratory usual restrictions on samples such as population from WEIRD[115] societies bias. In this sense, it contributes to new findings about specific contexts (e.g., geographically located, moment of the day) emotional valence changes in humans. Additionally to this space and time, demographics context is already in place comprising age and gender data useful for comparative studies (e.g., gender within age ranges). The research work developed to demonstrate the viability of this novel concept regarding age and gender comparative studies is discussed in Study (Chapter 4, p. 53).

[112] Such as TEQ for empathy self-appraisal as an affective process, or even longer questionnaires.

[113] Appendix A, p. 103

[114] Appendix B, p. 107

[115] Henrich et al., 2010

Regarding specific published contributions besides this dissertation, these are publicly available as (a) open source software[116] developed and shared in two Git repositories; and (b) scientific articles published and available in open science repositories, and also submitted to conferences.

[116] Refer to Appendix C, p. 115 for complete description.

1. CogA as a cognitive and affective library developed for SensAI. Includes integration, engineering, hacking, and development of sentiment analysis and also some natural language plug-in libraries from third-parties.
   `https://gitlab.com/nunoachenriques/coga`
2. SensAI as a mobile device sensing agent by means of an Android™ application.
   `https://gitlab.com/nunoachenriques/sensei`
3. SensAI Expanse as agent's expanded cognition and memory resources distributed in the Cloud encompassing AutoML and prediction capabilities served by a secure Web service.
   `https://gitlab.com/nunoachenriques/sensei-expanse`

4. VADER Sentiment Analysis in Java is an implementation of the VADER algorithm in Java. It started as a fork of the Java port[117] by Animesh Pandey of the NLTK VADER sentiment analysis module[118] written in Python from the original project[119] by the authors Hutto and Gilbert (2014). It is the same algorithm as an improved tool by extensive rewriting concerning relevant and required adaptations such as being Android™ ready.
`https://github.com/nunoachenriques/vader-sentiment-analysis`

5. Henriques, N. A. C., Coelho, H., & Garcia-Marques, L. (2019). SensAI+Expanse Adaptation on Human Behaviour Towards Emotional Valence Prediction.
   (a) Published (open science)
   `https://arxiv.org/abs/1912.10084`
   (b) Accepted as a regular paper (ADAPTIVE 2020)
   `https://www.iaria.org/conferences2020/ADAPTIVE20.html`

6. Henriques, N. A. C., Coelho, H., & Garcia-Marques, L. (2020). SensAI+Expanse Emotional Valence Prediction Studies with Cognition and Memory Integration.
   (a) Published (open science)
   `https://arxiv.org/abs/2001.09746`
   (b) Accepted as a regular paper (COGNITIVE 2020)
   `https://www.iaria.org/conferences2020/COGNITIVE20.html`

Furthermore, this work may contribute to future studies about the usefulness of this human-agent relationship type in order to improve health care. For instance, to diminish loneliness in humans by means of more companionship. Moreover, this work also contributes on the improvement of smartphone sensing use as a tool for research in psychology and medical sciences. In the near future, one may also devise contributions regarding the human-agent relationships and affective interactions. For instance, by making use of empathetic reactions measurement and evaluating those outcomes. This course of action may be used to verify and validate health status towards care improvement and significantly change the way humans live by themselves and also with daily present artificial entities.

[117] `https://github.com/apanimesh061/VaderSentimentJava`

[118] `https://www.nltk.org/_modules/nltk/sentiment/vader.html`

[119] `https://github.com/cjhutto/vaderSentiment`

5.3 Future

The smartphone sensing is currently a trend for continuous, non-invasive, and personalised health monitoring. Although equipped with several sensors yet still limited regarding biometric specifics such as the inability to check for specific molecules in the blood stream (e.g., cancer markers). Nevertheless, it is already accurate on detecting several pathologies and even different degrees of severity on specific ones regarding mental health. Furthermore, recent studies reveal the need for an up-to-date assessment regarding the effects of the agent's affective capabilities on long-term HAI towards better dyadic bonds. (Cornet & Holden, 2018; Greenes & Lorenzi, 1998; Place et al., 2017; Tsiourti, 2018; Vaidyam et al., 2019)

Further ahead, one may envision an extensive integration of distinct sensors for biometric data collection. This pervasive sensing may include data gathering at home such as in the toilet for body fluid's analysis, and also in the mirror for facial expressions and posture assessment. Moreover, currently available wearables (e.g., smart bands) with skin contact can gather instantaneous, direct, physiological data. This collection of individual data may be integrated and processed along with the data from smartphone sensors acquired by applications such as SensAI. (Batista et al., 2017; Nambiar et al., 2017; H. F. S. Gamboa, 2008; H. Gamboa et al., 2013; Santos et al., 2013)

Furthermore, the flexibility of any agent may depend on its cognition and memory capabilities, will to interact with the other, and the desire of sharing common emotions, sentiments and moods. Therefore, some improvements and modifications are thought to be useful of further exploration towards better scientific studies by an evolved SensAI+Expanse:

1. The human biometric-related data such as heartbeat rate is already in place and ready to be acquired. The access to this sensor data (if available in the device) only requires explicit activation (default values already set) in preferences, manifest, and Android™ application's main activity permissions checking, and then it will start collecting automatically. Nevertheless, additional work may be required to include all new sensor's data in the machine learning process. Moreover, the need for some engineering details such as all the application parameters being synced in Expanse will provide for the future ability to apply fine-tuning of SensAI in real-time and instantaneously (e.g., sensor data savings time interval). In order to properly achieve this integration SensAI should also evolve to embody a wearable device such as a wrist band thus improving the ability to sense physiological data.

2. The human's current physical activity type discovery may be improved by making use of metadata from geographic sources such as amenity[120] key values from OpenStreetMap. This heuristic-based approach may serve as a contingency solution for the smartphone sensing without a wearable complement (e.g., smartwatch). For instance, when engaged in an activity like swimming where usually the smartphone is kept in a locker thus away from the body. Although device's sensors may detect a still state yet a simple heuristic may infer the activity by geolocate places nearby and ask user for confirmation of the activity type (learning from there to avoid unnecessary interactions).

[120] `https://wiki.openstreetmap.org/wiki/Key:amenity?uselang=en`

3. A believable companion able to diminish loneliness of some humans by means of improved engagement such as detecting[121] the severity of negative affective states (e.g., depression) and take action, i.e., a full-featured conversational agent with adaptive empathy towards human care. Additionally, including a set of moral values to aid on making decisions such as taking action on some event. The agent will emerge with initial action restrictions, rules are implemented to minimize the error and deviations from the perceived affective state and moral guidelines. An example of a relevant rule is to only consider action from a perceived affective state if the likelihood threshold for that state is greater than 80%, i.e., if the agent is at least 80% sure of the affective state detected then it will act accordingly, otherwise, it will remain inactive. Moreover, agent's cognition should augment and new results will improve interaction (e.g., agent's affective state fine-tuning) and adaptive new behaviour such as the empathy score adaptation may use the learning error in an inverse non-proportional relation. Furthermore, SensAI may enabled classify options at some specific events for the human to grade its behaviour, and then learn from there including the results in new actions.

[121] Saeb et al., 2015; Provoost et al., 2019

4. The agent ability to recognise the human, and distinguish between different humans using the same device where it is embodied. Furthermore, to include the concept of identity or set of identities that encompass other entity (human or artificial). This will require extending demographic data to provide more detailed information such as level of education, region of origin, current residency, and more.

