Behind the Wizard’s Curtain

Designing and Developing Intelligent Systems for use in Educational Contexts.

by

Graham Parsonage

A thesis submitted in partial fulfilment for the requirements for the degree of Doctor of Philosophy at the University of Central Lancashire.

March 2024

Research Student Declaration Form

Type of Award: Doctor of Philosophy

School: Engineering and Computing

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution

I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work

No proof-reading service was used in the compilation of this thesis.

Signature of Candidate:

Print name: Graham Parsonage

Abstract

This thesis concerns the design and development of intelligent systems for use in educational contexts. The work presented took part either side of the Covid-19 lockdowns of 2020 and 2021 which profoundly affected its direction. The earlier chapters, two to four, describe research conducted prior to 2021 and consider system design from the perspective of the system stakeholders and how interface choices may impact on stakeholders' perceptions of a system's capabilities.

The latter part of the thesis, chapter six onwards, presents work conducted after the Covid-19 hiatus and is motivated largely by personal experience teaching remotely using video platforms such as MS Teams or Zoom. Stakeholders' trust and acceptance of the outputs from intelligent systems are a common theme throughout the work. Chapter 5 reviews the literature between 2019 and 2022 spanning either side of the Covid period.

Chapter 1 provides an introduction to the thesis outlining the approach taken, the research aims and objectives and the research contribution.

Chapter 2 provides background to the main concepts presented and discusses the field of Child-Computer Interaction (CCI) with a focus on its development as a discrete research discipline distinct from Human-Computer Interaction (HCI). The chapter also highlights some of the challenges faced when conducting research with children including ethical considerations.

It then presents an overview of Artificial Intelligence and some of its applications, followed by a brief history. It discusses Machine Learning based approaches that serve as support for the supervised learning implementations described in chapters 6, 7, and 8. Some of the building blocks of artificial neural networks, including feed-forward networks, backward propagation, and activation functions are also introduced. These ideas are further developed in Chapter 8 which describes an implementation that develops these concepts.

The chapter concludes by looking at similar work currently being con-

ducted in the field and notes that while other researchers are working on systems to automatically recognise engagement. The work described in the next chapters differs in its scope, intended target audience and methodology.

Chapter 3 considers the deployment of an intelligent system in an educational context that monitors children's behaviour during interaction with a computer or other digital technology and potentially makes an intervention if it identifies activity that may not be in the child's best interest. A model is proposed to inform the design of such a system based on the relationship between trust and acceptance. The Trust Acceptance Mapping Model (TAMM) is presented as a tool to indicate the likely success of the intelligent system design.

Chapter 4 explores how design choices regarding an IS's interface may affect both acceptance of its outputs and perceptions of its capabilities. Two studies are presented both of which introduce children to a Poppy Humanoid Robot. The first study examines how anthropomorphising the system may impact children's acceptance of its outputs. The children participating in the study perceived that a robot is able to learn while a computer is a rule based technology designed to perform well defined tasks.

In the second study the researcher introduces the Poppy robot in either "humanised" or "robot" form. In humanised form, the robot is referred to as she or Poppy and the children are asked to suggest things Poppy can learn to do. In robotised form, the robot is referred to as it or the robot and the children are asked to identify tasks it can be programmed to complete. The study finds that when the robot was introduced in humanised form, the children were more likely to attribute actions requiring learning or intelligence to it. When the robot was introduced in robot form, the children are more likely to attribute physical activities to it.

Chapter 5 presents a semi-systematic mapping review of the literature on HCI and CCI research related to AI. The terms HCI-AI and CCI-AI are used to describe the intersection between the disciplines. The AI taxonomy developed by AI Watch, the European Union's service "to monitor the development, uptake and impact of Artificial Intelligence", is used to classify and map the literature (Samoili et al., 2020).

In reviewing the literature, three approaches are adopted. Natural Language Processing (NLP) is used to perform semantic labelling of the research. The papers are classified by the researcher using the AI domain and subdomains described in the taxonomy. Finally, the research methods employed to produce the research are classified using the same AI taxonomy.

Chapter 6 presents PDLS, a peer observation approach to generate a labelled data set suitable for use in CCI research. The system is evaluated against the usability metrics of, effectiveness, efficiency, and satisfaction and is judged to be both efficient and satisfactory. Validation of its effectiveness is presented in Chapter 7. The CCI principle of Child Participation is central to the PDLS process, which generates labelled data in both a time and cost effective manner. Pupils were surveyed for their feelings on the accuracy of both their own and their peers' judgments on engagement status after completing the task and expressed their confidence in both these aspects.

It concludes by offering some thoughts that are intended to be helpful to other researchers who may wish to carry out similar studies and proposes the development of a data set that can be used as a resource for members of the CCI community who wish to undertake CCI research on emotion recognition or the application of computer vision to research with children.

Chapter 7 uses two methods to evaluate the accuracy or effectiveness of the PDLS. The first method uses the iMotions software to retrospectively analyse the video data generated. The second method employs expert reviewers to watch the videos captured by the pupils in the PDLS study and record engagement statuses independently of the original decisions.

Where there is an agreement between one or both of the reviewers and the observers original judgment, then the pupil observer's label is considered accurate. Where there is disagreement, then this is reviewed by the author with the goal of establishing the reasons for the inconsistency. The chapter concludes by discussing the strengths and weaknesses of the system and makes recommendations for its development and improvement.

Chapter 8 provides an overview of a Machine Learning based approach to implementing an engagement classifier for use with children in an educational context. The model described is a variant of a Recurrent Neural Network (RNN) called the Long Short-Term Memory (LSTM) Model and is selected for its ability to process sequences or cycles in the data. The output from the model is a binary classification which characterises the engagement level of the pupil completing the task as either engaged (1) or disengaged (0) and writes the classification to a video output.

In presenting the model the author acknowledges its limitations and it does not represent a production model but rather demonstrates the feasibility of the approach. Although the implementation displays the engagement classification to the video, this is not intended as a preference over the other potential interfaces considered in Chapter 4. As such, the ML model which provides the engine for the implementation of this IS could support multiple embodiments of the system.

Acknowledgements

To my wife and family for all their support and patience. We did it!

To all the people who gave me support and guidance:

My supervisory team, Professor Janet Read and Dr Matthew Horton

The research degree team at UClan, Dr Dan Fitton and Dr Gavin Sim

Sean Sturley at UWS who found me some space to work.

To the staff and pupils at the schools I worked at. It's no exaggeration to say I couldn't have completed it without you

Dr Simon Uttley and the staff and pupils at Blessed Hugh Farringdon Catholic School

Paul Edge and the staff and pupils at Ribblesdale High School

The staff and pupils at St Michaels C of E Kirkham

Special thanks to Steven Kay at Ribblesdale for all your help with the video studies.

List of Abbreviations

Table 1: List of Abbreviations

Abbreviation	Description
AI	Artificial Intelligence
ANN	Artificial Neural Network
AU	Action Unit
CCI	Child Computer Interaction
CCI-AI	Child Centred AI Research
CNN	Convolutional Neural Network
CRI	Child-Robot interaction
DL	Deep Learning
FACS	Facial Action Coding System
HCI	Human Computer Interaction
HCI-AI	Human Centred AI Research
HRI	Human-Robot Interaction
IDC	Conference on Interaction Design and Children
IS	Intelligent System
IxD	Interaction Design
LTSM	Long Term Short Memory
ML	Machine Learning
NLP	Natural Language Processing
RNN	Recurrent Neural Network
SNN	Spiking Neural Network
TAF	Trust Acceptance Framework
TAM	Technology Acceptance Model
TAMM	Trust Acceptance Mapping Model
UML	The Unified Modelling Language

Contents

List of Figures

List of Tables

Listings

Chapter 1

Introduction

1.1 Background and Context

I have been involved with children and education for more or less my entire life. Initially as a child, then as a parent, and since 2009 as an educator. During that period, which spans half a century, much has changed. My own education took place without any access to computers, in fact I did not use a computer until after I had finished my university education and entered the workplace. At school, the most complex technology we had access to was a digital calculator, and that was considered new-fangled. Education took place primarily in the classroom and was delivered by a teacher. If we were required to research a topic, then we were directed to a book.

My children's education took place over the period between the late 1990s and the early 2020s and was very different from my own. The Internet arrived, bringing with it opportunities and access to resources and materials at the press of a key that had not been available to previous generations. From the late 2000s with the advent of Massive Open Online Courses (MOOCs), education began to move out of the classroom and into online spaces (Palacios Hidalgo et al., 2020). In 2020, towards the end of my youngest child's education, the Covid-19 pandemic accelerated that process.

In the UK, the lockdowns of 2020 and 2021 required a change in how education was delivered. The forced closure of schools, colleges and universities fast forwarded the adoption of video conferencing platforms such as Zoom and MS Teams as the primary vehicle for teaching and learning (Tandon, 2021*a*). Education was forced out of the classroom and into distributed environments that presented both teachers and learners with a different set of challenges.

One of the key problems facing teachers was monitoring pupil engagement, particularly when a child either did not have access to a webcam or did not have it turned on. In a classroom, teachers routinely monitor their pupils for signs of disengagement, and the removal of the visual cues, that may prompt an intervention, affects the application of usual classroom practises and changes the classroom dynamic.

This research project straddled the Covid-19 lockdowns that profoundly influenced the output. Early work was carried out before the pandemic and was mainly concerned with the design of intelligent systems that would adapt their behaviour depending on the context and nature of an interaction with the child. Post-Covid-19 work was driven by my own experiences during lockdown delivering classes over MS Teams where often I found myself looking at tiles on a screen with no indication as to whether or not there was a student present behind the icon. Post-lockdown work focused on the design of intelligent systems to monitor children for signs of disengagement within an educational context.

The use of the term educational context as opposed to classroom is intentional. While mainstream school age education in the UK has largely returned to the classroom, formal learning still occurs in a wider context, and in such cases there may be a requirement for technology to support both the teacher and the learner to maximise the learning opportunity.

1.2 Thesis Structure

The structure outlined below covers the Pre/Post-Covid period described above, Chapter 2 is provided as background to both periods and Chapter 5 reviews the literature from 2019 - 2022 inclusive. Chapters 3 and 4 describe work conducted before the first lockdown in 2020 with the remainder of the work described conducted after the second lockdown ended in 2021.

Chapter 1

provides an introduction to the thesis outlining the approach taken, the research aims and objectives and the research contribution.

Chapter 2

provides background to the main concepts presented in this thesis. It discusses the field of Child-Computer Interaction (CCI) with a focus on the discipline's development as a discrete research discipline distinct from Human-Computer Interaction (HCI) and highlights some of the challenges faced when conducting research with children including ethical considerations.

It then presents an overview of Artificial Intelligence and some of its applications, followed by a brief history. It discusses Machine Learning based approaches that serve as support for the supervised learning implementations described in chapters 6, 7, and 8. It introduces some of the building blocks of artificial neural networks, including feed-forward networks, backward propagation, and activation functions. These ideas are further developed in Chapter 8 which describes an implementation that develops these concepts.

The chapter concludes by looking at similar work currently being conducted in the field and note that while other researchers are working on systems to automatically recognise engagement. The work described in the next chapters differs in its scope, intended target audience and methodology.

Chapter 3

considers the deployment of an intelligent system in an educational context that monitors children's behaviour during interaction with a computer or other digital technology and potentially makes an intervention if it identifies activity that may not be in the child's best interest. A model is proposed to inform the design of such a system based on the relationship between trust and acceptance. The Trust Acceptance Mapping Model (TAMM) is presented as a tool to indicate the likely success of the intelligent system design.

Chapter 4

explores how design choices regarding an IS's interface may affect both acceptance of its outputs and perceptions of its capabilities. Two studies are presented both of which introduce children to a Poppy Humanoid Robot. The first study examines how anthropomorphising the system may impact children's acceptance of its outputs. The children participating in the study perceived that a robot is able to learn while a computer is a rule based technology designed to perform well defined tasks.

In the second study the researcher introduces the Poppy robot in either "humanised" or "robot" form. In humanised form, the robot is referred to as she or Poppy and the children are asked to suggest things Poppy can learn to do. In robotised form, the robot is referred to as it or the robot and the children are asked to identify tasks it can be programmed to complete. The study finds that when the robot was introduced in humanised form, the children were more likely to attribute actions requiring learning or intelligence to it. When the robot was introduced in robot form, the children are more likely to attribute physical activities to it.

Chapter 5

presents a semi-systematic mapping review of the literature on HCI and CCI research related to AI. The terms HCI-AI and CCI-AI are used to describe the intersection between the disciplines. The AI taxonomy developed by AI Watch, the European Union's service "to monitor the development, uptake and impact of Artificial Intelligence", is used to classify and map the literature (Samoili et al., 2020).

In reviewing the literature, three approaches are adopted. Natural Language Processing (NLP) is used to perform semantic labelling of the research. The papers are classified by the researcher using the AI domain and subdomains described in the taxonomy. Finally, the research methods employed to produce the research are classified using the same AI taxonomy.

Chapter 6

presents PDLS, a peer observation approach to generate a labelled data set suitable for use in CCI research. The system is evaluated against the usability metrics, effectiveness, efficiency, and satisfaction and is judged to be both efficient and satisfactory. Validation of its effectiveness is presented in Chapter 7. The CCI principle of Child Participation is central to the PDLS process, which generates labelled data in both a time and cost effective manner. Pupils were surveyed for their feelings on the accuracy of both their own and their peers' judgments on engagement status after completing the task and expressed their confidence in both these aspects.

It concludes by offering some thoughts that are intended to be helpful to other researchers who may wish to carry out similar studies and proposes the development of a data set that can be used as a resource for members of the CCI community who wish to undertake CCI research on emotion recognition or the application of computer vision to research with children.

Chapter 7

uses two methods to evaluate the accuracy or effectiveness of the PDLS. The first method uses the iMotions software to retrospectively analyse the video data generated. The second method employs expert reviewers to watch the videos captured by the pupils in the PDLS study and record engagement statuses independently of the original decisions.

Where there is an agreement between one or both of the reviewers and the observers original judgment, then the pupil observer's label is considered accurate. Where there is disagreement, then this is reviewed by the author with the goal of establishing the reasons for the inconsistency. The chapter concludes by discussing the strengths and weaknesses of the system and makes recommendations for its development and improvement.

Chapter 8

provides an overview of a Machine Learning based approach to implementing an engagement classifier for use with children in an educational context. The model described is a variant of a Recurrent Neural Network (RNN) called the Long Short-Term Memory (LSTM) Model and is selected for its ability to process sequences or cycles in the data. The output from the model is a binary classification which characterises the engagement level of the pupil completing the task as either engaged (1) or disengaged (0) and writes the classification to a video output.

In presenting the model the author acknowledges its limitations and it does not represent a production model but rather demonstrates the feasibility of the approach. Although the implementation displays the engagement classification to the video, this is not intended as a preference over the other potential interfaces considered in Chapter 4. As such, the ML model which provides the engine for the implementation of this IS could support multiple embodiments of the system.

1.3 Research Methods and Approach

In their paper Research Contributions in Human-Computer Interaction, Wobbrock and Kientz (2016) identify seven main types of research contribution that are reproduced below and are useful in categorising the research methods and outputs from this thesis.

1. Empirical contributions. Data (qualitative or quantitative) collected through any of the methods described in this book: experimental design, surveys, focus groups, time diaries, sensors and other automated means, ethnography, and other methods.
2. Artifact contributions. The design and development of new artifacts, including interfaces, toolkits, and architectures, mock-ups, and “envisionments.” These artifacts, are often accompanied by empirical data about feedback or usage. This type of contribution is often known as HCI systems research, HCI interaction techniques, or HCI design prototypes.
3. Methodological contributions. New approaches that influence processes in research or practice, such as a new method, new application of a method, modification of a method, or a new metric or instrument for measurement.
4. Theoretical contributions. Concepts and models which are vehicles for thought, which may be predictive or descriptive, such as a framework, a design space, or a conceptual model.
5. Dataset contributions. A contribution which provides a corpus for the benefit of the research community, including a repository, benchmark tasks, and actual data.
6. Survey contributions. A review and synthesis of work done in a specific area, to help identify trends and specific topics that need more work. This type of contribution can only occur after research in a certain area has existed for a few years so that there is sufficient work to analyze.

7. Opinion contributions. Writings which seek to persuade the readers to change their minds, often utilizing portions of the other contributions listed above, not simply to inform, but to persuade.

This thesis draws on a number of these approaches. In common with much HCI and by extension CCI research, the work described in Chapter 3 is essentially empirical and draws on both quantitative and qualitative methods. Quantitative research is commonly used deductively to test a theory by testing the relationship between independent and dependent variables. Qualitative research methods are applied to the non-numeric data gathered and can often provide an insight into the relationships discovered using quantitative methods. The main vehicles for the research described are surveys and interviews. In designing and conducting empirical research, the author found the guidance in MacKenzie (2012)'s book, Human-Computer Interaction: An Empirical Research Perspective, an invaluable asset. Chapter 3 also generates a theoretical contribution, The Trust Acceptance Mapping Model.

Chapter 4 also employed qualitative techniques and to some extent Grounded Theory (Glaser and Strauss, 1967) where patterns are identified from within the data and then integrated into a theoretical framework.

Chapter 5 follows the guidance for conducting systematic and semi-systematic literature reviews contained in Booth et al. (2022)'s Systematic Approaches to a Successful Literature Review.

Chapters 6 and 7 are both methodological in describing a framework for generating labelled data and also present a system designed to make a dataset contribution.

Chapter 8 is best described as an artefact contribution in that in presents a ML model for classifying engagement in children in an educational context.

1.4 Research Aims and Objectives

The aim of this thesis is to explore the design and development of intelligent systems for deployment in an educational context. This section identifies the research objectives identified and questions considered when considering this task. For ease of reference the objectives and research questions considered are organised by chapter and presented below.

1.4.1 Chapter 3

Research Objectives

- RO1 Assess children's acceptance of the outputs from an intelligent systems.
- RO2 Assess adult's trust in the capabilities of an intelligent system to monitor and potentially make an intervention in a child's digital activity.
- RO3 Develop a model to analyse stakeholders' trust and acceptance in an intelligent system.

Research Questions

- RQ1 Are children more accepting of an intervention from a responsible adult than an intelligent system? (Chapter 3 Study 1).
- RQ2 To what extent do parents and caregivers trust technology to monitor their child's digital activity? (Chapter 3 Study 2).
- RQ3 To what extent do parents and caregivers trust technology to intervene in their child's digital activity? (Chapter 3 Study 2).
- RQ4 Would teachers trust an intelligent system to monitor pupils in their classroom for signs of disengagement and make appropriate interventions? (Chapter 3 Study 3).

1.4.2 Chapter 4

Research Objectives

RO4 Explore how the embodiment of a system affects users perceptions of its capabilities.

RO5 Establish whether the way in which an intelligent system is presented to children influences their view of its capabilities.

Research Questions

RQ5 How does anthropomorphising the system interface affect children's perceptions of its capabilities? (Chapter 4 Study 1).

RQ6 Does the way in which an intelligent system is presented to children influence their perceptions of its capabilities? (Chapter 4 Study 2).

1.4.3 Chapter 5

Research Objectives

RO6 To classify existing research in the areas of HCI and AI and CCI and AI using current taxonomies or definitions of the field of AI.

RO7 To classify existing research methods in the areas of HCI and AI and CCI and AI using current taxonomies or definitions of the field of AI.

RO8 To identify the main themes in the areas of HCI and AI and CCI and AI by evaluating the papers for semantic content.

Research Questions

RQ7 What are the prevalent research areas in the fields of HCI-AI and CCI-AI?

RQ8 What are the undeveloped research areas in the fields of HCI-AI and CCI-AI?

RQ9 What are the prevalent research methods in the fields of HCI-AI and CCI-AI?

RQ10 Are existing AI taxonomies sufficient to categorise research in the fields of HCI-AI and CCI-AI?

1.4.4 Chapter 6 and Chapter 7

Research Objectives

RO9 To develop a system for generating labelled video data reflecting children's engagement levels when completing a computerised task in an educational context.

RO10 To assess whether peer observation is an efficient, satisfactory and effective means of generating labelled data.

Research Questions

RQ11 Is peer observation an efficient method for generating labelled video data for use in identifying children's level of engagement with a computerised task in an educational context? (Chapter 6 Study 1).

RQ12 Is peer observation a satisfactory method for generating labelled video data for use in identifying children's level of engagement with a computerised task in an educational context?(Chapter 6 Study 2).

RQ13 Is peer observation an effective method for generating labelled video data for use in identifying children's level of engagement with a computerised task in an educational context? (Chapter 7.)

1.4.5 Chapter 8

Research Objectives

RO11 To Evaluate the feasibility of designing and developing an intelligent system for use in an educational context

Research Questions

RQ14 Can a machine learning based intelligent system be built to recognise and classify children's engagement within an educational context?

RQ15 What are the main challenges to building and deploying a ML based model to recognise and classify children's engagement within an educational context?

1.5 Research Contributions and Associated Publications

1.5.1 RC1 Conducting Research with Children

Chapter 2 provides a background to the challenges of conducting research with children and intelligent systems. Although this chapter does not present a study, much of the underpinning work was presented at the 4th International Conference on Human-Computer Interaction and User Experience in Indonesia, 2018. (Read et al., 2018).

1.5.2 RC2 Trust Acceptance Mapping Model (TAMM)

Chapter 3 presents the TAMM as a model to indicate the likely success of an intelligent system intended for deployment in a context where multiple groups of stakeholders may have diverse and potentially conflicting requirements. The model maps observed values for user acceptance and trust for stated use cases onto quadrants within a coordinate system referred to as the Trust Acceptance Framework (TAF). The top right quadrant can be regarded as the design goal where acceptance of system intervention and trust in the systems capability are both high. This work was presented at HCII 2023 (Parsonage et al., 2023*b*)

1.5.3 RC3 The Affect of Anthropomorphism on children's acceptance of the outputs of an intelligent system

Chapter 4 presents a study that assesses the affect of anthropomorphism on children's acceptance of the outputs from an intelligent system. The work concludes that children regard a humanoid robot as capable of learning whereas they regard a computer as a rule based system.

1.5.4 RC4 The Affect of Anthropomorphism on children's perceptions of the capabilities of an intelligent system

Chapter 4 presents a further study that assesses the affect of anthropomorphism on children's perceptions of the capabilities of an intelligent system. The way in which a Poppy Humanoid Robot is presented to the children is shown to affect the actions they perceive it will be able to perform. When the robot is humanised, the children are more likely to perceive that it is capable of learning. When it is presented as a robot that requires programming the children are more likely to envisage it performing physical activities. This work along with guidance for conducting experiments with children and robots was presented at IDC '20 (Parsonage et al., 2020).

1.5.5 RC5 A Semi-Systematic Mapping Review into AI-Centred HCI and AI-Centred CCI

Chapter 5 presents a semi-systematic review of research at the intersection between AI and HCI and AI and CCI. The review covers pertinent literature accepted for publication at the CHI 2019 - 2022 conferences and the IDC 2019 to 2022 conferences.

1.5.6 RC6 The Peer Data Labelling System (PDLS)

Chapter 6 and 7 present the PDLS, a novel and extensible approach to generating labelled data suitable for training supervised ML algorithms for use in CCI research and development. The novelty is in classifying one child's engagement using peer observation by another child, thus reducing the two stage process of detection and inference, common in emotion recognition, to a single phase. In doing so, this technique preserves context at the point of inference, reduces the time and cost of labelling data retrospectively and stays true to the CCI principle of keeping child-participation central to the design process. This work was accepted for publication at Interact 2023 (Parsonage et al., 2023*a*).

1.5.7 RC7 A Novel Machine Learning Based Approach to Classifying Engagement in Children in an Educational Context

Chapter 8 presents a ML based approach to identifying engagement in children in an education context.

1.6 Ethics

The research described in this thesis was approved by the University of Central Lancashire (UCLAN) STEMH Ethics Review Panel. DBS clearance was obtained in accordance with UK laws for working with children. Before the children were allowed to participate in the studies described in this thesis, full information was provided to the Head Teacher of the participating schools who gave their permission for the research studies to take place. Consent was then obtained from parents or carers.

The children participating in the studies were also informed of their purpose and given the option of whether or not to participate regardless of any prior consent from a responsible adult. Children were also informed of their right to withdraw their data after the studies were completed.

1.7 Summary

This chapter describes the motivation for this thesis and presents the objectives of the research and the questions it aims to answer. It sets out the research aims and contributions along with any associated publications. In doing so it aims to provide a roadmap to the approach taken to addressing at least some of the practicalities of creating machine learning based intelligent systems for use with children in an educational context.

Chapter 2

Children and Intelligent Systems - A Background

2.1 Introduction

This chapter is included for reference and provides a theoretical basis for subsequent chapters which discuss both children and intelligent systems. It gives an overview of the research disciplines HCI and CCI and introduces the field of AI and its applications, followed by a brief history. It summarises considerations for conducting research with children that are pertinent throughout the thesis and examines ML based approaches, which are relevant to the supervised learning implementations discussed in chapters 6, 7, and 8. The chapter introduces some of the basic components of artificial neural networks that are developed further in Chapter 8 which describes an implementation that uses them.

Within the field of computer vision, which is a core component of this research, supervised learning approaches tend to be less computationally expensive and more accurate than unsupervised approaches (Belgiu and Drăgut, 2014). They may also have a high training overhead and require reliable labelling of potentially considerable amounts of data (Bhavitha et al., 2017). Chapters 6, and 7 present a novel method for both generating and labelling data. The chapter concludes by briefly summarising the ongoing research at the intersection of CCI and AI. This is revisited in Chapter 5 which presents a semi-systematic review of research at the frontiers between HCI, CCI, and AI.

2.2 HCI, Interaction Design (IxD) and CCI

The Association for Computer Machinery (ACM) provides a working definition of HCI that includes the, "design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them" (Hewett et al., 1992). As Preece et al. (2019) remark, it concerns the interface between person and product rather than the implementation details, although the latter is clearly important.

From a computational perspective, the origins of HCI arguably have their roots in Sketchpad (Sutherland, 1964), an interface that allowed users to manipulate objects on a screen using a light pen. Sutherland's ideas, which formed the basis for the graphical interactive interface of the personal computer, were developed by Xerox PARC before being adopted by Macintosh and Microsoft (Myers, 1988) resulting in the windowed environments that will be familiar to anyone using a modern PC.

Clearly, HCI is not just about computers, but rather concerns the way in which humans interact with them. Modern HCI research champions multidisciplinary teams commonly including psychologists and educationalists as well as computer scientists. This approach extends to research on AI and HCI as noted by Antona et al. (2023) in their Special Issue on AI in HCI.

The relationship between HCI and IxD is not particularly well defined. For this thesis the explanation provided by Preece et al. (2019) is used, that is, IxD is essentially an extension of HCI and encompasses digital devices beyond the computer. The work described in these chapters involves interactions with technologies ranging from computers to robots and will use the terms HCI and IxD interchangeably. The different affordances offered by these technologies has resulted in diverse research groups such as Human-Robot Interaction (HRI) (Bartneck et al., 2020) and Chapter 4 considers children's interaction with robots.

It should also be noted that this thesis concerns the intersection between HCI, or more specifically CCI, and AI, an area in which Inkpen et al. (2019) note the HCI community has been quiet. The extent to which both the HCI and CCI communities have found their voice is explored in chapter 5. The terms HCI-AI and CCI-AI are used to describe AI research conducted through a HCI and CCI lens respectively.

CCI as a research discipline, related to but distinct from HCI, has its foundations in the 1990s, with the first Conference on Interaction Design and Children (IDC) taking place in the early 2000s. The distinction stems from the belief that children as a group are different from adults and should be treated differently. Read and Bekker (2011) identify the key differences between HCI and CCI as being the rate at which children change, the involvement of adults in children's interactions with technology, the context in which children use technology, and the underlying cultural and social assumptions about technology and children that determine what is good for children and what has value.

The rate at which children change is not a feature of this work, although there are changes in the approach to reflect different age groups. The participation of adults in the interaction of children with technology is discussed in Chapter 3 which deals with the design of software for multiple stakeholders. The context of children's use of technology is discussed in Chapter 6 which examines the role of context in the identification of child disengagement whilst the child interacts with a computerised task. Cultural assumptions are implicit in this work in that learning is considered as being inherently valuable.

2.3 Conducting Research with Children

As a starting point from which to consider how best to conduct research with children, it might be asked what constitutes a child? Perhaps surprisingly, according to United Kingdom law this is dependent on circumstance, starting from birth but with an upper limit ranging from 16 to 20 years (Thomson Reuter, 2023), dependent on the context in which the enquiry is made.

Most commonly the upper ceiling is applied on the 18th birthday (Gov UK, 2023) and the 18th birthday is also commonly used as the upper bound in CCI research. Lehnert et al. (2022) in their systematic review of the CCI field describe a child population ranging from 0 to 18 years old. Yarosh et al. (2011) noted in their own review of CCI that most of the studies within the IDC research community were conducted with children between 6 and 12 years old and noted the need to expand this base to account for the effect of age across the population.

Schapiro (1999) is clear in her distinction between childhood and adulthood, citing the paternalistic attitudes commonly exhibited toward children that would be entirely inappropriate if directed toward an adult. Children are then something more than miniature adults, and this should be reflected when conducting research that involves them. The concept that children have different requirements from adults is not new, with specialist children's literature authored as far back as the 17th century Tucker (2017).

The studies described in this thesis were conducted with children aged between 10 and 17 years spanning the last two years of the UK primary school system, the whole of the UK secondary school system, and the first year of UK Further Education system.

2.3.1 Child Development

Child Computer Interaction is greatly influenced by the work of three twentieth century thinkers; Jean Piaget, who developed the theory of constructivism, Seymour Papert who extended Piaget's ideas to develop constructionism, and the sociocultural theories of Lev Vygotsky. For a discussion of all three, see Ackermann (2001).

Central to Piaget's theory of constructivism is that children acquire or "construct" their own knowledge based on their own experiences. These knowledge structures make learning an individual process and children's learning is affected by their personal experiences rather than being something that is imparted to them or learnt by rote. Papert's constructionism extended Piaget's ideas arguing that children learn best when engaged in some form of collaborative creative process (Papert and Harel, 1991). The idea of child collaboration is a recurring theme in the CCI literature. Hourcade (2015) describes how inclusion in the design process allows children to become authors rather than consumers presenting the need to "deeply engage with stakeholders" as one of the "10 pillars of child-computer interaction".

Vygotsky and Cole (1978) argued that children construct their knowledge structures through social interaction. In what is now often described as a sociocultural approach to learning, children learn from their interactions with other people and tools (Cook and Cook, 2005) and knowledge is constructed socially rather than individually. Later developments in situativity theory (Greeno, 1998) emphasise the importance of context in learning and in particular how children interact with their environment. These ideas are discussed later in the thesis as the importance of context, as a vital component in identifying engagement in an educational situation, is developed.

2.3.2 Ethical Considerations

Ethics constitutes a fundamental part of any research study, but is particularly relevant to studies involving children who often do not have full control over the process. Despite this Van Mechelen et al. (2020) acknowledge that the CCI literature remains "underdeveloped in a number of areas including definition and theoretical basis, the reporting of formal ethical approval procedures and the extent to which design and particpation ethics is dealt with". In recent years, papers submitted to the IDC conference have included a mandatory section, Selection and Participation of Children, requiring authors to describe how children were selected, the consent processes followed, and how data sharing was communicated.

All studies described in this thesis were approved by the STEMH Ethics Review Panel of the University of Central Lancashire. Ethical approval was requested and granted in two phases, phase one granted in September 2017 covered the survey work, participatory studies and interviews whilst the second phase granted in May 2022 covered the video recording and processing of the data used to assess child engagement. Separate information packs for children, parents and schools were prepared describing the purpose of the study, the consent and withdrawal process, and protocols for data handling[1].

In line with the CHICI Group guidelines on best practise for conducting research with children (Read, 2023), child participants were informed before starting and at the end of the studies that they could choose whether to participate or withdraw their data from the study regardless of any prior consent. This included both prior personal consent and consent made on behalf of the child either by their parents or their school.

2.4 What are Intelligent Systems?

In section 2.3, an attempt was made to define what was meant by the description "child", at least in the context of this thesis. It would therefore seem appropriate to start this discussion on Intelligent Systems (IS) by attempting to define them. According to the Cambridge Dictionary, an "Intelligent System" is a "set of connected things or devices ... showing

[1] Documents pertaining to consent and data processing can be found at https://chici.org/studies/awa/

intelligence, or able to learn and understand” (Cambridge Dictionary, 2023).

This intelligence is often described as analogous to human behaviours (Shaw, 1998) and in the case of systems based on machine learning models (ML), the IS has the capacity to learn and adjust its behaviour accordingly. Professor John McCarthy sometimes referred to as the Father of Artificial Intelligence (AI) defined AI as,

> the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable (McCarthy, 2011).

Russell and Norvig (2020) break these behaviours into four approaches: acting humanly, thinking humanly, thinking rationally, and acting rationally. Acting humanly is the approach outlined by Turing (1950) in his seminal paper, The Imitation Game, that asks whether a computer could convince an interrogator that it was human. Thinking humanly, also referred to as the cognitive modelling approach, requires the computer to follow human cognitive processes and solve problems in the same manner as a human would. Thinking rationally is synonymous with approaches that apply the discipline of logic, and acting rationally assumes that the system will act to achieve the best outcome.

It is useful to note that AI-based systems do not always employ learning to reach their goal. Search algorithms such as breadth-first-search or uniform-cost-search are a case in point where look-forward techniques are employed to make informed decisions, but no learning takes place. In other words, unless some form of caching separate to the algorithm is employed, each run is carried out in isolation, and the process starts again. For many, IS have become synonymous with ML but this is not necessarily the case.

ML is a subset of AI (Kühl et al., 2020) where, rather than coding paths through data using traditional programming and logic constructs based on sequence, selection, and iteration, the model at the core of the program or

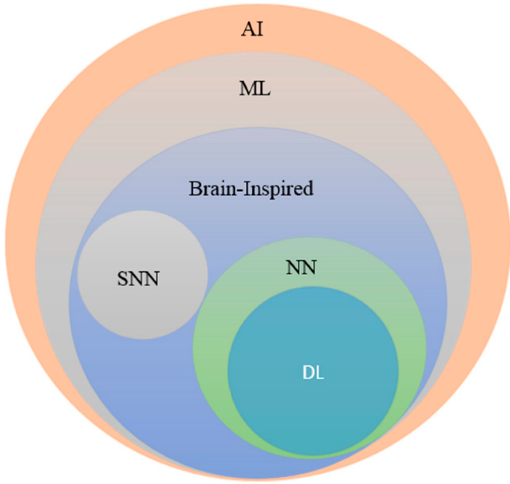


Figure 2.1: Taxonomy of AI (Alom et al., 2018)

system “learns” from the data upon which it is trained and then uses the “knowledge” of the domain to make predictions.

(Figure 2.1) shows this hierarchy with general AI techniques at the highest level and ML positioned within. ML based systems are further divided into those that employ brain-inspired cognitive techniques such as artificial neural networks (ANN), referred to as NN in Figure 2.1 and those that rely purely on statistical techniques such as regression. Spiking Neural Networks (SNN) are ANN that are designed more closely to mimic human cognitive processes (Ghosh-Dastidar and Adeli, 2009) whilst Deep Learning (DL) networks are ANN with one or more hidden inner layers. Examples of DL networks are Convolutional Neural Networks (CNN) which are employed in the engagement recognition system described in Chapter 8.

2.5 A Brief History

Russell and Norvig (2020) credit the beginnings of AI to the work of McCulloch and Pitts (1943) who applied propositional logic to neural events, Turing (1950)’s computational work and Hebb (1949)’s work on the links between neurons. The term AI was later coined by John McCarthy in

1956 at a workshop at Dartmouth College so at the time of writing AI research is eighty years old.

In 1957 Frank Rosenblatt proposed “The Perceptron” (Kanal, 2003) a single-layer neural network and in 1967 built the Mark 1 Perceptron, the first machine to be built around the perceptron algorithm. This early period of AI research showed great promise and led to inflated claims such as those made in 1957 by the economist and AI pioneer Herbert Simon that a machine could be built within ten years that would be a chess champion (Newell et al., 1957). It was actually forty years later that IBM’s Deep Blue Computer defeated Gary Kasparov, the then world chess champion (Miesel, 2011).

The reasons for this misplaced optimism were in part that the bounds of the tasks set for these early AI implementations were narrow or task-focused and did not scale well to larger problems. Before the development of the concept of NP-Completeness[2] researchers did not anticipate the intractable nature of the problems they were trying to solve as they increased in size. By the 1970s, the early optimism surrounding AI had been replaced by what became known as the First AI Winter, characterised by a period of reduced funding and activity. In the UK, the Lighthill Report on the state of AI research is considered a major factor in this period of relative inactivity (Agar, 2020).

The period covering the 1970s until the mid 1980s saw a move away from the previous general approach towards rule-based expert systems, which applied a large knowledge base to solve narrower problems. Funding for expert systems grew rapidly until the late 1980s when it became apparent that they were difficult to build and maintain and unsuitable for deployment in wider domains in part because of their inability to learn. The period from the late 1980s until the mid 1990s saw a reduction of funding for AI research and development and is generally referred to as the Second AI Winter.

[2]NP-Completeness incorporates the notion of hard problems, i.e., problems for which there is no known algorithmic solution with polynomial-time complexity (Nasar, 2016)

In the mid 1980s, with the application of backpropagation algorithms, ANN once again became a popular area of research increasing the range of AI applications. This period saw more research into machine learning and a change of emphasis from logic-driven expert systems towards systems based on probability and learning.

The advent of the Web, which helped vastly increase the volume of data available coupled with an increase in computational power, heralded the beginnings of the era of Big Data in the early 2000s. In 2011 IBM's Watson (Ferrucci, 2012) employed ML techniques to unstructured data to defeat two champions on the US game show Jeopardy. In 2016 DeepMind AlphaGo used deep neural networks to defeat a professional Go player. Go is known for its complexity which greatly exceeds that of chess (Bory, 2019).

At the time of writing, mid-2023, ChatGPT (Schulman et al., 2023) a machine learning model that uses reinforcement learning as a basis and features a conversational interface appears to be able to answer questions across multiple domains from global warming (Biswas, 2023) to education (Lo, 2023).

The Artificial Intelligence discussed in this thesis is largely centred upon machine learning techniques and the remainder of this discussion will reflect this.

2.6 Machine Learning

Typically, the ML algorithm employs either a statistical approach (Koller et al., 2007) such as regression or builds an artificial neural network (ANN) designed to mimic the workings of the human brain (Jain et al., 1996). The two approaches are not mutually exclusive, e.g. it is feasible although not necessarily desirable, to deploy an ANN to perform regression.

Where the ANN has multiple layers, it is called a deep learning network (DL) (Schmidhuber, 2015). Whatever approach the system designers adopt, the ML model requires a data source from which to “learn”.

This learning approach is generally used to classify ML techniques into two categories, supervised learning and unsupervised learning (Berry et al., 2019). Supervised learning techniques employ labelled data often called ground truth $\boldsymbol{T}$ where the value of the dependent variable is known. The ML model makes a prediction $(\hat{y})$ based on one or more independent variables $\boldsymbol{X}$ and compares $\hat{y}$ against the known value for the dependent variable (y).

Learning is conducted over a number of iterations or epochs with the goal of minimising the error or loss of the model, which is generally calculated as some form of the summation of the distance between $\hat{y}$ and y. Supervised learning approaches are further organised into two categories, regression (Montgomery et al., 2021) and classification (Singh et al., 2016). Regression models are used to predict a single continuous variable such as pass rate whilst classification models categorise data into two or more discrete classes such as engaged or disengaged.

Unsupervised learning is used to discover patterns in unlabelled data, usually resulting in the data being organised into groups (Celebi and Aydin, 2016). $\boldsymbol{T}$ does not exist within the data and the unsupervised model infers the patterns in the data without intervention (supervision).

Section 2.6.1 demonstrates the basis of the statistical ML approach using linear regression as an example.

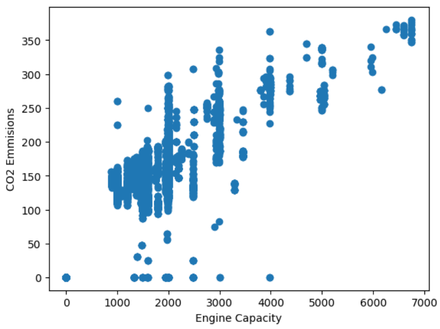
2.6.1 Supervised Learning - Regression

One of the simplest forms of prediction is linear regression, which builds a model to approximate the relationship between two or more variables. The following example uses data from the UK Vehicle Certification Agency[3] to predict CO2 emissions $(\hat{y})$ also known as the dependent variable from one or more variables $\boldsymbol{X}$ the independent variable. The data was selected as

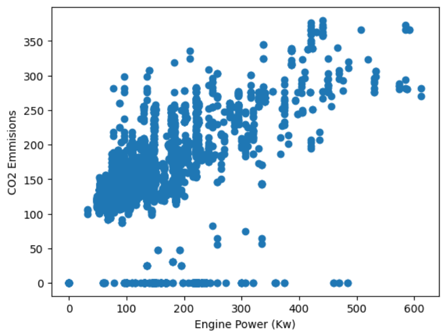
[3]https://carfueldata.vehicle-certification-agency.gov.uk/downloads/download.aspx?rg=latest

the pre-existing linear relationships between variables in the data make it suitable for illustrating linear regression.

Linear regression is only effective as a prediction method if linear relationships pre-exist in the data. A common method to establish this is to plot the data to establish a linear trend (Figure 2.2). In this case, CO2 emissions increase as both engine capacity and power rise, so linear regression is an acceptable approach.



(a) Engine Capacity vs CO2 Emissions



(b) Engine Power vs CO2 Emissions

Figure 2.2: Establishing an Existing Linear Relationship in the Data

Given the relationship between Engine Capacity $\boldsymbol{X}$ and the known values of CO2 emissions y, the aim of linear regression is to plot a line of best fit through the data that can be used to predict the values for CO2 emissions $\hat{y}$. The line of best fit minimises the loss or error between the known values for y and the predicted values $\hat{y}$. The line is calculated using the following equation:

$$y = mx + c \tag{2.1}$$

where m is the gradient and c is the y intercept. While this is useful for cases of simple linear regression where there are only two variables, for cases where there are more than one independent variable (multiple linear regression), it is common to rewrite the equation as:

$$\hat{y} = \theta_0 + \theta_1 x_1 \tag{2.2}$$

Where θ_0 is the intercept and θ_1 is the gradient of the fitting line. Multiple linear regression can then be expressed as:

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ...\theta_n x_n \tag{2.3}$$

or alternatively as a vector where $\boldsymbol{\theta^T}$ is the transpose vector:

$$\hat{y} = \boldsymbol{\theta^T X} \tag{2.4}$$

Given an example where CO2 emissions are predicted based on both Engine Capacity and Engine Power then:

$$\boldsymbol{\theta^T} = [\theta_0, \theta_1, \theta_2] \tag{2.5}$$

$$\boldsymbol{X} = \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} \tag{2.6}$$

where x_1 is the value of the engine capacity and x_2 is the value of the engine power. The first row of X is always assigned a value of 1 so the intercept or bias θ_0 remains unchanged when the vectors are multiplied. Note that rather than plotting a line of best fit as the data now has multiple dimensions, the line becomes a plane or hyperplane (Figure 2.3b).

There are a number of potential metrics used to minimise the loss of the model, and a commonly used method is the mean squared error (MSE), also known as the residual error. The mean of all residual errors shows how poorly the line fits the whole data set and can be calculated using the formula:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.7}$$

The aim is therefore to find values for θ_0 and θ_1 such that the MSE is minimised or, put another way, the model becomes optimised. The optimal gradient and intercept can be calculated directly from the dataset as follows:

$$\theta_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2} \tag{2.8}$$

$$\theta_0 = \bar{y} - \theta_1 \bar{x} \tag{2.9}$$

(Figure 2.3a) shows the fitted regression model for Engine Capacity and CO2 emissions produced by running the sklearn linear regression model[4]. A final useful metric is the R^2 score. R^2 is an indicator of how well the regression line accounts for the variation in the values of y. Scores range from 0 to 1 where a score of 0 means none of the values of y are on the line and a score of 1 that all the values of y are on the line. The relative squared error (RSE) is used to calculate the R^2 score:

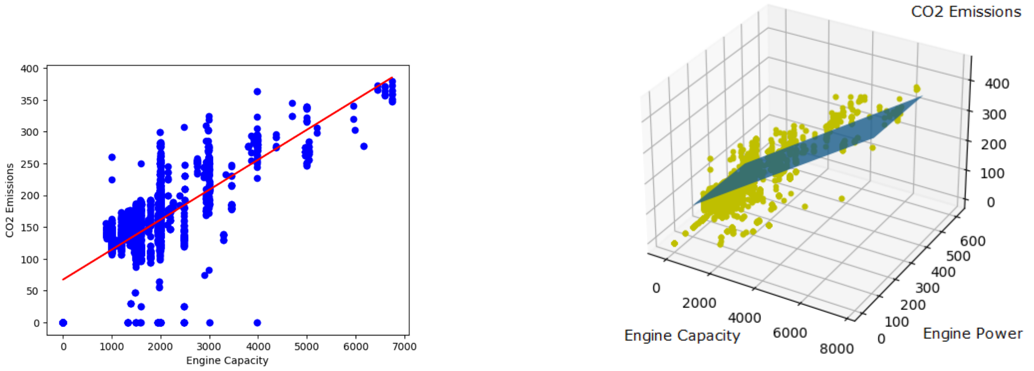
$$RSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{\sum_{i=1}^{n}(y_i - \bar{y}_i)} \tag{2.10}$$

$$R^2 = 1 - RSE \tag{2.11}$$

In the example given, the calculated R^2 is 0.6 indicating that 60% of the known values for y sit close to the line of best fit. Once trained, the model can then be used to predict values of CO2 emissions given an engine capacity. In this case, given an engine capacity of 2000, the regression model predicts CO2 emissions of 161.48.