The SensAI agent with its Cloud-connected Expanse augmentation of cognition and memory poses itself as one approach to artificial entities able to accompany human beings. This concept of companionship start as a simple, very much passive, non-invasive, and with a focus on emotional valence state prediction in spatial and temporal context. It also relies on a non-anthropomorphic artificial entity embodiment using a smartphone. As described above in the future perspectives, there is much more ahead to be done, and unforeseen possibilities may arise with further developments regarding new findings in all fields of Cognitive Science. Furthermore, the development of hybrid, knowledge-driven, reasoning-based (Marcus, 2020) cognitive architectures comprising sophisticated moral and identity-related capabilities, and able to be integrated in artificial agents will be of special interest.

Bibliography

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science*, *13*(1), 27–33. https://doi.org/10.1111/1467-9280.00405

Alves-Oliveira, P., Gomes, S., Chandak, A., Arriaga, P., Hoffman, G., & Paiva, A. (2019). Software architecture for YOLO, a creativity-stimulating robot. http://arxiv.org/abs/1909.10823

Alves-Oliveira, P., Sequeira, P., Melo, F. S., Castellano, G., & Paiva, A. (2019). Empathic Robot for Group Learning: A Field Study. *ACM Transactions on Human-Robot Interaction*, *8*(1), 1–34. https://doi.org/10.1145/3300188

Arbib, M. A., & Fellous, J.-M. (2004). Emotions: from brain to robot. *Trends in Cognitive Sciences*, *8*(12), 554–561. https://doi.org/10.1016/j.tics.2004.10.004

Asada, M. (2019). Artificial Pain: empathy, morality, and ethics as a developmental process of consciousness. *TOCAIS 2019 Towards Conscious AI Systems Symposium co-located with the Association for the Advancement of Artificial Intelligence 2019 Spring Symposium Series (AAAI SSS-19)*, *2287*. http://ceur-ws.org/Vol-2287/paper19.pdf

Ashton, M. C., & Lee, K. (2007). Empirical, Theoretical, and Practical Advantages of the HEXACO Model of Personality Structure. *Personality and Social Psychology Review*, *11*(2), 150–166. https://doi.org/10.1177/1088868306294907

Ashton, M. C., Lee, K., & de Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality Factors: A Review of Research and Theory. *Personality and Social Psychology Review*, *18*(2), 139–152. https://doi.org/10.1177/1088868314523838

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., Boies, K., & De Raad, B. (2004). A Six-Factor Structure of Personality-Descriptive Adjectives: Solutions From Psycholexical Studies in Seven Languages. *Journal of Personality and Social Psychology*, *86*(2), 356–366. https://doi.org/10.1037/0022-3514.86.2.356

Barford, K., Zhao, K., & Smillie, L. (2016). Integrating Interpersonal Traits: Congruence between the Big Five, HEXACO, and Interpersonal Circumplex. *Personality and Individual Differences*, *101*(October), 466. https://doi.org/10.1016/j.paid.2016.05.085

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social cognitive and affective neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsw154

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Barrett, L. F., Lewis, M., & Haviland-Jones, J. M. (2008). *Handbook of Emotions* (L. F. Barrett, M. Lewis, & J. M. Haviland-Jones, Eds.; 3rd ed.). The Guildford Press.

Batista, D., Silva, H., & Fred, A. (2017). Experimental Characterization and Analysis of the BITalino Platforms Against a Reference Device. *International Conf. of the IEEE Engineering in Medicine and Biology Society*, 2418–2421.

Bavelas, J. B., Black, A., Lemery, C. R., & Mullett, J. (1986). "I show how you feel": Motor mimicry as a communicative act. *Journal of Personality and Social Psychology*, *50*(2), 322–329. https://doi.org/10.1037/0022-3514.50.2.322

Bavelas, J. B., & Gerwing, J. (2007). Conversational Hand Gestures and Facial Displays in Face-to-Face Dialogue. In K. Fiedler (Ed.), *Social communication* (pp. 283–308). Psychology Press. https://doi.org/10.4324/9780203837702

Beedie, C., Terry, P., & Lane, A. (2005). Distinctions between emotion and mood. *Cognition & Emotion*, *19*(6), 847–878. https://doi.org/10.1080/02699930541000057

Bliss, T. V., & Cooke, S. F. (2011). Long-term potentiation and long-term depression: A clinical perspective. *Clinics*, *66*(SUPPL.1), 3–17. https://doi.org/10.1590/S1807-59322011001300002

Bota, P. J., Wang, C., Fred, A. L. N., & Placido Da Silva, H. (2019). A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals. *IEEE Access*, *7*, 140990–141020. https://doi.org/10.1109/access.2019.2944001

Brocklebank, S., Pauls, S., Rockmore, D., & Bates, T. C. (2015). A spectral clustering approach to the structure of personality: Contrasting the FFM and HEXACO models. *Journal of Research* *in Personality*, *57*, 100–109. https://doi.org/10.1016/j.jrp.2015.05.003

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., Vanderplas, J., Joly, A., Holt, B., & Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project, 1–15. http://arxiv.org/abs/1309.0238

Calefato, F., Lanubile, F., & Novielli, N. (2017). EmoTxt: A toolkit for emotion recognition from text. *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, *2018-Janua*(July), 79–80. https://doi.org/10.1109/ACIIW.2017.8272591

Carmona, P., Nunes, D., Raposo, D., Silva, D., Silva, J. S., & Herrera, C. (2015). Happy hour - improving mood with an emotionally aware application. *2015 15th International Conference on Innovations for Community Services (I4CS)*, 1–7. https://doi.org/10.1109/I4CS.2015.7294480

Carolis, B. D., Ferilli, S., Palestra, G., Redavid, D., Informatica, D., & Bari, U. (2016). Emotion-Recognition from Speech-based Interaction in AAL Environment. In S. Bandini, G. Cortellessa, & F. Palumbo (Eds.), *Proceedings of the artificial intelligence for ambient assisted living 2016* (p. 13). http://ceur-ws.org/Vol-1803/paper7.pdf

Castellano, G., Kessous, L., & Caridakis, G. (2008). Emotion Recognition through Multiple Modalities: Face, Body Gesture, Speech. *Affect and emotion in human-computer interaction* (pp. 92–103). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85099-1_8

Castellano, G., Leite, I., & Paiva, A. (2016). Detecting perceived quality of interaction with a robot using contextual features. *Autonomous Robots*. https://doi.org/10.1007/s10514-016-9592-y

Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., & McOwan, P. W. (2010). Affect recognition for interactive companions: Challenges and design in real world scenarios. *Journal on Multimodal User Interfaces*, *3*(1), 89–98. https://doi.org/10.1007/s12193-009-0033-5

Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., & McOwan, P. (2013). Multimodal affect modeling and recognition for empathic robot companions. *International Journal of Humanoid Robotics*, *10*(01), 1350010. https://doi.org/10.1142/S0219843613500102

Cevher, D., Zepf, S., & Klinger, R. (2019). Towards Multimodal Emotion Recognition in German Speech Events in Cars using Transfer Learning. `http://arxiv.org/abs/1909.02764`

Chicco, D. (2017). Ten quick tips for machine learning in computational biology. *BioData Mining*, *10*(1), 1–17. `https://doi.org/10.1186/s13040-017-0155-3`

Coelho, H., Trigo, P., & Costa, A. C. (2010). On the operationality of moral-sense decision making. *Advances in Social Simulation: 2010 2nd Brazilian Workshop on Social Simulation, BWSS 2010 - Co-located with Joint Conference SBIA/SBRN/JRI 2010*, 15–20. `https://doi.org/10.1109/BWSS.2010.13`

Cornet, V. P., & Holden, R. J. (2018). Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of Biomedical Informatics*, *77*, 120–132. `https://doi.org/10.1016/j.jbi.2017.12.008`

Cramer, H., Goddijn, J., Wielinga, B., & Evers, V. (2010). Effects of (in) accurate empathy and situational valence on attitudes towards robots. *HRI '10 Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, 141–142. `https://doi.org/10.1145/1734454.1734513`

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *43*(3), 245–265. `https://doi.org/10.1348/0144665031752934`

Da Rocha Costa, A. C., & Coelho, H. M. F. (2019). Interactional moral systems: A model of social mechanisms for the moral regulation of exchange processes in agent societies. *IEEE Transactions on Computational Social Systems*, *6*(4), 778–796. `https://doi.org/10.1109/TCSS.2019.2926950`

Damásio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain* (2nd ed.). ["O Erro de Descartes: Emoção, Razão e o Cérebro Humano", Círculo de Leitores, ISBN:9724211290].

Damásio, A. (2010). *Self Comes to Mind: Constructing the Conscious Brain* (1st ed.). ["O Livro da Consciência: A Construção do Cérebro Consciente", Círculo de Leitores, ISBN:9789896441203].

Davis, M. H. (1980). Interpersonal Reactivity Index: form. *JSAS Catalog of Selected Documents in Psychology*, *10*, 3. `https://doi.org/10.1037/t01093-000`

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. `https://doi.org/10.1037/0022-3514.44.1.113`

Denecke, K., Gabarron, E., Grainger, R., Konstantinidis, S. T., Lau, A., Rivera-Romero, O., Miron-Shatz, T., & Merolli, M. (2019). Artificial Intelligence for Participatory Health: Applications, Impact, and Future Implications. *Yearbook of Medical Informatics*. https://doi.org/10.1055/s-0039-1677902

de Waal, F. B. M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, *59*, 279–300. https://doi.org/10.1146/annurev.psych.59.103006.093625

Dias, J., & Paiva, A. (2005). Feeling and Reasoning: A Computational Model for Emotional Characters. *Epia 2005, 12th portuguese conference on artificial intelligence* (pp. 127–140). Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/11595014_13

Disney, A., Reynolds, J., Schuman, C. D., Klibisz, A., Young, A., & Plank, J. S. (2016). DANNA: A neuromorphic software ecosystem. *Biologically Inspired Cognitive Architectures*, *17*, 49–56. https://doi.org/10.1016/j.bica.2016.07.007

Doce, T., Dias, J., Prada, R., & Paiva, A. (2010). Creating Individual Agents through Personality Traits. *Intelligent Virtual Agents, 10th International Conference, IVA 2010*, *6356*, 257–264. https://doi.org/10.1007/978-3-642-15892-6_27

Duch, W. (2017). Why minds cannot be received, but are created by brains. *Scientia et Fides*, *5*(2), 171–198. https://doi.org/10.12775/SetF.2017.014

Ekman, P. (1994). Moods, Emotions, and Traits.

Ekman, P. (1999). Basic Emotions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470013494.ch3

Ekman, P. (2016). What Scientists Who Study Emotion Agree About. *Perspectives on Psychological Science*, *11*(1), 31–34. https://doi.org/10.1177/1745691615596992

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, *17*(2), 124–129. https://doi.org/10.1037/h0030377

Engen, H. G., & Singer, T. (2013). Empathy circuits. *Current opinion in neurobiology*, *23*(2), 275–282. https://doi.org/10.1016/j.conb.2012.11.003

Fayek, H., Lech, M., & Cavedon, L. (2015). Towards real-time Speech Emotion Recognition using deep neural networks. *2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 1–5. https://doi.org/10.1109/ICSPCS.2015.7391796

Frijda, N. (2008). The Psychologists' Point of View. *Handbook of emotions* (pp. 68–87). The Guildford Press.

Frijda, N. H., Mesquita, B., Sonnemans, J., & Van Goozen, S. (1991). The duration of affective phenomena or emotions, sentiments and passions. In K. T. Strongman (Ed.), *International review of studies of emotion* (pp. 187–223). Wiley. `https://www.researchgate.net/publication/255522683`

Gamboa, H., Silva, H., & Fred, A. (2013). HiMotion: a new research resource for the study of behavior, cognition, and emotion. *Multimedia Tools and Applications*. `https://doi.org/10.1007/s11042-013-1602-x`

Gamboa, H. F. S. (2008). *Multi-modal behavioral biometrics based on HCI and electrophysiology* (Doctoral dissertation April 2008). Universidade Técnica de Lisboa.

Garten, J., Kennedy, B., Hoover, J., Sagae, K., & Dehghani, M. (2018). Incorporating Demographic Embeddings Into Language Understanding. *Cognitive Science*, *43*(1), e12701. `https://doi.org/10.1111/cogs.12701`

Goertzel, B., Lian, R., Arel, I., de Garis, H., & Chen, S. (2010). A world survey of artificial brain projects, Part II: Biologically inspired cognitive architectures. *Neurocomputing*, *74*(1-3), 30–49. `https://doi.org/10.1016/j.neucom.2010.08.012`

Granatyr, J., Scalabrin, E. E., Osman, N., Dias, J., Nunes, M. A. S. N., Masthoff, J., Enembreck, F., Lessing, O. R., Sierra, C., & Paiva, A. M. (2017). The Need for Affective Trust Applied to Trust and Reputation Models. *ACM Computing Surveys*, *50*(4), 1–36. `https://doi.org/10.1145/3078833`

Greenes, R. A., & Lorenzi, N. M. (1998). Audacious Goals for Health and Biomedical Informatics in the New Millennium. *Journal of the American Medical Informatics Association*, *5*(5), 395–400. `https://doi.org/10.1136/jamia.1998.0050395`

Guo, Y. R., & Hoe-Lian Goh, D. (2016). Evaluation of affective embodied agents in an information literacy game. *Computers & Education*. `https://doi.org/10.1016/j.compedu.2016.09.013`

Harari, G. M., Müller, S. R., Aung, M. S. H., Rentfrow, P. J., Harari, G. M., Mu, S. R., Rentfrow, P. J., Müller, S. R., Aung, M. S. H., & Rentfrow, P. J. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences*, *18*, 83–90. `https://doi.org/10.1016/j.cobeha.2017.07.018`

Hegel, F., Spexard, T., Wrede, B., Horstmann, G., & Vogt, T. (2006). Playing a different imitation game: Interaction with an Empathic Android Robot. *2006 6th IEEE-RAS International Conference on Humanoid Robots*, 56–61. `https://doi.org/10.1109/ICHR.2006.321363`

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Heyes, C. (2018). Empathy is not in our genes. *Neuroscience & Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2018.11.001

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 216–225. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109

Inkster, B., Sarda, S., & Subramanian, V. (2018). An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth and uHealth*, *6*(11), e12106. https://doi.org/10.2196/12106

Jones, E., Oliphant, T., & Peterson, P. (2001). SciPy: Open Source Scientific Tools for Python. https://scipy.org/

Kaya, N., & Epps, H. H. (2004). Relationship Between Color and Emotion: a Study of College Students. *College Student Journal*, *38*(3), 396–405. https://nzdis.org/projects/attachments/299/colorassociation-students.pdf

Koyama, T. (2016). Ethical issues for social robots and the trust-based approach. *2016 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, 1–5. https://doi.org/10.1109/ARSO.2016.7736246

Lathia, N., Sandstrom, G. M., Mascolo, C., & Rentfrow, P. J. (2017). Happier People Live More Active Lives: Using Smartphones to Link Happiness and Physical Activity (R. A. Krukowski, Ed.). *PLOS ONE*, *12*(1), e0160589. https://doi.org/10.1371/journal.pone.0160589

Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (2012). Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings. *HRI '12 Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 367–374. http://dl.acm.org/citation.cfm?id=2157811

Leite, I., & Lehman, J. F. (2016). The Robot Who Knew Too Much. *Proceedings of the The 15th International Conference on Interaction Design and Children - IDC '16*, 379–387. https://doi.org/10.1145/2930674.2930687