Regression then uses one or more independent variables that may be continuous or discrete and predicts a single continuous dependent variable.

[4] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html



(a) Fitted Regression Model - Engine Capacity and CO2 Emissions

(b) Fitted Multiple Regression Model - Engine Capacity, Engine Power and CO2 Emissions

Figure 2.3: Simple and Multiple Regression Models

2.6.2 Supervised Learning - Classification

Whilst regression techniques predict a single continuous value, it is often the case that the ML model is required to classify the data. That is, predict to which class a case belongs, or alternatively, the probability that a case belongs to a given class. Chapter 6 describes a case in point where a child is observed and classified as either engaged or disengaged.

Classification is a supervised learning technique that categorises data into discrete classes. The model learns the relationship between a set of feature variables (the independent variable $\boldsymbol{X}$) and a target variable (the dependent variable y). The target variable is categorical with discrete values. There are a number of classification algorithms, e.g., k-nearest neighbours[5], here another form of regression called logistic regression is discussed.

[5] k-nearest neighbours employs labelled points to classify other labelled points based on their similarity to other cases

Logistic Regression

Logistic regression is similar to linear regression, but instead of trying to predict a continuous variable, it predicts a categorical variable. This categorical variable $\hat{y}$ may be binary (e.g., engaged or disengaged) or multiclass (e.g., highly engaged, engaged, disengaged, highly disengaged). Logistic regression predicts both the class of each case and the probability that a case belongs to a class.

For a binary classification where m are the features of $\boldsymbol{X}$ and n are the rows of $\boldsymbol{X}$:

$$\boldsymbol{X} \in \mathbb{R}^{mn} \tag{2.12}$$

$$y \in \{0, 1\} \tag{2.13}$$

$$\hat{y} = P(y = 1|x) \tag{2.14}$$

$$P(y = 0|x) = 1 - P(y = 1|x) \tag{2.15}$$

Logistic regression extends linear regression by using the sigmoid function as an activation function σ to convert the dependent variable returned by $\boldsymbol{\theta^T X}$ from a continuous value to a probability:

$$\sigma(\boldsymbol{\theta^T X}) = \frac{1}{1 + e^{-\theta^T X}} \tag{2.16}$$

$$\boldsymbol{\theta^T X} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ...\theta_n x_n \tag{2.17}$$

$$\sigma(\boldsymbol{\theta^T X}) = \sigma\theta_0 + \theta_1 x_1 + \theta_2 x_2 + ...\theta_n x_n \tag{2.18}$$

(Figure 2.4) shows how the sigmoid activation function σ converts the output from $\boldsymbol{\theta^T X}$ axis to a probability in the range 0 to 1.

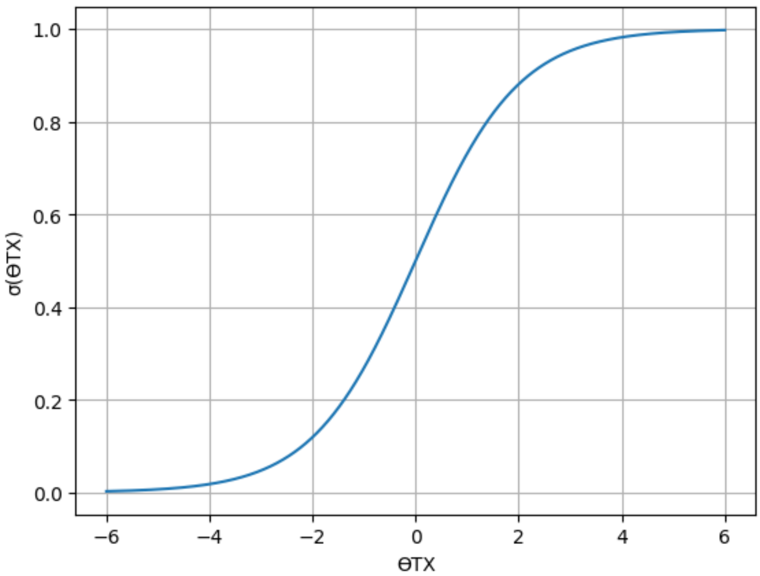


Figure 2.4: Sigmoid Activation Function

The question remains as to how to find the optimal value for θ. An approach for a binary classification may look something like this:

1. Initialise $\boldsymbol{\theta}$ with random values.
2. Calculate $\hat{y} = \sigma(\boldsymbol{\theta}^{\boldsymbol{T}}\boldsymbol{X})$. The output is the probability that the output belongs to the default class 1.
3. Compare $\hat{y}$ with the actual value y. The distance between $\hat{y}$ and y is the model error for this record, e.g., if the actual value is $y = 1$ and the predicted value is $\hat{y} = 0.6$, then the error for this record = 0.4.
4. Calculate the sum of the errors for all customer, $Cost = J(\boldsymbol{\theta})$
5. Adjust the value of $\boldsymbol{\theta}$ to minimise cost and repeat from step 2.

The error or cost for each record is therefore calculated and then summed to find the cost for the whole model, which is effectively the MSE.

$$Cost(\hat{y}, y) = \frac{1}{2}(\sigma(\boldsymbol{\theta^T X}) - y)^2 \tag{2.19}$$

$$J(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} Cost(\hat{y}, y) \tag{2.20}$$

The minimum point for the cost function indicates the best values for the model parameters, and at this point the model is said to be tuned.

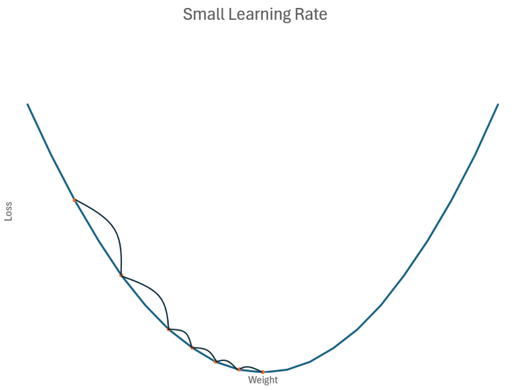
There are two general approaches to implementing this algorithm. One is mathematically using the derivative of the cost function. Finding the global minimum by this method is both complex and computationally expensive and it is more common to apply iterative algorithmic approaches such as gradient descent (Ruder, 2016). Such an approach is described below.

Model Optimisation - Gradient Descent

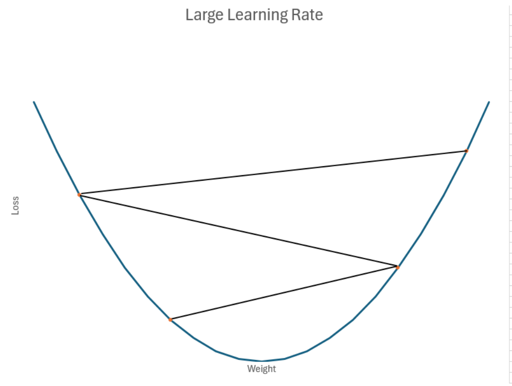
Gradient descent is analogous to moving down a slope or bowl (depending on the dimensionality of the model) until the bottom is reached. At this point, the cost is minimised and the weights for the model are optimal. The approach taken can be summarised as follows:

1. Select a random point on the slope.
2. Calculate the slope gradient at that point.
3. Move in the opposite direction to the slope to guarantee downward movement into the error curve.
4. The size of the gradient indicates the size of the step to take. A large gradient indicates the size of the step to take and indicates a greater distance from the minimum.

(Figure 2.5) demonstrates this approach and the importance of selecting a sensible value for the step (μ). If μ is too large, then the model may miss the optimal point, if it is too small, then the model may never converge, and the optimal point will not be found.



(a) Gradient Descent Small Learning Rate



(b) Gradient Descent Large Learning Rate

Figure 2.5: Gradient Descent Learning Rates

2.7 Artificial Neural Networks (ANN)

As discussed in Section 2.5 since the 1990s research into ANN has had a renaissance and applications which utilise this medium have greatly expanded. (Figure 2.6) shows the structure of the human neuron on which such systems are built.

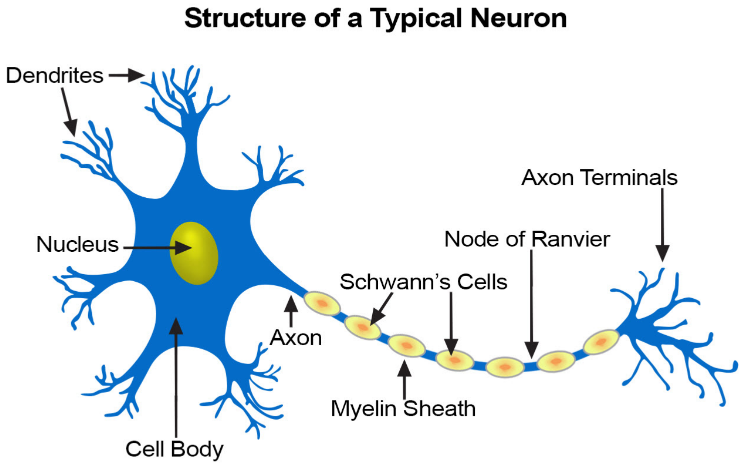


Figure 2.6: Anatomy of a Neuron - National Cancer Institute (2023)

The dendrites receive electrical impulses (data) from the terminals of other neurons and transit them to the cell body (soma), which is the area surrounding the nucleus. The nucleus processes the data and passes them to the axon, which carries the processed data to the terminal (synapses), where it becomes the input to other neurons. The brain learns by repeatedly activating certain neural connections that reinforce those connections.

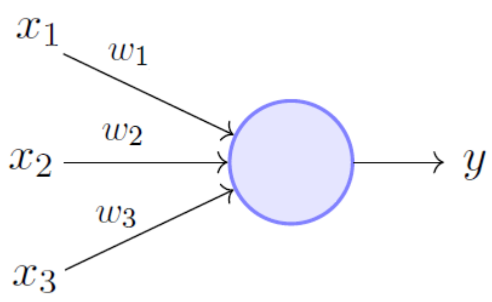


Figure 2.7: The Perceptron - Minsky and Papert (1969)

(Figure 2.7) illustrates the perceptron (see Section 2.5) an early ANN architecture that performed binary classification. The Perceptron takes a number of weighted binary inputs X and outputs a single binary output y. The weights w are applied to give relative importance to the input. The perceptron algorithm sets a threshold value which determines the value of y. If the summed weights of the inputs is greater than the threshold, it returns 1 otherwise it returns 0. Modern ANNs are not based on the perceptron, but it serves as a useful example to demonstrate the link between the anatomy of a human neuron (Figure 2.6) and the concept of an artificial neuron (Figure 2.7).

Data flows through an ANN using a process called forward propagation (Figure 2.8). z is the linear sum of input and weights with the bias applied. The bias is a constant that can be used to shift the output either positively or negatively. a is the output of the function and in the case of Figure 2.8 is the same as z, effectively a linear regression.

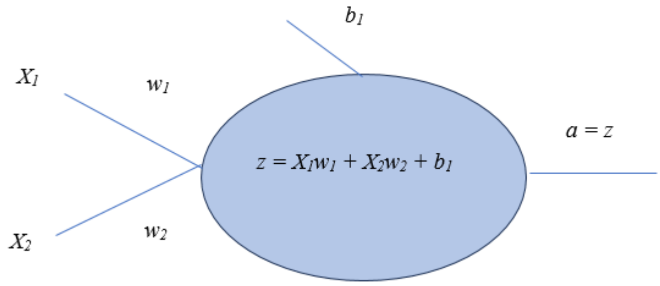


Figure 2.8: Forward Propagation through a network

Whereas linear regression always returns a continuous number; for classification problems where it is required to map the number into 2 or more discrete spaces, a method is required to map the linear value into a non-linear space. The answer is to apply an activation function such as the sigmoid function (Figure 2.9).

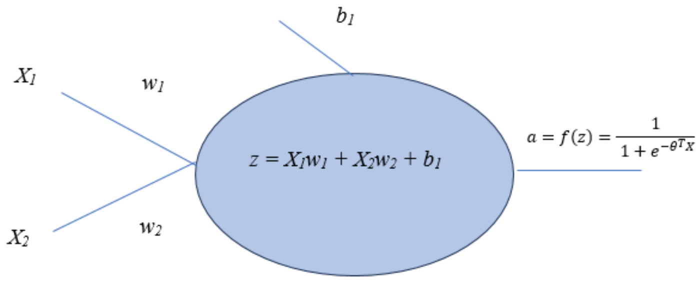


Figure 2.9: Forward Propagation using an activation function

(Figure 2.10) shows how the feed forward uses the activation function to convert the linear output from z into a non-linear value ranging between 0 and 1 giving a probability of the input belonging to a discrete class.

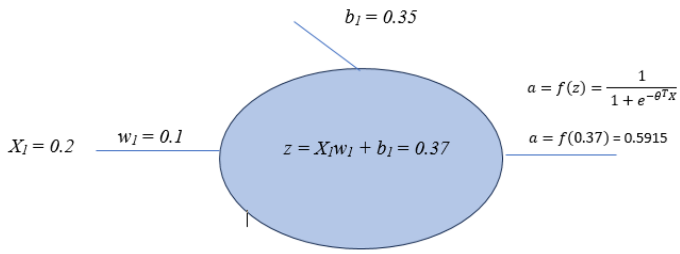


Figure 2.10: Forward Propagation Calculating the Probability

In cases where the ANN has more than one layer, the output of the first layer becomes the input of the next layer (Figure 2.11).

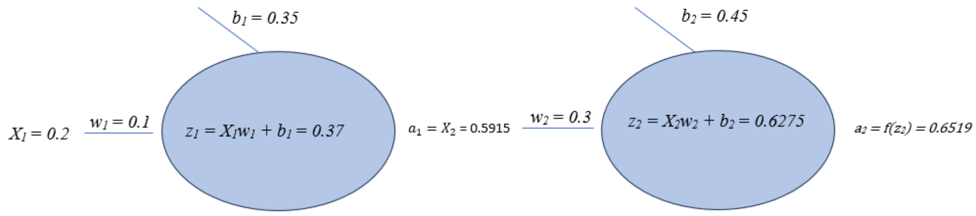
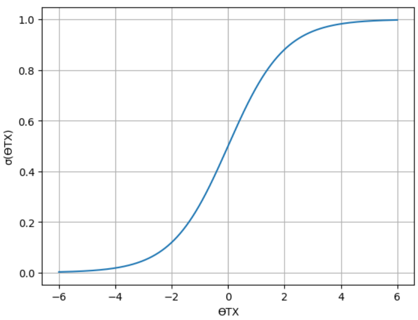


Figure 2.11: Forward Propagation with two layers

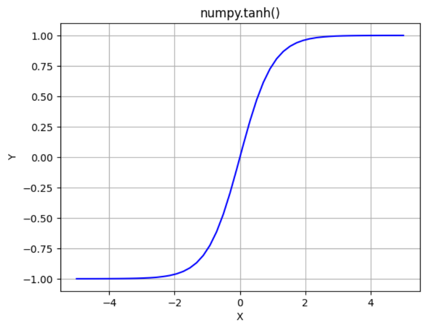
As discussed in Sections 2.6.1 and 2.6.2 the model is trained by optimising the weights and bias. This process is generally achieved by a method named backpropagation where the parameters are optimised from the back of the model to the front generally using the gradient-based algorithms described above.

2.7.1 Activation Functions

One of the problems that plagued the early implementations of ANN was the problem of vanishing gradients (Hochreiter, 1998). This was particularly true for deep networks employing backpropagation using the sigmoid function or derivatives such as tanh. One of the issues with the sigmoid function is it is not symmetrical around the origin, so the returns values are always positive, meaning that neurons will always fire creating a dense network with increased training times. Tanh addresses this by scaling the output from -1 to 1 and is symmetrical around the origin (Figure 2.12). It is however, still susceptible to the vanishing gradient problem.



(a) sigmoid - $a = \frac{1}{1+e^{-z}}$



(b) tanh $a = \frac{exp(z) - exp(-z)}{exp(z) + exp(-z)}$

Figure 2.12: Sigmoid and tanh Activation Functions

As training takes place, the inputs that are all between 0 and 1 are multiplied together becoming increasingly small, resulting in the gradients becoming smaller as progress is made backward into the network. As such, the neurons in the earlier layers of the network learn very slowly compared to the neurons in the later layers resulting in poor accuracy and longer training times. This is referred to as the vanishing gradient problem.

The sigmoid function has now been widely replaced by two activation functions, the rectified linear activation unit (ReLU) (Agarap, 2018) and Softmax (Bridle, 1990)). ReLU like sigmoid is non linear but does not activate for all neurons creating what is referred to as a sparse network. ReLU returns the value of the input or 0 if the input is negative (Figure 2.13). ReLU successfully overcomes the vanishing gradient problem.

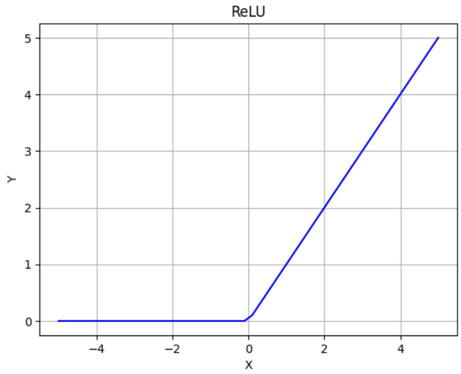


Figure 2.13: ReLU

Softmax is a form of sigmoid activation function:

$$a_i = \frac{e^{z_i}}{\sum_{k=1}^{n} e^{z_k}} \tag{2.21}$$

usually applied in the outer layer of a classification to calculate the probability that an output belongs to a given class.

2.8 Child Computer Interaction and Intelligent Systems

Having taken some time to introduce both the fields of Child Computer Interaction and ML based Inteligent Systems, it remains to discuss if and how they interact. Chapter 5 looks at the key trends at this intersection, this section reviews some of the work ongoing in the field.

One existing body of work focuses on Machine Learning education. In a recent systematic review of this field which considered ML education in mainstream American schools, Sanusi et al. (2023) identified just 43 conference papers and journal articles. As noted in Chapter 5, there is also a developing body of work on how children interact with social robots although much of this focusses on specific cases such as autism (Cabibihan et al., 2013), (Kim et al., 2013), ASD (Albo-Canals et al., 2018), (Vanderborght et al., 2012),or other healthcare contexts (Dawe et al., 2019).

Much of the work described in this thesis concerns deploying an IS to monitor children's engagement with a task. Huan (2017) identifies three methods for measuring engagement; self-reporting, observation and physiological methods. Self reporting requires the individual to assess their own engagement level and is not considered in this work as a feasible medium for generating the labelled data on which a system could be built.

Of more interest are approaches which combine one or both of the other two methods. Chapter 6 introduces an approach to identifying engagement in children using peer observation called the Peer Data Labelling System (PDLS). Hadfield et al. (2019) describe a deep learning approach to identifying engagement that has some commonality with the PDLS approach when children interact with a social robot. Their approach differs from the PDLS in that they employ multi-modal data rather than solely using computer vision algorithms and they focus on robot interaction whereas the focus here is on a broader range of tasks within an educational context.

There are similarities in that both Hadfield et al. (2019)'s approach and the PDLS apply an LSTM algorithm to process cycles in the data. However, they also apply three levels of classification, disengaged, partially engaged or fully engaged whereas the PDLS employs a binary classification, engaged or disengaged. Chapter 7 discusses some of the challenges in differentiating between these two states.

One of the key differences between an approach that requires multi-modal monitoring is the requirement for specialist sensors and monitoring equipment that increases its cost and reduces its scalability. Section 6.4.2 evaluates the efficiency of the PDLS judging it to be both cost effective and extensible.

Buono et al. (2023) have completed a study that uses facial analysis to identify engagement in undergraduate students conducting an online task. In common with the approach described here, they employ an LSTM implementation but they used the Facial Action Coding System (FACS) to label the data rather than the peer data labelling approach described in Chapters 6 and 7. FACS requires post-hoc processing and labelling of the data by experts whereas PDLS labels the data at source making it a cost effective alternative, particularly with the large volumes data common in computer vision tasks.

In a study with younger children, 5- 6 years, Yun et al. (2018) describe a CNN approach to recognising engagement using video data. Their approach differs from the one employed here in that they used experts to label the data rather than the peer approach described in Chapter 6. The efficiency of the PDLS is evaluated in Section 6.4.2 and it is argued that the PDLS is more cost effective way of labelling large volumes of data than post-hoc expert analysis.

There are then a small number of researchers working to apply machine learning techniques to develop intelligent systems and computer vision techniques for deployment with children. The work described here differs in that it extends the system design and development process to recognise the needs of the multiple stakeholders. It concentrates on developing systems suitable for use within mainstream secondary eduction and it introduces a novel approach to data generation and labelling that it suggests are key to the successful deployment of such systems.

2.9 Summary

This chapter provided background to the main concepts presented in this thesis. It discussed the field of Child-Computer Interaction (CCI) with a focus on the discipline's development as a discrete research discipline distinct from Human-Computer Interaction (HCI). It highlighted some of the challenges and ethical considerations faced when conducting research with children.

It then presented an overview and brief history of Artificial Intelligence and some of its applications. It discussed Machine Learning based approaches that serve as support for the supervised learning implementations described in Chapter 8 and the data the models are built on discussed in chapters 6 and 7. It goes on to introduce some of the building blocks of artificial neural networks, including feed-forward networks, backward propagation, and activation functions. These ideas are further developed in Chapter 8 which describes an implementation that develops these concepts.

The chapter concludes by looking at similar work currently being conducted in the field. It notes that while other researchers are working on systems to automatically recognise engagement, the work described in the next chapters differs in scope, target audience, and methodology .

Chapter 3

Designing Intelligent Systems

3.1 Introduction

This chapter considers the deployment of an intelligent system in an educational context that monitors children's behaviour during interaction with a computer or other digital technology and potentially makes an intervention if it identifies activity that may not be in the child's best interest. A model is proposed to inform the design of such a system based on the relationship between trust and acceptance. The Trust Acceptance Mapping Model (TAMM) is presented as a tool to indicate the likely success of the intelligent system design.

3.1.1 Stakeholders, Trust and Acceptance

Designing any system for children is likely to require satisfying the requirements of at least three groups of stakeholders (Hourcade, 2015):

- The learner (child)
- Parents or carers
- The education establishment (teachers)

For the system to be effective the child should be accepting of the systems outputs whilst parents and teachers need to trust the system to make effective and appropriate judgments. Trust (Holliday et al., 2016) and acceptance (Venkatesh, 2022) are core components in the successful adoption of most systems, but the potential for a stochastic intelligent system to change its output as it learns, potentially generating inconsistencies in its judgments, may make these goals harder to achieve (Glikson and Woolley, 2020).

Trust between humans is a complex and multifaceted concept supporting the belief that another will act with benevolence, integrity, predictability, and competence (Mcknight and Chervany, 2000). When evaluating or testing intelligent systems for trustworthiness, studies often identify competence (Waytz et al., 2014) and the transparency of the decision making process (Schmidt et al., 2020) as the primary exponents of trust. The latter, often referred to as AI Explainability is itself a rapidly expanding AI research area (Došilović et al., 2018), (Shin, 2021). This chapter focusses primarily on the system's competence.

3.1.2 Intelligent Systems and Educational Context

As noted in Chapter 1, the use of the term educational context rather than classroom is quite deliberate. Learning frequently takes place outside of the classroom (Brahimi and Sarirete, 2015) and education is now delivered on diverse and often distributed platforms. Massive Open Online Courses (MOOCs) popularised mass online education in 2008 (Palacios Hidalgo et al., 2020) and the Covid-19 lockdowns of 2020 and 2021 took education from all sectors out of the physical classroom and into virtual spaces offered by environments such as MS Teams and Zoom (Tandon, 2021*a*). This move online highlighted some key challenges, not least the issue of monitoring learners' engagement (Pokhrel and Chhetri, 2021*a*), (Oyedotun, 2020*a*) and online behaviour (Prathish et al., 2016).

Although mainstream school-level education in the UK has largely returned to the physical classroom, the pandemic has fast-forwarded the development and adoption of hybrid and blended learning pedagogical approaches Zhao and Watterston (2021*a*), highlighting the need for a tool that can help parents and teachers monitor and interpret children's interaction with content online, remotely, and in the classroom.

The context or location of teaching and learning also informs the scope and nature of the interaction between the system and the child. Within the classroom, teachers are likely to have a higher degree of control over the content presented to the child than the parent or caregiver may have within the home. Consequently, the focus is on monitoring for engagement within the classroom whilst examining a wider set of use cases that may face parents and carers within the wider educational context.

3.1.3 Monitoring Engagement

Engagement is widely considered a positive factor in, and an important driver of, children's attainment (Christenson et al., 2012*a*). Definitions of engagement range from a focus on interaction with a specific learning activity to a multidimensional approach that requires the child to engage at behavioural, affective, and cognitive levels (Groccia, 2018*a*). For the purpose of the studies described in this chapter, we consider engagement on task, namely a child's interaction with a computerised learning activity completed within an educational context. Furthermore, while some scholars conceptualise engagement and disengagement as related but separate phenomena (Fredricks et al., 2004*a*), within this context, engagement and disengagement are treated as the opposing ends of a single scale.

3.1.4 System Modelling

There are a huge number of modelling approaches and languages designed to support the system development lifecycle so the obvious question to ask is "why propose another one".

Many of the existing tools such as The Unified Modelling Language

(UML)[1]support the complete system lifecycle and attempt to synthesise multiple viewpoints and perspectives. The TAMM does not set out to do this, rather it focuses on stakeholder engagement as a method of assessing requirements from multiple view points. As such it has similarities with a genre of modelling often applied in resource and environmental management referred to as participatory modelling (Voinov and Bousquet, 2010). Participatory Modelling concentrates primarily on how people interact with models and data (Voinov et al., 2016) and the TAMM sets out to do the same.

The TAMM then is not intended to replace existing methodologies but to complement them. Returning to the UML which provides Use Case diagrams for capturing users interactions with the system (Rosenberg and Scott, 1999). The TAMM exists to check the sanity of the proposed use case rather than model how users interact with the systems interfaces for a given use case.

The first stage of the TAMM is to survey the system stakeholders, in this case, to ascertain the adults' trust and children's acceptance of the Intelligent Systems capabilities for a set of common uses cases. This process is described in Section 3.2. Section 3.4.1 describes how the survey results are plotted on a Cartesian framework called the Trust Acceptance Framework (TAF) to generate the Trust Acceptance Mapping Model (TAMM) which indicates the likely success of the proposed use case. In the studies described the survey data collected uses the same 1 to 10 scale. If different scales are used between surveys, then the data should be normalised before the TAMM is generated.

It should also be noted that there is a considerable body of work in the field of technology acceptance dating back to the 1980s and the Technology Acceptance Model (TAM) (Marangunić and Granić, 2015) that has been adopted by other frameworks such as UTAUT and UTAUT 2 (Chang, 2012). As Gansser and Reich (2021) note, none of these frameworks has AI as an "object of study for behavioral intention and use behavior".

[1]https://www.omg.org/spec/UML/

The TAMM is presented as novel approach to recording and mapping diverse stakeholder requirements within an intelligent system. While the examples presented analyse two variables, trust and acceptance, the use of a Cartesian coordinate system to visualise the model makes it suitable for analysing multiple dimensions of data.

3.2 Studies

Three studies were conducted with the stakeholders to ascertain their trust and acceptance of a theoretical intelligent system to be deployed in an educational context. Children were surveyed on the level of their acceptance of interventions made by adults and technology in their digital activity. Parents and caregivers were asked whether they would trust a technology to either monitor or act if their child was exposed to a given set of use cases. Teachers were interviewed as to how they would feel if a system to monitor pupil engagement was deployed in their classroom, specifically, their acceptance and trust in the system's outputs.

Study 1 surveyed children to assess their acceptance of intervention in a digital activity by either an adult or an intelligent system. Study 2 surveyed parents or caregivers to gain insight into their trust in an intelligent system to either monitor or intervene in a child's digital activity. For Study 3, teachers were interviewed to determine their attitudes towards the implementation in the classroom of a system that could monitor the children and make interventions if they showed signs of disengagement.

There were 4 research questions:

- R1 Are children more accepting of an intervention from a responsible adult than an intelligent system? (Study 1)
- R2 To what extent do parents and caregivers trust technology to monitor their child's digital activity? (Study 2)
- R3 To what extent do parents and caregivers trust technology to intervene in their child's digital activity? (Study 2)

R4 Would teachers trust an intelligent system to monitor pupils in their classroom for signs of disengagement and make appropriate interventions? (Study 3)

3.2.1 Participants

Study 1.

One hundred and twenty-nine children were recruited from a secondary school located in the UK. There were seventy-six females and fifty-three males. The age ranged from 11 to 17 years (mean = 13.27, SD = 1.462).

Study 2.

Twenty-seven parents or caregivers were recruited through social media and word of mouth. There were fourteen females and eleven males, two participants did not disclose their gender. For the age ranges, see Table 3.1. All participants participated voluntarily and no incentives were given.

Table 3.1: Adult Age Ranges

Age	Frequency	Percent
Age not Disclosed	1	3.7
25 - 34	2	7.4
35 - 44	15	55.6
45 - 54	6	22.2
55 - 64	3	11.1
Total	27	100.0

Although the participants were asked for their age, this data was not analysed further due to the relatively low number of responses.

Study 3.

Video interviews were conducted over two days with ten teachers, five male and five female from two UK secondary schools.

3.2.2 Apparatus

For the surveys carried out in Study 1 and Study 2, the data was collected remotely using a web-based interface. Paper versions of the survey were also made available. The survey software was developed using PHP and MySQL and hosted on an Apache web server. JQuery UI was used to implement the interactive user interface. The software was designed to be mobile responsive so that participants could complete the survey on smart phones and tablets as well as desktop PCs and laptops. Both studies combined questions using a Likert scale with values ranging from 1 to 10 where 1 indicated low consensus, e.g. low acceptance, and 10 indicated high consensus e.g. high acceptance.

Study 1 consisted of twenty-one questions. Twenty questions used the Likert scale and one question allowed the participant to enter free text. Study 2 consisted of thirty-two questions. Thirty questions used the Likert scale and two questions allowed the participant to enter free text. For a full set of questions see Appendix A.

3.2.3 Procedure

Study 1

Study 1 was password protected and only made available to the participating school. The head teacher completed consent forms to allow the children to participate in the survey. Additionally, children were given the option to opt out individually before submitting their data. The surveys were completed in a supervised environment using either the web-based form or the paper survey.

The survey was made up of two groups of ten questions. The first group of questions asked how accepting a child would be if a parent/caregiver or other adult intervened in their use of a digital technology for a given (this being the variable under examination) use case. The second group of questions asked, if a technology existed that could monitor the child's actions and take some action, how accepting would the child be of the intervention. The use case for the intervention was the same for each group of questions. The use cases specified in the survey were:

1. Safety
2. Security
3. Curiosity
4. Control
5. Task completion
6. Appropriateness
7. Enjoyment
8. Productivity
9. Learning
10. Economic (e.g. in game purchases).

For the first group, the question related to safety read:

> How accepting would you be if an adult took some action which effected your use of a digital technology because they were concerned about your safety?

The corresponding question for the second group read:

> How accepting would you be if the technology took some action which effected your use of a digital technology because it was trying to keep you safe?

The labels on the Likert scale ranged from not accepting to very accepting. The children were also asked to provide their age and gender. One additional question asked them to describe an occasion where an adult made an intervention related to their use of a digital technology and how they felt about it. The children were asked to complete all the questions and were able to navigate freely through the survey. The data was filtered prior to analysis so that only children who answered related questions across both groups of questions were included in the analysis for each pair of questions.

Study 2

Study 2 consisted of a group of ten questions and a further group of 20 questions and used the same use cases as Study 1. Two additional questions allowed the participant to enter free text. The first group of questions asked whether the participant would personally intervene in their child's use of a digital technology. The second group of twenty questions asked whether participants would trust a technology to either monitor or to take action if the child was exposed to one of the use cases.

For the first group, the question related to safety read:

> How much would concern for your child's safety or wellbeing influence whether you would intervene in their use of a digital technology?

The corresponding questions for the second group read:

> To what extent would you trust the technology to monitor your child's safety?
>
> To what extent would you trust the technology to take appropriate action when monitoring your child's safety?

Participants were also asked to provide their age and gender and the age of their children.

The survey also contained two free text questions which asked:

Q13 Please describe any other factors which influence your decision to monitor and intervene in your child's use of digital technologies.

Q14 If applicable please describe any occasion when you have intervened in your child's use of a digital technology and the impact of that action.

A full list of responses to Q13 and Q14 is available in Appendix D.

The data was filtered prior to analysis so that only participants who answered related questions across both groups were included in the analysis for each pair of questions.

Study 3

Study 3 focussed on teachers' attitudes towards engagement in the classroom. The teachers interviewed were asked:

S3.1 What do you understand by disengagement?

S3.2 What strategies do you employ to identify disengagement in the classroom?

S3.3 Can you describe the interventions (or range of interventions) you employ in the classroom to address disengagement?

S3.4 How trusting would you be of the technology to monitor the children for signs of disengagement?

S3.5 How trusting would you be of the technology to intervene if it identified disengagement?

S3.6 What concerns do you have about the deployment of such a technology?

When interviewing the teachers, the author was able to draw on prior personal experience as a secondary school teacher which aided in establishing rapport. This facilitated a more natural conversation and elicited responses that the teachers may not have given in a more formal setting.

3.3 Results

3.3.1 Study 1

A comparison of the median scores recorded for the children's acceptance of an intervention by either an adult or a technology indicates that the children scored them both within a single point on the scale for all the given use cases (Table 3.2). With the exception of interventions for curiosity and control, all the use cases were ranked $\geq$ to the mid-point of the scale with participants indicating mid to high acceptance of an intervention whether it originated from an adult or a technology.

Table 3.2: Children's Acceptance of Intervention

	Median Score	
Reason for Intervention	Adult Acceptance	Technological Acceptance
Safety	6	6
Security	6	6
Curiosity	5	4
Control	4	4
Complete	7	7
Appropriate	6	5
Enjoyment	7	6
Productivity	6	6
Learning	8	7
Financial	5	5

Mantel-Haenszel tests of trend[1] were conducted to understand whether there is an association between a child's level of acceptance of an intervention made by an adult and the level of acceptance of an intervention made by a technology for the same use case. The

[1] The Mantel-Haenszel test of trend is used to determine whether there is a linear trend (i.e., a linear relationship/association) between the two related ordinal variables that are represented in a crosstabulation table.

Mantel-Haenszel tests of trend showed a statistically significant linear association between the child's acceptance of intervention by an adult and their acceptance of intervention by a technology for all the use cases tested. Higher acceptance of an intervention by an adult was associated with higher acceptance of an intervention by the technology and vice versa.

Safety $\chi2(1) = 50.595$, p $<$ .01, r = .636

Security $\chi2(1) = 71.045$, p $<$ .01, r = .760

Curiosity $\chi2(1) = 55.229$, p $<$ .01, r = .673

Control $\chi2(1) = 60.285$, p $<$ .01, r = .697

Complete $\chi2(1) = 29.188$, p $<$ .01, r = .487

Appropriate $\chi2(1) = 47.795$, p $<$ .01, r = .618

Enjoyment $\chi2(1) = 33.561$, p $<$ .01, r = .520

Productivity $\chi2(1) = 46.352$, p $<$ .01, r = .614

Learning $\chi2(1) = 37.725$, p $<$ .01, r = .552

Financial $\chi2(1) = 39.694$, p $<$ .01, r = .573

A Pearson Partial Test of Correlation was used to establish the strength of the linear relationship between the variables and in all cases indicated a mid to strong positive correlation.[2]

[2]The Pearson test was recommended as the appropriate test to measure the strength

3.3.2 Study 2

Question 2 (Q2) of Study 2 asked participants to rank how often they intervened in their child's use of digital technologies such as computers or mobile devices whilst Question 15 (Q15) asked to what extent would participants trust the technology to monitor their child's everyday use of a digital technology?

A scatter plot of Q15 by Q2 (Figure 3.1) indicates that participants who ranked their frequency of intervention as low on the scale ranked their trust in the technology more highly than participants who indicated higher personal levels of intervention.

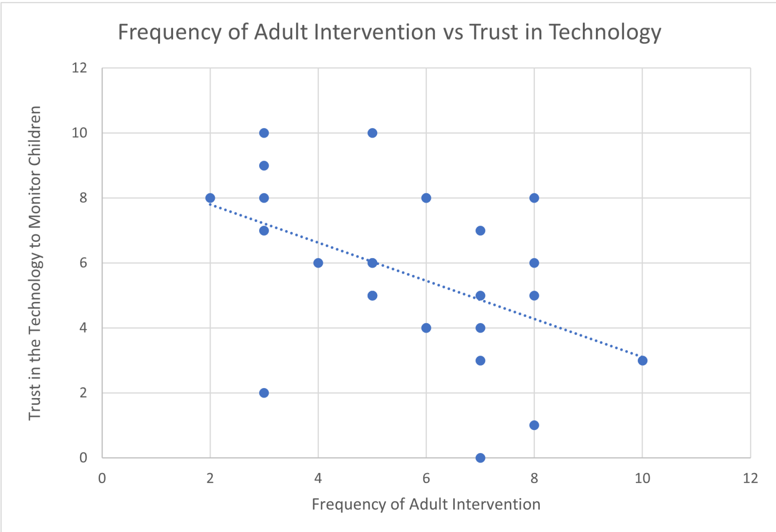


Figure 3.1: Trust in Technology by Adult Intervention

A Mantel-Haenszel test of trend was conducted to understand whether there is an association between how often adults intervene in their child's use of digital technologies and to what extent they would trust an agent to monitor their child's everyday use of a digital technology. The Mantel-Haenszel test of trend showed a statistically significant linear

of the correlation between the variable once a linear association had been established using a Mantel-Haenszel test of trend despite the data being ordinal and non-parametric. Spearman's rank-order correlation tests were also conducted and produced significant results in line with the results generated by the Pearson test.

Table 3.3: Top 5 Intervention Categories Ranked by Median Score

Adult Intervention		Technology Intervention	
Safety	10	Monitor Financial	7
Security	9	Monitor Appropriate	6
Appropriate	9	Take Action Financial	6
Financial	9	Monitor Safety	5.5
Help	7	Monitor Security	5.5

association between frequency of intervention and trust, $\chi 2(1) = 4.999$, $p < .05$, $r = -.447$. Adults who indicated higher intervention rates were associated with a lower trust of the agent and vice-versa.

Ranking the median scores for each of the use cases for personal intervention indicates that participants were more likely to intervene for reasons of safety, security, appropriate content and financial considerations (Table 3.3). These were also the use cases that participants indicated the highest trust in the technology to monitor or take action.

Questions 13 (Q13) and 14 (Q14) provided participants with the opportunity to elucidate further on the Likert responses. Q13 asked participants to: Describe any other factors which influence your decision to monitor and intervene in your child's use of digital technologies. Q14 asked: If applicable please describe any occasion when you have intervened in your child's use of a digital technology and the impact of that action. The answers provide a lens to further interpret the responses. Participants cite factors such as social media usage and online gaming where the child is interacting with a remote third party as reasons for intervention but also a desire to help and support the child in a digital activity. The full responses are presented in Appendix D.

In all the use cases tested there was a strong positive correlation between participants' trust in the technology to monitor children's activity and to make an appropriate intervention. The tested cases were, Safety, Security, Appropriateness of Accessed Content, Enjoyment, Financial Transactions, Productivity and Learning.

An analysis of the median scores indicates that these two facets were

Table 3.4: Trust in Technology to Monitor and Take Action

Reason for Intervention	Monitor	Take Action
Safety	5.5	5
Security	5.5	5
Appropriate	6	5
Enjoyment	2.5	3
Financial	7	6
Productivity	5	5
Learning	5	5

scored within a single point on the 10-point Likert scale employed (Table 3.4).

The Mantel-Haenszel test of trend showed a statically significant linear association between trust in the technology to monitor and trust in the technology to take action for all use cases.

Safety $\chi 2(1) = 14.444$, $p < .01$, $r = .811$

Security $\chi 2(1) = 3.383$, $p < .01$, $r = .933$

Appropriate $\chi 2(1) = 7.730$, $p < .01$, $r = .556$

Enjoyment $\chi 2(1) = 22.333$, $p < .01$, $r = .965$

Financial $\chi 2(1) = 21.530$, $p < .01$, $r = .910$

Productivity $\chi 2(1) = 18.562$, $p < .01$, $r = .862$

Learning $\chi 2(1) = 17.096$, $p < .01$, $r = .844$

Participants indicated a higher level of trust in the technology's capability to monitor a use case (R2) than to take appropriate action (R3). In all cases except for Question 23 and Question 24, the technology's capability to monitor and improve the child's enjoyment of an activity, the median value recorded was $\geq$ to the midpoint of the scale indicating at least a mid-level of trust in the technology's capabilities to perform the described roles.

3.3.3 Study 3

In answer to the question, What do you understand by disengagement? (S3.1), nine out of the ten teachers interviewed identified it as task focussed manifested by the children not completing the work they had been set. Eight teachers also identified behavioural traits as an indicator of disengagement. Teacher's remarked that, 'Disengagement starts off with them not doing the work' and 'not completing the work they should be focussing on at that time'. Behavioural indicators described were 'gazing into space', 'clicking pens', and 'not partaking in discussions'.

When asked about the strategies they employed to identify disengagement in the classroom (S3.2), nine out of ten of the teachers stated that the most important factor was knowing the child. All of the teachers interviewed deployed a range of classroom management techniques to keep the children on track (S3.3). The teachers routinely patrolled the classroom during lessons as well as utilising questioning techniques and short task durations to maintain pupil engagement.

When asked about their feelings regarding the deployment of a system to monitor engagement (S3.4) and make interventions (S3.5) only one of the teachers interviewed indicated that they would not accept the technology in their classroom. The other teachers indicated their acceptance subject to criteria, the most common of which was that the system outputs must be accurate and support the children's learning. Teachers also expressed their concern that such a system may be used as a monitoring tool to report on their personal effectiveness rather than as a educational aid (S3.6).

3.4 Discussion

For the designers of intelligent systems for use in an educational context there is a requirement to balance the need of three or more stakeholders, the child, the carer or parent, and the teacher. The role of each of the parties depends to some extent on the context in which the system is deployed. Within the classroom the child is the subject of the observation and is likely to have little control over the technology and software they are interacting with whilst the intelligent system monitors them. The technology and software are selected by the school and teacher and safeguards are in place to minimise any risk to the child's wellbeing. The teacher is present in the classroom, available to receive feedback from the intelligent system and can act accordingly.

In a context outside the classroom, the child is likely to have far more freedom in what they choose to interact with. The same level of safeguards present in the classroom are unlikely to be in place and the responsible adult may not be present in the room or even at the same location. The child remains the subject of the systems observation but the system has a dual role of both monitoring the child and also intervening in the child's interaction with the digital world. This study examines a non-exhaustive set of use cases that may occur during these interactions.

The process of training the intelligent system to recognise these use cases and its implementation or embodiment are deferred until later in the thesis. Rather, this chapter concentrates on the interplay between the child's acceptance of the system's outputs and the adults trust in their accuracy. That the children surveyed indicated a level of acceptance of an intervention by the technology $\geq$ to the midpoint on the scale in all but two of the use cases is indicative that they are at least comfortable with the theoretical system concept. It is also interesting that the children drew little distinction between an intervention from an adult and an intervention from the technology (R1). The highest scoring use case across both categories was learning which may bode well for deployment within an educational context.

Parents and carers appear on the whole to be less trusting of the system than the children are accepting and draw a bigger distinction between their personal judgments and the systems judgments. Even so, a level of trust was indicated for all but one use case, enjoyment $\geq$ to the midpoint on the scale. This is important as they may not be physically present at the time their child is interacting with a digital device, particularly as the child gets older. It is interesting to note that of the adults surveyed those who felt more inclined to personally intervene indicated less trust in the intelligent system than those who made fewer personal interventions.

The teachers interviewed were broadly supportive of the deployment of the intelligent system in the classroom with only one teacher expressing complete opposition to its deployment (R4). The context is of course important and the system may have more of a monitoring role to identify disengagement and alert the teacher to make an appropriate intervention. This would appear to be the best supported use case with the children indicating a mid to high level of acceptance of the systems output and the teachers prepared to accept the technology within their classrooms.

3.4.1 Trust Acceptance Mapping Model.

Within the wider educational context, it could be suggested that the system needs to balance the needs of both the children and the adult. The adult needs to feel sufficient trust in the system and the child needs to accept the systems interventions. This can be visualised by mapping levels of adult trust against levels of child acceptance for each of the use cases which we call the Trust Acceptance Framework (TAF). Placing trust along the x axis and acceptance on the y axis of a graph allows the data to be mapped as four quadrants (Figure 3.2) with the characteristics summarised below. The top right quadrant can be regarded as the design goal where acceptance of system intervention and trust in the systems capability are both high.