Likamwa, R., Liu, Y., Lane, N. D., & Zhong, L. (2013). MoodScope: Building a Mood Sensor from Smartphone Usage Patterns. *MobiSys '13 Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, 389–402. https://doi.org/10.1145/2462456.2464449

Lim, M. Y. (2012). Memory models for intelligent social companions. *Studies in Computational Intelligence*, *396*, 241–262. https://doi.org/10.1007/978-3-642-25691-2_10

Lin, Z., Xu, P., Winata, G. I., Liu, Z., Fung, P., Bay, C. W., & Kong, H. (2019). CAiRE: An End-to-End Empathetic Chatbot. http://arxiv.org/abs/1907.12108

Lind, M. N., Byrne, M. L., Wicks, G., Smidt, A. M., & Allen, N. B. (2018). The Effortless Assessment of Risk States (EARS) Tool: An Interpersonal Approach to Mobile Sensing. *JMIR Mental Health*, *5*(3), e10334. https://doi.org/10.2196/10334

Liu, X., & London, K. (2016). T.A.I: A Tangible AI Interface to Enhance Human-Artificial Intelligence (AI) Communication Beyond the Screen. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16*, *2*(1), 281–285. https://doi.org/10.1145/2901790.2901896

Lorenz, R., Hampshire, A., & Leech, R. (2017). Neuroadaptive Bayesian Optimization and Hypothesis Testing. *Trends in Cognitive Sciences*, *21*(3), 155–167. https://doi.org/10.1016/j.tics.2017.01.006

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2019). Explainable AI for Trees: From Local Explanations to Global Understanding. http://arxiv.org/abs/1905.04610

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *2017-Decem*(Section 2), 4766–4775.

Lüscher, C., & Malenka, R. C. (2012). NMDA Receptor-Dependent Long-Term Potentiation and Long-Term Depression (LTP/LTD). *Cold Spring Harbor Perspectives in Biology*, *4*(a005710), 1–15. https://doi.org/10.1101/cshperspect.a005710

Madden, T. J., Hewett, K., & Roth, M. S. (2000). Managing Images in Different Cultures: A Cross-National Study of Color Meanings and Preferences. *Journal of International Marketing*, *8*(4), 90–107. https://doi.org/10.1509/jimk.8.4.90.19795

Man, K., & Damasio, A. (2019). Homeostatically Motivated Intelligence for Feeling Machines. *TOCAIS 2019 Towards Conscious AI Systems Symposium co-located with the Association for the Advancement of Artificial Intelligence 2019 Spring Symposium*

Series (AAAI SSS-19), *2287*. http://ceur-ws.org/Vol-2287/short3.pdf

Marcus, G. (2020). The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence. (February). http://arxiv.org/abs/2002.06177

McCrae, R. R., & John, O. P. (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, *2*(11), 205. https://doi.org/10.21105/joss.00205

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th python in science conference* (pp. 51–56).

McRae, K., Ochsner, K. N., Mauss, I. B., Gabrieli, J. J. D., & Gross, J. J. (2008). Gender Differences in Emotion Regulation: An fMRI Study of Cognitive Reappraisal. *Group Processes & Intergroup Relations*, *11*(2), 143–162. https://doi.org/10.1177/1368430207088035

Melo, C. M. D., Carnevale, P., & Gratch, J. (2011). The Effect of Expression of Anger and Happiness in Computer Agents on Negotiations with Humans. In L. Sonenberg, P. Stone, K. Tumer, & P. Yolum (Eds.), *10th international conference on autonomous agents and multiagent systems aamas 2011* (pp. 937–944). IFAAMAS.

Minsky, M. (1961). Steps toward Artificial Intelligence. *Proceedings of the IRE*, *49*(1), 8–30. https://doi.org/10.1109/JRPROC.1961.287775

Morgan, A. (2016). Against compassion: in defence of a "hybrid" concept of empathy. *Nursing Philosophy*, 1–6. https://doi.org/10.1111/nup.12148

Nambiar, A., Bernardino, A., Nascimento, J. C., & Fred, A. (2017). Context-Aware Person Re-Identification in the Wild Via Fusion of Gait and Anthropometric Features. *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 973–980. https://doi.org/10.1109/FG.2017.121

Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in Virtual Agents and Robots. *ACM Transactions on Interactive Intelligent Systems*, *7*(3), 1–40. https://doi.org/10.1145/2912150

Palma, T. A., Santos, A. S., & Garcia-Marques, L. (2018). The future is now: the impact of present fluency in judgments about the future. *Memory*, *26*(2), 144–153. https://doi.org/10.1080/09658211.2017.1335328

Parisi, D., & Petrosino, G. (2010). Robots that have emotions. *Adaptive Behavior*, *18*(6), 453–469. https://doi.org/10.1177/1059712310388528

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pérez, J., Cerezo, E., Serón, F. J., & Rodríguez, L.-F. F. (2016). A cognitive-affective architecture for ECAs. *Biologically Inspired Cognitive Architectures*, *18*, 33–40. https://doi.org/10.1016/j.bica.2016.10.002

Perusquía-Hernández, M., Jáuregui, D. A. G., Cuberos-Balda, M., & Paez-Granados, D. (2019). Robot mirroring: A framework for self-tracking feedback through empathy with an artificial agent representing the self. http://arxiv.org/abs/1903.08524

Picard, R. W. (2003). Affective computing: challenges. *International Journal of Human-Computer Studies*, *59*(1-2), 55–64. https://doi.org/10.1016/S1071-5819(03)00052-1

Piwek, L., McKay, L. S., & Pollick, F. E. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition*, *130*(3), 271–277. https://doi.org/10.1016/j.cognition.2013.11.001

Place, S., Blanch-Hartigan, D., Rubin, C., Gorrostieta, C., Mead, C., Kane, J., Marx, B. P., Feast, J., Deckersbach, T., Pentland, A. “, Nierenberg, A., & Azarbayejani, A. (2017). Behavioral Indicators on a Mobile Sensing Platform Predict Clinically Validated Psychiatric Symptoms of Mood and Anxiety Disorders. *Journal of Medical Internet Research*, *19*(3), e75. https://doi.org/10.2196/jmir.6678

Plutchik, R. (2001). The Nature of Emotions. *American Scientist*, *89*(4), 344–350. https://doi.org/10.1511/2001.4.344

Polajnar, J., Dalvandi, B., & Polajnar, D. (2011). Does empathy between artificial agents improve agent teamwork? *Proceedings of the 10th IEEE International Conference on Cognitive Informatics and Cognitive Computing, ICCI*CC 2011*, 96–102. https://doi.org/10.1109/COGINF.2011.6016126

Potapov, A., & Rodionov, S. (2014). Universal empathy and ethical bias for artificial general intelligence. *Journal of Experimental & Theoretical Artificial Intelligence*, *26*(3), 405–416. https://doi.org/10.1080/0952813X.2014.895112

Provoost, S., Ruwaard, J., van Breda, W., Riper, H., Bosse, T., Breda, W. V., Riper, H., & Greenshaw, A. J. (2019). Validating Automated Sentiment Analysis of Online Cognitive Behavioral Therapy Patient Texts: An Exploratory Study. *Frontiers in*

Psychology, *10*(May), 1–12. https://doi.org/10.3389/fpsyg.2019.01065

Rana, R., Reilly, J., Jurdak, R., Hu, W., Li, X., & Soar, J. (2014). Affect Sensing on Smartphone - Possibilities of Understanding Cognitive Decline in Aging Population, 1–7. http://arxiv.org/abs/1407.5910

Ranjbartabar, H., Richards, D., Bilgin, A., & Kutay, C. (2019). First Impressions Count! The Role of the Human's Emotional State on Rapport Established with an Empathic versus Neutral Virtual Therapist. *IEEE Transactions on Affective Computing*, *PP*(100), 1–1. https://doi.org/10.1109/TAFFC.2019.2899305