Top Left - High Acceptance and Low Trust

The child sees value in a systems capability

The adult has little or no confidence in the system capabilities or features

Bottom Left - Low Acceptance and Low Trust

The child sees little or no value in a systems capability

The adult has little or no confidence in the system capabilities or features

Top Right - High Acceptance and High Trust

The child sees value in a systems capability

The adult has confidence in the system capabilities or features

Bottom Right - Low Acceptance and High Trust

The child sees little or no value in a systems capability

The adult has confidence in the system capabilities or features

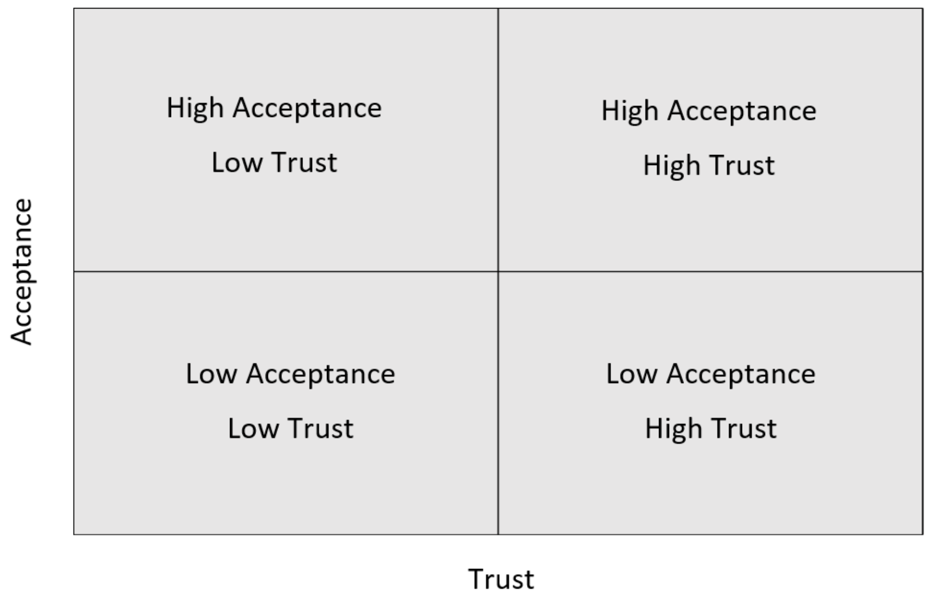


Figure 3.2: Trust Acceptance Framework (TAF)

As a baseline the adults personal inclination to intervene in their child's digital activity is mapped against the child's acceptance of the intervention

Figure 3.3: Trust Acceptance Mapping Model (TAMM) for Child Acceptance of Adult Intervention

onto the TAF. This constitutes the Trust Acceptance Mapping Model (TAMM) (Figure 3.3) which visualises the relationship between trust and acceptance for the given use cases for this configuration of the independent variables trust and acceptance.

More formally the dependent variable $trust\_acceptance$ is a 2-tuple (couple) (x, y) where x is the trust value for the specified use case and y is the acceptance value for the specified use case and x and y are bounded such that $\{1..10\} \Rightarrow \{x \in \mathbb{Z} : 1 \leq x \leq 10\}$. The $trust\_acceptance$ values generated for each of the use cases are plotted onto the TAMM as Cartesian coordinates.

All the values for $trust\_acceptance$ except for the use case control fall within or on the border of the High Acceptance High Trust quadrant indicating that the adults are trusting in their own ability to intervene and the children are accepting of the interventions.

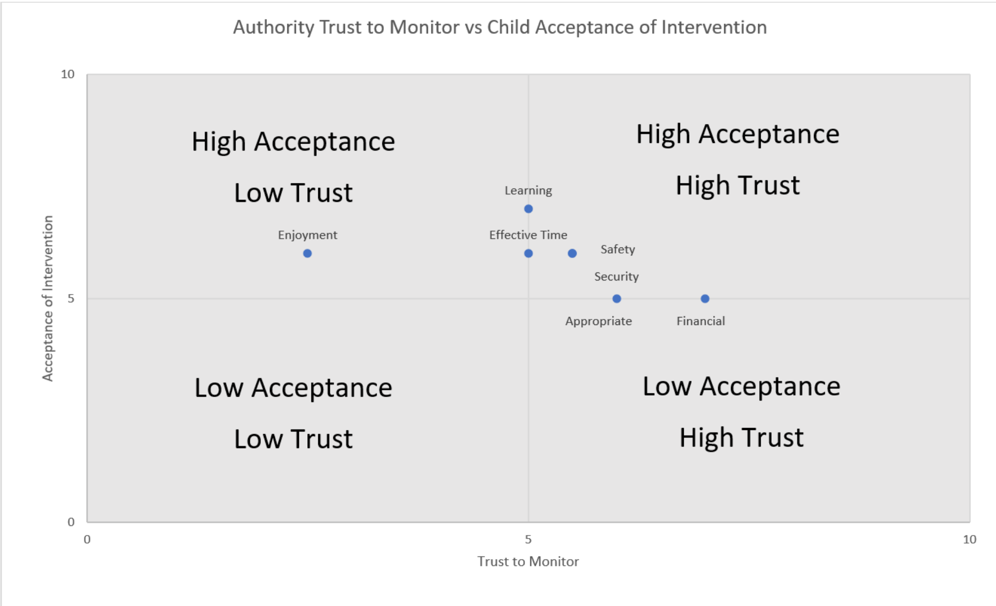


Figure 3.4: TAMM for the System as a Monitor

System as a Monitor.

Figure 3.4 maps the data from Studies 1 and 2 onto the trust acceptance framework where the adult expresses their level of trust in the systems monitoring of the child's behaviour and the child expresses their level of acceptance of the systems output. The learning, productivity (effective time), appropriate content, and financial use cases all sit on the edge of the High Acceptance and High Trust quadrant, with adults placing greater trust in the monitoring of financial transactions and monitoring content whilst the children are more accepting of learning and productivity.

From a system designers perspective, there is at least a consensus on the features on which a design can be based. The enjoyment use case sits in the High Acceptance and Low Trust quadrant indicating that whilst the children's acceptance of an intervention based on these grounds is high, the adult has little trust in the systems capabilities, and it is unlikely to be accepted if implemented.

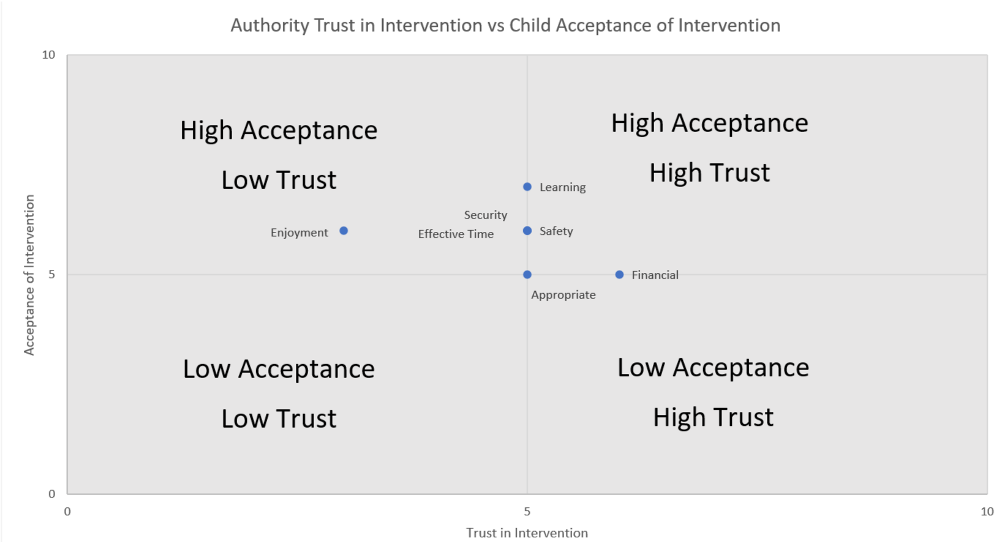


Figure 3.5: TAMM for the System as an Interventional Agent

System as an Interventional Agent.

Where the system is required to intervene in the child's digital interaction as opposed to just monitor it, it is less trusted by the adults to execute interventions and none of the data points fall within the High Acceptance High Trust quadrant (Figure 3.5). Clearly, this may have implications if there is no adult present to personally perform the intervention if the child is performing some action that may affect their wellbeing. This is an area for further work, and Chapter 4 assesses how the embodiment of the system affects stakeholders' trust and acceptance of its outputs.

3.5 Summary

This chapter explores child acceptance and adult trust in a theoretical intelligent system designed to monitor and potentially intervene in a child's interaction with a computer or other digital technology in an educational context. Three studies were conducted with the main stakeholders in the system, teachers, parents or caregivers, and children. The children were widely accepting of the interventions for the use cases presented. Parents and caregivers were more trusting of the system to monitor children's activity than in the potential of the system to make an appropriate intervention. Most of the teachers interviewed saw value in the deployment of the system in the classroom with the caveat that its outputs must be accurate. They also expressed concerns that rather than being deployed as a classroom aid, the IS would be used as a tool to monitor their performance.

It was identified that the system may have to play a different role depending on the educational context in which it is deployed. Within the school classroom, the teacher is present, and the child's interaction with any technology is closely controlled. Under these circumstances, it makes sense to deploy the system as a tool to monitor the children's behaviour and alert the teacher. The teachers interviewed were generally positive about the potential deployment of such a system in their classroom. In a wider educational context, there may be less control over the technology and software the child encounters, and the adult may not always be present. There is a conflict here between this increased risk to children and the decreased trust of parents and caregivers in the system to act as an interventional agent.

The Trust Acceptance Mapping Model is presented as a tool to indicate the likely success of the intelligent system design. Use cases which reside in the top right High Acceptance High Trust quadrant are likely to have a greater chance of adoption than those that fall in the other quadrants.

Chapter 4

The Effect of Design Choices on Children's Perceptions of a System's Capabilities

4.1 Introduction

Chapter 3 discussed factors influencing the design of intelligent systems. This chapter develops that theme firstly by presenting a study considering how implementation choices for an IS interface, i.e. the way in which it is embodied, may affect children's perceptions of its capabilities. The study took place in the secondary school that participated in the survey study described in chapter 3.2.1.

A second study conducted with younger children aged between 7 and 9 explores how researchers can influence the children's view of capabilities of a humanoid robot by the way they refer to it. This work was presented at IDC '20 (Parsonage et al., 2020) and presented guidelines for researchers working with children and robots.

Robots are a somewhat exceptional example of an intelligent agent often perceived to have capabilities that are not attributed to other digital technologies. As long ago as 1941, in the fictional work Runaround, Asimov (1941) set out his 3 laws governing robot behaviour which implicitly imply use cases not usually attributed to the humble PC. The durability of Asimov's laws is indicated by the lengthy list of academic papers that still cite them more than 80 years later (Decuypere et al., 2023), (Murphy, 2023). As such it should be no surprise that systems that deploy robots then are likely to be regarded by children as different to those that other digital technologies.

The human tendency to anthropomorphise is well documented, (for a wider reference see (Epley et al., 2007)) and it is not then surprising that this approach of attempting to make sense of our environment by superimposing our own behaviours and beliefs onto it should extend to our digital technologies. It is not uncommon for people to describe an interactive technology as, "having a mind of its own", particularly when the task to be accomplished is complex or the device is exhibiting unexpected behaviour.

Within the field of Human-Robot Interaction (HRI) there are a number of studies researching the effect of anthropomorphism on our perception of robots (Eyssel et al., 2011), (Złotowski et al., 2015) and also children's perception of robots (Tung, 2016). Research into natural interfaces has assessed the effect of different levels of anthropomorphism on an agent's persuasiveness (Khan and Sutcliffe, 2014) and the effectiveness of conversational agent interfaces such as Siri, OK Google and Alexa (Luger and Sellen, 2016). Researchers have even studied the effects of anthropomorphising consumer goods such as toasters (Burneleit and Hemmert, 2009) and kettles (Cowan et al., 2013).

What is less well researched is how the way in which children anthropomorphise differs from adults' motives for anthropomorphism. In their paper On Seeing Human: A Three-Factor Theory of Anthropomorphism (Epley et al., 2007), the authors propose that one of the factors that motivate both adults and children to anthropomorphise is effectance motivation, the motivation to explain and understand other agents. They hypothesise that children are motivated to anthropomorphise by the desire to explain or understand their environment whereas adults are more likely to seek control and predictability. This chapter examines whether how a robot is presented affects children's perceptions of its capabilities to what extent anthropomorphising a systems interface affects children's acceptance of its judgments.

One area where robots are increasingly utilised is education. Examples include robots supporting children in early years learning (Crompton et al., 2018), support for children with autism (Costa et al., 2015), (Pennisi et al., 2016) and the general integration of robots into the classroom (Edwards et al., 2016). This increased positioning of robots alongside children has stimulated and spawned a child-robot interaction (CRI) research community, notably a CRI workshop series beginning in 2015 (Child-Robot Interaction 2018) and a workshop at the 17th ACM Conference on Interaction Design and Children (IDC) (Charisi, 2018). Chapter 5 shows that CRI remains a prominent theme in the IDC research community.

This emergence of CRI as a research field and its links with the IDC and the wider CCI and HCI communities opens up opportunities to engage with children in designing future robots and in informing how such robots should look, act and behave. Including children in the design of technologies has long been a theme for the CCI community and the number of studies on how to engage with children in talking about robot design is increasing (Arnold et al., 2016), (Parsonage et al., 2020), (Alves-Oliveira et al., 2021).

4.2 Studies

4.2.1 Study 1

Study 1 was conducted with 20 year nine pupils from the same UK school that participated in the surveys detailed in chapter 3. Chapter 3 evaluated the children's acceptance of an intervention by an unspecified technology for a range of use cases. This study evaluates the affects of anthropomorphising the technology on the children's judgments. The children were introduced to a Poppy Humanoid Robot before they were asked about their acceptance of an intervention made either by a computer, robot, responsible adult or friend for a subset of the use cases presented in chapter 3.

4.2.2 Participants

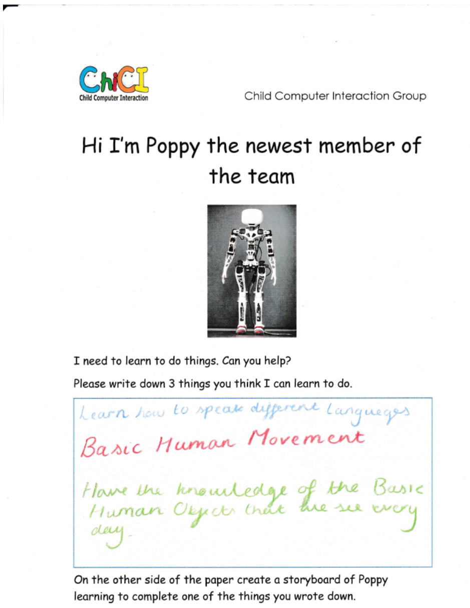
For logistical reasons, the children were split into two groups of ten. The first group was comprised of six boys and four girls and the second group was comprised of four boys and six girls. The children were all year 9 students aged between 13 and 14 years.

4.2.3 Apparatus

The robot used in the study was a Poppy Humanoid 3D printed humanoid robot designed to be used by educators, artists and scientists in a variety of medium (Lapeyre, 2018) (Figure 4.1). In the study the robot was not switched on but presented to the children sitting in a compliant state.

Figure 4.1: Poppy Humanoid Robot

A worksheet was prepared for the children to records their ideas about the robots's capabilities and for the children to create a storyboard of actions for the Poppy robot (Figure 4.2).



ChiCI
Child Computer Interaction
Child Computer Interaction Group

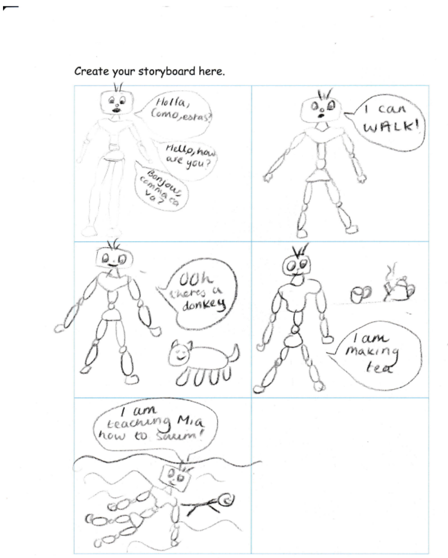
Hi I'm Poppy the newest member of the team

I need to learn to do things. Can you help?
Please write down 3 things you think I can learn to do.

Learn how to speak different languages
Basic Human Movement
Have the knowledge of the Basic Human Objects that we see every day.

On the other side of the paper create a storyboard of Poppy learning to complete one of the things you wrote down.

(a) Suggested Actions - Humanised



(b) Storyboard of Actions - Humanised

Figure 4.2: Updated Poppy Study with Storyboard

The use cases presented in the original survey were printed onto an A4 sheet of paper, one use case per sheet, along with an explanation and an accompanying example. For the safety use case (Figure 4.3), the children were told that, “The person or technology warns you that you are carrying out some action that could potentially cause you harm” and the example, “Giving your personal details to someone you do not know”.

Figure 4.3: Safety Use Case

The full set of use cases discussed are presented in (Table 4.1).

Table 4.1: Use Cases and Descriptions

Use Case	Description	Example
Safety	The person or technology warns you that you are carrying out some action that could potentially cause you harm.	Giving your personal details to someone you do not know
Security	The person or technology warns you that you are carrying out an action that could potentially result in damage to you or the system you are using.	Installing unsafe software or disabling security features on the software

Use Case	Description	Example
Curiosity	The person or technology is curious about what you are doing.	You take an action in a game and the technology or person asks you to explain why.
Helpfulness	The person or technology wants to help you complete a task.	The person or technology makes suggestions about actions you could take in order to complete a game or finish some work.
Control	The person or technology wants to take control over what you are doing	You have misspelt a word and the person or technology stops you and makes you correct it before continuing or they/it turn your music off because they believe it it distracting you from completing your homework.
Appropriate	The person or technology warns you that you are accessing inappropriate materials.	You are accessing material which has an 18+ restriction.
Enjoyment	The person or technology wants to make you happy.	The person or technology suggests an action or activity that they/it think you will enjoy such as an online game or physical activity.
Learning	The person or technology wants to improve your learning.	The person or technology suggests a resource they think will help you understand a problem.

Use Case	Description	Example
Productivity	The person or technology wishes to enhance your effectiveness.	The person or technology prompts you to take an action to help you achieve your goals such as going for a run to help you hit your fitness targets.
Economic	The person or technology believes you are carrying out an action that may cause you financial harm.	The person or technology prevents you from making multiple in game purchases.

Each pupil was issued with a set of four actor cards (Figure 4.4) to enable them to indicate their agreement with statements made by the researcher based upon the use cases. The four actors were:

1. Watching Robot
2. Computer
3. Parent
4. Friend

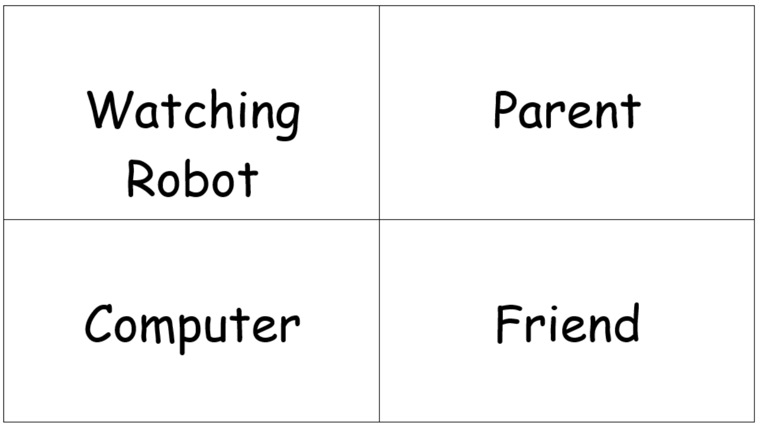


Figure 4.4: Actor Cards

4.2.4 Method

The researcher introduced the Poppy robot to both groups of children separately deliberately anthropomorphising it. The robot was referred to as she or Poppy and the children were asked to identify tasks “Poppy could learn how to do”. Both groups were introduced to the robot and completed the worksheet before they were asked to examine the use cases.

When examining the use cases, the researcher and pupils sat around a table and the A4 sheet detailing the use case for discussion was circulated to the pupils. The researcher answered any questions relating to the use case posed by the pupils and for each of the four actors in turn and then asked whether the students would be accepting of an intervention made by that actor for the given use case. Pupils who believed they would be accepting of the use case were asked to place their actor card on the table and the responses were tallied by the researcher. The researcher facilitated a discussion amongst the pupils who were asked to elaborate on their decisions.

4.2.5 Results

Due to time constraints the groups were only able to discuss six out of the ten use cases. The use cases not discussed were learning, economic, control and productivity. Curiosity was the only use case considered by both groups. The acceptance score for intervention by each actor for group 1 are shown in (Table 4.2) and group 2 (are shown in (Table 4.3). The use cases considered by group 1 were safety, security, curiosity, and helpfulness. The use cases considered by group 2 were curiosity, appropriateness, and enjoyment.

Table 4.2: Acceptance of Intervention by Actor - Group 1

Use Case	Computer	Watching Robot	Parent	Friend
Safety	6	2	9	7
Security	10	1	8	4
Curiosity	0	5	10	10
Helpfulness	9	9	7	10

Table 4.3: Acceptance of Intervention by Actor - Group 2

Use Case	Computer	Watching Robot	Parent	Friend
Curiosity	2	2	4	10
Appropriate	9	1	6	7
Enjoyment	6	5	10	8

The researcher facilitated a discussion amongst the children to allow them to expand on their judgements. The full data is presented in Appendix B

4.2.6 Discussion

In most cases the pupils differentiated between the capabilities of the robot and the computer. The pupils in Group 1 scored both the computer and parents higher for well-defined tasks such as safety and security. Pupils typically commented that the computer understood how to protect itself whilst parents were assumed to have their child's best interests at heart. During the discussion the children also remarked that the robot, "knew more than the computer".

Discussions indicated a high acceptance of rule-based intervention by the technology where the context of the intervention is understood and the children may have prior experience such as a web page white list. The children perceived that robots and computers are different whilst still accepting that they were both technologies. The perception was that robots were able to learn from their experiences and adjust their behaviour accordingly whilst computers are programmed and slavishly follow rules. Curiosity was the only use case discussed by both groups. Group 1 commented that, “computers can’t be curious, they just predict” whilst, “robots can learn”. Group 2 also believed that “robots can learn” and described the concept of the computer being curious as a, “terrifying idea”.

The children also expressed concerns about the rationale and motivation for the intervention. Security was a core concern, what was the motivation for the intervention by the technology and what will it do with the information it collected? An indication that whilst the children accorded relatively high levels of trust to the technology in the original survey there are still underlying concerns. The same concerns were not expressed about the human relationships. The children may not always want their parent to know what they are doing but they are comfortable with the motivation behind the parental intervention.

Interface choice is then an important consideration in designing ISs. The children’s perception that robot’s have a greater capability to learn than computers may impact on their acceptance of the judgments of an IS designed to monitor and intervene in their behaviour. In short, they may be more accepting of the outputs of a system with anthropomorphic affordances than one which they perceive as a rule based number cruncher.

Conversely when asked for their thoughts on what the IS might look like during the interviews described in Chapter 3, teachers opted for a standard computer interface. The system they envisaged was “built into the child’s laptop” possibly providing a “non-verbal warning or reminder” or “software that informs the teacher of [the pupils] disengagement”. This may simply reflect a practical approach and reaction on their behalf to the researcher’s suggestion of a robot let loose in their classroom. Nevertheless, there appears to be real differences between the system stakeholders, with the children perceiving the robot as able to learn over time while the teachers favour a computer or screen based implementation.

The second study described in this chapter also employs the Poppy humanoid robot and looks at ways in which researchers may inadvertently affect children's perceptions of the robot's capabilities by the way in which they present it.

4.2.7 Study 2

Read and MacFarlane (2006) highlight the danger of "suggestibility" or the influence, intentional or otherwise, that the researcher may hold on the child participant (Scullin and Ceci, 2001). Whilst their paper specifically referred to surveys, in this study the idea is extended to look at how suggestibility can influence children's perceptions of intelligent systems, in this case a humanoid robot offering anthropomorphic affordance (Norman, 2013). The study examines how the way in which the researchers present the robot to children affects their perceptions of its capabilities.

The following hypotheses are tested:

- H1 Introducing the robot as a robot would encourage participants to attribute predominantly physical actions to the robot.
- H2 Introducing the robot as human would encourage participants to attribute intelligence and emotional characteristics to the robot.

4.2.8 Method

Participants

Forty-three children were recruited from two local primary schools and attended the University laboratories on organised school trips on two separate occasions over a two-week period. The first group consisted of 8 boys and 20 girls aged 8 to 9. The second group consisted of 9 boys and 6 girls aged 7 to 8.

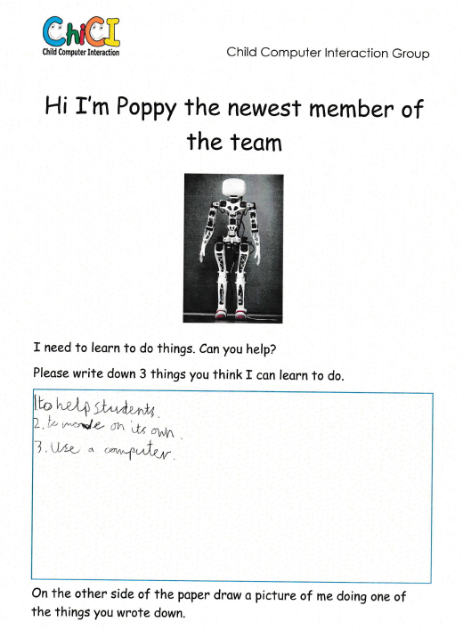
Apparatus

Two almost identical worksheets were prepared for the children to use in the data gathering activity. These were used to gather the ideas from the children and to provide a space for the children to draw the robot performing an activity. The worksheets differed only in how they referred to the robot as ’humanoid’ or ’robot’ (see Appendix C).

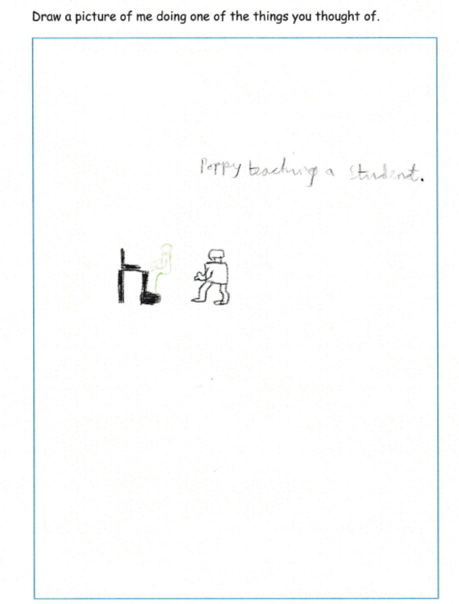
Procedure

Before the study and in accordance with the ethical procedures described above in Section 2.3.2, consent was obtained from the parents of the children. Additionally, the children had the opportunity to opt out at any time during the study if they did not wish to participate.

On arrival, the pupils were divided into groups of between 3 and 6 children by their accompanying teachers and subsequently attended the session in these groups. On entering the room, each group was introduced to the Poppy Humanoid in one of two ways. Either the robot was humanised and introduced as a member of the team (humanised condition) that wanted to learn, or the robot was introduced as a robot that required programming (robot condition). In the first case, the children were asked to write three things Poppy could learn to do (Figure 4.5). The researcher referred to the robot as she or Poppy. In the second case, Poppy was described as a robot that required programming, and the children were asked to write down three things the robot could be programmed to do (Figure 4.6). In both cases, the children were asked to draw a picture of the robot doing one of the things they had written down.

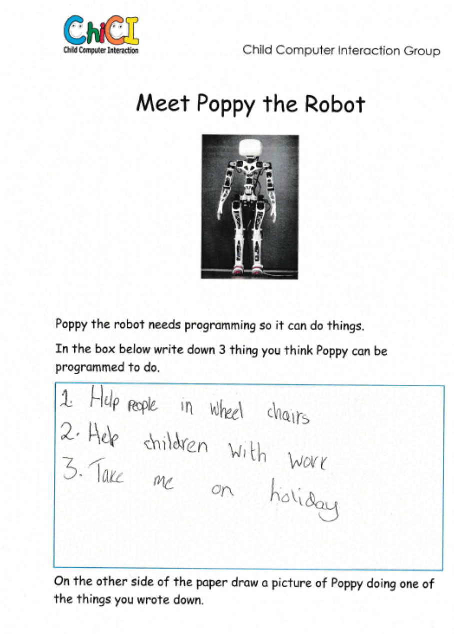


(a) Suggested Actions - Humanised

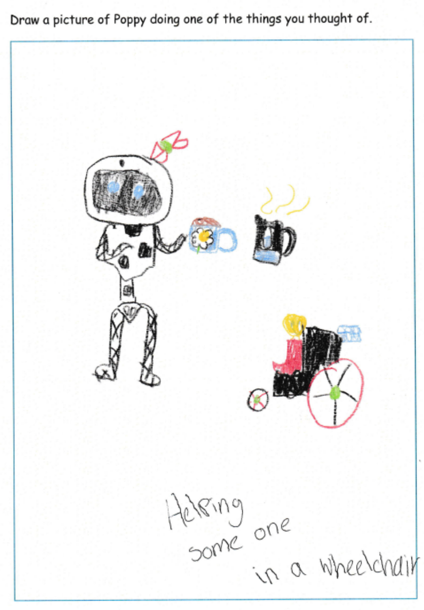


(b) Drawing of Actions - Humanised

Figure 4.5: Poppy presented - Humanised Condition



(a) Suggested Actions - Robot



(b) Drawing of Actions - Robot

Figure 4.6: Poppy presented - Robot Condition

Each group spent approximately 20 minutes on the task. After the researcher introduced the robot, the children had a short period of time to interact with it before completing one of the worksheets. The robot was not powered up. The researcher interacted with the children whilst they completed the worksheet and notes were made on comments the children made whilst completing the task. Twenty-two children completed the humanised worksheet and twenty-one children completed the robot worksheet.

Results

Thematic analysis was employed to analyse and define themes or action categories from the children's suggestions. Where an action did not fit into an existing category, a new action category was created. For each category of actions created, a definition was produced along with two examples to ensure that the suggestions were categorised correctly

(Table 4.4).

Table 4.4: Thematic Analysis

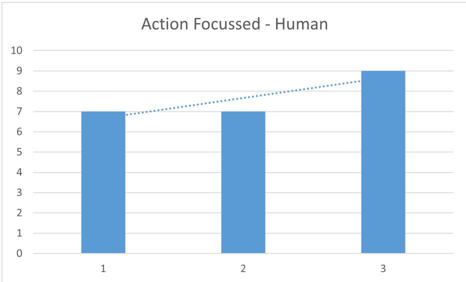
Action Category	Description
Action Focused	Completion of a physical action or sequence of actions such as walking or playing football.
Emotional	Exhibiting feelings towards self or others such as being happy or friendly.
Intelligence and Learning	Able to initiate or modify actions in the light of ongoing events such as driving a car or learning a language.
Assistive	Give help or support to carry out an action or task such as cleaning or helping with homework
Organic	Performing an action performed solely by a living entity such as breathing or dying.
Appearance	Make changes to physical features such as applying makeup or doing hair.

The actions suggested by the children were placed in the appropriate category and a count was taken (Table 4.5). Eighty-three individual ideas were generated for the human presentation and 104 for the robot presentation, totalling 187 suggestions provided by the children. Where the Poppy Humanoid was presented as human 39% of the actions the children generated were classified as Action Focused, 35% were classified as Intelligence and Learning and 14% as Assistive. Where the Poppy Humanoid was presented as a robot 71% of the actions generated were Action Focused, 14% Intelligence and Learning, and 14% Assistive. The Emotional and Appearance categories were unique to the human presentation. The Organic category had 1 suggestion per presentation.

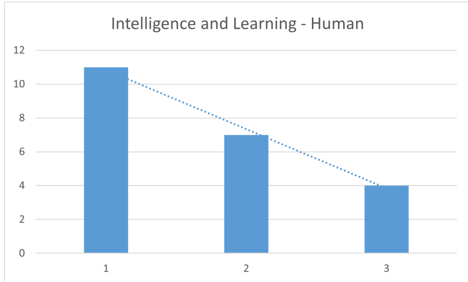
Table 4.5: Count of Actions by Theme

Action Category	Robot Introduced As	
	Human	Robot
Action Focussed	32	74
Emotional	6	0
Intelligence and Learning	29	14
Assistive	12	15
Organic	1	1
Appearance	3	0
Totals	83	104

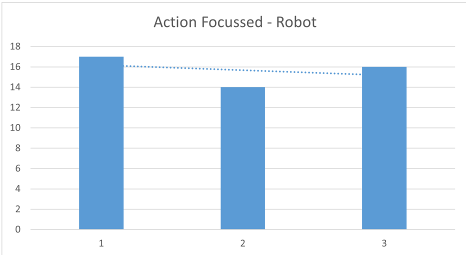
The first three actions suggested by each child were organised into action categories in the order they were suggested. Children were more likely to select an action categorised as requiring intelligence and learning as their first choice when the robot was introduced using the human condition. When the robot was introduced using the robot condition, the children were more likely to suggest an action categorised as action focussed (Figure 4.7).



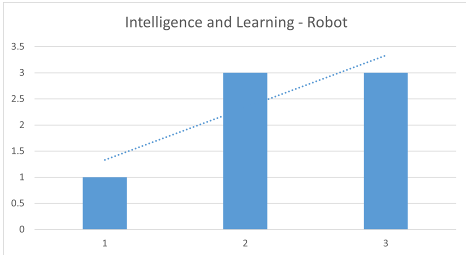
(a) Action Focussed - Humanised



(b) Intelligence and Learning - Humanised



(c) Action Focussed - Robot



(d) Intelligence and Learning - Robot

Figure 4.7: Children's Assigned Themes Ordered by Choice

(Table 4.6) shows the terms the children used to refer to the robot and indicates that children were more likely to attribute gender to the robot when it had been presented using the human condition. Most of the children who ascribed gender to the robot referred to it as she rather than he. This is unsurprising, as both the researcher and humanised worksheet referred to Poppy as she. Poppy is also considered a female name in western society.

Table 4.6: Poppy Humanoid Gender Assignment

Presentation Method	It	She/He	Poppy	None	Robot	Totals
Human	1	8	2	11	0	22
Robot	2	2	3	13	1	21

Two action categories (Figure 4.8), emotional and appearance, were only captured when the Poppy Humanoid was presented as human. The remaining action categories were common to both groups. The presence of these additional categories indicates that children's perceptions are influenced by the introductory method employed.

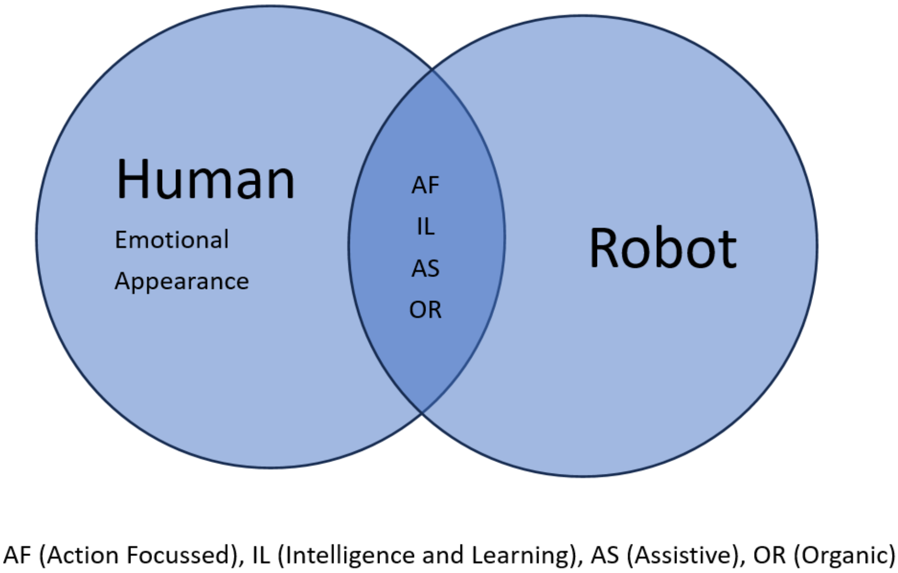


Figure 4.8: Assigned Action Categories

Discussion

The findings from the study suggest that the initial hypotheses presented are correct. When the robot was introduced using the robot condition, 71% of the suggestions the children made were classified as requiring completion of a physical action. Only 14% of the children suggested an action that required intelligence and learning and there were 0 suggestions in the emotional category. When the robot was introduced using the human condition, the number of suggestions for a physical action fell to 39%, while suggestions for actions requiring intelligence and learning increased to 35% with 7% of suggestions categorised as emotional.

It was not surprising that the use of a humanoid robot elicited suggestions for physical actions associated with human behaviour. The physical design of the robot provides clues to its persona and colours the child's perceptions of what it can achieve. This in conjunction with the child's personal experience produced a wide range of suggested activities. A child who had to use a wheelchair suggested that the robot could help people in wheelchairs, while children who enjoyed gymnastics suggested gymnastic-related activities.

The findings have limitations. The number of participants is small, and a larger sample would add weight to the findings. The study highlights global trends and further work is required to identify the root causes of the trend. It is assumed that the participants attributed characteristics to the robot based on the researchers' actions, however, the participants' views were not established prior to the study.

The main findings of this study are that when introducing a robot to children in a design session, the way the robot is introduced will have an impact on what the children offer in terms of requirements or ideas. With class-sized studies, children working in groups, and the possibility that they might 'copy' or 'share' ideas is unlikely to have much impact on the ideas or requirements generated.

4.3 Summary

This chapter considers how design choices regarding an IS's interface may affect both acceptance of its outputs and perceptions of its capabilities. Two studies were presented both of which introduced children to a Poppy Humanoid Robot. The first study examined how anthropomorphising the system may impact children's acceptance of its outputs. The children who participated perceived that the robot is able to learn while the computer is a rule based technology.

In the second study the researcher introduced the Poppy robot in either "humanised" or "robot" form. When the robot was introduced in humanised form, the children were more likely to attribute actions requiring learning or intelligence to it.

As noted in Chapter 3.5, teachers were broadly accepting of the introduction of a system to monitor engagement in their classroom while expressing a strong preference for a screen based interpretation for the interface. There is then a tension between the children's perception that the robot has a capability to learn, a facet they do not attribute to other digital technologies, and the teachers' preference for a PC based system. For a system design to be successful, it will need to accommodate these opposing viewpoints.

Chapter 5

A Review of HCI-AI and CCI-AI

5.1 Introduction

This chapter provides a semi-systematic mapping review of the literature on HCI and CCI research related to AI. The terms HCI-AI and CCI-AI are used to describe the intersection between the disciplines. These are catch-all terms used to describe this corpus and are distinct from Shneiderman's Human-Centered AI although there is clearly some crossover (Shneiderman, 2022).

An aim of the review was to classify existing research using current taxonomies or definitions of the field of AI. Any attempt to perform this task needs to reflect the diverse nature of the discipline and its practitioners, which to compile would be a major undertaking in itself. Instead, the AI taxonomy developed by AI Watch, the European Union's service "to monitor the development, uptake and impact of Artificial Intelligence", is used to classify and map the literature (Samoili et al., 2020).

In reviewing the literature, three approaches are adopted. First, Natural Language Processing (NLP) is used to perform semantic labelling of the research and second, the papers are classified using the AI domain and subdomains presented in Table 5.4. Finally, the research methods employed to produce the research were classified using the same AI taxonomy.

5.2 A Mapping Review of Current HCI and CCI Research relating to Artificial Intelligence

A mapping review utilising an adapted SALSA framework (Search, Assessment, Synthesis, and Analysis) (Grant and Booth, 2009) (Figure 5.1) was conducted with the aim of mapping existing HCI and CCI research against the taxonomy described in the JRC Technical Report, AI Watch Defining Artificial Intelligence (Samoili et al., 2020). Additionally, the research methods used by the researchers in producing their papers were identified with the aim of classifying the research approaches.

The appraisal and synthesis processes, commonly regarded as separate phases of the review process, were combined for this review. Quality control, in this case the appraisal of papers erroneously returned by the search, was a by product of the data classification which is part of the synthesis stage. A second stage of synthesis then took place to perform a semantic classification on each of the selected papers.

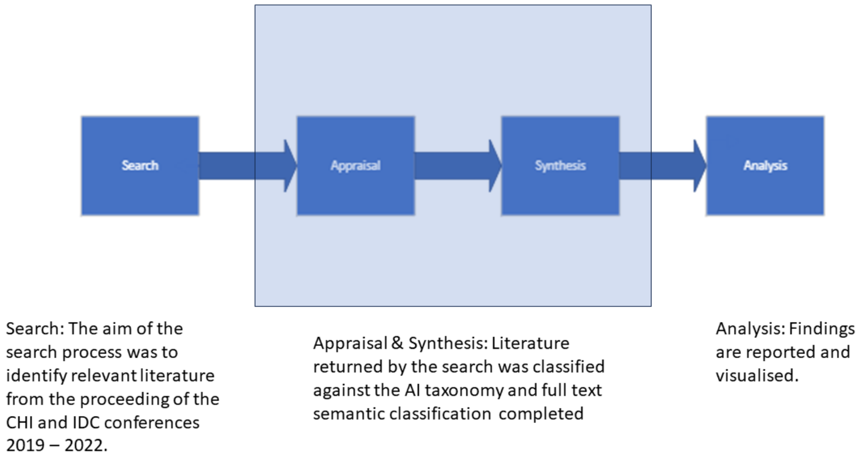


Figure 5.1: Adapted SALSA Framework

5.2.1 Method

Papers from ACM Special Interest Group on Computer-Human Interaction (CHI) conferences 2019 - 2022 inclusive and the ACM Interaction Design and Children (IDC) conferences 2019 - 2022 inclusive[1] were considered for inclusion in the review. The conferences, respectively, describe themselves as the "leading worldwide conference on HCI" (CHI, 2022) and the "premier conference on inclusive child-centered design, learning and interaction" (IDC, 2023*b*).

CHI papers were filtered using the online digital programme for each conference [2].The CHI conference is organised into sessions and papers were included for the sessions that were returned by querying the programme using the four keyphrases described in (Table 5.1):

(Figure 5.2) shows the process used for literature selection for the CHI '22 conference using the key phrase "Intelligent Systems". The search

[1]For further information on ACM conferences see https://www.acm.org/conferences

[2]The CHI conference programmes can be found online in the format https://programs.sigchi.org/chi/*yyyy*/search/content eg for CHI '20 the corresponding URL is https://programs.sigchi.org/chi/2020/search/content

Table 5.1: CHI 2019 - 2022 Search Phrases

	Search Phrase
1	Artificial Intelligence
2	AI
3	Machine Learning
4	Intelligent Systems

returned 5 sessions, 3 of which were paper sessions, one session was a special interest group (SIG) and the other a panel. The papers listed within each session were then located in the ACM Digital library conference proceedings (CHI, 2023*b*).

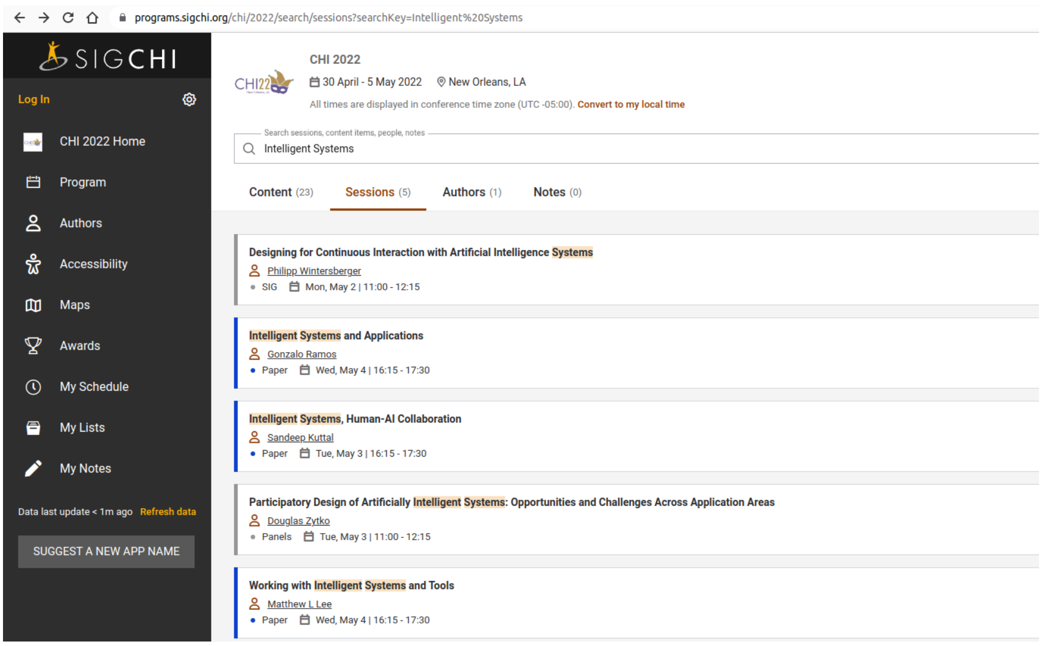


Figure 5.2: CHI Literature Search

(Figure 5.3) illustrates the four papers included in the review presented at the “Intelligent Systems and Applications” session in 2022.

Table 5.2: CHI Paper Breakdown 2019 - 2022

Year	Accepted Papers [3]	Sessions	Considered Papers	Included Papers (HCI-AI)	Paper Percentage
2019	703	423	22	20	3%
2020	760	467	27	26	3%
2021	747	359	58	37	5%
2022	637	414	32	30	5%

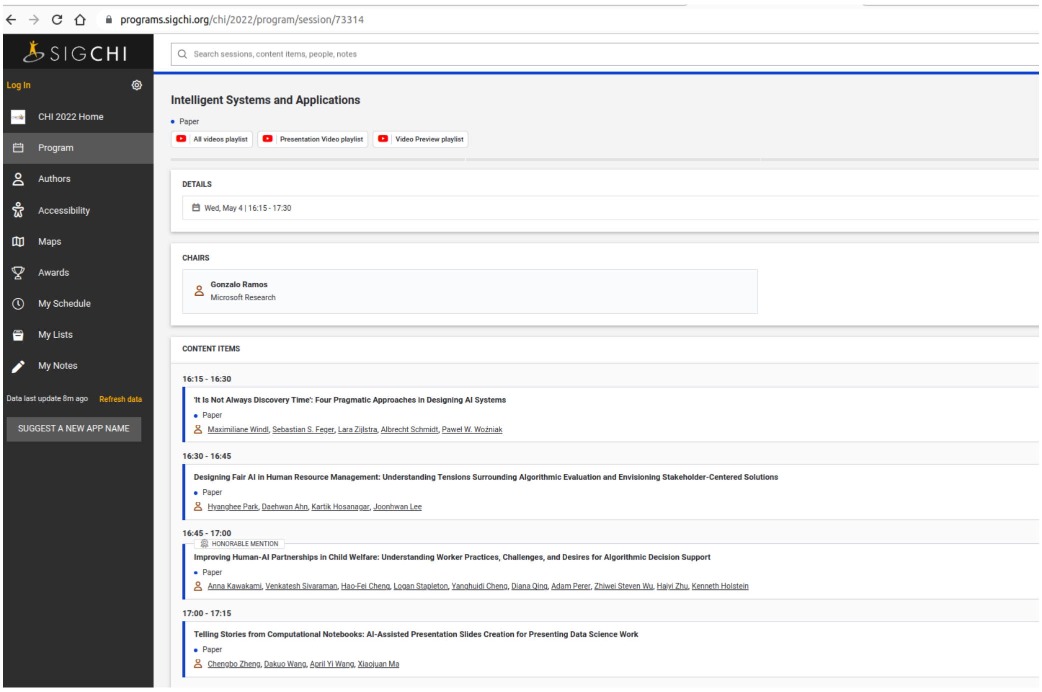


Figure 5.3: CHI Paper Search

(Table 5.2) summarises the total number of papers accepted for the conference (Accepted Papers) along with the number of papers that met the search criteria (Considered Papers) and the papers that form part of the analysis following the synthesis (Included Papers). Included papers are identified as HCI papers which fall within the AI domain and are referred to as HCI-AI studies or papers. During the 4 years, 113 papers out of a total of 2847 (4%) met the criteria for inclusion within the review.

The IDC conference is smaller than CHI and is not structured in the same way, notably there were no specific sessions matching the key phrase searches used to filter the CHI proceedings. Accordingly, full text searches using the same four key phrases (Table 5.1) were conducted against all articles included in the conference proceedings IDC (2023*a*). These included full papers, short papers, and works in progress (WIPs) and excluded workshops and the Doctoral Consortium. Papers matching one or more of the key phrases in their text were included in the review (Table 5.3). These papers referred to as CCI-AI papers amounted to 77 papers from a total of 261 accepted (30%).

Table 5.3: IDC Paper Breakdown 2019 - 2022

Year	Accepted Papers	Considered Papers	Included Papers (CCI-AI)	Paper Percentage
2019	41	13	7	17%
2020	81	23	16	20%
2021	77	38	28	36%
2022	62	37	26	41%

Review Synthesis and the AI Taxonomy

The process of selecting the papers required the author to assess the full text of each article for content associated with AI. AI is a broad and multifaceted field which, in their attempt to provide an operational definition for AI, Samoili et al. (2020) proposed a taxonomy along with a related set of keywords (Table 5.4).

This taxonomy is used here with the aim of mapping the state of existing AI research within the HCI (HCI-AI) and CCI (CCI-AI) research communities. The same taxonomy is then used to classify the research methods employed by the researchers in producing the contributing papers. Papers in which the research subject did not fit within the AI domains and the AI subdomains described in the taxonomy were discarded at this point and are not included in the further analysis described below. The full classification is included in (Appendix E).

Table 5.4: AI Watch Taxonomy of Artificial Intelligence Samoili et al. (2020)

AI domain	AI subdomain	Keyword	
	Knowledge representation;	case-based reasoning	inductive programming
		causal inference	information theory
	Automated reasoning;	causal models	knowledge representation & reasoning
Reasoning		common-sense reasoning	latent variable models
	Common sense reasoning	expert system	semantic web
		fuzzy logic	uncertainty in artificial intelligence
		graphical models	
	Planning and Scheduling;	bayesian optimisation	hierarchical task network
		constraint satisfaction	metaheuristic optimisation
Planning	Searching;	evolutionary algorithm	planning graph
		genetic algorithm	stochastic optimisation
	Optimisation	gradient descent	
		active learning	feature extraction
		adaptive learning	generative adversarial network
		adversarial machine learning	generative model
		adversarial network	multi-task learning
		anomaly detection	neural network
		artificial neural network	pattern recognition
		automated machine learning	probabilistic learning
		automatic classification	probabilistic model
		automatic recognition	recommender system
		bagging	recurrent neural network
Learning	Machine learning	bayesian modelling	recursive neural network
		boosting	reinforcement learning
		classification	semi-supervised learning
		clustering	statistical learning
		collaborative filtering	statistical relational learning
		content-based filtering	supervised learning
		convolutional neural network	support vector machine
		data mining	transfer learning
		deep learning	unstructured data
		deep neural network	unsupervised learning
		ensemble method	
		chatbot	natural language generation
		computational linguistics	machine translation
		conversation model	question an swering
Communication	Natural language processing	coreference resolution	sentiment analysis
		information extraction	text classification
		information retrieval	text mining
		natural language understanding	
		action recognition	object recognition
		face recognition	recognition technology
	Computer vision	gesture recognition	sensor network
		image processing	visual search
		image retrieval	
Perception		computational auditory scene	sound synthesis
		music information retrieval	speaker identification
	Audio processing	sound description	speech processing
		sound event recognition	speech recognition
		sound source separation	speech synthesis
		agent-based modelling	negotiation algorithm
		agreement technologies	network intelligence
	Multi-agent systems	computational economics	q-learning
		game theory	swarm intelligence
		intelligent agent	
Integration and Interaction		cognitive system	robot system
	Robotics and Automation	control theory	service robot
		human-ai interaction	social robot
		industrial robot	
	Connected and	autonomous driving	self-driving car
	Automated vehicles	autonomous system	unmanned vehicle
		autonomous vehicle	
		ai application	intelligence software
		ai benchmark	intelligent control
		ai competition	intelligent control system
Services	AI Services		

AI domain	AI subdomain	Keyword	
		ai software toolkit	intelligent hardware development
		analytics platform	intelligent software development
		big data	intelligent user interface
		business intelligence	internet of things
		central processing unit	machine learning framework
		computational creativity	machine learning library
		computational neuroscience	machine learning platform
		data analytics	personal assistant
		decision analytics	platform as a service
		decision support	tensor processing unit
		distributed computing	virtual environment
		graphics processing unit	virtual reality
AI Ethics and Philosophy	AI Ethics	accountability	safety
		explainability	security
		fairness	transparency
		privacy	
	Philosophy of AI	artificial general intelligence	weak artificial intelligence
		strong artificial intelligence	narrow artificial intelligence

5.2.2 Semantic Labelling and Natural Language Processing

As a second stage of synthesis, all the HCI-AI and CCI-AI papers identified for inclusion in the review were processed using two Natural Language Processing (NLP) algorithms, Term Frequency – Inverse Document Frequency (TF-IDF) (Ramos et al., 2003) and Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019).

TF-IDF uses three components, the Term Frequency (TF) which is defined as:

$$\frac{\text{The frequency of a word (w) within a text}}{\text{number of words in the text}}$$

Inverse Document Frequency (IDF), effectively the importance of a word in the text:

$$\log\left(\frac{\text{The number of sentences in the text}}{\text{The sentences including w}}\right)$$

TF-IDF, a score to measure the importance of w:

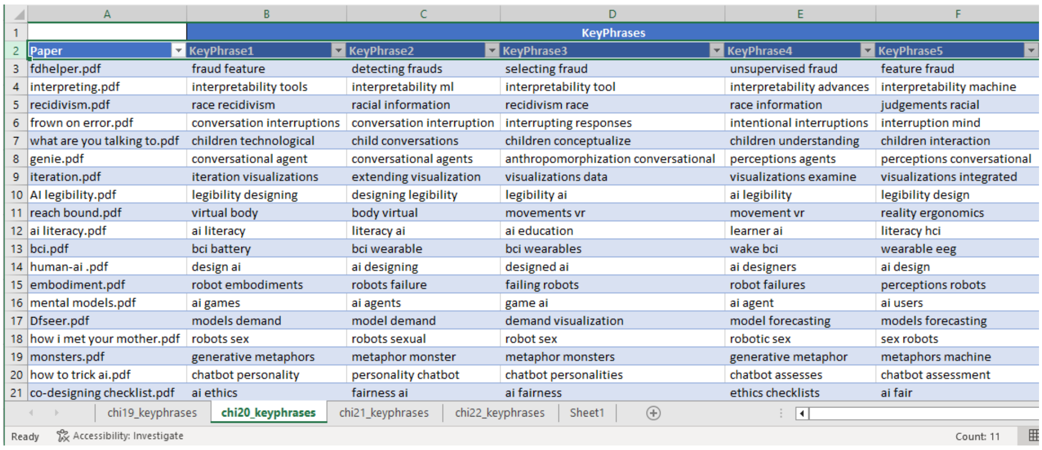
$$\text{TF x IDF}$$

BERT is an unsupervised NLP algorithm released by Google in 2018 that greatly improves performance in areas of NLP processing including semantic role labelling. The algorithm prevents the word currently being processed from assigning itself a meaning or from having a meaning independent of its context. Using a technique called masked language modelling the masked word is determined by BERT based solely on its context.

Each of the papers was downloaded from the ACM digital library in PDF format, converted to plaintext before being processed using the Python PyPDF2 library [4]. The TF-IDF implementation returned the top 5 keywords identified for each paper and the BERT implementation returned the top 5 two-word key phrase based on the full text. (Listing 5.1) shows the process for the keyphrases generated using the Python KeyBERT library [5] for the selected papers submitted to CHI '22.

The generated key phrases were then used to make a semantic classification of the papers. (Figure 5.4) shows the 2 word key phrases generated by the BERT algorithm for a subsection of the CHI '20 conference where column A is the file name and the subsequent 5 columns are the key phrases. A full listing of the key phrases generated by the BERT algorithm is available in (Appendix F).

[4] https://pypi.org/project/PyPDF2/
[5] https://pypi.org/project/keybert/



	A	B	C	D	E	F
1		KeyPhrases				
2	Paper	KeyPhrase1	KeyPhrase2	KeyPhrase3	KeyPhrase4	KeyPhrase5
3	fdhelper.pdf	fraud feature	detecting frauds	selecting fraud	unsupervised fraud	feature fraud
4	interpreting.pdf	interpretability tools	interpretability ml	interpretability tool	interpretability advances	interpretability machine
5	recidivism.pdf	race recidivism	racial information	recidivism race	race information	judgements racial
6	frown on error.pdf	conversation interruptions	conversation interruption	interrupting responses	intentional interruptions	interruption mind
7	what are you talking to.pdf	children technological	child conversations	children conceptualize	children understanding	children interaction
8	genie.pdf	conversational agent	conversational agents	anthropomorphization conversational	perceptions agents	perceptions conversational
9	iteration.pdf	iteration visualizations	extending visualization	visualizations data	visualizations examine	visualizations integrated
10	AI legibility.pdf	legibility designing	designing legibility	legibility ai	ai legibility	legibility design
11	reach bound.pdf	virtual body	body virtual	movements vr	movement vr	reality ergonomics
12	ai literacy.pdf	ai literacy	literacy ai	ai education	learner ai	literacy hci
13	bci.pdf	bci battery	bci wearable	bci wearables	wake bci	wearable eeg
14	human-ai .pdf	design ai	ai designing	designed ai	ai designers	ai design
15	embodiment.pdf	robot embodiments	robots failure	failing robots	robot failures	perceptions robots
16	mental models.pdf	ai games	ai agents	game ai	ai agent	ai users
17	Dfseer.pdf	models demand	model demand	demand visualization	model forecasting	models forecasting
18	how i met your mother.pdf	robots sex	robots sexual	robot sex	robotic sex	sex robots
19	monsters.pdf	generative metaphors	metaphor monster	metaphor monsters	generative metaphor	metaphors machine
20	how to trick ai.pdf	chatbot personality	personality chatbot	chatbot personalities	chatbot assesses	chatbot assessment
21	co-designing checklist.pdf	ai ethics	fairness ai	ai fairness	ethics checklists	ai fair

Figure 5.4: CHI '20 Keyphrases Generated using BERT

Listing 5.1: Keyphrase generation using BERT

```
import PyPDF2
from keybert import KeyBERT
import os

root_dir = 'papers'
sub_dir = 'chi22'
path = root_dir + '/' + sub_dir + '/'
f = open(sub_dir + "_keyphrases.csv", "w")
# Get all pdfs
with os.scandir(path) as entries:
    for entry in entries:
        # Read the pdf and convert to text
        try:
            pdfFileObj = open(path + entry.name, 'rb')
            pdfReader = PyPDF2.PdfFileReader(pdfFileObj)
            # print("Page Number:", pdfReader.numPages)
            num_pages = len(pdfReader.pages)
        except:
            print('Error:' + entry.name)

        paper = ''
        for i in range(0, num_pages):
            pageObj = pdfReader.pages[i]
            # extract text from page
            paper += pageObj.extractText()
        # Strip out the pdf line breaks
        paper = paper.replace('\n', ' ')
        kw_model = KeyBERT()
        keywords = kw_model.extract_keywords(paper, keyphrase_ngram_range=(1, 2))
        # print(keywords)
        keyphrases = entry.name + ', '
        for keyword in keywords:
            #print(keyword)
            keyphrases  += f'{keyword[0]}, '
        keyphrases = keyphrases[:-2] # strip the trailing comma and space
        print(keyphrases)
        f.write(keyphrases)
        f.write('\n')
f.close()
```

The TF-IDF algorithm ranks the relevance of words within a document by the frequency with which they occur. It is quick and easy to implement, but cannot provide an indication of semantic relevance (ie words that are conceptually related). Therefore, the keywords generated were not used further to identify themes in the papers. BERT, whilst significantly more complex and resource-hungry, provides the semantic relevance required to provide an indication of the paper's context. A Word Cloud based on the BERT keyphrases was generated for each conference. Each Word Cloud was restricted to a maximum of 50 words for clarity of display (Listing 5.2).

Listing 5.2: Word Cloud generation from BERT keyphrases

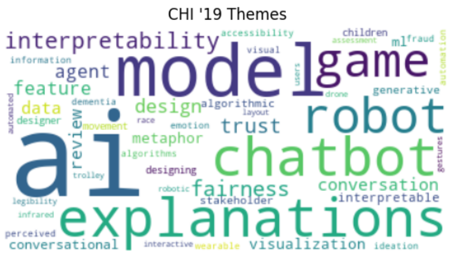
```
import numpy as np
import pandas as pd
from os import path
from PIL import Image
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
import itertools
import matplotlib.pyplot as plt
get_ipython().run_line_magic('', 'matplotlib inline')


# Load in the dataframe
df = pd.read_csv("chi19_keyphrases.csv", header=None)
#strip out the pdf filename
df.drop(df.columns[0], inplace=True, axis=1)
nested = df.values.tolist()
raw_keyphrases = list(itertools.chain(*nested))
keyphrases = list()
kp = ''
for keyphrase in raw_keyphrases:
    keyphrase = keyphrase.strip()
    keyphrase = keyphrase.replace(' ','~')
    #print(keyphrase)
    keyphrases.append(keyphrase)
    kp += f'{keyphrase} '
wordcloud = WordCloud(collocations = False, background_color = 'white', max_words=50).generate(kp)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.title("CHI '19 Themes")
plt.show()
```

5.2.3 Results

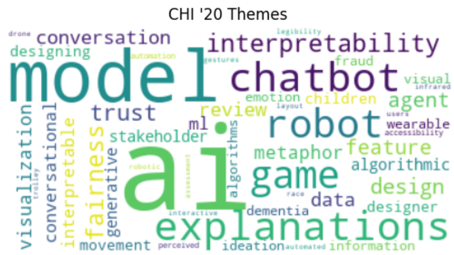
(Figure 5.5) shows the BERT analysis of the included HCI-AI papers as a Word Cloud and (Figure 5.6) shows the analysis of the CCI-AI papers. All CHI word clouds indicate that “AI” was an important theme of the HCI-AI papers. Other keywords common across the papers are “chatbot” and in three cases “robot”.

The prevalent themes in the IDC papers, (Figure 5.6) are focused around the core principles of the research group, namely “children” and “interaction”. There is little in the keyword analysis to indicate that AI is a growing research field within the CCI community and “AI” is not included in the top 50 keywords returned by the analysis for any of the years. This is despite the fact that a significantly higher percentage of accepted IDC articles (30%) compared to only 4% of CHI papers were classified as AI related.

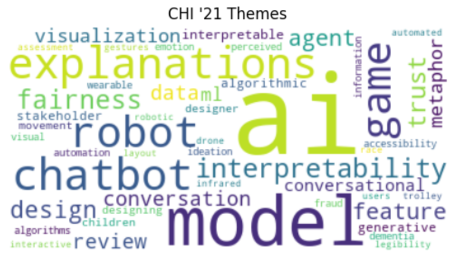
Robots are indicated as a prominent theme in CCI-AI research as are conversational agents. (Figure 5.7) groups the keywords for both CHI and IDC. Below the top level HCI-AI has perhaps the broader scope and indicates research interest in aspects such as AI Ethics including keywords such as “interpretability”, “fairness” and “trust” whereas CCI-AI is more focused on interaction with the child. It is worth emphasising that the research evaluated here is related to how people interact with AI based systems and so it should be expected that a thematic analysis would bring out these human factors.



(a) CHI 2019 HCI-AI Research Themes



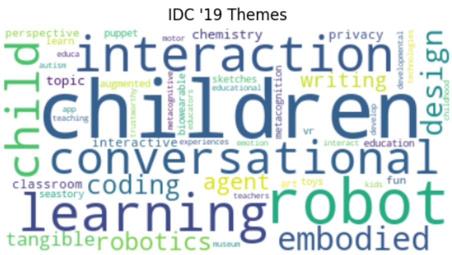
(b) CHI 2020 HCI-AI Research Themes



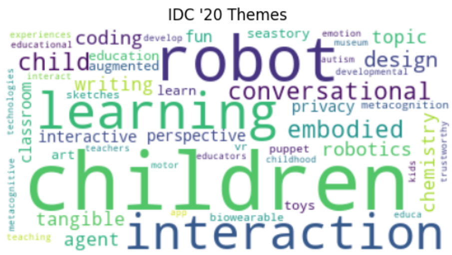
(c) CHI 2021 HCI-AI Research Themes

(d) CHI 2022 HCI-AI Research Themes

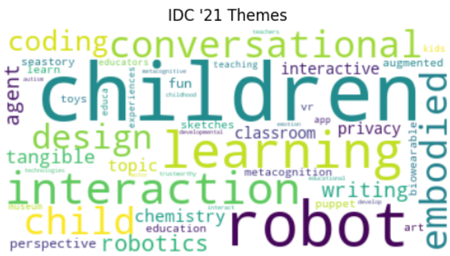
Figure 5.5: CHI 2019 - 2022 Research Themes by Conference



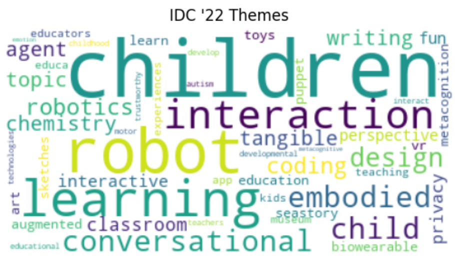
(a) IDC 2019 CCI-AI Research Themes



(b) IDC 2020 CCI-AI Research Themes

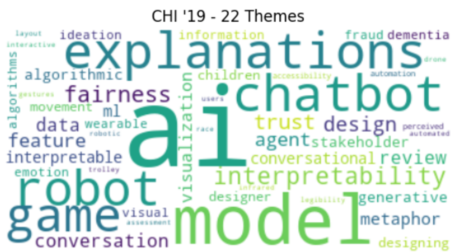


(c) IDC 2021 CCI-AI Research Themes

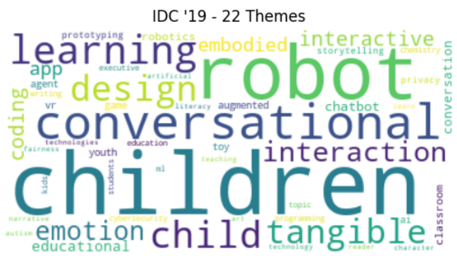


(d) IDC 2022 CCI-AI Research Themes

Figure 5.6: IDC 2019 - 2022 Research Themes by Conference



(a) CHI 2019-22 HCI-AI Research Themes



(b) IDC 2019-22 CCI-AI Research Themes

Figure 5.7: CHI - IDC 2019 - 2022 Research Themes Summary

Research Subject Classification CHI

The papers were then reviewed by the researcher and classified by subject using the domains and subdomains presented in the taxonomy. (Table 5.5) provides a breakdown of the subject areas covered by the AI related papers presented at CHI ’19 - CHI ’22 grouped by AI Domain and AI Subdomain. (Figure 5.8) gives a graphical view of the same data.

Table 5.5: CHI 2019 - 2022 HCI-AI Research Subject Classification

Subject Classification	**Papers by Conference**			
AI Domain	CHI ’19	CHI ’20	CHI ’21	CHI ’22
AI Ethics and Philosophy	4	14	16	3
Communication	2		2	8
Integration and Interaction		1	1	3
Learning	1		3	6
None AI	2	1	21	2
Perception	2			
Planning				
Reasoning				
Services	11	11	15	10
Totals	22	27	58	32

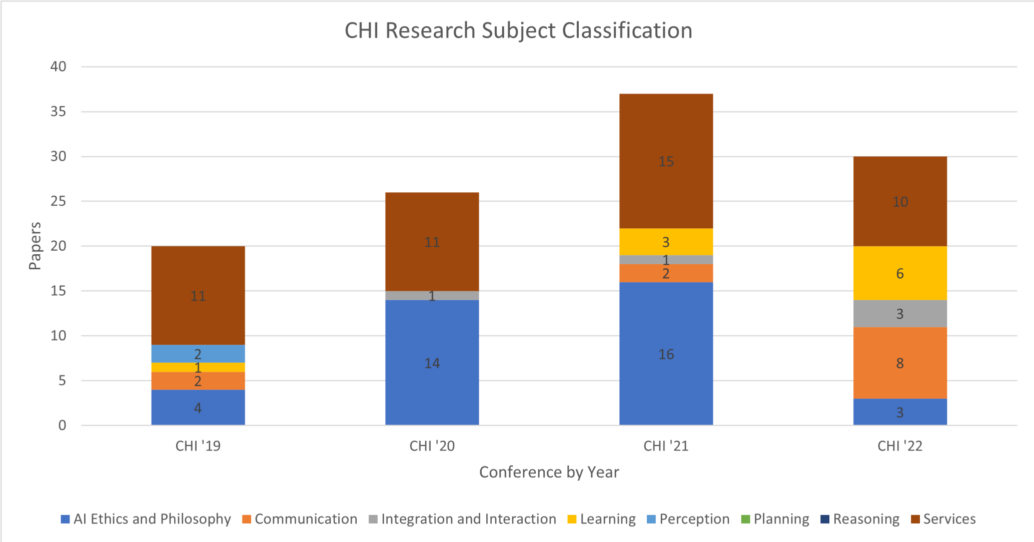


Figure 5.8: CHI 2019 - 2022 HCI-AI Research Subject Classification

In 2019, half of the articles considered for the review (eleven) are categorised as relating to AI Services. At keyword level, the majority of work is related to intelligent systems and interfaces (Chen et al., 2019), (Choi et al., 2019), (Constant and Levieux, 2019), (Jiang et al., 2019) but also work related to computational creativity (Guzdial et al., 2019), (McCormack et al., 2019), (Koch et al., 2019). Two papers described in the CHI Programme Sessions as AI related did not fit within the AI classification. Türkay and Adinolf (2019)'s paper included under the "Playing and AI" session examines collectable card games online and the community, whilst Phelan et al. (2019) included in the "Machine Learning and HCI" session reviews methods for statistical validation of research. Whilst this is undoubtedly applicable for AI researchers, the thrust of the paper discusses substituting Bayesian statistical approaches for more commonly used methods and does not directly address AI related research.

In 2020, fourteen papers are classified as belonging to the AI Ethics and Philosophy domain. In the AI Ethics subdomain, several papers examined the challenges faced in designing AI systems (Lindley et al., 2020), (Long and Magerko, 2020) while others looked at embodiment choices Troiano et al. (2020), Kontogiorgos et al. (2020) including anthropomorphism Kuzminykh et al. (2020). Dove and Fayard (2020)'s paper is unique in looking at some of the larger challenges of AI development framed in the context of the human-monster relationship and is the first paper to be classified under the AI Sub Domain “Philosophy of AI”.

AI Services was the second largest domain with researchers authoring content relating to data analytics (Sun, Li, Chen, Lee, Liu, Zhang, Huang, Shi and Xu, 2020), (Hohman et al., 2020), (Asai et al., 2020), intelligent user interfaces (Yan et al., 2020), (Xu and Warschauer, 2020*c*) and intelligent software development (Agarwal and Sivakumar, 2020), Cheema et al. (2020). Okuya et al. (2020) investigated the use of wall-sized displays as an aid to industrial design review and although this may have applications within the development of AI systems, it does not sit within the chosen taxonomy and is therefore classified as “None AI” and not considered within the analysis.

On first view CHI 2021 looks to contain significantly more AI related content than either CHI 2019 or CHI 2020, fifty-eight papers as opposed to twenty-two papers in 2019 and twenty-seven papers in 2020. On classification, twenty-one papers were found not to be AI related leaving 37 papers for consideration. The main reason for this was an expansion of the content within the programme sessions to reflect a broader grouping of the content, eg, Design Tools / Machine Learning / Fabrication / Visual Artefacts in Design Ideation (Figure 5.9). Therefore, a session-level keyword for “Machine Learning” returned results for unrelated papers examining fabrication (Miyatake et al., 2021), (Lakshmi et al., 2021). Not all the papers in this category are miscategorised, Lin and Brummelen (2021) are a notable exception writing about training for teachers wishing to embed AI in the curriculum. The taxonomy does not classify AI education, and as such, their paper is not included in the further analysis.

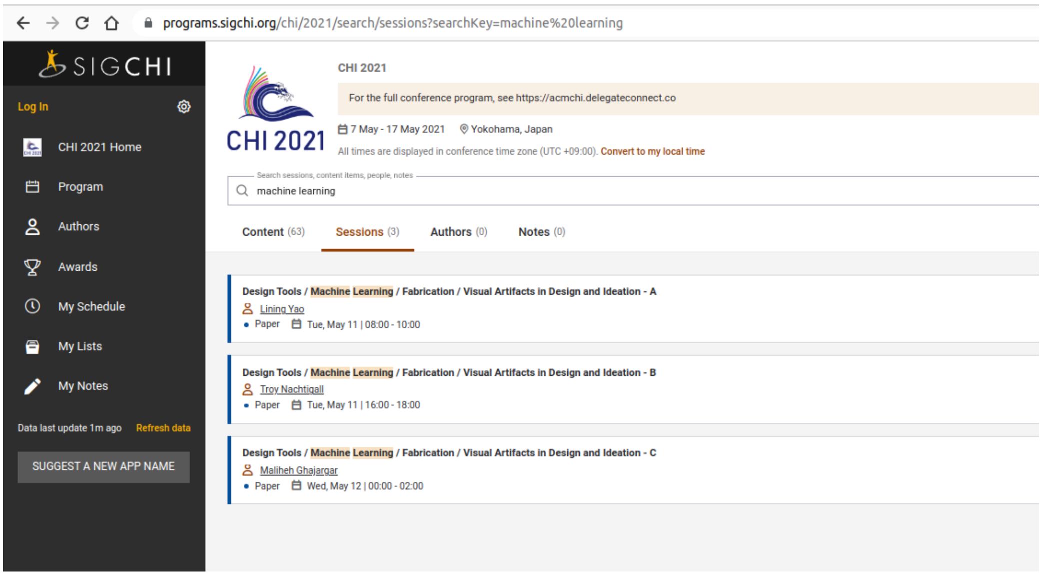


Figure 5.9: CHI Session Expansion

In 2022, thirty-two papers are listed in the proceedings as HCI-AI related. Two papers (Renom et al., 2022), (Pang et al., 2022) were classified as "None AI" topics. Notable is the increase in research classified as related to natural language processing, which accounts for 25% of the HCI-AI papers published in the proceedings in that year (Table 5.6).

Table 5.6: Natural Language Processing Breakdown 2019 - 2022

Year	Included Papers	Natural Language Processing Papers
2019	22	2
2020	27	0
2021	58	2
2022	32	8

Table 5.7: CHI 2019 - 2022 HCI-AI Implementation Method Classification

Implementation Classification	**Papers by Conference**			
AI Domain	CHI ’19	CHI ’20	CHI ’21	CHI ’22
AI Ethics and Philosophy	1			
Communication			2	1
Integration and Interaction	1	2	1	1
Learning	15	8	11	3
None AI	1	7	12	11
Perception		1		
Planning				
Reasoning				
Services	2	8	11	14
Sub Total	20	26	37	30
Not considered	2	1	21	2
Total	22	27	58	32

Research Implementation Classification CHI

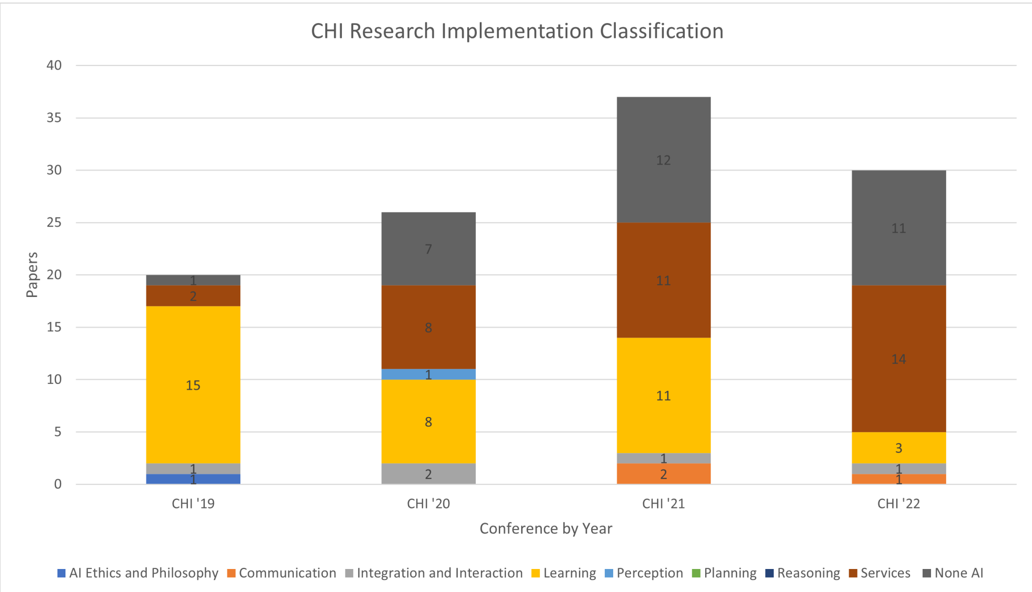


Figure 5.10: CHI 2019 - 2022 HCI-AI Implementation Method Classification

(Table 5.7) summarises the implementation techniques used by the researchers using the same taxonomy to perform the classification. (Figure

5.10) visualises the same information.

Most of the studies in 2019 (15) leveraged some form of machine learning, mostly supervised learning (Wang, Yang, Abdul and Lim, 2019) (Chen et al., 2019). McCormack et al. (2019) and Choi et al. (2019) used deep learning forms while a single paper used unsupervised techniques (Arakawa and Yakura, 2019). Innovatively Williams et al. (2019) used Popbots, a Robotics and Automation platform, to teach preschool children about AI concepts. Only three papers do not use AI as a basis for their work. Samson and Sumi (2019) use surveys to explore driver routing decisions. Phelan et al. (2019) developed R templates to help researchers implement Bayesian statistical techniques, and Türkay and Adinolf (2019) also employed surveys as their main data gathering technique.

In 2020, 19 of the 26 studies analysed used some form of AI as the basis for their findings. Of the studies that did not employ AI techniques, the majority used some form of mixed methodology employing surveys and online studies (Wang et al., 2020) as well as qualitative approaches (Dove and Fayard, 2020). Long and Magerko (2020) conducted a literature review as their main research medium. This is not to say that other papers did not position themselves within the literature! Rather, the literature was neither the main output nor the vehicle for the research. It should also be noted that several studies employ more than one method, one of which may be AI related, Kuzminykh et al. (2020) refer to their work as “qualitative multi-phase study” and use conversational agents as a medium. In such cases, the primary research method or vehicle is used to make the classification.

In 2021, discounting the 21 papers not considered as they fall outside the scope of this review, the largest single category “None AI” are the researchers who did not employ AI as part of their research methods. Researchers applied a variety of methods including qualitative studies (Hughes and Roy, 2021), surveys (Anik and Bunt, 2021), case studies Benjamin et al. (2021) and mixed methods (Hong et al., 2021). The subject classification for eight of the twelve studies was AI Ethics and Philosophy with papers reporting on aspects of the machine learning process such as morality (Lima et al., 2021), fairness (Park et al., 2021), (Cheng et al., 2021), and interpretability Suresh et al. (2021).

2022 produced 30 HCI-AI papers with eleven using some form of AI service or application as the basis on which to implement their research. Several researchers developed an AI application or hardware and evaluated some aspect of its performance or behaviour. Yan et al. (2022) produced smart eyewear and evaluated its impact on emotional health, while Zheng, Wang, Wang and Ma (2022) developed an AI-based application to aid in the preparation of data science visualisations and evaluated their effectiveness.

Research Subject Classification IDC

As is remarked in Section 5.2.3 the theme of the IDC papers is more concerned with the core subject matter of their research group, namely children and the way they interact with technologies. As also noted when describing the search method (see Section 5.2.1) for the review, IDC is significantly smaller than CHI and there are no specific AI related sessions within the conference. There is, however, a thread of CCI-AI running throughout the IDC proceedings and surprisingly pound for pound IDC has a larger AI representation than CHI.

(Table 5.8) summarises the CCI-AI research presented at IDC 2019 through to IDC 2022 and (Figure 5.11) provides a visualisation of the data.

Table 5.8: IDC 2019 - 2022 CCI-AI Research Subject Classification

Subject Classification	**Papers by Conference**			
AI Domain	IDC '19	IDC '20	IDC '21	IDC '22
AI Ethics and Philosophy			2	3
Communication	1	4	3	
Integration and Interaction	1	4	8	6
Learning	2	1	6	2
None AI	6	7	10	11
Perception				
Planning				
Reasoning				
Services	3	7	9	15
	13	23	38	37

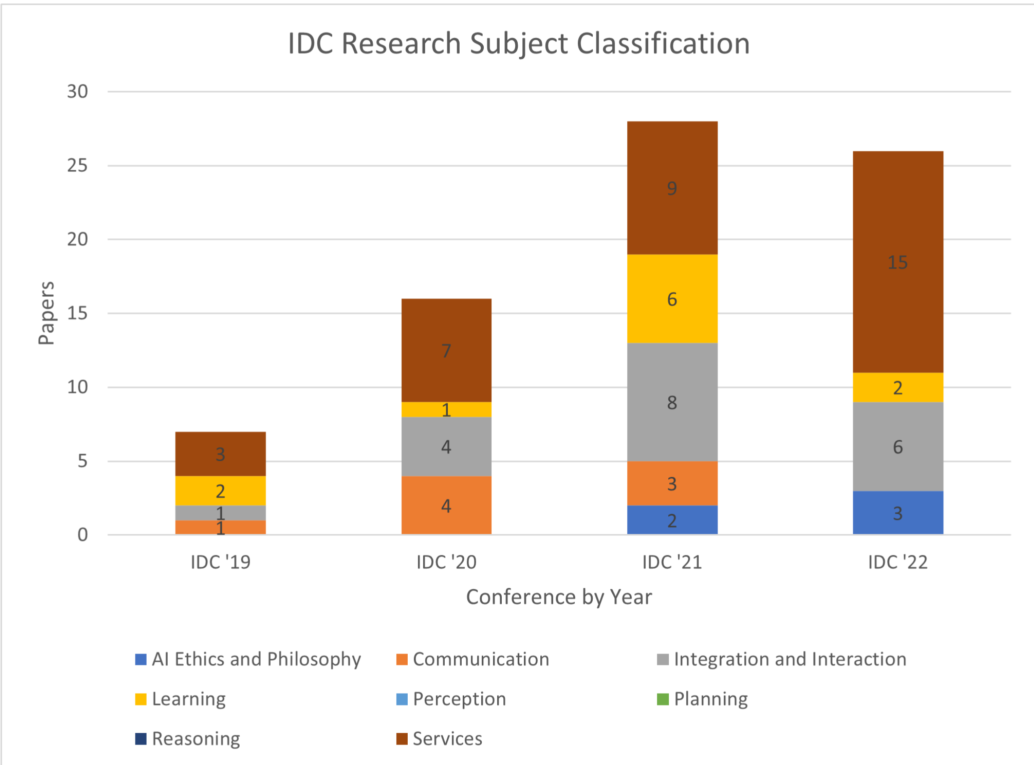


Figure 5.11: IDC 2019 - 2022 CCI-AI Research Subject Classification

Of the seven papers classified as CCI-AI at CHI 2019, three used an AI application as a significant part of their research. Kang et al. (2019) employed Augmented Reality AR as part of their research, while Sharma, Papavlasopoulou and Giannakos (2019) produced an application to explore the emotional state of children. Badillo-Urquiola et al. (2019) explored children's interaction with existing social media applications. How children interact with machine learning-based systems is explored by Fails et al. (2019) and Zimmermann-Niefield et al. (2019). In particular, there are no papers exploring AI Ethics and Philosophy in a CCI-AI context at IDC 2019 or IDC 2020 and only two papers at IDC 2021 (Charisi et al., 2021), (Melsión et al., 2021) and three papers at IDC 2022 (Zhao et al., 2022), (Escobar-Planas, 2022), (Antle et al., 2022).

AI applications are again well represented at IDC 2020 but perhaps of greater note are the papers classified under the top-level domain of "Integration and Interaction", which explore children's and in the case of van Ewijk et al. (2020) teachers interaction with social robots (Malinverni and Valero, 2020), (Cagiltay et al., 2020), (Boffi, 2020). Children's interaction with social robots is a well-established CCI research theme, and the author participated in a dedicated workshop at IDC 2018 on "Child Robot Interaction", which falls outside the scope of this review (Charisi, 2018)

Children's interaction with conversational agents is another recurring theme at the IDC conferences. This theme spans at least two AI domains, "Communication" specifically Natural Language Processing and "Services" specifically Intelligent User Interfaces. In reaching a classification, care was taken to identify the primary domain, but grey areas such as this require further work, namely, a second researcher to perform an independent classification of the data.

At IDC 2021 Petousi et al. (2021), Hiniker et al. (2021) and Motozawa et al. (2021) conducted research specifically related to how conversational agents affected children's perceptions, and this work has been classified as "Communication". On the contrary, at IDC 2022 Nguyen (2022*a*) employed a conversational agent as an intelligent interface, and this work is classified under “Services”. For a keyword-level analysis of all the papers considered in this review, see Appendix G

Research Implementation Classification IDC

(Table 5.9) summarises the CCI-AI implementation methods used at IDC 2019 to IDC 2022 and (Figure 5.12) provides a visualisation of the data.

Table 5.9: IDC 2019 - 2022 CCI-AI Research Implementation Classification

Implementation Classification	**Papers by Conference**			
AI Domain	IDC ’19	IDC ’20	IDC ’21	IDC ’22
AI Ethics and Philosophy				
Communication	1	4	3	
Integration and Interaction	1		8	4
Learning	1	1	6	3
None AI	2	8	1	6
Perception	1			
Planning				
Reasoning				
Services	1	3	10	13
Sub Total	7	16	28	26
Not considered	6	7	10	11
Total	13	23	38	37

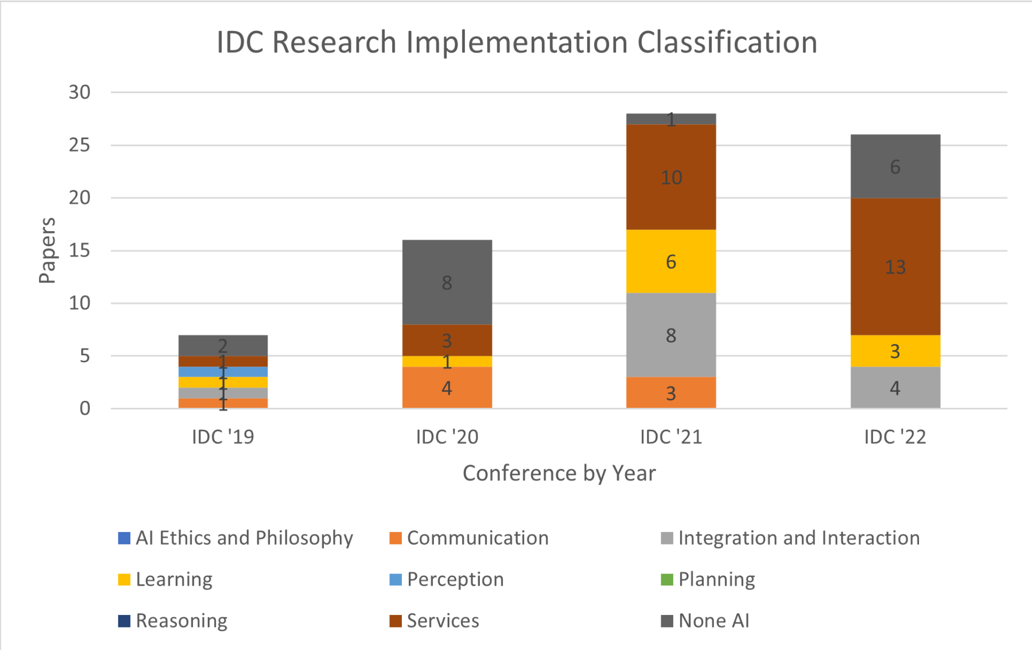


Figure 5.12: IDC 2019 - 2022 CCI-AI Research Implementation Classification

It should be noted that while “None AI” themes are excluded from the analysis of research themes for the review, “None AI” implementations are of course valid methods. Indeed, all the papers employ some aspect of research mode that is not AI based. Papers classed as “None AI” in this review, however, do not use AI in their research methods. The actual methods vary, El Shemy (2022) employ a literature review to look at Augmented Reality and AI related to autistic children. Both Ruan et al. (2020) and Boffi (2020) use “Wizard of Oz” techniques to simulate AI systems. Other writers use various techniques such as focus groups (van Ewijk et al., 2020) and studies (Shin and Holtz, 2020), (Yu and Roque, 2022). Cagiltay et al. (2020) use mixed methods, including a social robot in their work on “In-Home Robot”. In this case, the robot was not powered up, so the implementation is regarded as “None AI”.

Several studies fall into the category “Intergration and Interaction”. Most of these use robots to study children’s interactions with robots, e.g. (White et al., 2021), (Stower and Kappas, 2021), (Tolksdorf et al., 2021), (Cagiltay et al., 2020). The exceptions are Du and Breazeal (2022) who use a multi-agent system to explore “pedagogical agents” and Charisi et al. (2021) who explore fairness, classified under AI and ethics using social robots.

5.3 The Research Landscape

In both the HCI-AI and CCI-AI domains, there is a significant body of research into the interaction with AI based applications. This is of course to be expected from research groups whose prime aim is to study human interaction with technology-based systems. There is a greater emphasis on AI Ethics and Philosophy related research at CHI than IDC as a proportion of total papers. Conversely, IDC has a greater emphasis on Integration and Interaction than CHI predominantly consisting of research based around social robots. Both research groups have a small body of research around machine learning-based technologies.

Of equal interest is what is missing from the research landscape. There is minimal work at CHI around the domain of Perception (Arakawa et al., 2022), (Perusquia-Hernández et al., 2019) and none at IDC. This is particularly pertinent to the research presented in this thesis specifically the latter work which looks to employ computer vision techniques with children in order to classify engagement within the classroom. There is clearly an opportunity for research within this field.