Rashkin, H., Smith, E. M., Li, M., & Boureau, Y.-L. (2018). I Know the Feeling: Learning to Converse with Empathy. (1994), 1–13. http://arxiv.org/abs/1811.00207

Richards, D., Bilgin, A. A., & Ranjbartabart, H. (2018). Users' perceptions of empathic dialogue cues. *Proceedings of the 18th International Conference on Intelligent Virtual Agents - IVA '18*, 35–42. https://doi.org/10.1145/3267851.3267857

Riedl, R., Mohr, P. N. C., Kenning, P. H., Davis, F. D., & Heekeren, H. R. (2014). Trusting Humans and Avatars: A Brain Imaging Study Based on Evolution Theory. *Journal of Management Information Systems*, *30*(4), 83–114. https://doi.org/10.2753/MIS0742-1222300404

Romanycia, M. H. J., & Pelletier, F. J. (1985). What is a heuristic? *Computational Intelligence*, *1*(1), 47–58. https://doi.org/10.1111/j.1467-8640.1985.tb00058.x

Rukavina, S., Gruss, S., Hoffmann, H., & Traue, H. C. (2016). Facial Expression Reactions to Feedback in a Human-Computer Interaction—Does Gender Matter? *Psychology*, *07*(03), 356–367. https://doi.org/10.4236/psych.2016.73038

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *Journal of medical Internet research*, *17*(7), e175. https://doi.org/10.2196/jmir.4273

Sagar, M., Seymour, M., & Henderson, A. (2016). Creating connection with autonomous facial animation. *Communications of the ACM*, *59*(12), 82–91. https://doi.org/10.1145/2950041

Sakurai, Y., Ikegami, Y., Sakai, M., Fujikawa, H., Tsuruta, S., Gonzalez, A. J., Sakurai, E., Damiani, E., Kutics, A., Knauf, R., & Frati, F. (2019). VICA, a visual counseling agent for emotional distress.

Journal of Ambient Intelligence and Humanized Computing. https://doi.org/10.1007/s12652-019-01180-x

Santos, M., Fred, A., Silva, H., & Lourenço, A. (2013). Eigen Heartbeats for User Identification. *6th International Conference on Bio-inspired Systems and Signal Processing*, 1–5.

Sato, M., Poupyrev, I., & Harrison, C. (2012). Touché: Enhancing Touch Interaction on Humans, Screens, Liquids, and Everyday Objects. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, (100), 483–492. https://doi.org/10.1145/2207676.2207743

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion*, *23*(7), 1307–1351. https://doi.org/10.1080/02699930902928969

Schindler, S., & Kissler, J. (2016). People matter: Perceived sender identity modulates cerebral processing of socio-emotional language feedback. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2016.03.052

Schulz, E. (2017). *Towards a unifying theory of generalization* (Doctoral dissertation). University College London. http://discovery.ucl.ac.uk/id/eprint/1572581

Schwark, J. D. (2015). Toward a Taxonomy of Affective Computing. *International Journal of Human-Computer Interaction*, *31*(11), 761–768. https://doi.org/10.1080/10447318.2015.1064638

Shuman, V., & Scherer, K. R. (2015). Emotions, Psychological Structure of. *International encyclopedia of the social & behavioral sciences* (pp. 526–533). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.25007-1

Siddique, F. B., Kampman, O., Yang, Y., Dey, A., & Fung, P. (2017). Zara Returns: Improved Personality Induction and Adaptation by an Empathetic Virtual Agent. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics-System Demonstrations*, 121–126. https://doi.org/https://doi.org/10.18653/v1/P17-4021

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience & Biobehavioral Reviews*, *30*(6), 855–863. https://doi.org/10.1016/j.neubiorev.2006.06.011

Solera-Ureña, R., Moniz, H., Batista, F., Astudillo, R. F., Campos, J., Paiva, A., & Trancoso, I. (2016). Acoustic-Prosodic Automatic Personality Trait Assessment for Adults and Children. *Advances in Speech and Language Technologies for Iberian Languages: Third International Conference, IberSPEECH 2016, Lisbon, Portugal,* *November 23-25, 2016, Proceedings*, 192–201. https://doi.org/10.1007/978-3-319-49169-1_19

Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S.-F., & Pantic, M. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing*, *65*(11), 3–14. https://doi.org/10.1016/j.imavis.2017.08.003

Sorbello, R., Chella, A., Giardina, M., Nishio, S., & Ishiguro, H. (2016). An Architecture for Telenoid Robot as Empathic Conversational Android Companion for Elderly People. In E. Menegatti, N. Michael, K. Berns, & H. Yamaguchi (Eds.). Springer International Publishing. https://doi.org/10.1007/978-3-319-08338-4_68

Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale Development and Initial Validation of a Factor-Analytic Solution to Multiple Empathy Measures. *Journal of Personality Assessment*, *91*(1), 62–71. https://doi.org/10.1080/00223890802484381

Stueber, K. (2014). Empathy (E. N. Zalta, Ed.; Winter2014). http://plato.stanford.edu/archives/win2014/entries/empathy

Subramanian, R., Wache, J., Abadi, M. K., Vieriu, R. L., Winkler, S., & Sebe, N. (2018). ASCERTAIN: Emotion and Personality Recognition Using Commercial Sensors. *IEEE Transactions on Affective Computing*, *9*(2), 147–160. https://doi.org/10.1109/TAFFC.2016.2625250

Tay, B. T., Low, S. C., Ko, K. H., & Park, T. (2016). Types of humor that robots can play. *Computers in Human Behavior*, *60*, 19–28. https://doi.org/10.1016/j.chb.2016.01.042

Tipler, F. (2003). *A Física da Imortalidade — Cosmologia Moderna, Deus e a Ressurreição dos Mortos*. Editorial Bizâncio.

Trigeorgis, G., Ringeval, F., Brueckner, R., Marchi, E., Nicolaou, M. A., Schuller, B., & Zafeiriou, S. (2016). Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5200–5204. https://doi.org/10.1109/ICASSP.2016.7472669

Tsiourti, C. (2018). *Artificial agents as social companions: design guidelines for emotional interactions* (Doctoral dissertation). Université de Genève. https://doi.org/10.13097/archive-ouverte/unige:110600

Ugazio, G., & Ruff, C. (2017). Neural Control of Social Decisions. *Decision neuroscience* (pp. 233–245). Elsevier. https://doi.org/10.1016/B978-0-12-805308-9.00019-1

Vaes, J., Meconi, F., Sessa, P., & Olechowski, M. (2016). Minimal humanity cues induce neural empathic reactions towards non-human entities. *Neuropsychologia*, *89*, 132–140. https://doi.org/10.1016/j.neuropsychologia.2016.06.004

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *The Canadian Journal of Psychiatry*. https://doi.org/10.1177/0706743719828977

Vallverdù, J., Franzoni, V., & Milani, A. (2019). Errors, Biases and Overconfidence in Artificial Emotional Modeling. *IEEE/WIC/ACM International Conference on Web Intelligence on - WI '19 Companion*, 86–90. https://doi.org/10.1145/3358695.3361749

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science Engineering*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

Verduyn, P., Delaveau, P., Rotgé, J.-Y., Fossati, P., & Van Mechelen, I. (2015). Determinants of Emotion Duration and Underlying Psychological and Neural Mechanisms. *Emotion Review*, *7*(4), 330–335. https://doi.org/10.1177/1754073915590618

Vernon, D., Thill, S., & Ziemke, T. (2016). The Role of Intention in Cognitive Robotics. *Toward robotic socially believable behaving systems* (pp. 15–27). https://doi.org/10.1007/978-3-319-31056-5_3

Wada, K., Shibata, T., Saito, T., & Tanie, K. (2004). Effects of robot-assisted activity for elderly people and nurses at a day service center. *Proceedings of the IEEE*, *92*(11), 1780–1788. https://doi.org/10.1109/JPROC.2004.835378

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wood, R., Baxter, P., & Belpaeme, T. (2012). A review of long-term memory in natural and synthetic systems. *Adaptive Behavior*, *20*(2), 81–103. https://doi.org/10.1177/1059712311421219

Yalçin, Ö. N. (2018). Modeling Empathy in Embodied Conversational Agents. *Proceedings of the 2018 on International Conference on Multimodal Interaction - ICMI '18*, 546–550. https://doi.org/10.1145/3242969.3264977

Yalçin, Ö. N. (2019). Evaluating Empathy in Artificial Agents. http://arxiv.org/abs/1908.05341

Zhao, M., Adib, F., & Katabi, D. (2016). Emotion Recognition using Wireless Signals. *MobiCom'16, October 03 - 07, 2016: The 22nd Annual International Conference on Mobile Computing and Networking*. https://doi.org/10.1145/2973750.2973762

Zhu, R., Wang, Z., Ma, Z., Wang, G., & Xue, J. H. (2018). LRID: A new metric of multi-class imbalance degree based on likelihood-ratio test. *Pattern Recognition Letters*, *116*, 36–42. https://doi.org/10.1016/j.patrec.2018.09.012

Appendix A

Development Tools

This appendix refers to the tools used for the development and system integration of SensAI and SensAI Expanse, i.e., SensAI+Expanse platform. The whole project uses free software[122] ranging from the Debian GNU/Linux operating system to the tools, programming languages, specific libraries and frameworks used for the development.

[122] https://en.wikipedia.org/wiki/Free_software

A.1 Application Programming

The references described are the ones used more frequently and relevant to facilitate the development. Several other tools such as BASH Script Unix shell and command language for scripting helpers were used where fit. R and R Studio where used at initial stages and discarded afterwards.

Android Studio Integrated development environment (IDE) for designing, programming, building, testing, debugging and profiling applications on Android™ devices. Main language used is Java although several other such as Extensible Markup Language (XML) based are included.

Used for SensAI and CogA development.

`https://developer.android.com/studio/`

PyCharm IDE for coding and testing Python language-based applications. Also for development using other languages such as Structured Query Language (SQL) and Procedural Language/-PostgreSQL (PL/pgSQL) for the RDBMS repository.

Used for SensAI Expanse development.

`https://www.jetbrains.com/pycharm/`

A.2 Documentation Production

The references described are the relevant ones used to facilitate the production of several documents. Additionally, to the referenced below, Google Docs and Sheets (`https://www.google.com/docs/about/`) were also used for *ad hoc* writings in a digital notebook.

TeXstudio An integrated writing environment for creating LaTeX-based documents.

Used for all technical and scientific writings including this dissertation.

`https://www.texstudio.org/`

Google Drawings Is a free, Web-based diagramming software developed by Google. It allows users to collaborate and work together in real time and also share access to view or comment.

Used for some diagrams included in this dissertation.

`https://en.wikipedia.org/wiki/Google_Drawings`

A.3 Servers and Services

The references described are the ones used for the system integration. Several other tools such as BASH Script Unix shell and command language for scripting helpers were used where fit.

Apache HTTP Server An open-source HTTP server for modern operating systems. A secure, efficient and extensible server that provides HTTP services in sync with the current HTTP standards.

Used for SensAI+Expanse system integration: secure mobile application Web access to the agent expanded resources using HTTP over TLS.

`https://httpd.apache.org/`

Flask A micro framework for Web development in Python. Integrates Web service deployment and repository data access in agile manner, i.e., resources saving, less code, more performance.

Used for SensAI+Expanse system integration: connect Web service and data repository access.

`https://pypi.org/project/Flask/`

PostgreSQL A powerful, open source object-relational database system that uses and extends the SQL language combined with many features that safely store and scale the most complicated data workloads.

Used for SensAI+Expanse system integration: data repository for all agents.

https://www.postgresql.org/

WSO2 Enterprise Integrator A middleware in Java to integrate Web service deployment and repository data access in agile manner, i.e., less code, more specification and automatic deployment.

(**DEPRECATED since Expanse version 0.2. See Flask**) Used during Expanse version 0.1 for SensAI+Expanse system integration: connect Web service and data repository access.

https://wso2.com/integration/

A.4 Trademarks

Android is a trademark of Google LLC.

Appendix B

Study

B.1 Tools

The tools used for the experimental procedure supporting SensAI+Expanse. The concept of publishing the application available publicly for current majority of mobile devices is applied. This conceptual choice helps to avoid the laboratory usual restrictions on samples such as people from WEIRD[123] societies bias. Moreover, the development and deployment effort is reduced to a minimum for a reasonable diverse[124] and large number of potential human subjects.

Android™ is chosen because (a) it is a free and open source operating system; and (b) holds (2018Q3) a market share majority of 86.8%[125]. Targeting the application to platform version of at least 6.0 up to 9 equals 71% of all Android™ devices[126]. Therefore, 61.628% of the global market, i.e., more than 1800 M humans from 3000 M[127] potential subjects.

Google Play The publicly accessible store where SensAI is published.

Used for SensAI application distribution.

`https://play.google.com/store/apps/details?id=net.nunoachenriques.sensei`

Google Play Console The tool to publish SensAI app. Have features that help improve quality and more by reporting several statistics over time intervals. Moreover, features ways of selecting and inviting users for initial testing with a restricted group.

Used for SensAI release publishing, restricted group testing, statistics gathering to improve quality (e.g, application crash or not responding reports).

`https://developer.android.com/distribute/console/`

[123] Henrich et al., 2010

[124] SensAI was already installed by users from ten countries and four continents (Africa, America, Asia, Europe)

[125] `https://www.idc.com/promo/smartphone-market-share/os`

[126] `https://developer.android.com/about/dashboards/`

[127] `https://newzoo.com/insights/articles/newzoos-2018-global-mobile-market-report-insights-into-the-worlds-3-billion-smartphone-users/`

Jupyter Notebook Web-based user interface for Project Jupyter. Using over a Jupyter server where the kernel runs and loads SensAI Expanse developed modules in Python for data extraction, transformation, and loading. The notebook is a file merging documentation text (Markdown format) and Python code with data and graphics output in line.

Used for SensAI Expanse data analysis, statistics and charts production. Also, for machine learning modules prototyping and fine-tuning.

`https://jupyter.org/`

All the source code including libraries and frameworks used as processing and visualisation tools for data analysis from results is referenced in Appendix C, p. 115.

B.2 Parameters

This section presents more in-depth details for each entity model adaptation with a table for each estimator with duration and score followed by relevant parameter auto discovered by machine learning in pipeline process. A table with population data common to every estimator learning process is shown first. This (Table B.1, p. 109) includes data shape (number of samples and features), Likelihood Ratio Imbalance Degree (LRID)[128], and class counts of each classification (negative, neutral, positive).