There are no papers included in the review classified as belonging to the “Reasoning” or “Planning” AI Domains. These areas are categorised by more theoretical concepts such as “fuzzy logic” or “stochastic optimisation”, and it is perhaps not surprising that they are not represented in this corpus of work which looks at the application of the AI fields as opposed to the theory that underpins them.

Of more interest is that only two papers were classified in the subdomain “Philosophy of AI” at CHI (Dove and Fayard, 2020), (Zhu et al., 2021) and no IDC papers. This is perhaps surprisingly low given the current media coverage (Valance, 2023), (Vallance, 2023) predicting apocalyptic consequences as the big technology companies push on with developing AI based chatbots such as ChatGPT (Schulman et al., 2023) which move us closer towards implementing strong artificial intelligence. However, a search for “chatbot” in the CHI ’23 programme (CHI, 2023*a*) - which is not within the scope of this review generated twenty-three hits as opposed to eleven hits in 2021 and nine in 2022 so it is realistic to hypothesise that this may well be a developing trend in future work.

Finally, it is noted that Samoili’s taxonomy, (Samoili et al., 2020) in most cases had sufficient coverage to satisfactorily classify the corpus of papers which constitute this review. There are some omissions at keyword level in the “AI Ethics” subdomain where some pertinent concepts are not represented, Namely interpretability (Poursabzi-Sangdeh et al., 2021), morality (Lima et al., 2021) and of pertinence to this thesis, trust (Jiang et al., 2021).

5.4 Limitations of the Review

In presenting this review as semi-systematic there is a need to explore its limitations. Only papers from the CHI and IDC conferences have been considered for inclusion. While these are self-styled as the leading conferences in their fields it must be noted that there are other conferences and journals publishing in these areas.

While all the papers from IDC were considered for review, CHI papers were filtered based upon the conference’s own pre-classification criteria and as such relevant work not meeting those criteria may have been omitted.

Finally, research classifications were made solely by the author of this thesis. In doing so it should be acknowledged that a second opinion would have been beneficial, particularly where work could have been classified in multiple categories.

5.5 Summary

This chapter describes a semi-systematic review of the HCI-AI and CCI-AI papers presented at CHI ’19 to CHI ’22 and IDC ’19 to IDC ’22 inclusive. The papers were evaluated for their semantic content and then both the research output and the research methods were classified using the taxonomy produced by Samoili et al. (2020).

Of significance to this thesis is the lack or research into the development of intelligent systems that use computer vision algorithms with children.

Chapter 6

The Peer Data Labelling System (PDLS)

6.1 Introduction

Chapters 3 and 4 present design considerations when creating an IS to be used by children. This chapter outlines the importance of data when building an IS to be deployed in an educational context. It presents a novel and extensible approach to generating labelled data suitable for training supervised machine learning algorithms for use in Child Computer Interaction (CCI) research and development, called the Peer Data Labelling System (PDLS). PDLS is evaluated against the usability metrics, effectiveness, efficiency, and satisfaction and is judged to be both efficient and satisfactory. A further analysis of its effectiveness is discussed in Chapter 7.

It concludes by offering some thoughts that are intended to be helpful to other researchers who may wish to carry out similar studies and propose the development of a data set that can be used as a resource for members of the CCI community who wish to undertake CCI research on emotion recognition or the application of computer vision to research with children.

In the UK alone there are more than 9 million children in school classrooms engaged in learning activities (*Schools, pupils and their characteristics, academic year 2021/22*, 2022). Learning is a complex process that relies on many different factors including the teacher's skill in maintaining pupils' attention to their learning activities so that they complete any set tasks. Traditionally, teachers walked around the class and kept an eye on what the pupils were doing, but as technology came into classrooms, new approaches were possible. Dating from the early 20th century (Léon, 1962), language laboratories were one of such innovations, where students sat in booths with earphones on and accessed content from a console controlled by the teacher who could listen to their responses and monitor their progression. More recently, this idea was extended to other systems, for example, where a teacher could see the screen of a pupil who was working on a task.

As children use more technology in the classroom, it becomes enticing to consider what the computer might be able to do independently to keep a child engaged on a task. With web cams on most devices, and with sophisticated computer vision technology, it should be possible to "watch" children, monitor their progress and intervene when they disengage. For this computational task to be successful, the computer or system needs to be able to "see" a child's face and movement and identify if he/ she is engaged on task or not; this requires a trained recogniser and, by extension, a suitable data set to perform the training.

For a system to be successful in "watching" children, there must be acceptance from the children in the system's ability to monitor their behaviour and trust in the system's ability to differentiate between engagement and disengagement Parsonage et al. (2023*b*). In this study, we explore these two facets of such a system by first considering the extent to which pupils can assist in the design of such a system and then exploring whether pupils would accept the system.

As indicated in Chapter 2, in CCI research it is common for children to participate in design activities. In this study, we 'employ' children as labellers of data by using their expertise to decide if a peer is engaged on task or not. This is considered to be a novel approach to assist in training a recogniser. This chapter includes reflections on the approach taken, survey findings that suggest pupils would accept such a method, and the beginning of a data set that others in the CCI community can use and develop.

This chapter proposes a novel and extensible approach to generating labelled data suitable for training supervised ML algorithms for use in CCI research and development, called the Peer Data Labelling System (PDLS). The novelty is in classifying one child's engagement using peer observation by another child, thus reducing the two-stage process of detection and inference common in emotion recognition to a single phase. In doing so, this technique preserves context at the point of inference, reduces the time and cost of labelling data retrospectively, and stays true to the CCI principle of keeping child participation central to the design process. The approach is evaluated using the usability metrics of effectiveness, efficiency, and satisfaction.

6.1.1 Learning and Engagement

Pupil engagement is widely considered a positive factor and an important driver of pupil attainment (Christenson et al., 2012*b*). There is a corpus of work dating back at least to Ralph Tyler's work in the 1930s linking time on task to attainment (Tyler, 2013), (Fisher and Berliner, 1985). Definitions of engagement range from a focus on interaction with a specific learning activity to a multidimensional approach requiring the pupil to engage at behavioural, affective and cognitive levels (Groccia, 2018*b*). The latter multidimensional approach extends the context of engagement beyond the immediate task, incorporating activities and interaction beyond the boundaries of the classroom into the home and the wider community (Reschly and Christenson, 2012). In this way, pupils can be engaged at one or more levels while simultaneously disengaging at another level (Trowler, 2010). This study considers engagement on task, namely a pupil's interaction with a computerised learning activity completed within a school classroom. Furthermore, while some scholars conceptualise engagement and disengagement as related but separate phenomena (Fredricks et al., 2004*b*), within this context, engagement and disengagement are treated as opposing ends of a single scale.

Education is now delivered over diverse and often distributed platforms and the Covid-19 lockdowns of 2020 and 2021 took education from all sectors out of the physical classroom and into virtual spaces offered by environments such as MS Teams and Zoom (Tandon, 2021*b*). Such environments provide their own challenges with regard to monitoring pupil engagement (Pokhrel and Chhetri, 2021*b*), (Oyedotun, 2020*b*), as in many cases, where pupils have no access to a webcam, the environments remove the visual cues that teachers rely on to monitor whether pupils are on task. In many online learning environments, academics and teachers frequently find themselves faced with a wall of tiled images with little indication of whether there is a human presence behind the facade. Although mainstream school level education across the world has largely returned to the physical classroom, the pandemic has fast-forwarded the development and adoption of hybrid and blended learning pedagogical approaches (Zhao and Watterston, 2021*b*) creating new requirements for tools and techniques that can help teachers monitor and interpret the level of engagement with academic tasks both online and in the classroom.

6.1.2 Recognising Children's Engagement

From an early age children are able to discern how another's face represents a mood or attitude. Cues like smiles are used early on by very young children to detect how their parents are feeling, and by early childhood children can recognise when their parents or friends are angry, sad, or excited by their facial expressions and demeanour. The study of children's understanding of emotions based on facial expressions and other stimuli is a well-researched field (Gross and Ballif, 1991). Writing in 2013 Widen (2013) identified 452 articles published on children's understanding of facial expressions. In the literature, there is support for the argument that children's ability to discern emotion begins in early childhood and develops through adolescence and into adulthood (Durand et al., 2007), (Gao and Maurer, 2009), (Malsert et al., 2020). Children are also able to differentiate between contexts of expressions; for example, they can understand that a parent crying at a TV drama is not the same as one crying following an injury (Pollak et al., 2009). Hence, context is an important factor in the accuracy of children's recognition and classification of emotion (Theurel et al., 2016).

Engagement and disengagement are recognisable from the way the learner focuses on a task. While not exactly the same as an emotion, the sign of engagement is typically a focus on a task with little head movement, with eyes facing the task in hand and a lack of distraction (Alkabbany et al., 2019). There have been several ML-based approaches which utilise the link between engagement classification and emotion classification and analysis (Sharma et al., 2022) (Shen et al., 2022). Within the literature, relatively few of these concern work performed specifically with children and the author is unaware of any that utilise children's own ability to identify emotional states.

6.1.3 Methods of Emotion Recognition

A popular and established system for emotion recognition is the Facial Action Coding System (FACS) (Ekman and Friesen, 1978). Originally developed in the 1970s, but still used today, FACS breaks down facial expressions into combinations of muscle movements called Action Units (AU) (Cohn et al., 2007). One drawback to FACS is the considerable training required, which at the time of writing is estimated by the Paul Ekman Group to be between 50 and 100 hours (Ekman, 2020). Additionally, for the large corpus of videos or images required to train a ML model, the time required for a group of trained practitioners to retrospectively label the data is likely to render such an approach impractical.

An alternative approach commonly used both in academia and commercially is to automate the emotion classification process using algorithms such as AFFDEX (McDuff et al., 2016), (Bishay et al., 2022) or FACET (Littlewort et al., 2011). There are several studies (Stöckli et al., 2018), (Dupré et al., 2020) that attempt to validate the comparative effectiveness and performance of the algorithms. Software such as iMotions (iMotions, 2022) can combine facial expression analysis with other sensors such as eye tracking or an electroencephalography (EEG) to combine a range of insights into the human emotional state (Kulke et al., 2020). Although the algorithmic approach clearly has the potential to save considerable time compared to the retrospective analysis by human experts outlined above and can be used to perform real-time analysis, there is concern that current emotion recognition systems are less accurate than their human counterparts when employed on children Bryant and Howard (2019). Here it is argued that a real-time evaluation and classification at point of capture performed by child observers has the potential to offer significant benefits over either approach.

6.1.4 Existing Data Sets for Machine Learning that Include Children

A search for existing data sets featuring children, that are suitable for use in behavioural studies, indicates that specialised child-centered data sets are relatively scarce. Princeton University Library have curated a directory of databases containing face stimulus sets available for use in behavioural studies of which just four are specific to children (*Databases (A-Z) - face image databases - research guides at Princeton University*, 2022). The most substantial database The Child Affective Facial Expressions Set (CAFE) (LoBue and Thrasher, 2015) features around 1200 pictures of children aged 2 to 8. There are three other databases listed (Negrão et al., 2021), (Khan et al., 2019), (Webb et al., 2018) all of which are relatively specialised and small, particularly compared to more generalised image data sets such as ImageNet (Deng et al., 2009). This lack of material restricts the options for CCI researchers looking for data as a starting point on which to train their models. At a time where a growing number of academic studies are exploring ML based systems and intelligent interfaces both within the CCI/ IDC community (Rubegni et al., 2022*a*), (Dietz et al., 2022*a*), (Nguyen, 2022*b*) and the wider HCI community (Kim et al., 2022*a*), (Jasim et al., 2022*a*) this chapter presents an approach to data labelling that makes child participation intrinsic not only to the development of the system but to the core of the system's outputs.

6.2 Studies

6.2.1 Pilot Study

Before commencing the full studies, a small pilot study was carried out in a primary school in the UK. The researcher observed a class of Year 5 pupils for a morning to try to observe visual clues to children becoming disengaged from their lessons. The concept behind the study was to aid in the collection and preparation of video data to be used to train a machine learning model described below.

The observation highlighted several interesting points. First, disengagement is temporal, children's attention drifts, but is not necessarily lost. In other words, a child can become distracted without becoming disengaged. Second, children exhibit visual clues which seem to indicate, at least to the untrained observer, that their attention is elsewhere. During the observation, one member of the class seemed to be paying no attention to the teacher. She was looking around the room and playing with things on her desk. As soon as the teacher asked a question, her hand was first up to answer it, which she successfully did. This temporal aspect indicates that disengagement is not a frozen moment in time, but the result of an ongoing sequence. This influences the development choices described in Chapter 8 that use a suitable model for processing sequential data. The disjuncture between the visual clues given by the pupil and the fact that they were clearly paying attention is an indicator, at least to the researcher, that care should be taken not to make judgments in isolation, hence the emphasis on context that runs through this chapter.

6.2.2 Main Studies

Two studies were conducted between June and October 2022. The aim of the first study was to generate a body of video data that captured the engagement status of the pupils while they completed a computerised task in a classroom environment. Labels for the engagement status of the pupil completing the task were recorded synchronously by peer observation, effectively reducing the two-stage operation of detection and inference to a single-stage operation while maintaining context during inference and in a time and resource effective manner.

The principle of child participation was central to the design of the study, with each child contributing both to the body of data and to the data labelling. The second study assessed the pupils' experience of, and confidence in, the data labelling process and a theoretical system based on its output. In addition to adding to the copus of data the second study asked four research questions:

- R1 What is the level of confidence of the pupils in the ability of their peers to assess their engagement status whilst completing a task?
- R2 What is the level of confidence of the pupils in their own ability to assess the engagement status of their peers while performing a task?
- R3 How accepting would pupils be if a system were used in the classroom to monitor their level of engagement?
- R4 To what degree would the pupils trust the system to identify disengagement?

Both studies were conducted at Ribblesdale High School, Clitheroe, Lancashire, UK with the first study taking place between the 17th June and

8th July 2022 and the second between the 3rd and 14th October 2022. In both cases the studies took place in Computer Science lessons and were supervised by the Head of Computer Science. The pupils used the same web-based interface for both studies. For the first study, the pupils' peer judgements were logged, and for the second study, the children completed the questionnaire (see Apparatus) after completing the logging process. Each child was allocated 15 minutes to engage with the online material while being observed, after which the roles were reversed. The supervisor was instructed not to intervene if he identified a child as disengaged during their assigned time, but to allow the peer observer to record their judgment. Ribblesdale High School was selected as the host school for the studies, as it is a Microsoft Training Academy and its pupils routinely use IT as part of their learning experience. Each pupil at the school is assigned a Microsoft Surface Pro which was used to both deliver the online content and capture and label the video footage.

6.2.3 Participants

Forty-five pupils from Ribblesdale High School participated in the studies. Twenty-two children, (12 boys and 10 girls) aged 11 to 15 years, participated in the first study, and a further twenty-three children, (10 boys, 13 girls) aged 11 to 12 years, participated in the second study. Before the study began, written consent was obtained from the school, parents or caregivers, and the pupils. The pupils were also advised that they could withdraw their data after completing the task regardless of any prior consent given by themselves or third parties. No incentives or rewards were offered to the children who participated in the study. All the children recruited had their timetabled Computer Science lessons with the Head of Computer Science, who supervised all the sessions. Additionally, the study was designed so that pupils who did not participate in the study were still able to participate in the lesson by completing the same on-line task without being observed or recorded.

6.2.4 Apparatus

Three artefacts were prepared for the studies, the first was a website of material about cryptography. This was designed by the researcher and was intended to be something that would be new to the pupils. The material was designed in conjunction with the Head of Computer Science at the school, who also supervised the study and complements the school's Computer Science curriculum. It consisted of an introduction to cryptography and cryptanalysis interspersed with interactive quizzes and encryption and decryption activities using an online Caesar Cipher wheel, which allows pupils to test their understanding of the material presented (Figure 6.1). The material was designed to support at least 15 minutes of activity, which was the time allocated to each pupil to interact with the cryptography webpage and was deemed suitable by the teachers for children aged 11 to 15 years.

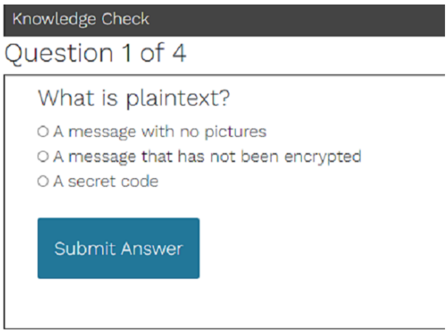


Figure 6.1: A quiz question from the online task

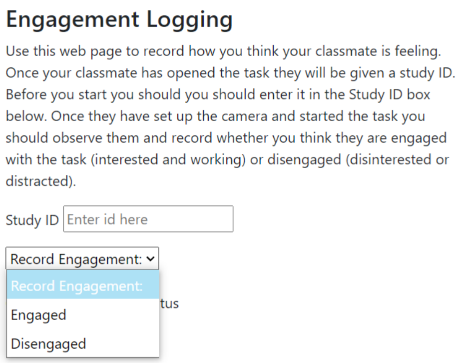


Figure 6.2: Logging Engagement

The second artefact was an online form that allowed the (pupil) observer to log the engagement level of the pupil completing the cryptography task. Using the form, the observer recorded the engagement level as a binary value; engaged (interested and working) or disengaged (disinterested or distracted). When the observer felt that the learner had changed their engagement category, they recorded the updated value (Figure 6.2).

The final artefact, used only in the second study, was a short questionnaire.

Pupils completed the questionnaire to gauge their feelings about the logging process. Pupils were asked:

1. How accurately they thought their classmate had judged their engagement level whilst completing the task
2. How accurately they thought they had judged their classmate's engagement level whilst completing the task
3. How accepting they would be if a system was utilised in the classroom to monitor their engagement level
4. To what degree would they trust the system to identify disengagement
5. The action the system should take if it identified disengagement.

For questions 1 to 4, the pupils were given a Likert scale ranging from 1 - 10 to rate their responses, where 1 was equivalent to low and 10 was equivalent to high. For instance, for Question 1, a recorded score of 1 would indicate that the pupil thought the accuracy of their classmate's judgement of their engagement level was low, whilst a score of 10 would indicate a perceived high accuracy of judgment.

The video capture of the children who completed the cryptography task (artefact 1) was implemented using the Open Source WebRTC API webrtc.org (2023). Video files were created in .webm format on the pupils' Surface Pro rather than on the server, minimising data transfer over the school network, increasing security and ensuring the pupils' ownership of their data until they agreed for the supervisor to collect their video file. The logging data (Artefact 2) was written to a MariaDB database on the web server.

6.2.5 Procedure

Pupils worked in pairs in the school classroom using their Microsoft Surface Pro machines connected to the school network. Each pair took turns as the learner and observer, switching roles half way through the study. The learner completed the online cryptography task on their Surface Pro. The observer was placed so that they could watch the learner complete the task, but could not see their Surface Pro screen (Figure 6.3). The placement of the children was intended to ensure that both learners were given the same learning experience.

Figure 6.3: Peer Observation of the task

In addition to ensuring that the learner was positioned so that they were centrally placed in the video, where possible the supervisor arranged the pupils in the classroom in a way that would avoid interference such as other faces in the video background when capturing the footage. Before the pupils commenced the study the supervisor explained the different components of the task and the logging process and answered any questions. The importance of the logging process was emphasised to the children as having the same importance as the computerised task. The supervisor checked that the cameras were configured correctly and that the children had correctly identified and shared the study ID.

On commencement of the online task, the learner was directed to a webpage at https://chici.org/studies/awa/. On accessing the page, the system allocated each learner a unique ID which they shared with the observer. This ID was used both to anonymise the data in the study and, in the case of the first study, to synchronise the recorded video data with the data logged by the observer. When the observation was complete, the learner and observer switched roles and the process was repeated in reverse.

The study supervisor explained the process of configuring the Surface Pro webcam to capture their video data while they completed the task and explained to the pupils the need to share their unique Study ID with their observer before accessing the online material (Figure 6.4). When the learner was ready to start the study, they selected the Start Recording button, which opened up the Cryptography task in a new browser tab. The webcam continued to run in a separate browser tab until the task was completed.

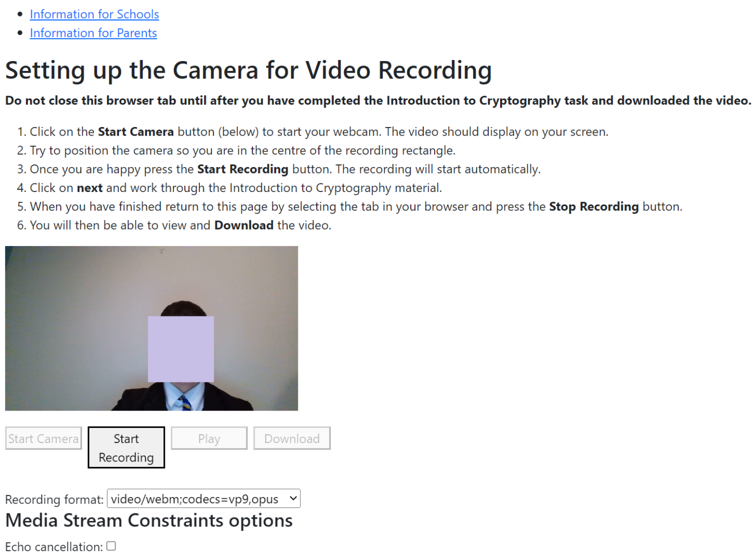


Figure 6.4: Configuring the study

Once the learner had started the task, the observer began recording their engagement status using the Engagement Logging form. The server generated a Unix timestamp when the pupil selected the Start Recording button. Unix timestamps were also generated from the same server when the observer recorded a change in engagement status through the Engagement Logging form. Timestamps were recorded in a MariaDB database using the unique Study ID as a key. On completion of the task, the pupil selected the Stop Recording button which generated a final Unix Timestamp. The pupil then selected the Download button to save the video to the hard drive of their Surface Pro in the format <study id>.webm. On completion of the task, the supervisor ensured that the learner noted their unique study ID and the videos were transferred by the supervisor from the pupils' Surface Pros to Toshiba Canvio Portable Storage USB devices encrypted using Bitlocker Drive Encryption for transport prior to processing by the research team.

For the second study, after completing the same online task described in Study 1, the supervisor introduced the pupils to the paper-based questionnaire (Artefact 3). The pupils were asked to complete the questions and record any other observations about the study. Most children did not provide additional information. After completion the surveys were collected by the supervisor and forwarded to the research team along with the video data for analysis.

6.2.6 Data Processing and Cleaning

Study 2 Video Data

The data collection process for Study 1 and Study 2 both used the same server and software. The school updated its firewall settings in the intervening period between the studies and while video files were generated for Study 2, no timestamps were written to the database. The video data for Study 2 was therefore discarded as no labels could be generated for the engagement status. The surveys the pupils completed were still valid as they completed the PDLS process and are discussed later.

Study 1 Video Data

Study 1 produced 22 videos of which 17 were usable. 2 videos were discarded as they had audio but no image frames and 3 videos were complete, but they had no engagement statuses recorded. The remainder of this section describes the process followed to partition and perform a binary classification of the video data, engaged or disengaged.

Unix timestamps were used to record changes in engagement status. Each timestamp represents the number of seconds elapsed since January 1st 1970 and calculating the difference between the recorded values in the database facilitated partitioning the video files into engaged and disengaged subfiles.

Given a video with an ID of 196 and recorded starting and ending timestamps of 1655818818 and 1655819111 it can be calculated that the learner started the study at 14:40:18 GMT+0100 (British Summer Time) on Tuesday June 21, 2022 and ended it at 14:45:11.

Each time the observer updated the learner's engagement status using the logging interface a timestamp was recorded on the database along with the new status (Table 6.1). In this case, the observer has logged the starting status 2 seconds before the learner started the task.

Table 6.1: Recorded Timestamps for Study 196

ID	Timestamp	Status
196	1655818816	ENGAGED
196	1655818893	DISENGAGED
196	1655818897	ENGAGED
196	1655818932	DISENGAGED
196	1655818936	ENGAGED

As part of the data cleansing process, the starting timestamp is adjusted by +2 from 1655818816 to 1655818818 to align it with the beginning of the video. The video end timestamp generated when the learner stops recording is appended to the data as the final status in the table which allows the video to be partitioned so that each observation has a start and end time. This is shown in Table 6.2 with the adjusted start time for the first recorded status.

Table 6.2: Processed Timestamps for Study 196

Start	End	Status
1655818818	1655818892	ENGAGED
1655818893	1655818896	DISENGAGED
1655818897	1655818931	ENGAGED
1655818932	1655818935	DISENGAGED
1655818936	1655819111	ENGAGED

The aim of this exercise is to split each video into sections based on the engagement status generated by the observer. The videos were cut into parts using Python wrappers for the Open Source FFmpeg video processing library, and the new videos renamed by appending an incremental suffix. They were then written to separate directories, engaged or disengaged dependent on the engagement status recorded by the observer. Study ID 196 resulted in the following video files, see Table 6.3.

Table 6.3: Videos Generated for Study 196

Directory	Filename	Duration in Seconds
Engaged	196_1.mp4	75
Disengaged	196_2.mp4	4
Engaged	196_3.mp4	35
Disengaged	196_4.mp4	3
Engaged	196_5.mp4	176

6.3 Results

The 17 usable videos and engagement logs yielded 2 hours, 33 minutes and 48 seconds of video, of which 2 hours, 27 minutes and 32 seconds have labels generated from the pupil logs. The video footage was standardised at 25 frames per second resulting in 221,300 1280 x 720 labelled JPEG images. The observers logged 57 instances of an engaged status totalling 2 hours, 12 minutes and 33 seconds yielding 198,825 labelled images. Forty-four instances of a disengaged status were recorded totalling 14 minutes and 59 seconds, yielding 22,475 images (Tables 6.4, 6.5).

Table 6.4: Breakdown of logged Data

ID	Duration	Logged Duration	Engaged Frequency	Engaged Duration	Disengaged Frequency	Disengaged Duration
166	00:07:02	00:05:30	1	00:05:30	0	00:00:00
171	00:18:13	00:18:06	2	00:18:03	1	00:00:03
172	00:12:51	00:12:33	10	00:12:08	9	00:00:25
173	00:20:13	00:19:52	3	00:11:02	2	00:08:50
196	00:04:52	00:04:48	3	00:04:42	2	00:00:06
212	00:12:07	00:11:18	13	00:09:48	13	00:01:30
213	00:09:00	00:08:44	8	00:06:40	7	00:02:04
219	00:04:00	00:03:27	5	00:02:57	5	00:00:30
231	00:02:54	00:02:48	1	00:02:48	0	00:00:00
237	00:06:55	00:06:41	2	00:06:03	1	00:00:38
238	00:07:39	00:07:29	1	00:07:29	0	00:00:00
239	00:16:51	00:16:40	2	00:16:10	1	00:00:30
242	00:04:51	00:04:29	1	00:04:29	0	00:00:00
243	00:07:18	00:07:06	1	00:07:06	0	00:00:00
244	00:11:33	00:11:21	2	00:11:06	2	00:00:15
245	00:02:24	00:02:08	1	00:02:08	0	00:00:00
246	00:05:05	00:04:32	1	00:04:24	1	00:00:08
Totals	02:33:48	02:27:32	57	02:12:33	44	00:14:59

Table 6.5: Image Generation from Processed Videos

Status	Logged Duration	Seconds (s)	Frames per Second (fps)	Images (s x fps)
Engaged	02:12:33	7953	25	198825
Disengaged	00:14:59	899	25	22475

The pupils were allocated 15 minutes each to complete the task so as to fit in with the school's lesson duration, although the majority did not use all their time. The time spent on the task ranged from 2 minutes and 24 seconds to 20 minutes and 13 seconds (M = 09:03, SD = 05:26). The logged duration ranged in time from 2 minutes and 8 seconds to 19 minutes and 52 seconds (M = 08:41, SD = 05:28). Six minutes and 16 seconds of video were discarded, as they had no logging status. The majority of the discarded data, i.e the difference between Duration and Logged duration, see Table 6.4, occurred at the beginning of the videos in the period after the learner had started the video camera generating the starting timestamp and before the observer recorded their first engagement status.

When logging the data, observers adjudged that the learners were engaged in the task for 2 hours, 12 minutes and 33 seconds, ≈90% of the logged time, and disengaged for 14 minutes and 59 seconds, ≈10% of the logged time. The number of statuses recorded was distributed more evenly, with 57 of the 101 statuses, ≈56% logged as engaged and 44, ≈44% as disengaged. The average duration of an engagement instance was 2 minutes and 20 seconds, and the average duration of the learners' disengagement was 20 seconds. The frequency of the recorded data ranged from a single recording of engaged for Study ID 166 to to 26 recorded statuses for Study ID 212 ($M = 3.35$, $SD = 3.6$).

In addition, 22 questionnaires from the second study were completed, the results of which are presented in Appendix H and summarised below. the children were asked:

1. How accurately they thought their classmate had judged their engagement level whilst completing the task
2. How accurately they thought they had judged their classmate's engagement level whilst completing the task
3. How accepting they would be if a system was utilised in the classroom to monitor their engagement level
4. To what degree would they trust the system to identify disengagement

Of the 22 pupils who completed the study, 21 answered all the questions and 1 pupil only answered question 1. Based on a scale of 1 to 10 where 1 indicates low agreement and 10 indicates high agreement, on average the pupils ranked their classmates accuracy at judging their personal engagement at 8.5/10 ($SD = 1.87$). They ranked their own judgment of their classmate engagement status higher at 9/10 ($SD = 1.58$). They were less accepting of the system's deployment 7/10 ($SD = 2.1$) and capability to identify disengagement 7/10 ($SD = 2.4$).

6.4 Discussion

6.4.1 Evaluating the Usability of the Process

This chapter describes PDLS, a novel and extensible approach to generating labelled data suitable for training supervised ML algorithms for use in CCI research and development. Here it is evaluated using the usability metrics effectiveness, efficiency and satisfaction outlined in ISO 9241-11 *ISO - International Organization for Standardization. ISO 9241-11:2018(en) Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts* (2018).

6.4.2 Efficiency

PDLS is judged to be a time and cost-efficient system that compares favourably with the other options considered. FACS coding by human experts requires extensive training and has a considerable time and cost overhead. A relatively small study such as this can generate significant volumes of data, 284,800 images, that require labelling prior to use, which is challenging at best and impractical at worst for human experts to code in a time and cost effective manner. PDLS labels the data at the point of capture using peer judgments and does not require post-hoc analysis of the data.

Another alternative for labelling the data is to use algorithmic implementations such as AFFDEX and products that implement them such as iMotions. iMotions can be configured to perform evaluations in real time but is considerably more costly than PDLS which requires no specialist equipment other than a laptop and a camera, both of which are relatively low cost and freely available. PDLS is also highly extensible and suitable for gathering and labelling data concurrently.

iMotions can also be used to retrospectively label the data. This algorithmic approach can process large quantities of data quickly, but still has the same resource and cost implications. Additionally, there is little in the literature on the application of algorithmic approaches to Facial Action Coding for children, for a review, see Martinzez Martinez (2019).

6.4.3 Satisfaction

Pupils indicated their satisfaction with both their own and their peers effectiveness in reaching a classification and the potential of a system built on the study data to make effective judgments. They expressed high confidence in their own ability to accurately measure the level of engagement of their classmate (R2). They were only marginally less positive about the ability of their classmate to assess their own engagement levels (R1). The children were also asked how accepting they would be if a system was deployed to monitor their level of engagement in the classroom and how trusting they would be in the accuracy of its judgements. Children were neutral to accepting in ranking the proposed system (R3) and its predictions (R4) with both scores lower than their confidence in their own and their peers' ability.

6.4.4 Effectiveness

Evaluating the effectiveness of PDLS is challenging, however, the initial signs appear promising. The children's judgments appear to be consistent, and there are few outliers in the data indicating that the classifications are cohesive and the children are measuring the same phenomena.

On the surface a 90% engagement rate with a short task carried out in a classroom with a teacher present sounds sensible. Whilst it cannot be said with certainty that the children's judgments are correct, a random sample of ten of the 44 videos that were classified as disengaged indicates that in the majority of cases the learner is exhibiting behaviour which may be associated with disengagement or distraction from the task (Table 6.6). Certainly their focus often appears to be elsewhere.

Table 6.6: Characteristics of Children's observations of disengagement

Study ID clip	Observation of Behaviour
171_2	The learner appears distracted and looks away from the screen
172_4	The learner is laughing
173_4	The learner is talking and hits out at someone off camera
196_2	The learner is laughing and appears distracted
212_12	The learner is smiling and scratching their head
212_24	The learner is smiling but appears to be working
213_6	The learner is smiling and scratching their ear
213_10	The learner is smiling and looks away from the screen in parts but appears to be working
219_1	The learner is talking and looking away from the screen
237_2	The learner appears to be working but is holding a conversation unrelated to the task

The exception may be video 212 where although the learner appeared amused by something there is no obvious indication that they were not engaged. Study 212 had the most statuses recorded across both categories, (26 for a logged duration of 11 minutes and 18 seconds), or one every 26 seconds on average with an average duration of $\approx$7 seconds for each logging of disengagement. As such it is feasible that the observer's judgements were not in line with the other children.

This is clearly not conclusive and the full validation process followed is discussed in Chapter 7.

6.4.5 A Child-Centred Process

The final stated objective was to keep true to the CCI principle of keeping child participation central to the design process. In using the children's own classifications to generate the data set, they become central not just to the design process but also to the operation of a system built using that data set. They are in effect judging themselves. Figure 6.5 illustrates this process. The pupils' first classify each other's level of engagement in the classroom using the PDLS method. The labelled data is then used by the system to learn about engagement, this learning process is entirely dependent on the children's classifications. Once operational, the system monitors the children in the classroom and uses what it has learnt from them to classify their engagement level.

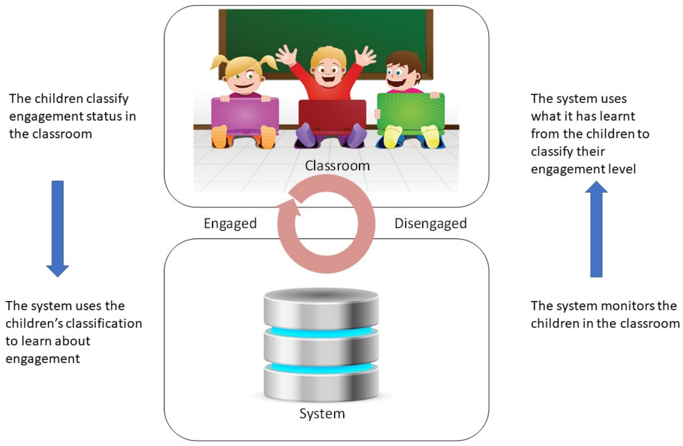


Figure 6.5: Child Participation Model

PDLS not only uses the pupils' judgment to label the data, but by the very nature of the supervised machine learning process, their participation and input will form the basis of future system development and deployment.

6.4.6 Data Bias and Authenticity

Data bias is a recurrent theme in ML literature (Mehrabi et al., 2021), (Jiang and Nachum, 2020) and beyond. In the UK in 2020 there was uproar that the algorithm designed to predict exam results was unfair and disadvantaged students from certain demographics, resulting in teachers predicting grades (Coughlan, 2020). As IS become increasingly embedded into society, it is an inherent responsibility of designers and developers to ensure that the decisions made by the technology are fair. When making this point, it should be noted that the data collected for this study is produced from a single computerised task in one school and the output from any ML model built based on this data will reflect these limitations.

To address these limitations, further studies should reflect children's diverse backgrounds increasing the scope of the data set and therefore the quality of the judgments produced by ML models trained upon it. In addition, the scope and circumstance of the observed tasks can be extended to provide new context to the observations. Whilst the work described involved a computerised task and webcam it is feasible that judgments could be recorded of children completing more traditional activities which do not involve computers.

Regardless of the medium, in order to maintain context, PDLS should be deployed in the pupils' usual educational environment, ideally supervised by a teacher as opposed to under laboratory conditions with researchers. In the classroom, pupils are in a familiar environment and are less likely to be distracted by unknown or strange circumstances, and the context under which peer judgments are made is less likely to be affected. Therefore, the data collected should result in a more accurate model and better decisions.

There is of course a downside to this approach, whilst the equipment is comparatively simple and unlikely to fail schools have their own IT systems and configurations, of which it can be easy to fall foul. Prior to the second study, the participating school adjusted their firewall settings which blocked the database writes, with the result that the second set of videos were not labelled and could not be used for training. The research team must also cede control of the experiment to the supervisor, so careful briefing is required to ensure the results of any recordings are fit for purpose.

6.5 Summary

This chapter presented PDLS, a peer observation approach to generate a labelled data set suitable for use in CCI research. The system is evaluated against the usability metrics, effectiveness, efficiency, and satisfaction and is judged to be both efficient and satisfactory. Validation of its effectiveness is presented in Chapter 7. The CCI principle of Child Participation is central to the PDLS process, which generates labelled data in both a time and cost effective manner. Pupils were surveyed for their feelings on the accuracy of both their own and their peers' judgments on engagement status after completing the task and expressed their confidence in both these aspects.

It concludes by offering some thoughts that are intended to be helpful to other researchers who may wish to carry out similar studies and propose the development of a data set that can be used as a resource for members of the CCI community who wish to undertake CCI research on emotion recognition or the application of computer vision to research with children.

Chapter 7

Assessing the Effectiveness of the PDLS

7.1 Introduction

Chapter 6 described PDLS, a system for generating labelled data suitable for training supervised ML algorithms. PDLS was assessed to be both efficient and satisfactory relative to the alternatives considered. This chapter evaluates the accuracy or effectiveness of the system. For a ML model to make accurate predictions it requires accurate data on which to train. Poor quality input data results in poor quality outputs often referred to as garbage in, garbage out (GIGO) systems.

Two separate methods were employed to establish the accuracy of PDLS. The first used the iMotions software to retrospectively analyse the video data generated from the study described in section 6.2.6 with the aim of triangulating the output from the PDLS and the software. The second method employed expert reviewers to watch the videos captured by the pupils' laptops in their entirety and record engagement statuses independently of the original decisions. Where there is a agreement between one or both of the reviewers and the observers original judgment, then the pupil observer's label is considered accurate. Where there is disagreement, then this is reviewed by the author with the goal of establishing the reasons for the inconsistency. The chapter concludes by discussing the strengths and weaknesses of the system and makes recommendations for its development and improvement.

7.2 Assessing Accuracy using iMotions

The iMotions software was used to perform post hoc verification of the labels generated by PDLS. The video data from the first study was input into iMotions, which was configured to perform emotion analysis using the Affectiva AFFDEX algorithm (Bishay et al., 2022) which reports on a range of emotions including engagement.

Before analysis by the software, the video data was standardised at 25 frames per second and further broken down into 3 second clips for ease of processing. Attempts to process the full-length videos caused the software to fail. The software was unable to perform an engagement classification on many of the frames indicated by 0 in the report. In other cases, it generated percentage scores that fluctuate from one end of the scale (0 to 100) to the other over very short timescales (Table 7.1).

Table 7.1: Segment of the iMotions Engagement Report using Affectiva AFFDEX

Timestamp (*ms*)	Engagement Rating $\%$
14252	0.13487616181373596
14298	0.098771192133426666
14344	0
14390	0.23590019345283508
14436	0
14482	0.85175901651382446
14528	60.243587493896484
14575	98.717597961425781
14621	0
14667	0
14713	0
14759	98.322868347167969
14805	13.619963645935059

In the given example, which covers a period of just over half a second (553 *ms*) the software reported results ranging from no classification to engagement levels ranging from $< 1\%$ to $> 98\%$. The most likely cause of the fluctuations are noise in the data but these scores clearly differ from the children's judgements which were more consistent and longer in duration. This may also reflect, at least in part the preservation of the context in which the original judgments were made.

It should be noted that iMotions allows the user to calibrate the software where analysis is performed at the point of capture, but the retrospective validation used did not allow for this, which may have affected the results. The iMotions software also supports a multimodal approach to classification, offering additional tools such as electroencephalogram (EEG), electromyography (EMG), and electrocardiogram ECG which extend its analytical capabilities beyond facial coding. However, such features require a range of different sensors to capture the data, which also makes such an approach unfeasible for capturing and labelling the data on the scale required in a classroom.

Using iMotions to validate this data set was not considered viable due to both the fluctuations in the data and the large number of instances where the software was not able to make a classification. The second approach to classifying the data employed expert reviewers who viewed the same video footage.

7.3 Assessing Accuracy using Reviewers

7.3.1 Participants

Two members of the research project, Reviewer 1 (R1) and Reviewer 2 (R2) watched the 17 labelled videos taken from the first study which took place between June 17th and July 8th, 2022. Chapter 6 Section 6.2.6 describes how the videos were selected.

As described in Chapter 6 a pupil observer watched another pupil learner complete a computerised task and used a simple web interface to record their level of engagement with it. Using the same interface, the reviewers judged the level of engagement of the learner as engaged or disengaged. Once both reviewers had independently reviewed the footage the author reviewed the results to establish where there were differences between the reviewers' and the pupils' judgments.

Both reviewers are qualified educators currently working in Higher Education with significant experience of working with children of this age group. R1 was a secondary school teacher before moving to higher education. The author is also a qualified educator who has also previously worked as a secondary school teacher.

The reviewers were familiar with the project and the definitions of engagement outlined in Chapter 6 and in particular the notion of engagement on task utilised in this work. The reviewers received no additional training to support them in identifying engagement or disengagement.

7.3.2 Apparatus

A web interface was developed that presented the reviewers with a list of videos generated by the original studies (Figure 7.1).

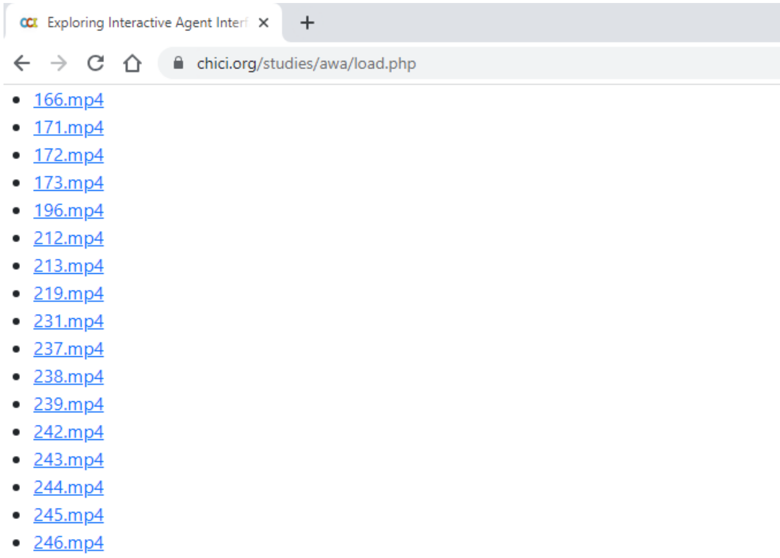


Figure 7.1: Selecting a Study for Validation

On selecting a study link, the page loaded, showing the video recording from the study and the logging interface. The reviewers selected their ID from the Validator drop-down list and the Study ID was prepopulated. On selecting the Play Video button, the video started, and the reviewers used the Record Engagement drop-down list to select the pupil's engagement status as either engaged or disengaged. This is the same method that the pupils used to generate the statuses when they observed the original study (Figure 7.2).

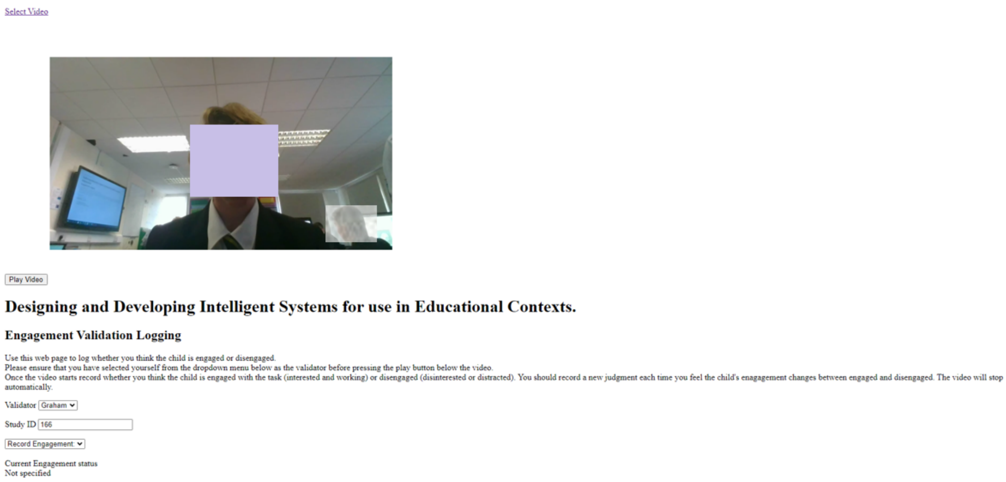


Figure 7.2: Interface for Validating a Study

On selecting Play Video, a timestamp was generated capturing the video start time and written to the database. When the video ended, a timestamp was automatically generated representing the end of the video (Table 7.2).

Table 7.2: Video Start and End Times - Reviewer 1

uid	study_id	start_time	end_time
R1	166	1686554794	1686555222
R1	171	1686555249	1686556345
R1	172	1686562958	1686563733
R1	173	1686563761	1686564978
R1	196	1688315013	1688315310
R1	212	1688315329	1688316061
R1	213	1688316199	1688316743
R1	219	1688316760	1688317003
R1	231	1688317025	1688317203
R1	237	1688317224	1688317642
R1	238	1688317662	1688318123
R1	239	1688318138	1688319151
R1	242	1688319186	1688319480

uid	study_id	start_time	end_time
R1	243	1688322608	1688323050
R1	244	1688323067	1688323763
R1	245	1688323779	1688323926
R1	246	1688323942	1688324251

When the reviewers logged a change in engagement status a timestamp was generated and written to the database along with the engagement status. (Table 7.3) shows this for Study ID 166.

Table 7.3: Generating Engagement Timestamps

uid	status	id	time
R1	ENGAGED	166	1686554798
R1	DISENGAGED	166	1686554814
R1	ENGAGED	166	1686554820
R1	DISENGAGED	166	1686554876
R1	ENGAGED	166	1686554879
R1	DISENGAGED	166	1686554919
R1	ENGAGED	166	1686554921
R1	DISENGAGED	166	1686555046
R1	ENGAGED	166	1686555049
R1	DISENGAGED	166	1686555124
R1	ENGAGED	166	1686555126
R1	DISENGAGED	166	1686555200
R1	ENGAGED	166	1686555203

From this, the start and end times for each observed engagement status were derived (Table 7.4).

Table 7.4: Engagement Statuses - Study 166

Start	End	Study ID	Status
1686554798	1686554813	166	ENGAGED
1686554814	1686554819	166	DISENGAGED
1686554820	1686554875	166	ENGAGED
1686554876	1686554878	166	DISENGAGED
1686554879	1686554918	166	ENGAGED
1686554919	1686554920	166	DISENGAGED
1686554921	1686555045	166	ENGAGED
1686555046	1686555048	166	DISENGAGED
1686555049	1686555123	166	ENGAGED
1686555124	1686555125	166	DISENGAGED
1686555126	1686555199	166	ENGAGED
1686555200	1686555202	166	DISENGAGED
1686555203	1686555222	166	ENGAGED

Finally, the offset into the video was calculated, and the starting point and duration of each period were calculated. This process is shown for Study id 166 for R1. It shows that R1 logged the first engaged status at 4 seconds. R1 did not see a change in status for 16 seconds, at which point a disengaged status was recorded for a period of 6 seconds, and so forth (Table 7.5).

Table 7.5: Completed Validation Timings (Study 166)

Start:	4	Duration:	16	Status:	ENGAGED
Start:	20	Duration:	6	Status:	DISENGAGED
Start:	26	Duration:	56	Status:	ENGAGED
Start:	82	Duration:	3	Status:	DISENGAGED
Start:	85	Duration:	40	Status:	ENGAGED
Start:	125	Duration:	2	Status:	DISENGAGED
Start:	127	Duration:	125	Status:	ENGAGED
Start:	252	Duration:	3	Status:	DISENGAGED
Start:	255	Duration:	75	Status:	ENGAGED
Start:	330	Duration:	2	Status:	DISENGAGED
Start:	332	Duration:	74	Status:	ENGAGED
Start:	406	Duration:	3	Status:	DISENGAGED
Start:	409	Duration:	19	Status:	ENGAGED

This process was repeated for both the reviewers and the original pupil observations for all studies.

7.3.3 Results

The data collected from both the original studies and the validation of the videos is presented in Appendix I.

To understand how the reviewers and pupils perceived disengagement over time, the frequency of the period of recorded disengagement for all the studies was derived. Table 7.6 summarises the time span of the disengagement observations, ranging between one and twenty seconds, grouped by the observers, and Figure 7.3 plots the data.

R1 has recorded 14 instances of disengagement with a duration of 2 seconds and 26 instances of disengagement with a duration of 3 seconds. In the original studies, the pupil observers only recorded a single instance with a duration of 2 seconds and 6 instances of disengagement lasting 3 seconds. R2 recorded six instances for both periods.

Table 7.6: Comparative Observation of Disengagement by Duration over all Videos

	Frequency of Observation		
Duration (secs)	R1	R2	Pupil Observer
1	0	1	0
2	14	6	1
3	26	6	6
4	8	4	9
5	3	0	5
6	8	2	2
7	5	1	5
8	6	6	2
9	2	1	2
10	1	1	2
11	0	1	3
12	0	1	2
13	1	2	1
14	0	1	1
15	1	0	0
16	1	0	1
17	0	1	0
18	0	0	0
19	0	1	0
20	0	0	0

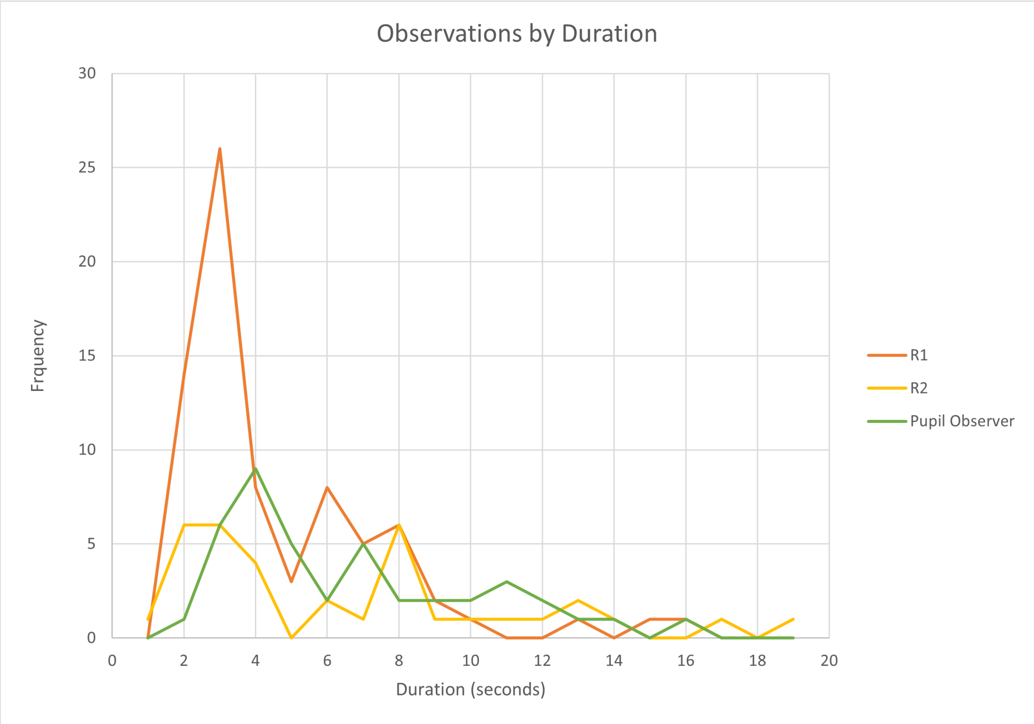
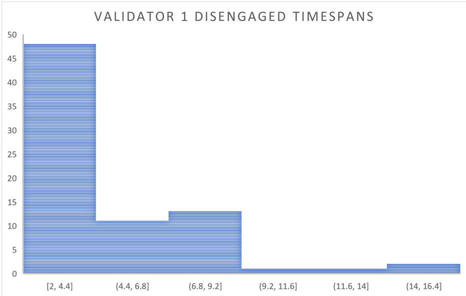


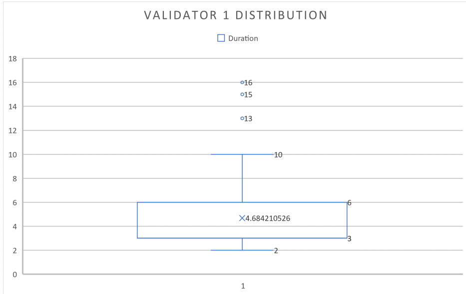
Figure 7.3: Comparative Observations of Disengagement by Duration over all Videos

Figure 7.4 shows the distribution for all the videos. Two outliers from the Pupil Observer data for Study 173 with durations of 334 and 198 seconds have been omitted to aid in visualising the data shown on graph 7.4f. Box plots indicate the distribution of the data around the median with the box bounding 50% of the data lying between the 1st and 3rd quartile, also known as the interquartile range. In all three plots, the mean value is higher than the median and the data is said to be positively skewed. Data points falling outside the whiskers are referred to as outliers and, indicate logged periods of disengagement that are not in line with the other observations. In this case, they are all much greater than the other observed durations.

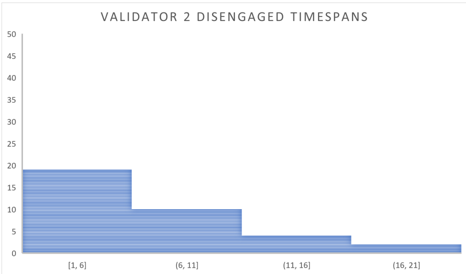
The median values logged for the duration of disengagement are 3 seconds for R1, 6 seconds for R2, and 7 seconds for the pupil observer. The mean values are 4.68 seconds for R1, 6.63 seconds for R2, and 20 seconds for the pupil observers.



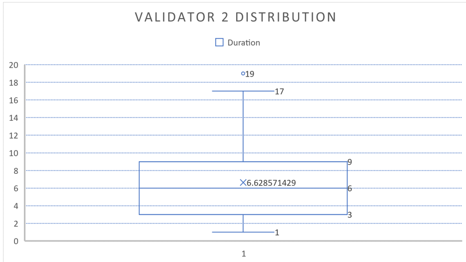
(a) R1 Disengaged Time spans



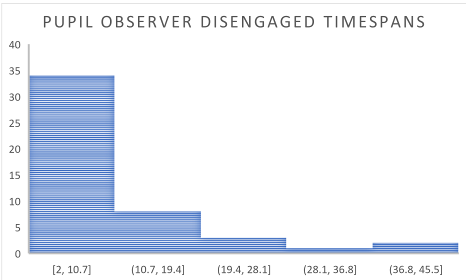
(b) R1 Distribution



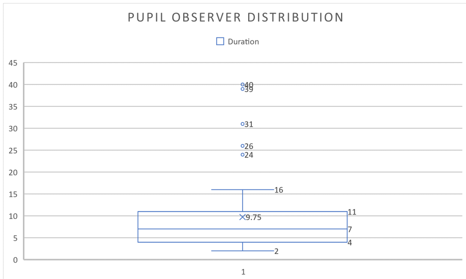
(c) R2 Disengaged Time spans



(d) R2 Distribution



(e) Pupil Observer Disengaged Time spans



(f) Pupil Observer Distribution

Figure 7.4: Observed Disengagement by Duration

7.3.4 Normalised Results

The comparative observations of disengagement shown in Figure 7.3 suggest that R1 was more inclined to record short periods of disengagement lasting two or three seconds, which were not observed in the original study or by R2. On reviewing the data, the author identified that often the subject appeared temporarily distracted and immediately returned to the task, and as such did not disengage. To remove these anomalies from the data set, all the logged values with duration ≤ 3 seconds were discarded.

The author also reviewed the other instances of recorded disengagement that fell outside the interquartile ranges in the data. In Study 173, the Pupil Observer recorded periods of disengagement lasting 469 and 198 seconds, which deviate significantly from the other instances recorded, so this study was discarded. Likewise in Study 213 the Pupil Observer recorded periods of 26 seconds and 40 seconds of disengagement that were not validated by R1 or R2. Other than some minor distraction, the author could not discern disengagement lasting for these time spans. Study 213 was discarded.

Studies 237 and 239 also have long periods of disengagement recorded by the Pupil Observer of 39 seconds and 31 seconds, respectively, which were not supported by the reviewers. In Study 237 the subject has a brief conversation with the teacher but continues to work, and in Study 239 the subject appears confused by the task and says “I don’t get it” but continues to work. Both studies are discarded.

Table 7.7 summarises the revised time span of the disengagement observations that range between 1 and 20 seconds after removing the anomalies and Figure 7.5 plots the data.

Table 7.7: Normalised Comparative Observation by Duration

	Revised Frequency of Observation		
Duration (secs)	R1	R2	Pupil Observer
1	0	0	0
2	0	0	0
3	0	0	0
4	6	3	9
5	3	0	4
6	7	2	3
7	3	0	2
8	4	5	2
9	2	1	1
10	1	1	1
11	0	1	1
12	0	1	0
13	1	2	1
14	0	1	1
15	1	0	0
16	1	0	1
17	0	1	0
18	0	0	0
19	0	1	1
20	0	0	0

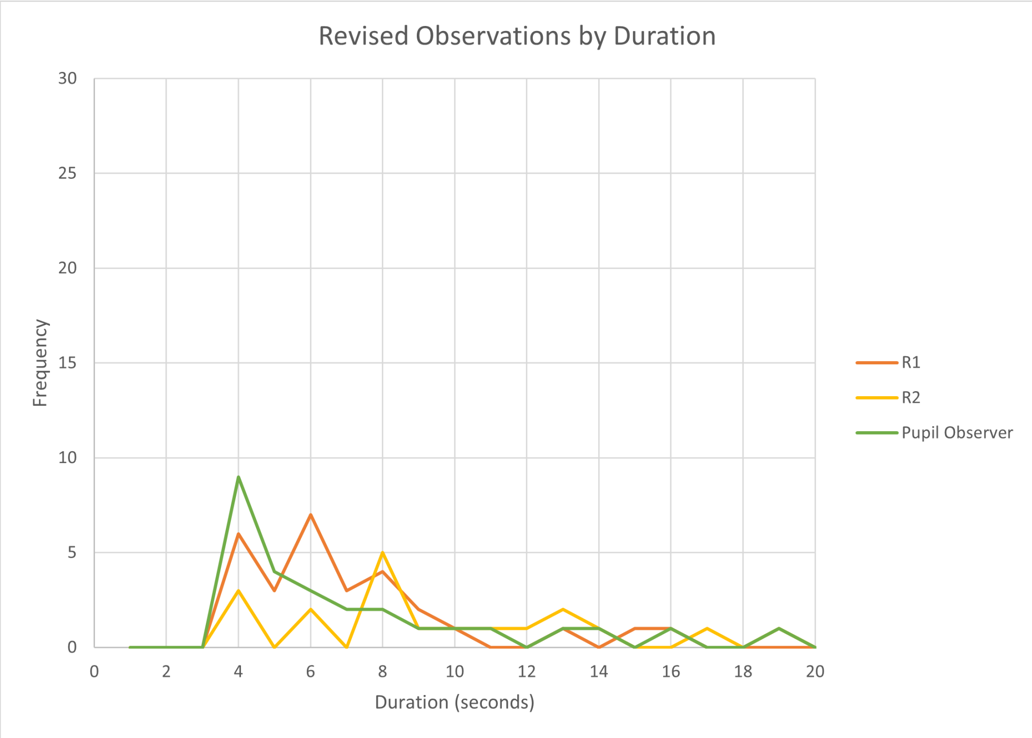


Figure 7.5: Normalised Comparative Observations by Duration

Table 7.8 and Table 7.9 summarise the data normalisation process. Eighty-six instances of logged disengaged statuses have been discarded, reducing the count from 161 to 86 occurrences. Most of these come from the validation exercise, with 47 instances discarded from R1 and 16 from R2. From the original study, 23 instances of disengagement were discarded, 7 of which had a short duration of 2 or 3 seconds.

The removal of the outliers from the study data has reduced the mean duration from 20 seconds to 7.26 seconds, while discarding the observations with a short duration has decreased the median value from 7 to 6 seconds.

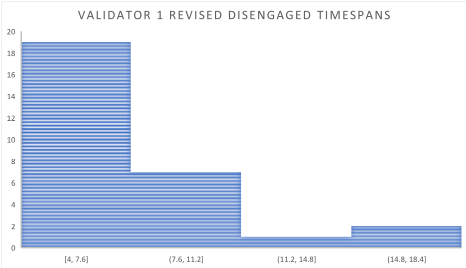
Table 7.8: Summarised Disengagement by Duration

	R1	R2	Pupil Observer
Count	76	35	50
Average	4.68	6.63	20
STD	2.93	4.60	52.76
Median	3	6	7

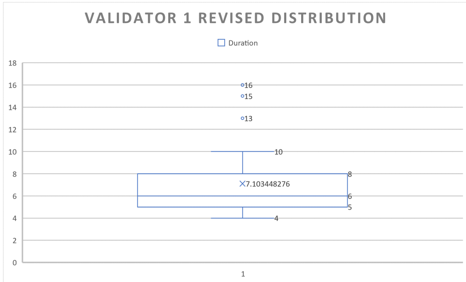
Table 7.9: Normalised Summarised Disengagement by Duration

	R1	R2	Pupil Observer
Count	29	19	27
Average	7.10	9.58	7.26
STD	3.08	4.15	4.03
Median	6	8	6

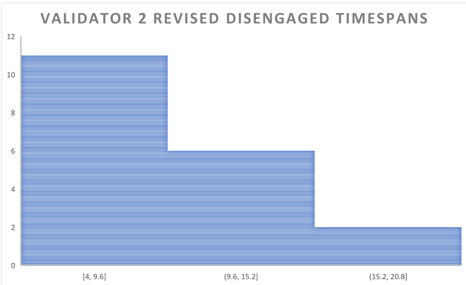
Figure 7.6 shows the normalised distribution for the entire data set.



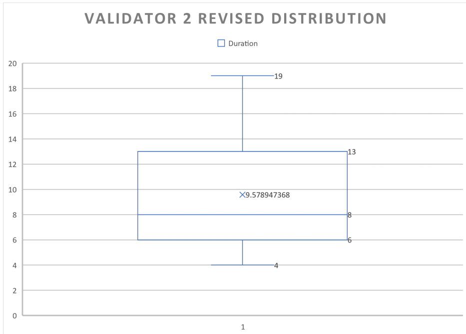
(a) R1 Revised Disengaged Time spans



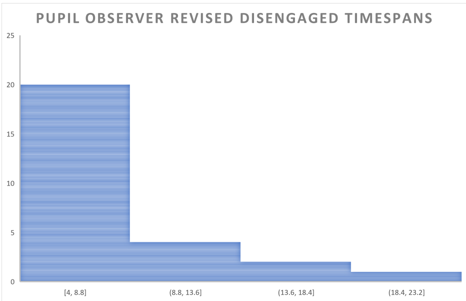
(b) R1 Revised Distribution



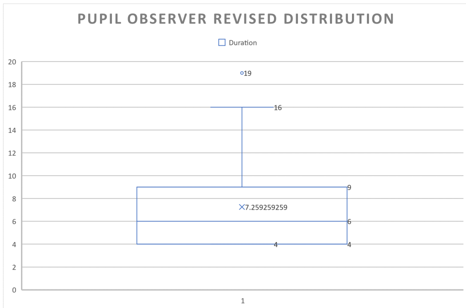
(c) R2 Normalised Disengaged Time spans



(d) R2 Normalised Distribution



(e) Pupil Observer Normalised Disengaged Time spans



(f) Pupil Observer Normalised Distribution

Figure 7.6: Normalised Observed Disengagement by Duration

7.3.5 Data Validation

Table 7.10 summarises the normalised observation of disengagement for all the included studies.

Table 7.10: Normalised Instances of Observations of Disengagement

Study ID	Pupil Observer	Validator 1	Validator 2
166	No Disengaged Status		
171	526 – 530 seconds	525 – 538 seconds	529 – 537 seconds
172	150 – 154 seconds	150 – 159 seconds	150 – 154 seconds
	313 – 319 seconds	312 – 316 seconds	
	345 – 350 seconds		
	405 – 409 seconds	400 – 404 seconds	397 – 408 seconds
196	75 – 79 seconds	74 – 80 seconds	69 – 79 seconds
	114 – 118 seconds	115 – 121 seconds	
212	24 – 31 seconds	20 – 35 seconds	
	37 – 46 seconds		
	72 – 77 seconds		
	85 – 90 seconds		
	133 – 147 seconds		
	158 – 163 seconds	160 – 166 seconds	
	185 – 192 seconds		
	211 – 217 seconds		
	231 – 235 seconds		
	301 – 311 seconds		
	357 – 373 seconds	355 – 363 seconds	353 – 361 seconds
	399 – 403 seconds	398 – 403 seconds	
	457 – 468 seconds		
219	24 – 31 seconds	21 – 37 seconds	19 – 38 seconds
	93 – 105 seconds	105 – 112 seconds	104 – 112 seconds
	133 – 138 seconds		
	172 – 180 seconds		
238	No Disengaged Status		

Study ID	Pupil Observer	Validator 1	Validator 2
242	No Disengaged Status		
243	No Disengaged Status		
244	627 – 631 seconds 681 – 694 seconds		
245	No Disengaged Status		
246	297 – 305 seconds		294 – 307 seconds

Study 212 appears to be different from the other studies in terms of the frequency of the observations. On review, there was little evidence to support the Pupil Observer's observations of disengagement. During the instance logged starting at 37 seconds, the learner looks amused and may be slightly distracted, but appears to continue working on the task. The Pupil Observer logs two instances of disengagement between 72 and 90 seconds each lasting 5 seconds where the learner continues to work. Likewise with the other observations it appears that the pupil may be amused by some external event which may be generated by the observer but does not appear to disengage from the task.

Discarding the observations from Study 212 leaves 11 studies and yields 14 instances of disengagement logged with 9 validated or 64% and a duration of 80 seconds. During the same period, the pupil observer recorded 3530 seconds of engagement, indicating that the learners were disengaged for just over 2% of the time while performing the task set. Table 7.11 summarises the validated disengagement data and indicates that the pupils were disengaged for just 53 seconds during the study.

Table 7.11: Validated Disengagement

Study ID	Validated Disengagement
171	4 seconds
172	4 seconds 6 seconds 4 seconds
196	4 seconds 4 seconds
219	7 seconds 12 seconds
246	8 seconds

7.4 PDLS Effectiveness

What then does this say about the effectiveness of the PDLS process? First, the majority of the data labelled during the process is verified by the reviewers. That is to say, in this study the default state of the learners is engaged and the process is accurate and effective in identifying engagement.

Identifying disengagement is more problematic, and the review process identified weaknesses in the process. Reviewers were divided on what constitutes disengagement in this context. R1 recorded a high number of instances of 2 or 3 seconds that R2 and the pupils did not log as disengagement.

R1 appears to have recorded a status when the pupil is temporarily distracted and the challenge here is that because the process is conducted in real time there is no way of knowing how long the distraction will last. The approach taken here was to discard the data, but an alternative approach could be to ignore small durations of disengagement and change the label to engaged. The latter approach would clearly help the efficiency of the PDLS process. Either way, training for the observers to aid consistency in drawing the distinction between temporary distraction and disengagement from task before they participate may create more consistent observations and improve the effectiveness of the process.

The reviewers indicated that the interface to record the engagement status was difficult to use and that the feedback from the software was not clear enough. The current system used a drop-down box to toggle between classifications, and this can be replaced by a clicker, which would facilitate recording statuses without having to concentrate on the screen. The textual on-screen feedback as to the current engagement level can be replaced with a larger graphic to help participants quickly identify the current recorded status.

7.5 Summary

Chapter 6 introduced the PDLS, a system for generating labelled data using peer observation. In this chapter, the effectiveness of the process is evaluated using both the iMotions software and human reviewers. The iMotions software did not produce consistent classifications and major amendments to the PDLS would be required if that validation route was pursued. The review process found that the pupil observers and reviewers reached consensus in classifying most of the data as engaged. Recognising disengagement is more challenging, and further work is required to ensure that there is more consistency in what the participants recognise as engagement. Several changes are proposed for the software interface prior to further studies to support more efficient recordings.

Chapter 8

Classifying Disengagement

8.1 Introduction

Chapter 2 discusses the principles of Machine Learning (ML) based systems and provides the theoretical basis for the work described here. This chapter applies the material to the creation of a ML based intelligent system trained to recognise disengagement in pupils whilst completing an educational task within a classroom environment. The process for both generating data and the task observed is described in Chapter 6.

The model described here is a variant of a Recurrent Neural Network (RNN) called the Long Short-Term Memory (LSTM) Model (Yu et al., 2019) and is selected for its ability to process sequences in data. The Model is pre-trained using the Inception-v3 Convolutional Neural Network (CNN)[1] using the ImageNet[2] dataset in a process referred to as transfer learning. The output from the model is a binary classification which characterises the engagement level of the pupil completing the task as either engaged (1) or disengaged (0) and writes the classification to the video output.

Due to the high demands placed on computer hardware when training such a model, it was trained using the Google Colab platform[3] and used a

[1] https://cloud.google.com/tpu/docs/inception-v3-advanced
[2] https://www.image-net.org
[3] https://colab.research.google.com

Tensor Processing Unit (TPU) hardware accelerator and a High-RAM runtime shape (Jouppi et al., 2017) to maximise performance.

8.2 Limitations of the Implementation

The model described here is intended to indicate the feasibility of developing an IS using a ML model for deployment in an educational context. It is not presented as a production model and the author acknowledges that much work remains to be done before the implementation could be fully tested.

Labelled data, or the lack of it, presents the major challenge. This thesis presents the PDLS as a potential solution to this but at the time of writing there is not enough data to build a reliable classification model. The loss of the video data from the second study conducted in October 2022 has not helped the situation, see section 6.2.6, but even had that been included, two small studies would not capture the diversity required to build a fully working model.

The data verification process described in Chapter 7 reduced the data labelled as disengaged further so in producing this implementation the choice was made to use the unvalidated data from the 17 studies described in section 6.2.6. This allowed for the model to be demonstrated but with the proviso that its output should be treated with care.

8.3 Convolutional Neural Networks (CNN)

The first stage of building the model is to pre-train it using a CNN. CNNs are a form of Deep Learning (DL) model that can be applied to a wide range of computer vision tasks where the model is required to perform some form of image classification (Gu et al., 2018). For an in-depth description of the components and workings of a CNN see Albawi et al. (2017).

More generally the convolution process generally written as $z = x * k$ where x is the input vector, effectively a series of pixels of size n and k is the kernel or filter of size l used to extract features from the input. Russell and Norvig (2020) define the convolution process as:

$$z_i = \sum_{j=1}^{l} k_j xj + i - (l+1)/2 \cdot \tag{8.1}$$

where for each position i of output z the dot product is calculated between k and the portion of x centred on x_i with width l.

CNNs are highly effective at automating feature extraction from the input image by applying a series of filters or kernels that each detect a different feature of the image. This is particularly useful as it removes the requirement for manual labelling of the data. Generally edges are detected in the first layers of the network and then primitive shapes followed by more detailed features. The features operate at higher levels of abstraction as the layers progress from input to output and the kernels are optimised during the training process as the model learns (Figure 8.1).

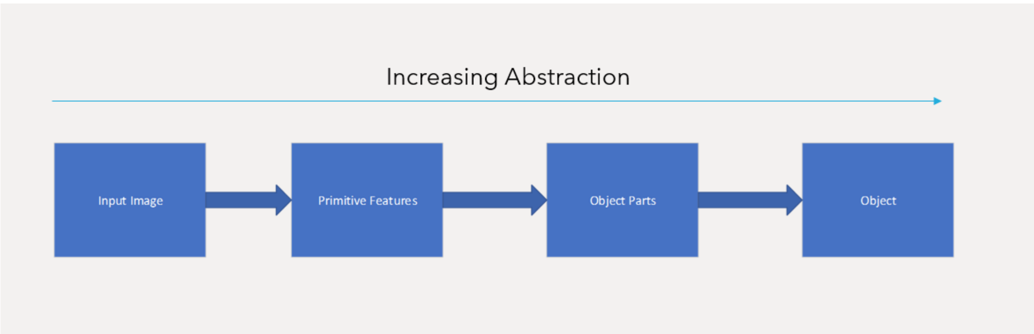


Figure 8.1: High Level Overview of the Convolution Process

Another advantage of CNNs over other Artificial Neural Network (ANN) architectures is that the CNN can be trained to recognise objects in images regardless of their position. In this aspect the CNN is described as spatially invariant (Russell and Norvig, 2020). Rather than fully connecting all the pixels in the input image to the first hidden layer, a cost of n^2, the CNN works with localised regions of the input which greatly reduces the computational cost of training. CNNs then are computationally more efficient than alternative fully connected architectures such as the Multi Layer Perceptron (MLP) (Noriega, 2005), are spatially invariant and can automate the process of feature extraction when applied to image classification. This implementation uses a pretrained version of the Inception-v3 CNN[4] which is then further trained on the labelled image files generated from the video data.

8.3.1 Transfer Learning

Chapter 2 introduces a form of learning, named supervised learning, where the model learns from labelled data and compares a predicted value y_hat against the known value y in order to learn. Such an approach clearly requires a suitable library of labelled images on which the model can be trained which may not be feasible. Chapter 6 notes the lack of labelled data suitable for use in behavioural studies with children. An alternative approach is to adopt an unsupervised approach where the model discovers patterns within the data without labels most commonly clustering the data in some way.

The method applied here utilises a third approach called transfer learning where knowledge learnt from one domain is transferred to another resulting in the requirement for less labelled data and quicker learning times. A human analogy may be that a person who has learnt to play a string instrument such as a violin may well find it easier to learn to play another string instrument such as a viola or cello than someone who has no previous experience. Transfer learning has been shown to be effective when utilised with the CNN architecture (Shaha and Pawar, 2018), (Shin et al., 2016). Effectively the weights from the pre-trained network are used as the starting point for the new model which is then trained using the new data and optimised using techniques such as gradient descent (see Chapter 2).

[4]https://keras.io/api/applications/inceptionv3/

8.4 Recurrent Neural Networks (RNN)

The CNN whilst efficient and effective at image classification particularly when pre-trained on existing data is not ideal as a model for classifying disengagement. Section 2.7 describes the flow of data through a feed forward network of which the CNN is an example. Feed forward networks are acyclic, effectively they handle each input in isolation where the input is independent of the output which makes them unsuitable for modelling data that possesses a temporal aspect such as the disengagement data described in Chapter 6. The RNN architecture allows cycles in that they use the output from the previous step as the input to the current one making them suitable for processing sequences in the data where there are dependencies between points in the data (Figure 8.2).

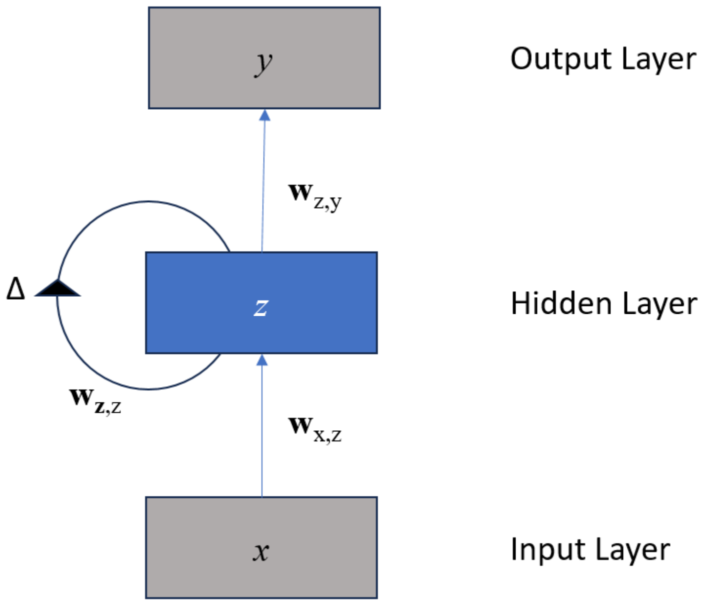


Figure 8.2: The Basic RNN structure adapted from Russell and Norvig (2020)

Section 2.7.1 describes the vanishing gradient problem which resulted in poor accuracy and long training times for ANNs.The RNN is affected not only by vanishing gradients but also by exploding gradients where the inputs become exponentially larger as they are multiplied together to much the same effect.

The implementation described in this chapter uses a variant of a RNN called the LSTM. LSTMs utilise a memory cell which is passed between timesteps at each stage rather than multiplying the output by the weight matrix (Russell and Norvig, 2020) and are effective for problems which require training over longer sequences and are less susceptible to vanishing and exploding gradients (Manaswi and Manaswi, 2018).

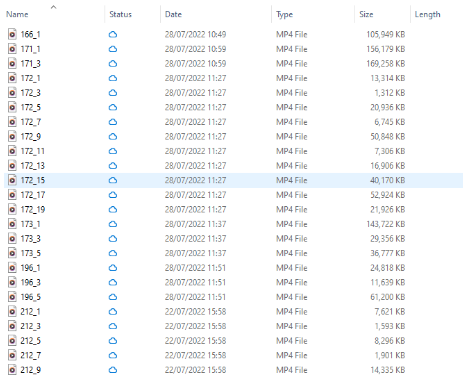
8.5 Implementation

This implementation predominantly uses the Python Keras[5] library as a wrapper for TensorFlow[6] to generate the LSTM and CNN models used to perform the classification. Due to the high level of processing required, training took place on the Google Colab platform.

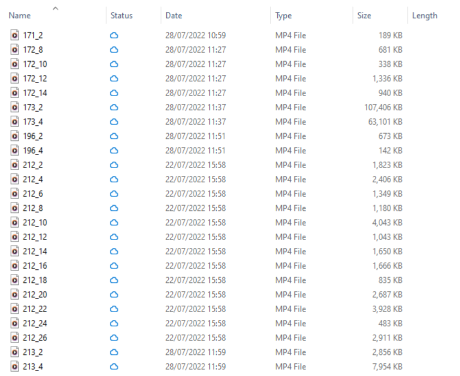
8.5.1 Data Preprocessing

The implementation uses the labelled data generated by the PDLS process described in Chapter 6. The original study saved the videos in .webm format[7] which were discovered to contain incomplete file metadata, in particular the video durations were missing. The durations were restored by converting the .webm files to .mp4 using the ffmpeg video processing utility. The videos were then split into engaged and disengaged segments based on the PDLS labels and written to separate directories (Figure 8.3).

[5]https://keras.io/
[6]https://www.tensorflow.org/
[7]https://www.webmproject.org/



(a) Engaged Video Files



(b) Disengaged Video Files

Figure 8.3: Segmented Video Files

To simplify batch loading to the model the .mp4 files were further segmented into durations of 3 seconds and standardised at 25 frames per second so that each video clip consisted of 75 .jpeg images (Figure 8.4).

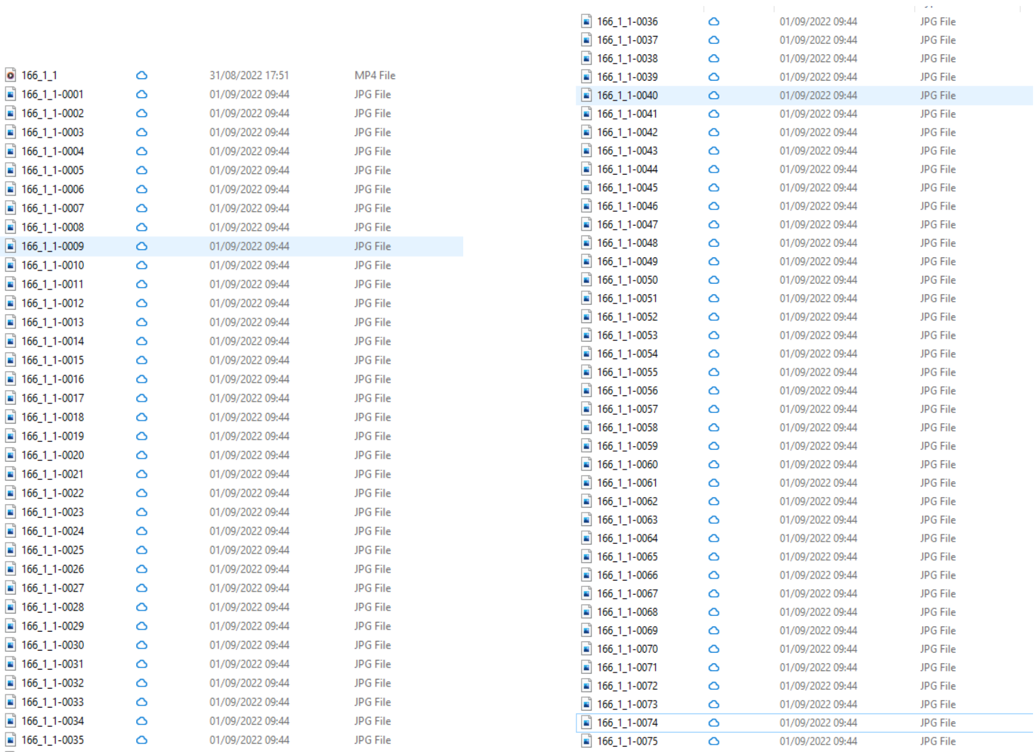


Figure 8.4: The first 3 second engaged segment of Study 166

The final stage of the data preprocessing phase was to divide the data into training and testing buckets with 70% of the data used for training the model and 30% reserved for testing. (Table 8.1).

Table 8.1: Data Breakdown by jpeg

	Engaged	Disengaged
Training	104,644	11,572
Testing	44,799	4,890

8.5.2 Model Training

The model uses the Inception v3 CNN (Szegedy et al., 2016) which is pre-trained on the ImageNet dataset to iteratively extract features from each jpeg file. Google describe Inception v3 as “an image recognition model that has been shown to attain greater than 78.1% accuracy on the ImageNet dataset (Google, 2023).” The model is then further trained using the labelled images generated by the pupil study (Chapter 6). The features extracted are saved to file creating a sequence file for each video. The process of feature extraction converts the raw data into numerical features which are then saved to file and used to train the model while preserving the original data (MathWorks, 2023).

These features are then passed to the LSTM network in order to perform the classification for each video. When training the model, a technique known as early stopping[8] which allows the developer to specify an arbitrarily high number of training epochs and stop early if classification accuracy has not improved is employed. In this case, early stopping was set at 20 epochs with error on the test data (val_loss) minimised on epoch 35 and training halting after 55 epochs. The best weights for each video are written to file and used to perform the classification.

[8] https://keras.io/api/callbacks/early_stopping/

8.5.3 Video Classification

The video classification process is similar to the training process. Python provides two main libraries for image manipulation, Pillow[9] and OpenCV[10]. Here, the OpenCV library is used to process the video frames and write the classification to a new video file. Each file is read frame by frame and feature extraction performed by the trained Inception v3 model to create a sequence object. The sequence object is passed to the predict method of the trained LSTM model which performs a binary classification of engaged (1) or disengaged (0) which is written to the top left corner of the output file. Figure 8.5 shows a pupil judged to be engaged with the task by the model.

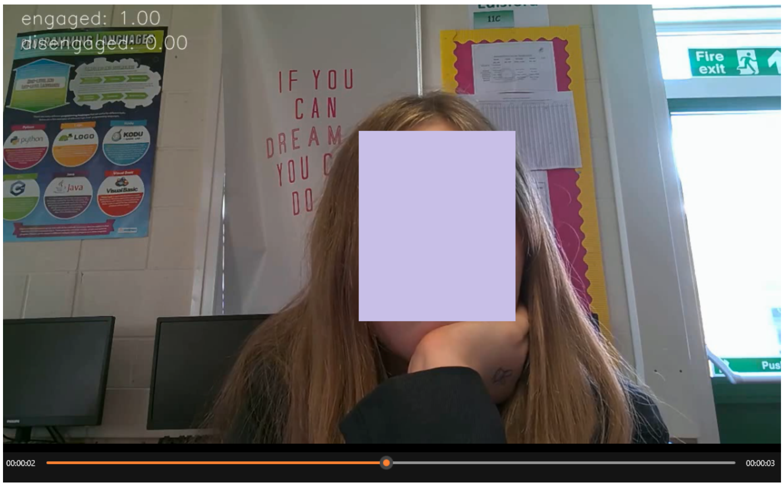


Figure 8.5: Engagement Classification

Writing the engagement status to the video is an implementation choice to demonstrate the process and not indicative of a preference in approach. Chapter 4 discusses alternative interface paradigms that the model could support all of which are viable embodiments of the IS. As such the model and the interface should be considered as distinct but connected components of the overall system.

[9]https://python-pillow.org/
[10]https://opencv.org/

8.6 Summary

This chapter provides an overview of an approach to implementing an engagement classifier for use with children in an educational context. The model described is a variant of a Recurrent Neural Network (RNN) called the Long Short-Term Memory (LSTM) Model and was selected for its ability to process sequences or cycles in the data. The output from the model is a binary classification which characterises the engagement level of the pupil completing the task as either engaged (1) or disengaged (0) and writes the classification to the video output.

In presenting the model the author acknowledges its limitations and it does not represent a production model but rather demonstrates the feasibility of the approach. Although the implementation displays the engagement classification to the video, this is not intended as a preference over other potential interfaces. As such the ML model which provides the engine for the implementation of this IS could support multiple embodiments of the system.

Chapter 9

Conclusion

9.1 Introduction

This chapter provides a summary of the work presented in this thesis. It revisits the research questions presented in Chapter 1.4 and assesses the extent to which they were answered. It goes on to outline proposed future research and development work for designing and developing intelligent systems for use in educational contexts.

9.2 Research Summary

The thesis set out to explore the challenge of designing and developing intelligent systems for use in educational contexts. The term educational context was used as opposed to classroom in acknowledgment that learning commonly takes place outside the traditional school building. The work and studies presented involved participants from the UK key stage 2 and key stage 3 age groups, in this case ages 7 to 15 inclusive.

The work completed spanned the period either side of the Covid-19 lockdowns in 2020 and 2021 and this affected the direction of the work. Work conducted prior to the pandemic is more general and discusses IS in general while work completed after the pandemic was influenced by the author’s own experience of teaching in an online environment during lockdown and is focussed on learner engagement.

The research recognises the diversity of the stakeholders in such a system and presents a modelling technique to balance their disparate needs named the Trust Acceptance Mapping Model (TAMM). It assesses the effects of anthropomorphising the system interface on children’s acceptance of an IS’s outputs and reviews how the research team may influence children’s perspectives in the way that they present a system interface.

A semi-systematic literature review was completed to explore the main research themes and methods employed in conducting HCI-AI and CCI-AI research. It identified few research contributions within the CHI and IDC research communities involving children and the computer vision techniques that are explored in the latter work conducted.

To address the lack of data available to CCI researchers employing computer vision techniques with children, the PDLS is presented. The PDLS is a novel peer labelling technique suitable for labelling video data at source. The PDLS is evaluated using the metrics efficiency, satisfaction, and effectiveness.

The work concludes by describing the implementation of a ML based classifier that monitors children for signs of disengagement when completing a computerised task. The system is trained on data generated by the PDLS and employs a LTSM model in recognition of the temporal nature of disengagement.

9.3 Answers to the Research Questions

9.3.1 RQ1 asked Are children more accepting of an intervention from a responsible adult than an intelligent system?

A comparison of the median scores recorded for the children's acceptance of an intervention by either an adult or a technology found that the children scored them both within a single point on the scale for all the given use cases. As such the children surveyed made little distinction between an intervention from an adult and an intervention by the IS.

9.3.2 RQ2 To what extent do parents and caregivers trust technology to monitor their child's digital activity? RQ3 To what extent do parents and caregivers trust technology to intervene in their child's digital activity?

Parents and caregivers indicated a higher level of trust in the technology's capability to monitor a use case than to take appropriate action. An analysis of the median scores indicated that these two facets were scored within a single point on the 10-point Likert scale employed. Parents and caregivers were generally neutral about the technologies capability to both monitor and intervene.

9.3.3 RQ4 Would teachers trust an intelligent system to monitor pupils in their classroom for signs of disengagement and make appropriate interventions?

When asked about their feelings regarding the deployment of a system to monitor engagement and make interventions only one of the teachers interviewed indicated that they would not accept the technology in their classroom. The other teachers indicated their acceptance subject to criteria, the most common of which was that the system outputs must be accurate and support the children's learning.

9.3.4 RQ5 How does anthropomorphising the system interface affect children's perceptions of its capabilities?

The perception of the children was that a humanised robot was able to learn from its experiences and adjust its behaviour accordingly whilst computers are programmed and slavishly follow rules.

9.3.5 RQ6 Does the way in which an intelligent system is presented to children influence their perceptions of its capabilities?

When the robot was introduced using the robot condition, 71% of the suggestions the children made were classified as requiring completion of a physical action. Only 14% of the children suggested an action that required intelligence and learning and there were 0 suggestions in the emotional category. When the robot was introduced using the human condition, the number of suggestions for a physical action fell to 39%, while suggestions for actions requiring intelligence and learning increased to 35% with 7% of suggestions categorised as emotional. It follows that the way in which the

IS is presented does influence children’s perception of its capabilities.

9.3.6 RQ7 What are the prevalent research areas in the fields of HCI-AI and CCI-AI?

In both the HCI-AI and CCI-AI domains, there is a significant body of research into the interaction with AI based applications. There is a greater emphasis on AI Ethics and Philosophy related research at CHI than IDC as a proportion of total papers. Conversely, IDC has a greater emphasis on Integration and Interaction than CHI predominantly consisting of research based around social robots. Both research groups have a small body of research around machine learning-based technologies.

9.3.7 RQ8 What are the undeveloped research areas in the fields of HCI-AI and CCI-AI?

There is minimal work at CHI around the domain of Perception and none at IDC. This is particularly pertinent to the research presented in this thesis specifically the latter work which looks to employ computer vision techniques with children in order to classify engagement within the classroom.

There are no papers included in the review classified as belonging to the Reasoning or Planning AI Domains. These areas are categorised by more theoretical concepts such as fuzzy logic or stochastic optimisation, and it is perhaps not surprising that they are not represented in this corpus of work which looks at the application of the AI fields as opposed to the theory that underpins them.

Only two papers were classified in the subdomain Philosophy of AI at CHI and no IDC papers.