[128] Zhu et al. (2018)

Table B.1: Data common to every estimator process for each population entity

	n_tuples	n_features	lrid	class_count (-1, 0, 1)
0	5382	124	−0.24	[576, 2216, 1513]
1	6822	86	−0.82	[304, 4112, 1041]
2	29149	96	−1.28	[217, 0, 23101]
3	15144	70	−0.16	[0, 3671, 8443]
4	1741	24	−0.55	[0, 204, 1188]
5	691	28	−1.53	[19, 25, 508]
6	200119	257	−0.47	[10816, 55240, 94039]
7	186130	282	−1.10	[13517, 10356, 125031]
8	427	46	−0.36	[0, 72, 269]
9	28278	131	−1.90	[81, 627, 21914]
10	1795	77	−0.04	[353, 536, 547]
11	13435	41	−1.07	[0, 399, 10349]
12	20392	65	−1.91	[299, 155, 15859]
13	6931	25	−0.53	[676, 1070, 3798]
14	5130	15	−0.34	[376, 1612, 2116]
15	11605	22	−0.02	[2861, 3722, 2701]
16	4371	56	−1.12	[18, 716, 2762]
17	950	8	−0.64	[25, 268, 467]
18	7203	159	−0.42	[475, 1923, 3364]
19	29026	94	−0.82	[203, 7914, 15103]
20	37834	44	−0.45	[2248, 17651, 10368]
21	1126	9	−0.21	[0, 246, 654]
22	9885	91	−0.60	[0, 1052, 6856]
23	2396	23	−0.59	[64, 784, 1068]
24	5192	121	−0.53	[419, 932, 2802]
25	10027	34	−0.91	[313, 1516, 6192]
26	3964	165	−0.50	[306, 2075, 790]
27	2342	7	−0.85	[78, 388, 1407]
28	1860	6	−0.06	[670, 406, 412]
29	19726	124	−0.58	[884, 4786, 10110]
30	360	5	−0.25	[0, 73, 215]

Table B.2: Dummy duration, score and relevant parameters of each entity model

	duration	score	strategy
0	00:00:00	0.37	stratified
1	00:00:02	0.61	stratified
2	00:00:02	0.98	stratified
3	00:00:02	0.58	stratified
4	00:00:00	0.75	stratified
5	00:00:01	0.81	stratified
6	00:00:02	0.47	stratified
7	00:00:02	0.72	stratified
8	00:00:02	0.65	stratified
9	00:00:02	0.95	stratified
10	00:00:02	0.34	stratified
11	00:00:03	0.93	stratified
12	00:00:02	0.95	stratified
13	00:00:02	0.52	stratified
14	00:00:02	0.42	stratified
15	00:00:00	0.34	stratified
16	00:00:00	0.67	stratified
17	00:00:00	0.51	stratified
18	00:00:02	0.43	stratified
19	00:00:02	0.55	stratified
20	00:00:00	0.46	stratified
21	00:00:01	0.62	stratified
22	00:00:00	0.78	stratified
23	00:00:00	0.49	stratified
24	00:00:02	0.53	stratified
25	00:00:03	0.65	stratified
26	00:00:01	0.52	stratified
27	00:00:02	0.60	stratified
28	00:00:00	0.36	stratified
29	00:00:02	0.50	stratified
30	00:00:00	0.54	stratified

Table B.3: Logistic Regression duration, score and relevant parameters of each entity model

	duration	score	max_iter	solver	penalty	C
0	00:00:12	0.69	10000	liblinear	l1	0.1
1	00:03:21	0.79	10000	saga	l1	5.7
2	00:09:22	0.99	10000	saga	l2	0.3
3	00:03:14	0.79	10000	saga	l2	0.8
4	00:00:21	0.97	10000	saga	l2	6.0
5	00:00:11	0.96	10000	liblinear	l1	10.0
6	00:04:35	0.59	10000	liblinear	l1	10.0
7	00:05:44	0.84	10000	liblinear	l2	4.7
8	00:00:08	0.95	10000	liblinear	l2	8.0
9	00:00:32	0.98	10000	liblinear	l1	5.4
10	00:02:33	0.85	10000	saga	l1	10.0
11	00:02:07	0.97	10000	saga	l1	0.8
12	00:16:04	0.98	10000	saga	l1	10.0
13	00:02:15	0.67	10000	saga	l2	10.0
14	00:01:04	0.57	10000	saga	l1	3.0
15	00:01:46	0.48	10000	saga	l2	0.1
16	00:04:44	0.84	10000	saga	l1	9.8
17	00:00:08	0.89	10000	liblinear	l1	1.1
18	00:00:13	0.63	10000	liblinear	l2	2.3
19	00:21:25	0.68	10000	saga	l2	10.0
20	00:13:25	0.58	10000	saga	l2	0.4
21	00:00:08	0.76	10000	liblinear	l2	0.2
22	00:05:09	0.89	10000	saga	l1	9.1
23	00:00:51	0.74	10000	saga	l2	0.1
24	00:00:14	0.73	10000	liblinear	l1	6.3
25	00:04:39	0.82	10000	saga	l1	2.3
26	00:00:13	0.77	10000	liblinear	l2	1.8
27	00:00:40	0.85	10000	saga	l1	6.7
28	00:00:29	0.75	10000	saga	l2	0.1
29	00:00:30	0.65	10000	liblinear	l2	2.7
30	00:00:04	1.00	10000	liblinear	l2	2.4

Table B.4: Extreme Gradient Boosting duration, score and relevant parameters of each entity model

	duration	score	max_depth	n_estimators	learning_rate	gamma	max_delta_step	booster
0	00:01:04	0.91	2	100	1.000 00	0.000 001	0	gbtree
1	00:01:05	0.89	4	100	0.276 47	0.000 004	1	gbtree
2	00:00:40	0.99	3	100	0.000 93	0.000 223	1	gbtree
3	00:00:45	0.91	4	100	0.629 90	0.000 007	0	gbtree
4	00:00:10	0.98	5	100	1.000 00	0.000 001	1	gbtree
5	00:00:10	1.00	5	100	1.000 00	0.000 001	1	gbtree
6	01:09:44	0.68	5	100	1.000 00	0.000 001	0	gbtree
7	00:13:40	0.84	2	100	0.000 01	0.795 198	0	gbtree
8	00:00:01	0.99	4	100	0.246 15	5.384 447	1	gbtree
9	00:05:31	0.98	5	100	1.000 00	0.000 001	0	gbtree
10	00:00:23	0.97	5	100	0.402 83	0.093 924	0	gbtree
11	00:00:25	0.97	5	100	1.000 00	0.000 001	0	gbtree
12	00:00:29	0.98	3	100	0.000 03	0.003 075	1	gbtree
13	00:00:28	0.92	4	100	1.000 00	0.000 001	1	gbtree
14	00:00:20	0.93	3	100	0.719 88	0.000 011	1	gbtree
15	00:00:40	0.77	3	100	1.000 00	0.000 001	1	gbtree
16	00:00:30	0.90	5	100	1.000 00	0.000 001	1	gbtree
17	00:00:11	1.00	2	100	0.126 15	0.010 789	0	gbtree
18	00:02:00	0.91	5	100	1.000 00	0.000 001	0	gbtree
19	00:04:30	0.74	5	100	1.000 00	0.000 001	1	gbtree
20	00:03:22	0.74	5	100	1.000 00	0.020 153	0	gbtree
21	00:00:08	0.89	4	100	0.819 10	0.000 001	1	gbtree
22	00:00:37	0.93	4	100	1.000 00	0.000 001	1	gbtree
23	00:00:16	0.95	5	100	1.000 00	0.000 461	0	gbtree
24	00:01:12	0.89	5	100	0.245 30	0.000 001	1	gbtree
25	00:00:49	0.89	5	100	1.000 00	0.000 017	1	gbtree
26	00:01:02	0.94	3	100	1.000 00	0.011 838	1	gbtree
27	00:00:13	0.98	4	100	0.198 42	2.829 842	0	gbtree
28	00:00:11	0.97	5	100	0.137 76	0.000 259	0	gbtree
29	00:03:57	0.77	5	100	1.000 00	0.000 001	1	gbtree
30	00:00:00	1.00	4	100	0.457 13	0.460 825	0	gbtree