9.3.8 RQ9 What are the prevalent research methods in the fields of HCI-AI and CCI-AI?

The majority of the studies reviewed at CHI used either some form of machine learning model or AI application as a vehicle for their studies. A number of papers used techniques classified as None-AI such as surveys or literature reviews. A similar pattern existed at IDC although fewer studies applied ML and a larger proportion used some form of robot.

9.3.9 RQ10 Are existing AI taxonomies sufficient to categorise research in the fields of HCI-AI and CCI-AI?

Samoili's taxonomy had sufficient coverage to satisfactorily classify the corpus of papers which constituted the review. There are some omissions at keyword level in the AI Ethics subdomain where some pertinent concepts are not represented. Namely interpretability, morality and of pertinence to this thesis, trust.

9.3.10 RQ11 Is peer observation an efficient method for generating labelled video data for use in identifying children's level of engagement with a computerised task in an educational context?

PDLS is judged to be a time and cost-efficient system that compares favourably with the other options considered. PDLS labels the data at the point of capture using peer judgments and does not require post-hoc analysis of the data.

9.3.11 RQ12 Is peer observation a satisfactory method for generating labelled video data for use in identifying children's level of engagement with a computerised task in an educational context?

Pupils indicated their satisfaction with both their own and their peers effectiveness in reaching a classification and the potential of a system built on the study data to make effective judgments. They expressed high confidence in their own ability to accurately measure the level of engagement of their classmate. They were only marginally less positive about the ability of their classmate to assess their own engagement levels.

9.3.12 RQ13 Is peer observation a effective method for generating labelled video data for use in identifying children's level of engagement with a computerised task in an educational context?

The majority of the data labelled during the process is verified by the reviewers. The default state of the learners is engaged and the process is accurate and effective in identifying engagement.

Identifying disengagement is more problematic, and the review process identified weaknesses in the PDLS process.

9.3.13 RQ14 Can a machine learning based intelligent system be built to recognise and classify children's engagement within an educational context?

Chapter 8 describes a ML based system that has been built to classify engagement based around the LTSM model. Other examples of similar implementations are described in Chapter 2.

9.3.14 RQ15 What are the main challenges to building and deploying a ML based model to recognise and classify children's engagement within an educational context?

There are a number of challenges to building and deploying such a model. From a technical perspective, the ML process requires significant hardware power to process the raw video footage. Of greater note is the current lack of data required to train the model. More work is required to both streamline the data labelling process and build a corpus of data that can accurately classify data from diverse users.

9.4 Limitations of this Work

In presenting this work it must be noted that its findings are based on a series of small studies. All the work reported took place in state schools within the UK. No selection of participants took place within those groups and it included children from multiple ethic groups.

The data generated by the PDLS study came from a single UK secondary school that as a Microsoft Academy, has an established IT infrastructure where children routinely used computers as part of their lessons which may have impacted on their approach to the computerised task. However, as the task was developed specifically for the study, it is not believed this will have greatly affected the results.

Whilst a ML based model has been built as an engagement classifier, considerably more work is required to assess its effectiveness. In recognition that many of the problems lie with the quality of the input data and in particular the lack of disengaged data, that task is deferred for further work.

9.5 Further Work

Further work will concentrate on the development of the PDLS process. In particular the accuracy of the data labelling process which the author intends to support with better briefing for participants and an improved interface for logging. It is intended that a library of tasks will be developed to better support data collection. In doing so it is hoped that a corpus of data will be built to support other researchers who wish to carry out ML based tasks with children.

As ML models develop over time it is expected that the system implementation details will change and thorough verification of the systems outputs is required if it is shown to be fit for purpose. Of particular interest are the implementation choices around the interface given the preferences of the main stakeholders and it will be exciting to see how that develops.

9.6 Closing Remarks

This work addresses the challenges of designing and developing intelligent systems for use in educational contexts and covers the full software development lifecycle. The research project took place either side of the Covid-19 lockdowns which greatly influenced its direction.

Much of the work described in these pages is supported and validated by peer reviewed publications. Chapter 2 discussed the challenges of conducting work in this arena and many of the ideas were presented at the 4th International Conference on Human-Computer Interaction and User Experience in Indonesia 2018 (Read et al., 2018).

Chapter 3 considered design implications for IS and presents the Trust Acceptance Mapping Model (TAMM) as an approach to verifying user requirements where a system has multiple stakeholders. This work was presented at HCII 2023 (Parsonage et al., 2023*b*).

Chapter 4 looked at implementation choices for the system and in particular how anthropomorphism affects children's expectations of the capabilities of an IS. This work alongside guidelines for conducting experiments with children and robots was published at IDC ’20 (Parsonage et al., 2020).

Chapter 5 considered the literature covering the intersection between HCI, CCI, and AI referred to as HCI-AI and CCI-AI. It aims to both identify current research trends and position this work within the body of existing work.

Chapters 6 and 7 introduced the Peer Data Labelling System (PDLS) as a novel approach to labelling engagement data suitable for training a ML based IS. This work was published at Interact 2023 (Parsonage et al., 2023*a*). Chapter 8 presents an ML based approach to building a system that utilises data gathered by use of the PDLS.

There are two themes developed throughout the thesis that are worthy of special mention. The first is trust which was deemed so important a model was developed around it, see Chapter 3. The second is engagement on which the post Covid-19 work described in Chapters 6 - 8 focussed. Finally, in no way should the work described be considered as complete. Rather, a staging point has been reached and foundations laid for future work. Much remains to be done; particularly, around developing the PDLS and the data generated to design and develop ML based intelligent systems for use in educational contexts.

Bibliography

Ackermann, E. (2001), ‘Piaget’s constructivism, papert’s constructionism: What’s the difference’, *Future of learning group publication* **5**(3), 438.

Agar, J. (2020), ‘What is science for? the lighthill report on artificial intelligence reinterpreted’, *The British Journal for the History of Science* **53**(3), 289–310.

Agarap, A. F. (2018), ‘Deep learning using rectified linear units (relu)’, *arXiv preprint arXiv:1803.08375* .

Agarwal, M. and Sivakumar, R. (2020), ‘Charge for a whole day: Extending battery life for bci wearables using a lightweight wake-up command’, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Agostinelli, F., Mavalankar, M., Khandelwal, V., Tang, H., Wu, D., Berry, B., Srivastava, B., Sheth, A. and Irvin, M. (2021), ‘Designing children’s new learning partner: collaborative artificial intelligence for learning to solve the rubik’s cube’, pp. 610–614.

Aki Tamashiro, M., Van Mechelen, M., Schaper, M.-M. and Sejer Iversen, O. (2021), ‘Introducing teenagers to machine learning through design fiction: an exploratory case study’, pp. 471–475.

Albawi, S., Mohammed, T. A. and Al-Zawi, S. (2017), Understanding of a convolutional neural network, *in* ‘2017 international conference on engineering and technology (ICET)’, Ieee, pp. 1–6.

Albo-Canals, J., Martelo, A. B., Relkin, E., Hannon, D., Heerink, M., Heinemann, M., Leidl, K. and Bers, M. U. (2018), ‘A pilot study of the kibo robot in children with severe asd’, *International Journal of Social Robotics* **10**, 371–383.

Alkabbany, I., Ali, A., Farag, A., Bennett, I., Ghanoum, M. and Farag, A. (2019), Measuring student engagement level using facial information, *in* ‘2019 IEEE International Conference on Image Processing (ICIP)’, IEEE, pp. 3337–3341.

Almjally, A., Howland, K. and Good, J. (2020), ‘Investigating children’s spontaneous gestures when programming using tuis and guis’, pp. 36–48.

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Van Esesn, B. C., Awwal, A. A. S. and Asari, V. K. (2018), ‘The history began from alexnet: A comprehensive survey on deep learning approaches’, *arXiv preprint arXiv:1803.01164* .

AlOmar, E. A., Aljedaani, W., Tamjeed, M., Mkaouer, M. W. and El-Glaly, Y. N. (2021), ‘Finding the needle in a haystack: On the automatic identification of accessibility user reviews’, *Proceedings of the 2021 CHI conference on human factors in computing systems* pp. 1–15.

Alves-Oliveira, P., Arriaga, P., Paiva, A. and Hoffman, G. (2021), Children as robot designers, *in* ‘Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction’, pp. 399–408.

Andrade, A., Dionisio, M. and Nisi, V. (2022), ‘Seastory: An interactive narrative using collaborative features’, pp. 633–636.

Anik, A. I. and Bunt, A. (2021), ‘Data-centric explanations: Explaining training data of machine learning systems to promote transparency’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Antle, A. N., Murai, Y., Kitson, A., Candau, Y., Dao-Kroeker, Z. M.-T. and Adibi, A. (2022), ‘“there are a lot of moral issues with biowearables”...

teaching design ethics through a critical making biowearable workshop', pp. 327–340.

Antona, M., Margetis, G., Ntoa, S. and Degen, H. (2023), 'Special issue on ai in hci', *International Journal of Human-Computer Interaction* pp. 1–4. **URL:** *https://www.tandfonline.com/action/journalInformation?journalCode=hihc20*

Arakawa, R. and Yakura, H. (2019), 'Rescue: A framework for real-time feedback on behavioral cues using multimodal anomaly detection', *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Arakawa, R., Yakura, H. and Kobayashi, S. (2022), 'Vocabencounter: Nmt-powered vocabulary learning by presenting computer-generated usages of foreign words into users' daily lives', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–21.

Arnold, L., Lee, K. J. and Yip, J. C. (2016), 'Co-designing with children: An approach to social robot design', *ACM Human-Robot Interaction (HRI)* .

Asai, K., Fukusato, T. and Igarashi, T. (2020), 'Integrated development environment with interactive scatter plot for examining statistical modeling', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–7.

Ashktorab, Z., Dugan, C., Johnson, J., Pan, Q., Zhang, W., Kumaravel, S. and Campbell, M. (2021), 'Effects of communication directionality and ai agent differences in human-ai interaction', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–15.

Asimov, I. (1941), 'Three laws of robotics', *Asimov, I. Runaround* **2**.

Bachynskyi, M. and Müller, J. (2020), 'Dynamics of aimed mid-air movements', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–12.

Badillo-Urquiola, K., Smriti, D., McNally, B., Golub, E., Bonsignore, E. and

Wisniewski, P. J. (2019), ‘Stranger danger! social media app features co-designed with children to keep them safe online’, pp. 394–406.

Bae, S., Yang, R., Gyory, P., Uhr, J., Szafir, D. A. and Do, E. Y.-L. (2021), ‘Touching information with diy paper charts & ar markers’, pp. 433–438.

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T. and Weld, D. (2021), ‘Does the whole exceed its parts? the effect of ai explanations on complementary team performance’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–16.

Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M. and Šabanović, S. (2020), *Human-robot interaction: An introduction*, Cambridge University Press.

Belgiu, M. and Drăgut, L. (2014), ‘Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery’, *ISPRS Journal of Photogrammetry and Remote Sensing* **96**, 67–75.

Benjamin, J. J., Berger, A., Merrill, N. and Pierce, J. (2021), ‘Machine learning uncertainty as a design material: a post-phenomenological inquiry’, *Proceedings of the 2021 CHI conference on human factors in computing systems* pp. 1–14.

Bennett, C. L., Gleason, C., Scheuerman, M. K., Bigham, J. P., Guo, A. and To, A. (2021), ‘“it’s complicated”: Negotiating accessibility and (mis) representation in image descriptions of race, gender, and disability’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–19.

Berry, M. W., Mohamed, A. and Yap, B. W. (2019), *Supervised and unsupervised learning for data science*, Springer.

Bhaduri, S., Biddy, Q. L., Bush, J., Suresh, A. and Sumner, T. (2021), ‘3dnst: A framework towards understanding children’s interaction with tinkercad and enhancing spatial thinking skills’, pp. 257–267.

Bhavitha, B. K., Rodrigues, A. P. and Chiplunkar, N. N. (2017), 'Comparative study of machine learning techniques in sentimental analysis', *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)* pp. 216–221.

Bilstrup, K.-E. K., Kaspersen, M. H., Lunding, M. S., Schaper, M.-M., Van Mechelen, M., Tamashiro, M. A., Smith, R. C., Iversen, O. S. and Petersen, M. G. (2022), 'Supporting critical data literacy in k-9 education: three principles for enriching pupils' relationship to data', pp. 225–236.

Bishay, M., Preston, K., Strafuss, M., Page, G., Turcot, J. and Mavadati, M. (2022), 'Affdex 2.0: A real-time facial expression analysis toolkit', *arXiv preprint arXiv:2202.12059* .

Biswas, S. S. (2023), 'Potential use of chat gpt in global warming', *Annals of biomedical engineering* **51**(6), 1126–1127.

Boffi, L. (2020), 'Ding-dong: the storybell and its wizard', pp. 421–425.

Boggust, A., Hoover, B., Satyanarayan, A. and Strobelt, H. (2022), 'Shared interest: Measuring human-ai alignment to identify recurring patterns in model behavior', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–17.

Booth, A., Sutton, A., Clowes, M. and Martin-St James, M. (2022), *Systematic approaches to a successful literature review*, third edition. edn, SAGE, Los Angeles.

Bory, P. (2019), 'Deep new: The shifting narratives of artificial intelligence from deep blue to alphago', *Convergence* **25**(4), 627–642.

Brahimi, T. and Sarirete, A. (2015), 'Learning outside the classroom through moocs', *Computers in Human Behavior* **51**, 604–609.
URL: *http://dx.doi.org/10.1016/j.chb.2015.03.013*

Bridle, J. S. (1990), 'Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition', pp. 227–236.

Bruns, M., Ossevoort, S. and Petersen, M. G. (2021), 'Expressivity in interaction: A framework for design', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Bryant, D. and Howard, A. (2019), A comparative analysis of emotion-detecting ai systems with respect to algorithm performance and dataset diversity, *in* 'Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society', pp. 377–382.

Bunian, S., Li, K., Jemmali, C., Harteveld, C., Fu, Y. and El-Nasr, M. S. S. (2021), 'Vins: Visual search for mobile user interface design', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Buono, P., De Carolis, B., D'Errico, F., Macchiarulo, N. and Palestra, G. (2023), 'Assessing student engagement from facial behavior in on-line learning', *Multimedia Tools and Applications* **82**(9), 12859–12877.

Burneleit, E. and Hemmert, F. (2009), 'Living interfaces : The impatient toaster', *International Conference on Tangible, Embedded and Embodied Interaction* pp. 21–22.

Bäuerle, A., Ángel Alexander Cabrera, Hohman, F., Maher, M., Koski, D., Suau, X., Barik, T. and Moritz, D. (2022), 'Symphony: Composing interactive interfaces for machine learning', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Cabibihan, J.-J., Javed, H., Ang, M. and Aljunied, S. M. (2013), 'Why robots? a survey on the roles and benefits of social robots in the therapy of children with autism', *International journal of social robotics* **5**, 593–618.

Cabrera, L., Maloney, J. H. and Weintrop, D. (2019), 'Programs in the palm of your hand: How live programming shapes children's interactions with physical computing devices', pp. 227–236.

Cagiltay, B., Ho, H.-R., Michaelis, J. E. and Mutlu, B. (2020), 'Investigating family perceptions and design preferences for an in-home robot', pp. 229–242.

Cagiltay, B., Michaelis, J., Sebo, S. and Mutlu, B. (2022), 'Exploring children's preferences for taking care of a social robot', pp. 382–388.

Cagiltay, B., White, N. T., Ibtasar, R., Mutlu, B. and Michaelis, J. (2022), 'Understanding factors that shape children's long term engagement with an in-home learning companion robot', pp. 362–373.

Cai, W., Jin, Y. and Chen, L. (2022), 'Impacts of personal characteristics on user trust in conversational recommender systems', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Cambo, S. A. and Gergle, D. (2022), 'Model positionality and computational reflexivity: Promoting reflexivity in data science', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–19.

Cambridge Dictionary (2023), 'Intelligent system'. Accessed: 27th June 2023.
URL: *https://dictionary.cambridge.org/example/english/intelligent-system*

Celebi, M. E. and Aydin, K. (2016), *Unsupervised learning algorithms*, Springer.

Chang, A. (2012), 'Utaut and utaut 2: A review and agenda for future research', *The Winners* **13**(2), 10–114.

Charisi, V. (2018), 'Session details: Child-robot interaction'.
URL: *https://doi.org/10.1145/3247769*

Charisi, V., Imai, T., Rinta, T., Nakhayenze, J. M. and Gomez, R. (2021), 'Exploring the concept of fairness in everyday, imaginary and robot scenarios: A cross-cultural study with children in japan and uganda', pp. 532–536.

Chatain, J., Ramp, V., Gashaj, V., Fayolle, V., Kapur, M., Sumner, R. W. and Magnenat, S. (2022), 'Grasping derivatives: Teaching mathematics through embodied interactions using tablets and virtual reality', pp. 98–108.

Cheema, N., Frey-Law, L. A., Naderi, K., Lehtinen, J., Slusallek, P. and Hämäläinen, P. (2020), 'Predicting mid-air interaction movements and fatigue using deep reinforcement learning', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Chen, F., Xia, K., Dhabalia, K. and Hong, J. I. (2019), 'Messageontap: a suggestive interface to facilitate messaging-related tasks', *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Chen, J. X., Vitale, F. and McGrenere, J. (2021), 'What happens after death? using a design workbook to understand user expectations for preparing their data', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Chen, L., Takashima, K., Fujita, K. and Kitamura, Y. (2021), 'Pinpointfly: An egocentric position-control drone interface using mobile ar', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Cheng, H.-F., Stapleton, L., Wang, R., Bullock, P., Chouldechova, A., Wu, Z. S. S. and Zhu, H. (2021), 'Soliciting stakeholders' fairness notions in child maltreatment predictive systems', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–17.

Cheung, V., Antle, A. N., Sarker, S., Fan, M., Fan, J. and Pasquier, P. (2020), 'Techniques for augmented-tangibles on mobile devices for early childhood learning', pp. 589–601.

CHI (2022), 'Acm chi conference on human factors in computing systems'. Accessed: 5th June 2023.
URL: *https://chi2022.acm.org/*

CHI (2023*a*), 'Chi '23 search'. Accessed: 5th June 2023.
URL: *https://programs.sigchi.org/chi/2023/search/content*

CHI (2023*b*), 'Chi conference - proceedings'. Accessed: 5th June 2023. **URL:** *https://dl.acm.org/conference/chi/proceedings*

Chiou, Y.-M., Barnes, T., Jelenewicz, S. M., Mouza, C. and Shen, C.-C. (2021), 'Teacher views on storytelling-based cybersecurity education with social robots', pp. 508–512.

Choi, M., Park, C., Yang, S., Kim, Y., Choo, J. and Hong, S. R. (2019), 'Aila: Attentive interactive labeling assistant for document classification through attention-based deep neural networks', *Proceedings of the 2019 CHI conference on human factors in computing systems* pp. 1–12.

Christenson, S., Reschly, A. L., Wylie, C. et al. (2012*a*), *Handbook of research on student engagement*, Vol. 840, Springer.

Christenson, S., Reschly, A. L., Wylie, C. et al. (2012*b*), *Handbook of research on student engagement*, Vol. 840, Springer.

Chung, J. J. Y., Kim, W., Yoo, K. M., Lee, H., Adar, E. and Chang, M. (2022), 'Talebrush: sketching stories with generative pretrained language models', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–19.

Cila, N. (2022), 'Designing human-agent collaborations: Commitment, responsiveness, and support', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–18.

Cimolino, G. and Graham, T. C. N. (2022), 'Two heads are better than one: A dimension space for unifying human and artificial intelligence in shared control', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–21.

Cohn, J. F., Ambadar, Z. and Ekman, P. (2007), 'Observer-based measurement of facial expression with the facial action coding system', *The handbook of emotion elicitation and assessment* **1**(3), 203–221.

Constant, T. and Levieux, G. (2019), 'Dynamic difficulty adjustment impact on players' confidence', *Proceedings of the 2019 CHI conference on human factors in computing systems* pp. 1–12.

Cook, J. L. and Cook, G. (2005), *Child development: Principles and perspectives.*, Pearson Education New Zealand.

Corbett, E. and Dantec, C. L. (2021), 'Designing civic technology with trust', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–17.

Costa, S., Lehmann, H., Dautenhahn, K., Robins, B. and Soares, F. (2015), 'Using a humanoid robot to elicit body awareness and appropriate physical interaction in children with autism', *International journal of social robotics* **7**, 265–278.

Coughlan, S. (2020), 'Why did the a-level algorithm say no?'. Accessed on 06.01.2023.
URL: *https://www.bbc.co.uk/news/education-53787203*

Cowan, B. R., Bowers, C. P., Beale, R. and Pinder, C. (2013), 'The stroppy kettle: an intervention to break energy consumption habits. in chi'13 extended abstracts on human factors in computing systems (pp. bt - 1485e1490).'.

Crompton, H., Gregory, K. and Burke, D. (2018), 'Humanoid robots supporting children's learning in an early childhood setting', *British Journal of Educational Technology* **49**(5), 911–927.

Cumbo, B. J. and Iversen, O. S. (2020), 'Cci in the wild: Designing for environmental stewardship through children's nature-play', pp. 335–348.

Câmara Olim, S. M., Nisi, V. and Rubegni, E. (2022), 'Periodic fable augmenting chemistry with technology, characters and storytelling', *Interaction Design and Children* pp. 123–136.

Dai, J. and Moffatt, K. (2021), 'Surfacing the voices of people with dementia: strategies for effective inclusion of proxy stakeholders in qualitative research', *Proceedings of the 2021 CHI conference on human factors in computing systems* pp. 1–13.

Dang, H., Mecke, L. and Buschek, D. (2022), 'Ganslider: How users control generative models for images using multiple sliders with and without feedforward information', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–15.

Dao-Kroeker, Z., Kitson, A., N. Antle, A., Murai, Y. and Adibi, A. (2021), 'Designing biotech ethics cards: Promoting critical making during an online workshop with youth', pp. 450–455.

Databases (A-Z) - face image databases - research guides at Princeton University (2022). Accessed 0n 18.01.2023.
URL: *https://libguides.princeton.edu/facedatabases*

Dawe, J., Sutherland, C., Barco, A. and Broadbent, E. (2019), 'Can social robots help children in healthcare contexts? a scoping review', *BMJ paediatrics open* **3**(1).

Decuypere, M., Alirezabeigi, S., Grimaldi, E., Hartong, S., Kiesewetter, S., Landri, P., Masschelein, J., Piattoeva, N., Ratner, H., Simons, M. et al. (2023), 'Laws of edu-automation? three different approaches to deal with processes of automation and artificial intelligence in the field of education', *Postdigital Science and Education* **5**(1), 44–55.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L. (2009), Imagenet: A large-scale hierarchical image database, *in* '2009 IEEE conference on computer vision and pattern recognition', Ieee, pp. 248–255.

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019), 'Bert: Pre-training of deep bidirectional transformers for language understanding', *arXiv preprint arXiv:1810.04805* .
URL: *https://github.com/tensorflow/tensor2tensor*

Dietz, G., King Chen, J., Beason, J., Tarrow, M., Hilliard, A. and Shapiro, R. B. (2022*a*), Artonomous: Introducing middle school students to

reinforcement learning through virtual robotics, *in* ‘Interaction Design and Children’, IDC ’22, Association for Computing Machinery, New York, NY, USA, p. 430–441.
URL: *https://doi.org/10.1145/3501712.3529736*

Dietz, G., King Chen, J., Beason, J., Tarrow, M., Hilliard, A. and Shapiro, R. B. (2022*b*), ‘Artonomous: Introducing middle school students to reinforcement learning through virtual robotics’, pp. 430–441.

DiPaola, D., Payne, B. H. and Breazeal, C. (2020), ‘Decoding design agendas: an ethical design activity for middle school students’, pp. 1–10.

Došilović, F. K., Brčić, M. and Hlupić, N. (2018), Explainable artificial intelligence: A survey, *in* ‘2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)’, IEEE, pp. 0210–0215.

Dove, G. and Fayard, A.-L. (2020), ‘Monsters, metaphors, and machine learning’, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–17.

Druga, S., Ball, T. and Ko, A. (2022), ‘How families design and program games: a qualitative analysis of a 4-week online in-home study’, pp. 237–252.

Druga, S. and Ko, A. J. (2021), ‘How do children’s perceptions of machine intelligence change when training and coding smart programs?’, pp. 49–61.

Du, X. and Breazeal, C. (2022), ‘Exploring changes in special education teachers’ attitudes and design belief towards pedagogical agents in co-designing with children’, pp. 472–478.

Dupré, D., Krumhuber, E. G., Küster, D. and McKeown, G. J. (2020), ‘A performance comparison of eight commercially available automatic classifiers for facial affect recognition’, *PloS one* **15**(4), e0231968.

Durand, K., Gallay, M., Seigneuric, A., Robichon, F. and Baudouin, J.-Y.

(2007), ‘The development of facial emotion recognition: The role of configural information’, *Journal of experimental child psychology* **97**(1), 14–27.

Edwards, A., Edwards, C., Spence, P. R., Harris, C. and Gambino, A. (2016), ‘Robots in the classroom: Differences in students’ perceptions of credibility and learning between “teacher as robot” and “robot as teacher”’, *Computers in Human Behavior* **65**, 627–634.

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O. and Weisz, J. D. (2021), ‘Expanding explainability: Towards social transparency in ai systems’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–19.

Ekman, P. (2020), ‘Facial action coding system’.
URL: *https://www.paulekman.com/facial-action-coding-system/*

Ekman, P. and Friesen, W. V. (1978), ‘Facial action coding system’, *Environmental Psychology & Nonverbal Behavior* .

El Shemy, I. (2022), ‘Language learning with mobile augmented reality and artificial intelligence for children with autism spectrum disorder’, pp. 658–661.

Elbeleidy, S., Rosen, D., Liu, D., Shick, A. and Williams, T. (2021), ‘Analyzing teleoperation interface usage of robots in therapy for children with autism’, pp. 112–118.

Epley, N., Waytz, A. and Cacioppo, J. T. (2007), ‘On seeing human: A three-factor theory of anthropomorphism.’, *Psychological Review* **114**, 864–886.
URL: *http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.114.4.864*

Eriksson, E., Baykal, G. E. and Torgersson, O. (2022), ‘The role of learning theory in child-computer interaction-a semi-systematic literature review’, pp. 50–68.

Escobar-Planas, M. (2022), 'Towards trustworthy conversational agents for children', pp. 693–695.

Ettehadi, O., Anderson, F., Tindale, A. and Somanath, S. (2021), 'Documented: Embedding information onto and retrieving information from 3d printed objects', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–11.

Eyssel, F., Kuchenbrandt, D. and Bobinger, S. (2011), 'Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism', *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on* pp. 61–67.

Fails, J. A., Pera, M. S., Anuyah, O., Kennington, C., Wright, K. L. and Bigirimana, W. (2019), 'Query formulation assistance for kids: What is available, when to help & what kids want', pp. 109–120.

Ferreira, P. S. (2022), '"quero fazer": A game featured platform for early childhood informal education', pp. 665–667.

Ferrucci, D. A. (2012), 'Introduction to "this is watson"', *IBM Journal of Research and Development* **56**(3.4), 1–1.

Fisher, C. W. and Berliner, D. C. (1985), *Perspectives on instructional time*, Addison-Wesley Longman Limited.

Fredricks, J. A., Blumenfeld, P. C. and Paris, A. H. (2004*a*), 'School engagement: Potential of the concept, state of the evidence', *Review of Educational Research* **74**, 59–109.

Fredricks, J. A., Blumenfeld, P. C. and Paris, A. H. (2004*b*), 'School engagement: Potential of the concept, state of the evidence', *Review of educational research* **74**(1), 59–109.

Fuhrmann, T., Ahmed, D. I., Arikson, L., Wirth, M., Miller, M. L., Li, E., Lam, A., Blikstein, P. and Riedel-Kruse, I. (2021), 'Scientific inquiry in middle schools by combining computational thinking, wet lab experiments, and liquid handling robots', pp. 444–449.

Gagan, I. T. M., Matias, M. A. M. E., Tan, I., Vinco, C. M., Ong, E. and Resurreccion, R. (2022), 'Designing a virtual talking companion to support the social-emotional learning of children with asd', pp. 464–471.

Gansser, O. A. and Reich, C. S. (2021), 'A new acceptance model for artificial intelligence with extensions to utaut2: An empirical study in three segments of application', *Technology in Society* **65**, 101535.

Gao, X. and Maurer, D. (2009), 'Influence of intensity on children's sensitivity to happy, sad, and fearful facial expressions', *Journal of experimental child psychology* **102**(4), 503–521.

Gero, K. I., Ashktorab, Z., Dugan, C., Pan, Q., Johnson, J., Geyer, W., Ruiz, M., Miller, S., Millen, D. R., Campbell, M. et al. (2020), 'Mental models of ai agents in a cooperative game setting', *Proceedings of the 2020 chi conference on human factors in computing systems* pp. 1–12.

Ghosh-Dastidar, S. and Adeli, H. (2009), 'Spiking neural networks', *International journal of neural systems* **19**(04), 295–308.

Glaser, B. G. and Strauss, A. L. (1967), 'The discovery of grounded theory: strategies for qualitative research (grounded theory)', *Taylor & Francis eBooks DRM Free Collection* .

Glikson, E. and Woolley, A. W. (2020), 'Human trust in artificial intelligence: Review of empirical research', *Academy of Management Annals* **14**, 627–660.

Google (2023), 'Advanced guide to inception v3'. Accessed on 9th August 2023.
URL: *https://cloud.google.com/tpu/docs/inception-v3-advanced*

Gordon, M. L., Zhou, K., Patel, K., Hashimoto, T. and Bernstein, M. S. (2021), 'The disagreement deconvolution: Bringing machine learning performance metrics in line with reality', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Gov UK (2023), 'Case management guidance - definitions - guidance -

gov.uk'. Accessed: 26th June 2023.
URL:
https://www.gov.uk/guidance/case-management-guidance/definitions

Grace, K., Finch, E., Gulbransen-Diaz, N. and Henderson, H. (2022), 'Q-chef: The impact of surprise-eliciting systems on food-related decision-making', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Grant, M. J. and Booth, A. (2009), 'A typology of reviews: an analysis of 14 review types and associated methodologies', *Health information and libraries journal* **26**, 91–108.

Gray, C. M., Santos, C., Bielova, N., Toth, M. and Clifford, D. (2021), 'Dark patterns and the legal requirements of consent banners: An interaction criticism perspective', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–18.

Greeno, J. G. (1998), 'The situativity of knowing, learning, and research.', *American psychologist* **53**(1), 5.

Groccia, J. E. (2018*a*), 'What is student engagement?', *New directions for teaching and learning* pp. 11–20.
URL: *https://onlinelibrary.wiley.com/doi/10.1002/tl.20287*

Groccia, J. E. (2018*b*), 'What is student engagement?', *New directions for teaching and learning* **2018**(154), 11–20.

Gross, A. L. and Ballif, B. (1991), 'Children's understanding of emotion from facial expressions and situations: A review', *Developmental review* **11**(4), 368–398.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. et al. (2018), 'Recent advances in convolutional neural networks', *Pattern recognition* **77**, 354–377.

Guneysu Ozgur, A., Majlesi, A. R., Taburet, V., Meijer, S., Leite, I. and Kuoppamäki, S. (2022), 'Designing tangible robot mediated co-located games to enhance social inclusion for neurodivergent children', pp. 536–543.

Guo, J., Guo, J., Yang, C., Wu, Y. and Sun, L. (2021), 'Shing: A conversational agent to alert customers of suspected online-payment fraud with empathetical communication skills', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–11.

Gürbüzsel, İ., Göksun, T. and Coşkun, A. (2022), 'Eliciting parents' insights into products for supporting and tracking children's fine motor development', pp. 544–550.

Guzdial, M., Liao, N., Chen, J., Chen, S.-Y., Shah, S., Shah, V., Reno, J., Smith, G. and Riedl, M. O. (2019), 'Friend, collaborator, student, manager: How design of an ai-driven game level editor affects creators', *Proceedings of the 2019 CHI conference on human factors in computing systems* pp. 1–13.

Hadash, S., Willemsen, M. C., Snijders, C. and IJsselsteijn, W. A. (2022), 'Improving understandability of feature contributions in model-agnostic explainable ai tools', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–9.

Hadfield, J., Chalvatzaki, G., Koutras, P., Khamassi, M., Tzafestas, C. S. and Maragos, P. (2019), A deep learning approach for multi-view engagement estimation of children in a child-robot joint attention task, *in* '2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)', IEEE, pp. 1251–1256.

Han, X., Zhou, M., Turner, M. J. and Yeh, T. (2021), 'Designing effective interview chatbots: Automatic chatbot profiling and design suggestion generation for chatbot debugging', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–15.

Hebb, D. O. (1949), *The Organization of Behavior: A Psychological Theory*, Wiley New York.

Hewett, T. T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G. and Verplank, W. (1992), *Curricula for Human-Computer Interaction ACM Special Interest Group on Computer-Human Interaction Curriculum Development Group ADVANCING HUMAN CAPABILITIES THROUGH INFORMATION TECHNOLOGY*, Association of Computing Machinery (ACM).

Hiniker, A., Wang, A., Tran, J., Zhang, M. R., Radesky, J., Sobel, K. and Hong, S. R. (2021), 'Can conversational agents change the way children talk to people?', pp. 338–349.

Ho, H.-R., Cagiltay, B., White, N. T., Hubbard, E. M. and Mutlu, B. (2021), 'Robomath: Designing a learning companion robot to support children's numerical skills', pp. 283–293.

Ho, M.-L., Ratakonda, D. k. and Fails, J. A. (2022), 'Kidconnect vr: Technology to stay connected', pp. 645–648.

Hochreiter, S. (1998), 'The vanishing gradient problem during learning recurrent neural nets and problem solutions', *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **6**(02), 107–116.

Hohman, F., Head, A., Caruana, R., DeLine, R. and Drucker, S. M. (2019), 'Gamut: A design probe to understand how data scientists understand machine learning models', *Proceedings of the 2019 CHI conference on human factors in computing systems* pp. 1–13.

Hohman, F., Wongsuphasawat, K., Kery, M. B. and Patel, K. (2020), 'Understanding and visualizing data iteration in machine learning', *Proceedings of the 2020 CHI conference on human factors in computing systems* pp. 1–13.

Holliday, D., Wilson, S. and Stumpf, S. (2016), 'User trust in intelligent systems: A journey over time', *Proceedings of the 21st International Conference on Intelligent User Interfaces* .
URL: *http://dx.doi.org/10.1145/2856767.2856811*

Hong, M. K., Fourney, A., DeBellis, D. and Amershi, S. (2021), 'Planning for natural language failures with the ai playbook', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–11.

Hope Currin, F., Diederich, K., Blasi, K., Dale Schmidt, A., David, H., Peterman, K. and Hourcade, J. P. (2021), 'Supporting shy preschool children in joining social play', pp. 396–407.

Hope, T., Tamari, R., Hershcovich, D., Kang, H. B., Chan, J., Kittur, A. and Shahaf, D. (2022), 'Scaling creative inspiration with fine-grained functional aspects of ideas', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–15.

Hourcade, J. P. (2015), 'Child-computer interaction', *Self, Iowa City, Iowa* .

Hu, K., Bakker, M. A., Li, S., Kraska, T. and Hidalgo, C. (2019), 'Vizml: A machine learning approach to visualization recommendation', *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* pp. 1–12.

Huan, R. (2017), Evaluating engagement in digital narratives from facial data, *in* 'Proceedings of the 19th ACM International Conference on Multimodal Interaction', pp. 652–655.

Huan, R. and Brewster, S. A. (2021), 'Designing an engaging story-stem taken from the mcast test', pp. 18–31.

Hughes, M. A. and Roy, D. (2021), 'Keeper: A synchronous online conversation environment informed by in-person facilitation practices', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

IBM (n.d.), 'What is gradient descent?'. Accessed: 1st July 2023.
URL: *https://www.ibm.com/topics/gradient-descent*

IDC (2023*a*), 'Idc conference - proceedings'. Accessed: 6th June 2023.
URL: *https://dl.acm.org/conference/idc/proceedings*

IDC (2023*b*), 'Idc – june 19-23, 2023'. Accessed: 5th June 2023.
URL: *https://idc.acm.org/2023/*

Im, H. and Rogers, C. (2021), ‘Draw2code: Low-cost tangible programming for creating ar animations’, pp. 427–432.

iMotions (2022), ‘Unpack human behavior’.
URL: *https://imotions.com/*

Inkpen, K., Chancellor, S., De Choudhury, M., Veale, M. and Baumer, E. P. (2019), ‘Where is the human? bridging the gap between ai and hci’, pp. 1–9.

ISO - International Organization for Standardization. ISO 9241-11:2018(en) Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts (2018). Accessed 0n 18.01.2023.
URL: *https://www.iso.org/obp/ui/*

Jain, A. K., Mao, J. and Mohiuddin, K. M. (1996), ‘Artificial neural networks: A tutorial’, *Computer* **29**, 31–44.

Jasim, M., Collins, C., Sarvghad, A. and Mahyar, N. (2022*a*), Supporting serendipitous discovery and balanced analysis of online product reviews with interaction-driven metrics and bias-mitigating suggestions, *in* ‘Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems’, CHI ’22, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3491102.3517649*

Jasim, M., Collins, C., Sarvghad, A. and Mahyar, N. (2022*b*), ‘Supporting serendipitous discovery and balanced analysis of online product reviews with interaction-driven metrics and bias-mitigating suggestions’, *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* .
URL: *https://doi.org/10.1145/3491102.3517649*

Jiang, H. and Nachum, O. (2020), Identifying and correcting label bias in machine learning, *in* ‘International Conference on Artificial Intelligence and Statistics’, PMLR, pp. 702–712.

Jiang, W., Sarsenbayeva, Z., van Berkel, N., Wang, C., Yu, D., Wei, J., Goncalves, J. and Kostakos, V. (2021), ‘User trust in assisted decision-making using miniaturized near-infrared spectroscopy’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–16.

Jiang, Y., Du, R., Lutteroth, C. and Stuerzlinger, W. (2019), ‘Orc layout: Adaptive gui layout with or-constraints’, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* pp. 1–12.

Jo, Y., Kim, M. and Han, K. (2019), ‘How do humans assess the credibility on web blogs: Qualifying and verifying human factors with machine learning’, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* pp. 1–12.

Jones, S. T., Perez, M., Lee, S. P., Furuichi, K. and Worsley, M. (2019), ‘Facilitation in an intergenerational making activity: How facilitative moves shift across traditional and digital fabrication’, pp. 237–245.

Jouppi, N. P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., Bates, S., Bhatia, S., Boden, N., Borchers, A. et al. (2017), In-datacenter performance analysis of a tensor processing unit, *in* ‘Proceedings of the 44th annual international symposium on computer architecture’, pp. 1–12.

Kanal, L. N. (2003), *Perceptron*, John Wiley and Sons Ltd., GBR, p. 1383–1385.

Kang, S., Norooz, L., Bonsignore, E., Byrne, V., Clegg, T. and Froehlich, J. E. (2019), ‘Prototypar: Prototyping and simulating complex systems with paper craft and augmented reality’, pp. 253–266.

Kang, Y., Sun, Z., Wang, S., Huang, Z., Wu, Z. and Ma, X. (2021), ‘Metamap: Supporting visual metaphor ideation through multi-dimensional example-based exploration’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–15.

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H. and Vaughan, J. W. (2020), ‘Interpreting interpretability: understanding data scientists’ use of interpretability tools for machine learning', *Proceedings of the 2020 CHI conference on human factors in computing systems* pp. 1–14.

Key, C., Browne, F., Taylor, N. and Rogers, J. (2021), 'Proceed with care: Reimagining home iot through a care perspective', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–15.

Khan, R. A., Crenn, A., Meyer, A. and Bouakaz, S. (2019), 'A novel database of children's spontaneous facial expressions (liris-cse)', *Image and Vision Computing* **83**, 61–69.

Khan, R. F. and Sutcliffe, A. (2014), 'Attractive agents are more persuasive', *International Journal of Human-Computer Interaction* **30**, 142–150.

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R. and Scassellati, B. (2013), 'Social robots as embedded reinforcers of social behavior in children with autism', *Journal of autism and developmental disorders* **43**, 1038–1049.

Kim, T. S., Choi, D., Choi, Y. and Kim, J. (2022*a*), Stylette: Styling the web with natural language, *in* 'Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems', CHI '22, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3491102.3501931*

Kim, T. S., Choi, D., Choi, Y. and Kim, J. (2022*b*), 'Stylette: Styling the web with natural language', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* .
URL: *https://doi.org/10.1145/3491102.3501931*

Kirabo, L., Carter, E. J., Barry, D. and Steinfeld, A. (2021), 'Priorities, technology, & power: Co-designing an inclusive transit agenda in kampala, uganda', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–11.

Koch, J., Lucero, A., Hegemann, L. and Oulasvirta, A. (2019), 'May ai? design ideation with cooperative contextual bandits', *Proceedings of the*

2019 CHI Conference on Human Factors in Computing Systems pp. 1–12.

Kocielnik, R., Amershi, S. and Bennett, P. N. (2019), 'Will you accept an imperfect ai? exploring designs for adjusting end-user expectations of ai systems', *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Koller, D., Friedman, N., Džeroski, S., Sutton, C., McCallum, A., Pfeffer, A., Abbeel, P., Wong, M.-F., Meek, C., Neville, J. et al. (2007), *Introduction to statistical relational learning*, MIT press.

Kontogiorgos, D., Waveren, S. V., Wallberg, O., Pereira, A., Leite, I. and Gustafson, J. (2020), 'Embodiment effects in interactions with failing robots', *Proceedings of the 2020 CHI conference on human factors in computing systems* pp. 1–14.

Kulke, L., Feyerabend, D. and Schacht, A. (2020), 'A comparison of the affectiva imotions facial expression analysis software with emg for identifying facial expressions of emotion', *Frontiers in Psychology* **11**, 329.

Kuzminykh, A., Sun, J., Govindaraju, N., Avery, J. and Lank, E. (2020), 'Genie in the bottle: Anthropomorphized perceptions of conversational agents', *Proceedings of the 2020 CHI conference on human factors in computing systems* pp. 1–13.

Kühl, N., Goutier, M., Hirt, R. and Satzger, G. (2020), 'Machine learning in artificial intelligence: Towards a common understanding', *arXiv preprint arXiv:2004.04686* .

Lai, V., Carton, S., Bhatnagar, R., Liao, Q. V., Zhang, Y. and Tan, C. (2022), 'Human-ai collaboration via conditional delegation: A case study of content moderation', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–18.

Lakshmi, U., Hofmann, M., Mack, K., Hudson, S. E., Mankoff, J. and Arriaga, R. I. (2021), 'Medical maker response to covid-19: Distributed manufacturing infrastructure for stopgap protective equipment', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Lambton-Howard, D., Simpson, E., Quimby, K., Kharrufa, A., Ng, H. H. M., Foster, E. and Olivier, P. (2021), 'Blending into everyday life: designing a social media-based peer support system', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Langer, M., Hunsicker, T., Feldkamp, T., König, C. J. and Grgić-Hlača, N. (2022), '"look! it'sa computer program! it's an algorithm! it's ai!": does terminology affect human perceptions and evaluations of algorithmic decision-making systems?', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–28.

Lapeyre, M. (2018), 'Poppy projectopen source platform for th creation, use and sharing of interactive 3d printed robots'. Accessed on 3rd July 2023. **URL:** *https://www.poppy-project.org/en/*

Lechelt, S., Rogers, Y. and Marquardt, N. (2020), 'Coming to your senses: promoting critical thinking about sensors through playful interaction in classrooms', pp. 11–22.

Lee-Cultura, S., Sharma, K., Cosentino, G., Papavlasopoulou, S. and Giannakos, M. (2021), 'Children's play and problem solving in motion-based educational games: Synergies between human annotations and multi-modal data', pp. 408–420.

Lee-Cultura, S., Sharma, K., Papavlasopoulou, S., Retalis, S. and Giannakos, M. (2020), 'Using sensing technologies to explain children's self-representation in motion-based educational games', pp. 541–555.

Lee, M. H., Siewiorek, D. P., Smailagic, A., Bernardino, A. and i Badia, S. (2021), 'A human-ai collaborative approach for clinical decision making on rehabilitation assessment', *Proceedings of the 2021 CHI conference on human factors in computing systems* pp. 1–14.

Lehnert, F. K., Niess, J., Lallemand, C., Markopoulos, P., Fischbach, A. and

Koenig, V. (2022), ‘Child–computer interaction: From a systematic review towards an integrated understanding of interaction design methods for children’, *International Journal of Child-Computer Interaction* **32**, 100398.

Lemmer, S. J., Song, J. Y. and Corso, J. J. (2021), ‘Crowdsourcing more effective initializations for single-target trackers through automatic re-querying’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Léon, P. R. (1962), ‘Laboratoire de langues et correction phonétique’.

Lewis, B. and Venkatasubramanian, K. (2021), ‘“i... got my nose-print. but it wasn’t accurate”: How people with upper extremity impairment authenticate on their personal computing devices’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Li, F. M., Chen, D. L., Fan, M. and Truong, K. N. (2021), ‘“i choose assistive devices that save my face” a study on perceptions of accessibility and assistive technology use conducted in china’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Li, Y., Balogh, S., Luttringer, A., Weber, T., Mayer, S. and Hussmann, H. (2022), ‘Designing a tangible interface to “force” children collaboration’, pp. 576–582.

Liang, C., Proft, J., Andersen, E. and Knepper, R. A. (2019), ‘Implicit communication of actionable information in human-ai teams’, *Proceedings of the 2019 CHI conference on human factors in computing systems* pp. 1–13.