Table B.5: TensorFlow Keras MLP duration, score and relevant parameters of each entity model

	duration	score	units	dropout_rate	batch_size	epochs
0	00:13:19	0.91	500	0.10	19	100
1	00:18:55	0.87	400	0.10	16	100
2	00:17:47	0.99	300	0.39	127	100
3	00:34:42	0.91	500	0.10	16	100
4	00:04:03	0.98	100	0.31	93	100
5	00:00:36	1.00	500	0.27	34	100
6	06:48:39	0.68	300	0.10	128	100
7	07:42:24	0.86	500	0.10	128	100
8	00:01:16	0.99	500	0.26	44	100
9	01:05:01	0.98	500	0.23	87	100
10	00:04:08	0.97	200	0.36	31	100
11	00:26:16	0.97	100	0.27	54	100
12	00:49:03	0.98	500	0.10	16	100
13	00:15:37	0.92	500	0.17	86	100
14	00:11:54	0.92	200	0.10	16	100
15	00:23:54	0.76	300	0.10	16	100
16	00:10:29	0.92	400	0.32	34	100
17	00:03:02	1.00	400	0.32	52	100
18	00:19:19	0.89	400	0.35	38	100
19	01:00:53	0.74	400	0.10	16	100
20	01:27:00	0.73	500	0.10	16	100
21	00:01:54	0.88	200	0.45	128	100
22	00:23:58	0.93	300	0.21	21	100
23	00:05:07	0.93	200	0.10	16	100
24	00:15:51	0.90	300	0.10	16	100
25	00:20:39	0.88	300	0.10	127	100
26	00:11:58	0.94	300	0.10	16	100
27	00:06:17	0.98	500	0.11	20	100
28	00:04:38	0.97	400	0.10	128	100
29	00:46:18	0.76	300	0.10	16	100
30	00:00:14	1.00	300	0.25	97	100

Appendix C

Source Code

The project source code and the agent itself is publicly accessible in the following Git repositories:

CogA Cognitive and affective library developed and used in SensAI.

`https://gitlab.com/nunoachenriques/coga`

The Empathetic agent (EMPA)[129] package delivers affective, natural language and other useful libraries to support an artificial agent towards empathetic interaction. Plug-ins developed and included:

[129] `https://gitlab.com/nunoachenriques/coga/tree/master/empa`

1. Apertium Language Translator EMPA Plug-in
   `https://gitlab.com/nunoachenriques/coga/tree/master/empa-plugin-apertiumlanguagetranslator`
2. Optimaize Language Detector EMPA Plug-in
   `https://gitlab.com/nunoachenriques/coga/tree/master/empa-plugin-optimaizelanguagedetector`
3. VADER Sentiment Analysis EMPA Plug-in
   `https://gitlab.com/nunoachenriques/coga/tree/master/empa-plugin-vadersentimentanalysis`

SensAI The mobile agent as an Android™ application.

`https://gitlab.com/nunoachenriques/sensei`

Includes third-party free software libraries for Android™ (Java) besides the ones already integrated by CogA plug-ins:

- `MPAndroidChart`
  "A powerful Android chart view / graph view library, supporting line- bar- pie- radar- bubble- and candlestick charts as well as scaling, dragging and animations."
  `https://github.com/PhilJay/MPAndroidChart`
- `Twitter4J`
  "Twitter4J is an open-sourced, mavenized and Google App

Engine safe Java library for the Twitter API which is released under the Apache License 2.0."
`https://github.com/Twitter4J/Twitter4J`

- `Twitter Kit for Android`
  "Twitter Kit is a multi-module gradle project containing several Twitter SDKs including TweetComposer, TwitterCore, and TweetUi. Twitter Kit is designed to make interacting with Twitter seamless and efficient."
  `https://github.com/twitter-archive/twitter-kit-android`

SensAI Expanse The agent's expanded cognition and memory resources distributed in the Cloud.
`https://gitlab.com/nunoachenriques/sensei-expanse`

The Expanse includes third-party free software libraries and frameworks as Python packages. Below is a description of the ones required for a timely and accurate development towards proper data science including AutoML, efficient energy use and also easy explainable by feature importance inspection.

- `bokeh`
  "Bokeh is an interactive visualization library for modern web browsers. It provides elegant, concise construction of versatile graphics, and affords high-performance interactivity over large or streaming datasets. Bokeh can help anyone who would like to quickly and easily make interactive plots, dashboards, and data applications."
  `https://bokeh.org/`

- `hdbscan`[130]
  "The hdbscan library is a suite of tools to use unsupervised learning to find clusters, or dense regions, of a dataset. The primary algorithm is HDBSCAN* as proposed by Campello, Moulavi, and Sander. The library provides a high performance implementation of this algorithm, along with tools for analysing the resulting clustering."
  `https://hdbscan.readthedocs.io/`

130 McInnes et al. (2017)

- `matplotlib`[131]
  "Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms."
  `https://matplotlib.org/`

131 Hunter (2007)
`v3.1.1 10.5281/zenodo.3264781`

- `numpy`[132]
  "NumPy is the fundamental package for scientific computing with Python. It contains among other things: (a) a powerful N-dimensional array object; (b) sophisticated (broadcasting)

132 van der Walt et al. (2011)

functions; (c) tools for integrating C/C++ and Fortran code; (d) useful linear algebra, Fourier transform, and random number capabilities"
`https://numpy.org/`

- `pandas`[133]
"pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language."
`https://pandas.pydata.org/`

- `psycopg2`
"Psycopg is the most popular PostgreSQL database adapter for the Python programming language. Its main features are the complete implementation of the Python DB API 2.0 specification and the thread safety (several threads can share the same connection). It was designed for heavily multi-threaded applications that create and destroy lots of cursors and make a large number of concurrent INSERT or UPDATE."
`https://github.com/psycopg/psycopg2`

- `scikit-learn`[134]
"Machine Learning in Python: (a) Simple and efficient tools for data mining and data analysis; (b) Accessible to everybody, and reusable in various contexts; (c) Built on NumPy, SciPy, and matplotlib; (d) Open source, commercially usable - BSD license"
`https://scikit-learn.org/`

- `scikit-optimize`[135]
"Scikit-Optimize, or skopt, is a simple and efficient library to minimize (very) expensive and noisy black-box functions. It implements several methods for sequential model-based optimization. skopt is reusable in many contexts and accessible."
`https://scikit-optimize.github.io/`

- `scipy`[136]
"The SciPy library is one of the core packages that make up the SciPy stack. It provides many user-friendly and efficient numerical routines such as routines for numerical integration, interpolation, optimization, linear algebra and statistics."
`https://scipy.org/`

- `shap`[137]
"SHAP (SHapley Additive exPlanations) is a unified approach to explain the output of any machine learning model. SHAP connects game theory with local explanations, uniting several previous methods [...] and representing the only possible consistent and locally accurate additive feature attribution

133 McKinney (2010)

134 Buitinck et al. (2013) and Pedregosa et al. (2011)

135 `v0.5.2 10.5281/zenodo.1207017`

136 Jones et al. (2001)

137 Lundberg and Lee (2017) and Lundberg et al. (2019)

method based on expectations [...]”
`https://github.com/slundberg/shap`

- `tensorflow`
  “TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications.”
  `https://www.tensorflow.org/`
- `xgboost`
  “XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solve many data science problems in a fast and accurate way. The same code runs on major distributed environment (Hadoop, SGE, MPI) and can solve problems beyond billions of examples.”
  `https://xgboost.readthedocs.io/`

VADER Sentiment Analysis in Java Implementation of VADER algorithm in Java. Started as a fork of the Java port[138] by Animesh Pandey of the NLTK VADER sentiment analysis module[139] written in Python from the original project[140] by the authors Hutto and Gilbert (2014). It is the same algorithm as an improved tool by extensive rewriting with relevant adaptations such as Android™ ready.
`https://github.com/nunoachenriques/vader-sentiment-analysis`

[138] `https://github.com/apanimesh061/VaderSentimentJava`
[139] `https://www.nltk.org/_modules/nltk/sentiment/vader.html`
[140] `https://github.com/cjhutto/vaderSentiment`