Liao, M. and Sundar, S. S. (2021), ‘How should ai systems talk to users when collecting their personal information? effects of role framing and self-referencing on human-ai interaction’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Liao, M., Sundar, S. S. and Walther, J. B. (2022), ‘User trust in recommendation systems: A comparison of content-based, collaborative and demographic filtering', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Lima, G., Grgić-Hlača, N. and Cha, M. (2021), 'Human perceptions on moral responsibility of ai: A case study in ai-assisted bail decision-making', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–17.

Lin, G. C., Schoenfeld, I., Thompson, M., Xia, Y., Uz-Bilgin, C. and Leech, K. (2022), '" what color are the fish's scales?" exploring parents' and children's natural interactions with a child-friendly virtual agent during storybook reading', pp. 185–195.

Lin, P. and Brummelen, J. V. (2021), 'Engaging teachers to co-design integrated ai curriculum for k-12 classrooms', *Proceedings of the 2021 CHI conference on human factors in computing systems* pp. 1–12.

Lindley, J., Akmal, H. A., Pilling, F. and Coulton, P. (2020), 'Researching ai legibility through design', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J. and Bartlett, M. (2011), The computer expression recognition toolbox (cert), *in* '2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)', IEEE, pp. 298–305.

Lo, C. K. (2023), 'What is the impact of chatgpt on education? a rapid review of the literature', *Education Sciences* **13**(4), 410.

LoBue, V. and Thrasher, C. (2015), 'The child affective facial expression (cafe) set: Validity and reliability from untrained adults', *Frontiers in psychology* **5**, 1532.

Long, D. and Magerko, B. (2020), 'What is ai literacy? competencies and design considerations', *Proceedings of the 2020 CHI conference on human factors in computing systems* pp. 1–16.

Long, D., McKlin, T., Weisling, A., Martin, W., Blough, S., Voravong, K. and

Magerko, B. (2020), ‘Out of tune: discord and learning in a music programming museum exhibit’, pp. 75–86.

Louie, R., Gergle, D. and Zhang, H. (2022), ‘Affinder: Expressing concepts of situations that afford activities using context-detectors’, *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–18.

Lovato, S. B., Piper, A. M. and Wartella, E. A. (2019), ‘Hey google, do unicorns exist? conversational agents as a path to answers to children’s questions’, pp. 301–313.

Lu, Z. and Yin, M. (2021), ‘Human reliance on machine learning models when performance feedback is limited: Heuristics and risks’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–16.

Luger, E. and Sellen, A. (2016), ‘”like having a really bad pa” the gulf between user expectation and experience of conversational agents’, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI ’16* pp. 5286–5297.
URL: *http://dl.acm.org/citation.cfm?doid=2858036.2858288*

Lyons, H., Wijenayake, S., Miller, T. and Velloso, E. (2022), ‘What’s the appeal? perceptions of review processes for algorithmic decisions’, *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–15.

Mack, K., McDonnell, E., Jain, D., Wang, L. L., Froehlich, J. E. and Findlater, L. (2021), ‘What do we mean by “accessibility research”? a literature survey of accessibility papers in chi and assets from 1994 to 2019’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–18.

MacKenzie, I. S. (2012), ‘Human-computer interaction: An empirical research perspective’.

Madaio, M. A., Stark, L., Vaughan, J. W. and Wallach, H. (2020), ‘Co-designing checklists to understand organizational challenges and opportunities around fairness in ai’, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Mahmood, A., Fung, J. W., Won, I. and Huang, C.-M. (2022), ‘Owning mistakes sincerely: Strategies for mitigating ai errors’, *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–11.

Maldonado, H. and Zekelman, A. (2019), ‘Designing tangible abcs: Fröbel’s sticks and rings for the 21st century’, pp. 326–333.

Malinverni, L. and Valero, C. (2020), ‘What is a robot? an artistic approach to understand children’s imaginaries about robots’, pp. 250–261.

Mallari, K., Inkpen, K., Johns, P., Tan, S., Ramesh, D. and Kamar, E. (2020), ‘Do i look like a criminal? examining how race presentation impacts human judgement of recidivism’, *Proceedings of the 2020 Chi conference on human factors in computing systems* pp. 1–13.

Malsert, J., Palama, A. and Gentaz, E. (2020), ‘Emotional facial perception development in 7, 9 and 11 year-old children: The emergence of a silent eye-tracked emotional other-race effect’, *Plos one* **15**(5), e0233008.

Manaswi, N. K. and Manaswi, N. K. (2018), ‘Rnn and lstm’, *Deep Learning with Applications Using Python: Chatbots and Face, Object, and Speech Recognition With TensorFlow and Keras* pp. 115–126.

Mansi, G., Kim, S. R. and Roberts, J. (2022), ‘Ready, set, art: Technology needs and tools for remote k-2 art education’, pp. 150–184.

Marangunić, N. and Granić, A. (2015), ‘Technology acceptance model: a literature review from 1986 to 2013’, *Universal access in the information society* **14**, 81–95.

Marconi, A., Schiavo, G., Massa, P., Mencarini, E. and Deppieri, G. (2021), ‘From sustainable mobility to good deeds: Supporting school participation during covid-19 emergency through a playful education platform', pp. 80–86.

Martinez, A. M. (2019), 'The promises and perils of automated facial action coding in studying children's emotions.', *Developmental Psychology* **55**(9), 1965.

MathWorks (2023), 'Feature extraction'. Accessed on 24thSeptember 2023.
URL: *https://uk.mathworks.com/discovery/feature-extraction.html*

Mayer, S., Reinhardt, J., Schweigert, R., Jelke, B., Schwind, V., Wolf, K. and Henze, N. (2020), 'Improving humans' ability to interpret deictic gestures in virtual reality', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

McCarthy, J. (2011), 'What is ai'. Accessed: 27th June 2023.
URL: *http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html*

McCormack, J., Gifford, T., Hutchings, P., Rodriguez, M. T. L., Yee-King, M. and d'Inverno, M. (2019), 'In a silent way: Communication between ai and improvising musicians beyond sound', *Proceedings of the 2019 chi conference on human factors in computing systems* pp. 1–11.

McCulloch, W. S. and Pitts, W. (1943), 'A logical calculus of the ideas immanent in nervous activity', *The bulletin of mathematical biophysics* **5**, 115–133.

McDermott, T., Robson, J., Winters, N. and Malmberg, L.-E. (2022), 'Mapping the changing landscape of child-computer interaction research through correlated topic modelling', pp. 82–97.

McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J. and Kaliouby, R. e. (2016), Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit, *in* 'Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems', pp. 3723–3726.

McEwan, M., Phillips, C., Wyeth, P. and Johnson, D. (2020), 'Puppy island:

Theory-driven design of a serious game for young children with cystic fibrosis', pp. 532–540.

Mcknight, D. H. and Chervany, N. L. (2000), 'Trust and distrust definitions : One bite at a time', *Proceedings of the workshop on Deception, Fraud, and Trust in Agent Societies held during the Autonomous Agents Conference: Trust in Cyber-societies, Integrating the Human and Artificial Perspectives* pp. 27–54.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2021), 'A survey on bias and fairness in machine learning', *ACM Computing Surveys (CSUR)* **54**(6), 1–35.

Melsión, G. I., Torre, I., Vidal, E. and Leite, I. (2021), 'Using explainability to help children understandgender bias in ai', pp. 87–99.

Michaelis, J. E. and Mutlu, B. (2019), 'Supporting interest in science learning with a social robot', pp. 71–82.

Miesel, D. (2011), 'Deep blue'. Accessed: 27th June 2023.
URL: *https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/*

Minsky, M. and Papert, S. (1969), 'An introduction to computational geometry', *Cambridge tiass., HIT* **479**(480), 104.

Mirnig, A. G. and Meschtscherjakov, A. (2019), 'Trolled by the trolley problem: on what matters for ethical decision making in automated vehicles', *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* pp. 1–10.

Miyatake, M., Narumi, K., Sekiya, Y. and Kawahara, Y. (2021), 'Flower jelly printer: Slit injection printing for parametrically designed flower jelly', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–10.

Molina, M. D., Sundar, S. S., Rony, M. M. U., Hassan, N., Le, T. and Lee, D. (2021), 'Does clickbait actually attract more clicks? three clickbait studies you must read', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–19.

Montgomery, D. C., Peck, E. A. and Vining, G. G. (2021), *Introduction to linear regression analysis*, John Wiley & Sons.

Morales-Navarro, L., Thompson, N., Kafai, Y., Shaw, M. and Pinkard, N. (2022), 'Reimagining and co-designing with youth an hour of code activity for critical engagement with computing', pp. 288–296.

Motozawa, M., Murakami, Y., Pituxcoosuvarn, M., Takasaki, T. and Mori, Y. (2021), 'Conversation analysis for facilitation in children's intercultural collaboration', pp. 62–68.

Murphy, R. R. (2023), 'A retrospective of isaac asimov at 102 and his three laws', *Science Robotics* **8**(74), eadg3178.

Myers, B. A. (1988), 'Human-computer interaction technology', *Interactions* **5**, 44–54.

Nasar, A. A. (2016), 'The history of algorithmic complexity', *The Mathematics Enthusiast* **13**(3), 217–242.

National Cancer Institute (2023), 'Seer training modules, anatomy and physiology'. Accessed: 1st July 2023.
URL: *https://www.training.seer.cancer.gov/anatomy/nervous/tissue.html*

Negrão, J. G., Osorio, A. A. C., Siciliano, R. F., Lederman, V. R. G., Kozasa, E. H., D'Antino, M. E. F., Tamborim, A., Santos, V., de Leucas, D. L. B., Camargo, P. S. et al. (2021), 'The child emotion facial expression set: a database for emotion recognition in children', *Frontiers in psychology* **12**, 666245.

Newell, A., Shaw, J. C. and Simon, H. A. (1957), Empirical explorations of the logic theory machine: a case study in heuristic, *in* 'Papers presented at the February 26-28, 1957, western joint computer conference: Techniques for reliability', pp. 218–230.

Nguyen, H. (2022*a*), 'Examining teenagers' perceptions of conversational agents in learning settings', pp. 374–381.

Nguyen, H. (2022*b*), Examining teenagers' perceptions of conversational agents in learning settings, *in* 'Interaction Design and Children', IDC '22, Association for Computing Machinery, New York, NY, USA, p. 374–381.
URL: *https://doi.org/10.1145/3501712.3529740*

Noguchi, Y. and Tanaka, F. (2020), 'Omoy: a handheld robotic gadget that shifts its weight to express emotions and intentions', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Noriega, L. (2005), 'Multilayer perceptron tutorial', *School of Computing. Staffordshire University* **4**(5), 444.

Norman, D. (2013), *The design of everyday things: Revised and expanded edition*, Basic books.

Nouwen, M. and Duflos, M. (2021), 'Tiktok as a data gathering space: the case of grandchildren and grandparents during the covid-19 pandemic', pp. 498–502.

Okuya, Y., Gladin, O., Ladevèze, N., Fleury, C. and Bourdot, P. (2020), 'Investigating collaborative exploration of design alternatives on a wall-sized display', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–12.

Oyedotun, T. D. (2020*a*), 'Sudden change of pedagogy in education driven by covid-19: Perspectives and evaluation from a developing country', *Research in Globalization* **2**, 100029.

Oyedotun, T. D. (2020*b*), 'Sudden change of pedagogy in education driven by covid-19: Perspectives and evaluation from a developing country', *Research in Globalization* **2**, 100029.

Palacios Hidalgo, F. J., Huertas Abril, C. A. and Parra, M. G. (2020), 'Moocs: Origins, concept and didactic applications: A systematic review of the literature (2012–2019)', *Technology, Knowledge and Learning* **25**, 853–879.

Pan, L., Yu, C., Li, J., Huang, T., Bi, X. and Shi, Y. (2022), 'Automatically generating and improving voice command interface from operation sequences on smartphones', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–21.

Pang, Y., Reinecke, K. and Just, R. (2022), 'Apéritif: Scaffolding preregistrations to automatically generate analysis code and methods descriptions', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–15.

Panigutti, C., Beretta, A., Giannotti, F. and Pedreschi, D. (2022), 'Understanding the impact of explanations on advice-taking: a user study for ai-based clinical decision support systems', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–9.

Papert, S. and Harel, I. (1991), 'Situating constructionism', *constructionism* **36**(2), 1–11.

Park, H., Ahn, D., Hosanagar, K. and Lee, J. (2021), 'Human-ai interaction in human resource management: Understanding why employees resist algorithmic evaluation at workplaces and how to mitigate burdens', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–15.

Park, H., Ahn, D., Hosanagar, K. and Lee, J. (2022), 'Designing fair ai in human resource management: Understanding tensions surrounding algorithmic evaluation and envisioning stakeholder-centered solutions', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–22.

Parsonage, G., Horton, M. and Read, J. (2023*a*), The peer data labelling system (pdls). a participatory approach to classifying engagement in the classroom, *in* 'IFIP Conference on Human-Computer Interaction', Springer, pp. 224–233.

Parsonage, G., Horton, M. and Read, J. (2023*b*), Trust acceptance mapping-designing intelligent systems for use in an educational context, *in* 'International Conference on Human-Computer Interaction', Springer, pp. 34–50.

Parsonage, G., Horton, M. and Read, J. C. (2020), 'Designing experiments for children and robots', pp. 169–174.

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S. and Pioggia, G. (2016), 'Autism and social robotics: A systematic review', *Autism Research* **9**(2), 165–183.

Perusquia-Hernández, M., Ayabe-Kanamura, S., Suzuki, K. and Kumano, S. (2019), 'The invisible potential of facial electromyography: a comparison of emg and computer vision when distinguishing posed from spontaneous smiles', *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* pp. 1–9.

Petousi, D., Katifori, A., McKinney, S., Perry, S., Roussou, M. and Ioannidis, Y. (2021), 'Social bots of conviction as dialogue facilitators for history education: Promoting historical empathy in teens through dialogue', pp. 326–337.

Phelan, C., Hullman, J., Kay, M. and Resnick, P. (2019), 'Some prior (s) experience necessary: Templates for getting started with bayesian analysis', *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* pp. 1–12.

Pokhrel, S. and Chhetri, R. (2021*a*), 'A literature review on impact of covid-19 pandemic on teaching and learning', *Higher Education for the Future* **8**, 133–141.

Pokhrel, S. and Chhetri, R. (2021*b*), 'A literature review on impact of covid-19 pandemic on teaching and learning', *Higher Education for the Future* **8**(1), 133–141.

Pollak, S. D., Messner, M., Kistler, D. J. and Cohn, J. F. (2009), 'Development of perceptual expertise in emotion recognition', *Cognition* **110**(2), 242–247.

Potapov, K. and Marshall, P. (2020), 'Lifemosaic: Co-design of a personal informatics tool for youth', pp. 519–531.

Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Vaughan, J. W. W. and Wallach, H. (2021), 'Manipulating and measuring model interpretability', *Proceedings of the 2021 CHI conference on human factors in computing systems* pp. 1–52.

Prange, A., Barz, M., Heimann-Steinert, A. and Sonntag, D. (2021), 'Explainable automatic evaluation of the trail making test for dementia screening', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–9.

Prathish, S., S., A. N. and Bijlani, K. (2016), 'An intelligent system for online exam monitoring', *2016 International Conference on Information Science (ICIS)* pp. 138–143.

Preece, J., Sharp, H. and Rogers, Y. (2019), *Interaction design: beyond human-computer interaction*, 5 edn, Wiley and Sons.

Ramos, J. et al. (2003), 'Using tf-idf to determine word relevance in document queries', *Proceedings of the first instructional conference on machine learning* **242**, 29–48.

Read, J. C. (2023), 'Ethics chici child interaction'. Accessed: 27th June 2023.
URL: *https://chici.org/ethics/*

Read, J. C. and Bekker, M. M. (2011), 'The nature of child computer interaction', *BCS-HCI '11 Proceedings of the 25th BCS Conference on Human-Computer Interaction* pp. 163–170.

Read, J. C. and MacFarlane, S. (2006), 'Using the fun toolkit and other survey methods to gather opinions in child computer interaction', pp. 81–88.

Read, J. C., Sharma, S. and Parsonage, G. (2018), Childlike computing: Systems that think like humans and act like children, *in* 'Proceedings of the 4th International Conference on Human-Computer Interaction and User Experience in Indonesia, CHIuXiD'18', pp. 109–112.

Read, J., Fitton, D. and Horton, M. (2021), 'Capturing and considering idea development in school pupils' design activities', pp. 147–152.

Read, J., Sim, G., Horton, M. and Fitton, D. (2022), 'Reporting back in hci work with children', pp. 517–522.

Renom, M. A., Caramiaux, B. and Beaudouin-Lafon, M. (2022), 'Exploring technical reasoning in digital tool use', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–17.

Reschly, A. L. and Christenson, S. L. (2012), 'Moving from "context matters" to engaged partnerships with families', *Journal of Educational and Psychological Consultation* **22**(1-2), 62–78.

Richardson, B., Garcia-Gathright, J., Way, S. F., Thom, J. and Cramer, H. (2021), 'Towards fairness in practice: A practitioner-oriented rubric for evaluating fair ml toolkits', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Rodrigues, P., Mitic, M., Krammer, I., Dias, J., Prada, R. and Schrank, B. (2022), 'The hub: Designing an interactive social space for pre-adolescents' well-being', pp. 551–556.

Rosenberg, D. and Scott, K. (1999), *Use case driven object modeling with UML*, Springer.

Ross, A., Chen, N., Hang, E. Z., Glassman, E. L. and Doshi-Velez, F. (2021), 'Evaluating the interpretability of generative models by interactive reconstruction', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–15.

Ruan, S., He, J., Ying, R., Burkle, J., Hakim, D., Wang, A., Yin, Y., Zhou, L., Xu, Q., AbuHashem, A. et al. (2020), ‘Supporting children’s math learning with feedback-augmented narrative technology’, pp. 567–580.

Ruan, X., Palansuriya, C. and Constantin, A. (2022), ‘Real-time feedback based on emotion recognition for improving children’s metacognitive monitoring skill’, pp. 672–675.

Rubegni, E., Malinverni, L. and Yip, J. (2022*a*), “don’t let the robots walk our dogs, but it’s ok for them to do our homework”: Children’s perceptions, fears, and hopes in social robots., *in* ‘Interaction Design and Children’, IDC ’22, Association for Computing Machinery, New York, NY, USA, p. 352–361.
URL: *https://doi.org/10.1145/3501712.3529726*

Rubegni, E., Malinverni, L. and Yip, J. (2022*b*), ‘“don’t let the robots walk our dogs, but it’s ok for them to do our homework”: children’s perceptions, fears, and hopes in social robots.’, pp. 352–361.

Ruder, S. (2016), ‘An overview of gradient descent optimization algorithms’, *arXiv preprint arXiv:1609.04747* .

Russell, S. J. and Norvig, P. (2020), *Artificial Intelligence: a modern approach*, 4 edn, Pearson Education.

Samoili, S., Cobo, M. L., Gómez, E., Prato, G. D., Martínez-Plumed, F. and Delipetrev, B. (2020), ‘Ai watch. defining artificial intelligence. towards an operational definition and taxonomy of artificial intelligence’.

Samson, B. P. V. and Sumi, Y. (2019), ‘Exploring factors that influence connected drivers to (not) use or follow recommended optimal routes’, *Proceedings of the 2019 CHI conference on human factors in computing systems* pp. 1–14.

Sanoubari, E., Muñoz Cardona, J. E., Mahdi, H., Young, J. E., Houston, A. and Dautenhahn, K. (2021), ‘Robots, bullies and stories: a remote co-design study with children’, pp. 171–182.

Santos, K.-A., Ong, E. and Resurreccion, R. (2020), ‘Therapist vibe:

children's expressions of their emotions through storytelling with a chatbot', pp. 483–494.

Sanusi, I. T., Oyelere, S. S., Vartiainen, H., Suhonen, J. and Tukiainen, M. (2023), 'A systematic review of teaching and learning machine learning in k-12 education', *Education and Information Technologies* **28**(5), 5967–5997.

Saquib, N., Kazi, R. H., yi Wei, L., Mark, G. and Roy, D. (2021), 'Constructing embodied algebra by sketching', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–16.

Sasaki Otani, M. E. (2022), '"track-track: Let's follow the cat!" reflecting on children's biometric data processing through a micro puppet show', pp. 629–632.

Schapiro, T. (1999), 'What is a child?', *Ethics* **109**(4), 715–738.
URL: *https://doi.org/10.1086/233943*

Schloss, I., O. Bailey, J. and Tripathi, S. (2021), '"i'm in his belly!": Children's responses to different types of characters in virtual reality', pp. 43–48.

Schmidhuber, J. (2015), 'Deep learning in neural networks: An overview', *Neural networks* **61**, 85–117.

Schmidt, P., Biessmann, F. and Teubner, T. (2020), 'Transparency and trust in artificial intelligence systems', *Journal of Decision Systems* **29**, 260–278.
URL: *https://doi.org/10.1080/12460125.2020.1819094*

Schools, pupils and their characteristics, academic year 2021/22 (2022). Accessed 0n 18.01.2023.
URL: *https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics*

Schulman, J., Zoph, B., Kim, C., Hilton, J., Menick, J., Weng, J., Uribe, J. F. C., Fedus, L., Metz, L., Pokorny, M., Lopes, R. G., Zhao, S., Vijayvergiya, A., Sigler, E., Perelman, A., Voss, C., Heaton, M., Parish, J., Cummings, D., Nayak, R., Balcom, V., Schnurr, D., Kaftan, T., Hallacy, C., Turley, N., Deutsch, N., Goel, V., Ward, J., Konstantinidis, A., Zaremba, W., Ouyang, L., Bogdonoff, L., Gross, J., Medina, D., Yoo, S., Lee, T., Lowe, R., Mossing, D., Huizinga, J., Jiang, R., Wainwright, C., Almeida, D., Lin, S., Zhang, M., Xiao, K., Slama, K., Bills, S., Gray, A., Leike, J., Pachocki, J., Tillet, P., Jain, S., Brockman, G., Ryder, N., Paino, A., Yuan, Q., Winter, C., Wang, B., Bavarian, M., Babuschkin, I., Sidor, S., Kanitscheider, I., Pavlov, M., Plappert, M., Tezak, N., Jun, H., Zhuk, W., Pong, V., Kaiser, L., Tworek, J., Carr, A., Weng, L., Agarwal, S., Cobbe, K., Kosaraju, V., Power, A., Polu, S., Han, J., Puri, R., Jain, S., Chess, B., Gibson, C., Boiko, O., Parparita, E., Tootoonchian, A., Kosic, K. and Hesse, C. (2023), 'Introducing chatgpt'. Accessed: 5th June 2023. **URL:** *https://openai.com/blog/chatgpt*

Scullin, M. H. and Ceci, S. J. (2001), 'A suggestibility scale for children', *Personality and Individual Differences* **30**(5), 843–856.

Seraj, M., Große, C. S., Autexier, S. and Drechsler, R. (2019), 'Smart homes programming: Development and evaluation of an educational programming application for young learners', pp. 146–152.

Shaha, M. and Pawar, M. (2018), Transfer learning for image classification, *in* '2018 second international conference on electronics, communication and aerospace technology (ICECA)', IEEE, pp. 656–660.

Sharma, K., Papavlasopoulou, S. and Giannakos, M. (2019), 'Joint emotional state of children and perceived collaborative experience in coding activities', pp. 133–145.

Sharma, K., Papavlasopoulou, S., Lee-Cultura, S. and Giannakos, M. (2021), 'Information flow and children's emotions during collaborative coding: A causal analysis', pp. 350–362.

Sharma, P., Joshi, S., Gautam, S., Maharjan, S., Khanal, S. R., Reis, M. C., Barroso, J. and de Jesus Filipe, V. M. (2022), Student engagement detection using emotion analysis, eye tracking and head movement with machine learning, *in* 'International Conference on Technology and Innovation in Learning, Teaching and Education', Springer, pp. 52–68.

Sharma, S., Kallioniemi, P., Hakulinen, J., Keskinen, T. and Turunen, M. (2019), 'Exploring globally inclusive online collaboration for indian and finnish schoolchildren', pp. 153–160.

Shaw, I. S. (1998), 'What is an intelligent system?', *Fuzzy Control of Industrial Systems: Theory and Applications* pp. 1–3.

Shen, J., Yang, H., Li, J. and Cheng, Z. (2022), 'Assessing learning engagement based on facial expression recognition in mooc's scenario', *Multimedia Systems* **28**(2), 469–478.

Shi, W., Wang, X., Oh, Y. J., Zhang, J., Sahay, S. and Yu, Z. (2020), 'Effects of persuasive dialogues: testing bot identities and inquiry strategies', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Shin, D. (2021), 'The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable ai', *International Journal of Human-Computer Studies* **146**, 102551.

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D. and Summers, R. M. (2016), 'Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning', *IEEE transactions on medical imaging* **35**(5), 1285–1298.

Shin, J. Y. and Holtz, B. (2020), 'Identifying opportunities and challenges: how children use technologies for managing diabetes', pp. 495–507.

Shinohara, K., McQuaid, M. and Jacobo, N. (2021), 'The burden of survival: How doctoral students in computing bridge the chasm of inaccessibility', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Shneiderman, B. (2022), *Human-centered AI*, Oxford University Press.

Silva, C., Prandi, C., Nunes, N. J. and Nisi, V. (2020), ‘Blue whale street art as a landmark: extracting landmarks from children’s cognitive maps for the design of locative systems’, pp. 602–613.

Simko, L., Chin, B., Na, S., Saluja, H. K., Zhu, T. Q., Kohno, T., Hiniker, A., Yip, J. and Cobb, C. (2021), ‘Would you rather: A focus group method for eliciting and discussing formative design insights with children’, pp. 131–146.

Sin, J., Franz, R. L., Munteanu, C. and Neves, B. B. (2021), ‘Digital design marginalization: New perspectives on designing inclusive interfaces’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–11.

Singh, A., Thakur, N. and Sharma, A. (2016), ‘A review of supervised machine learning algorithms’, *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)* pp. 1310–1315.

Song, K. W. and Paulos, E. (2021), ‘Unmaking: Enabling and celebrating the creative material of failure, destruction, decay, and deformation’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–12.

Soni, N., Aloba, A., Morga, K. S., Wisniewski, P. J. and Anthony, L. (2019), ‘A framework of touchscreen interaction design recommendations for children (tidrc) characterizing the gap between research evidence and design practice’, pp. 419–431.

Spitale, M., Silleresi, S., Cosentino, G., Panzeri, F. and Garzotto, F. (2020), ‘” whom would you like to talk with?” exploring conversational agents for children’s linguistic assessment’, pp. 262–272.

Stefanidi, E., Korozi, M., Leonidis, A., Arampatzis, D., Antona, M. and Papagiannakis, G. (2021), ‘When children program intelligent environments: Lessons learned from a serious ar game’, pp. 375–386.

Stöckli, S., Schulte-Mecklenbeck, M., Borer, S. and Samson, A. C. (2018), 'Facial expression analysis with affdex and facet: A validation study', *Behavior research methods* **50**(4), 1446–1460.

Stower, R. and Kappas, A. (2021), 'Cozmonaots: Designing an autonomous learning task with social and educational robots', pp. 542–546.

Sun, D., Feng, Z., Chen, Y., Wang, Y., Zeng, J., Yuan, M., Pong, T.-C. and Qu, H. (2020), 'Dfseer: A visual analytics approach to facilitate model selection for demand forecasting', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Sun, J., Li, Y., Chen, C., Lee, J., Liu, X., Zhang, Z., Huang, L., Shi, L. and Xu, W. (2020), 'Fdhelper: assist unsupervised fraud detection experts with interactive feature selection and evaluation', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–12.

Suresh, H., Gomez, S. R., Nam, K. K. and Satyanarayan, A. (2021), 'Beyond expertise and roles: A framework to characterize the stakeholders of interpretable machine learning and their needs', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–16.

Sutherland, I. E. (1964), 'Sketchpad a man-machine graphical communication system*', *Proceedings of the SHARE design automation workshop* pp. 6–329.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016), Rethinking the inception architecture for computer vision, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 2818–2826.

Tandon, U. (2021*a*), 'Factors influencing adoption of online teaching by school teachers: A study during covid-19 pandemic', *Journal of Public Affairs* **21**, e2503.

Tandon, U. (2021*b*), ‘Factors influencing adoption of online teaching by school teachers: A study during covid-19 pandemic’, *Journal of Public Affairs* **21**(4), e2503.

Theurel, A., Witt, A., Malsert, J., Lejeune, F., Fiorentini, C., Barisnikov, K. and Gentaz, E. (2016), ‘The integration of visual context information in facial emotion recognition in 5-to 15-year-olds’, *Journal of experimental child psychology* **150**, 252–271.

Thomas, X., Powell, L., Polsley, S., Ray, S. and Hammond, T. (2022), ‘Identifying features that characterize children’s free-hand sketches using machine learning’, pp. 529–535.

Thomson Reuter (2023), ‘Practical law’. Accessed: 26th June 2023.
URL: *https://uk.practicallaw.thomsonreuters.com/7-501-0825?*

Tigwell, G. W. (2021), ‘Nuanced perspectives toward disability simulations from digital designers, blind, low vision, and color blind people’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–15.

Tisza, G., Sharma, K., Papavlasopoulou, S., Markopoulos, P. and Giannakos, M. (2022), ‘Understanding fun in learning to code: A multi-modal data approach’, pp. 274–287.

Tolksdorf, N. F., Viertel, F. E., Crawshaw, C. E. and Rohlfing, K. J. (2021), ‘Do shy children keep more distance from a social robot? exploring shy children’s proxemics with a social robot or a human’, pp. 527–531.

Troiano, G. M., Wood, M. and Harteveld, C. (2020), ‘” and this, kids, is how i met your mother”: Consumerist, mundane, and uncanny futures with sex robots’, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–17.

Trowler, V. (2010), ‘Student engagement literature review’, *The higher education academy* **11**(1), 1–15.

Tsai, C.-H., You, Y., Gui, X., Kou, Y. and Carroll, J. M. (2021), ‘Exploring and promoting diagnostic transparency and explainability in online symptom checkers', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–17.

Tseng, T., Murai, Y., Freed, N., Gelosi, D., Ta, T. D. and Kawahara, Y. (2021), 'Plushpal: Storytelling with interactive plush toys and machine learning', pp. 236–245.

Tucker, H. F. (2017), *Growing Up: Childhood*, 1 edn, Blackwell,.

Tung, F.-W. (2016), 'Child perception of humanoid robot appearance and behavior', *International Journal of Human-Computer Interaction* **32**, 493–502.
URL: *http://www.tandfonline.com/doi/full/10.1080/10447318.2016.1172808*

Turing, A. M. (1950), 'Mind', *Mind* **59**(236), 433–460.

Tyler, R. W. (2013), Basic principles of curriculum and instruction, *in* 'Curriculum Studies Reader E2', Routledge, pp. 60–68.

Türkay, S. and Adinolf, S. (2019), 'Friending to flame: How social features affect player behaviours in an online collectible card game', *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* pp. 1–12.

Uzor, S., Jacques, J. T., Dudley, J. J. and Kristensson, P. O. (2021), 'Investigating the accessibility of crowdwork tasks on mechanical turk', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Vacca, R., DesPortes, K., Tes, M., Silander, M., Amato, A., Matuk, C. and Woods, P. J. (2022), 'What do you meme? students communicating their experiences, intuitions, and biases surrounding data through memes', pp. 212–224.

Valance, C. (2023), 'Artificial intelligence could lead to extinction, experts

warn - bbc news'. Accessed: 5th June 2023.
URL: *https://www.bbc.co.uk/news/uk-65746524*

Vallance, C. (2023), 'Powerful artificial intelligence ban possible, government adviser warns - bbc news'. Accessed: 5th June 2023.
URL: *https://www.bbc.co.uk/news/technology-65779181*

Van Brummelen, J., Tabunshchyk, V. and Heng, T. (2021), '"alexa, can i program you?": Student perceptions of conversational artificial intelligence before and after programming alexa', pp. 305–313.

van Ewijk, G., Smakman, M. and Konijn, E. A. (2020), 'Teachers' perspectives on social robots in education: an exploratory case study', pp. 273–280.

Van Mechelen, M., Baykal, G. E., Dindler, C., Eriksson, E. and Iversen, O. S. (2020), '18 years of ethics in child-computer interaction research: A systematic literature review', pp. 161–183.

Vanderborght, B., Simut, R., Saldien, J., Pop, C., Rusu, A. S., Pintea, S., Lefeber, D. and David, D. O. (2012), 'Using the social robot probo as a social story telling agent for children with asd', *Interaction Studies* **13**(3), 348–372.

Venkatasubramanian, K., Skorinko, J. L. M., Kobeissi, M., Lewis, B., Jutras, N., Bosma, P., Mullaly, J., Kelly, B., Lloyd, D., Freark, M. et al. (2021), 'Exploring a reporting tool to empower individuals with intellectual and developmental disabilities to self-report abuse', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Venkatesh, V. (2022), 'Adoption and use of ai tools: a research agenda grounded in utaut', *Annals of Operations Research* pp. 1–12.

Voinov, A. and Bousquet, F. (2010), 'Modelling with stakeholders', *Environmental modelling & software* **25**(11), 1268–1281.

Voinov, A., Kolagani, N., McCall, M. K., Glynn, P. D., Kragt, M. E., Ostermann, F. O., Pierce, S. A. and Ramu, P. (2016), 'Modelling with stakeholders–next generation', *Environmental Modelling & Software* **77**, 196–220.

Voulgari, I., Zammit, M., Stouraitis, E., Liapis, A. and Yannakakis, G. (2021), 'Learn to machine learn: designing a game based approach for teaching machine learning to primary and secondary education students', pp. 593–598.

Vygotsky, L. S. and Cole, M. (1978), *Mind in society: Development of higher psychological processes*, Harvard university press.

Völkel, S. T., Haeuslschmid, R., Werner, A., Hussmann, H. and Butz, A. (2020), 'How to trick ai: Users' strategies for protecting themselves from automatic personality assessment', *Proceedings of the 2020 CHI conference on human factors in computing systems* pp. 1–15.

Wan, X., Zhou, X., Ye, Z., Mortensen, C. K. and Bai, Z. (2020), 'Smileycluster: supporting accessible machine learning in k-12 scientific discovery', pp. 23–35.

Wang, D., Andres, J., Weisz, J. D., Oduor, E. and Dugan, C. (2021), 'Autods: Towards human-centered automation of data science', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–12.

Wang, D., Prabhat, S. and Sambasivan, N. (2022), 'Whose ai dream? in search of the aspiration in data annotation.', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–16.

Wang, D., Yang, Q., Abdul, A. and Lim, B. Y. (2019), 'Designing theory-driven user-centric explainable ai', *Proceedings of the 2019 CHI conference on human factors in computing systems* pp. 1–15.

Wang, Q., Ming, Y., Jin, Z., Shen, Q., Liu, D., Smith, M. J., Veeramachaneni, K. and Qu, H. (2019), 'Atmseer: Increasing transparency and controllability in automated machine learning', *Proceedings of the 2019 CHI conference on human factors in computing systems* pp. 1–12.

Wang, R., Harper, F. M. and Zhu, H. (2020), ‘Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences’, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Waytz, A., Heafner, J. and Epley, N. (2014), ‘The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle’, *Journal of Experimental Social Psychology* **52**, 113–117.

Webb, R., Ayers, S. and Endress, A. (2018), ‘The city infant faces database: A validated set of infant facial expressions’, *Behavior research methods* **50**(1), 151–159.

webrtc.org (2023), ‘Webrtc’.
URL: *https://webrtc.org/*

Wentzel, J., d’Eon, G. and Vogel, D. (2020), ‘Improving virtual reality ergonomics through reach-bounded non-linear input amplification’, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–12.

White, N. T., Cagiltay, B., Michaelis, J. E. and Mutlu, B. (2021), ‘Designing emotionally expressive social commentary to facilitate child-robot interaction’, pp. 314–325.

Widen, S. C. (2013), ‘Children’s interpretation of facial expressions: The long path from valence-based to specific discrete categories’, *Emotion Review* **5**(1), 72–77.

Williams, R., Park, H. W. and Breazeal, C. (2019), ‘A is for artificial intelligence: the impact of artificial intelligence activities on young children’s perceptions of robots’, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* pp. 1–11.

Windl, M., Feger, S. S., Zijlstra, L., Schmidt, A. and Wozniak, P. W. (2022), “it is not always discovery time’: Four pragmatic approaches in designing ai systems', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–12.

Wobbrock, J. O. and Kientz, J. A. (2016), 'Research contributions in human-computer interaction', *interactions* **23**(3), 38–44.

Worsley, M. (2022), 'Pe++: Exploring opportunities for connecting computer science and physical education in elementary school', pp. 590–595.

Xin, D., Wu, E. Y., Lee, D. J.-L., Salehi, N. and Parameswaran, A. (2021), 'Whither automl? understanding the role of automation in machine learning workflows', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–16.

Xu, Y. and Warschauer, M. (2020*a*), 'A content analysis of voice-based apps on the market for early literacy development', pp. 361–371.

Xu, Y. and Warschauer, M. (2020*b*), 'Exploring young children's engagement in joint reading with a conversational agent', pp. 216–228.

Xu, Y. and Warschauer, M. (2020*c*), 'What are you talking to?: Understanding children's perceptions of conversational agents', *Proceedings of the 2020 CHI conference on human factors in computing systems* pp. 1–13.

Yadollahi, E., Couto, M., Dillenbourg, P. and Paiva, A. (2022), 'Do children adapt their perspective to a robot when they fail to complete a task?', pp. 341–351.

Yan, Y., Yu, C., Zheng, W., Tang, R., Xu, X. and Shi, Y. (2020), 'Frownonerror: Interrupting responses from smart speakers by facial expressions', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* pp. 1–14.

Yan, Z., Wu, Y., Zhang, Y. and Chen, X. (2022), 'Emoglass: an end-to-end ai-enabled wearable platform for enhancing self-awareness of emotional health', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–19.

Yang, Q., Steinfeld, A., Rosé, C. and Zimmerman, J. (2020), 'Re-examining whether, why, and how human-ai interaction is uniquely difficult to design', *Proceedings of the 2020 chi conference on human factors in computing systems* pp. 1–13.

Yarosh, S., Radu, I., Hunter, S. and Rosenbaum, E. (2011), Examining values: an analysis of nine years of idc research, *in* 'Proceedings of the 10th International Conference on Interaction Design and Children', pp. 136–144.

Yeh, S.-F., Wu, M.-H., Chen, T.-Y., Lin, Y.-C., Chang, X., Chiang, Y.-H. and Chang, Y.-J. (2022), 'How to guide task-oriented chatbot users, and when: A mixed-methods study of combinations of chatbot guidance types and timings', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–16.

Yim, J. B., Ko, D. and Lee, W. (2021), 'Romi: Screen peripheral companion for mediating children's healthy screen use', pp. 456–460.

Yin, M., Vaughan, J. W. and Wallach, H. (2019), 'Understanding the effect of accuracy on trust in machine learning models', *Proceedings of the 2019 chi conference on human factors in computing systems* pp. 1–12.

Yoo, M., Odom, W. and Berger, A. (2021), 'Understanding everyday experiences of reminiscence for people with blindness: Practices, tensions and probing new design possibilities', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–15.

Yu, J. and Roque, R. (2022), 'Young children's perceptions of coding and implications', pp. 448–451.

Yu, Y., Si, X., Hu, C. and Zhang, J. (2019), 'A review of recurrent neural networks: Lstm cells and network architectures', *Neural computation* **31**(7), 1235–1270.

Yun, W.-H., Lee, D., Park, C., Kim, J. and Kim, J. (2018), 'Automatic recognition of children engagement from facial video using convolutional neural networks', *IEEE Transactions on Affective Computing* **11**(4), 696–707.

Zaheer, N., Ahmad, O. U., Ahmed, A., Khan, M. S. and Shabbir, M. (2021), 'Semour: A scripted emotional speech repository for urdu', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–12.

Zarei, N., Quek, F., Chu, S. L. and Brown, S. A. (2021), 'Towards designing enactment-scaffolded narrative authoring tools for elementary-school children', pp. 387–395.

Zarei, N., Quek, F., Chu, S. L., Chan, A. and Howell, J. (2022), 'Designing interactive contextual cues for children's video-stimulated writing', pp. 196–205.

Zhang, C., Zhou, Z., Wu, J., Hu, Y., Shao, Y., Liu, J., Hu, Y., Ying, F. and Yao, C. (2021), 'Bio sketchbook: an ai-assisted sketching partner for children's biodiversity observational learning', pp. 466–470.

Zhang, E. and Banovic, N. (2021), 'Method for exploring generative adversarial networks (gans) via automatically generated image galleries', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–15.

Zhang, L., Liu, L., Ying, W., Huang, M., Yao, C. and Ying, F. (2021), 'Modhera: A modular kit for parents to take care babies', pp. 547–551.

Zhao, J., Duron, B. and Wang, G. (2022), 'Koala hero: Inform children of privacy risks of mobile apps', pp. 523–528.

Zhao, Y. and Watterston, J. (2021*a*), 'The changes we need: Education post covid-19', *Journal of Educational Change* **22**, 3–12.

Zhao, Y. and Watterston, J. (2021*b*), 'The changes we need: Education post covid-19', *Journal of Educational Change* **22**(1), 3–12.

Zheng, C., Wang, D., Wang, A. Y. and Ma, X. (2022), 'Telling stories from computational notebooks: Ai-assisted presentation slides creation for presenting data science work', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–20.

Zheng, Q., Tang, Y., Liu, Y., Liu, W. and Huang, Y. (2022), 'Ux research on conversational human-ai interaction: A literature review of the acm digital library', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–24.

Zhou, X., Li, K., Munawar, A. M. and Bai, Z. (2021), 'Scaffolding design to bridge the gaps between machine learning and scientific discovery for k-12 stem education', pp. 604–609.

Zhu, J., Villareale, J., Javvaji, N., Risi, S., Löwe, M., Weigelt, R. and Harteveld, C. (2021), 'Player-ai interaction: What neural network games reveal about ai as play', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–17.

Zhu, X., Zhang, X., Chen, Z., Dong, Z., Gu, Z. and Chang, D. (2022), 'The trusted listener: The influence of anthropomorphic eye design of social robots on user's perception of trustworthiness', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–13.

Zimmermann-Niefield, A., Polson, S., Moreno, C. and Shapiro, R. B. (2020), 'Youth making machine learning models for gesture-controlled interactive media', pp. 63–74.

Zimmermann-Niefield, A., Turner, M., Murphy, B., Kane, S. K. and Shapiro, R. B. (2019), 'Youth learning machine learning through building models of athletic moves', pp. 121–132.

Złotowski, J., Proudfoot, D., Yogeeswaran, K. and Bartneck, C. (2015), 'Anthropomorphism: opportunities and challenges in human–robot interaction', *International journal of social robotics* **7**, 347–360.

Appendix A

Surveys

Child Survey Questions

Which digital technologies do you use? (Please check all that apply)	☐ PC or laptop ☐ Smart Phone ☐ Tablet ☐ Games Console ☐ Internet Access ☐ Other (Please indicate below)

The questions below ask how you would feel if a parent/carer or other adult took an action related to your use of a digital technology. For example, they may stop you using your smart phone if they were concerned for your safety or suggest you look at a webpage to help you with your homework.

For each question mark on the scale how you would feel about accepting the adults action if the reason for it was explained to you. For example, question 1 is asking:

> How accepting would you be if an adult took some action which effected your use of a digital technology because they were concerned about your safety?

		Not accepting --------------- Very accepting
1	Concern for your safety.	0----1----2----3----4----5----6----7----8----9----10
2	Concern for your security.	0----1----2----3----4----5----6----7----8----9----10
3	Curiosity about what you were doing.	0----1----2----3----4----5----6----7----8----9----10
4	To control what you were doing.	0----1----2----3----4----5----6----7----8----9----10
5	To help you complete a task.	0----1----2----3----4----5----6----7----8----9----10
6	To check that the content you were viewing was appropriate.	0----1----2----3----4----5----6----7----8----9----10
7	To make sure you were enjoying yourself.	0----1----2----3----4----5----6----7----8----9----10
8	To help you get the most benefit from your time spent.	0----1----2----3----4----5----6----7----8----9----10
9	To help you learn.	0----1----2----3----4----5----6----7----8----9----10
10	To stop you buying things such as in game purchases.	0----1----2----3----4----5----6----7----8----9----10
11	Describe an occasion when an adult took an action related to your use of a digital technology and how you felt about it.	

Figure A.1: Child Survey Page 1

If a technology existed that could monitor your actions and take some action. How accepting would you be of the technology if the actions it took were for the reasons given. For example, question 12 is asking,

> How accepting would you be if the technology took some action which effected your use of a digital technology because it was trying to keep you safe?

		Not accepting --------------- Very accepting
12	Concern for your safety.	0----1----2----3----4----5----6----7----8----9----10
13	Concern for your security.	0----1----2----3----4----5----6----7----8----9----10
14	Curiosity about what you were doing.	0----1----2----3----4----5----6----7----8----9----10
15	To control what you were doing.	0----1----2----3----4----5----6----7----8----9----10
16	To help you complete a task.	0----1----2----3----4----5----6----7----8----9----10
17	To check that the content you were viewing was appropriate.	0----1----2----3----4----5----6----7----8----9----10
18	To make sure you were enjoying yourself.	0----1----2----3----4----5----6----7----8----9----10
19	To help you get the most benefit from your time spent.	0----1----2----3----4----5----6----7----8----9----10
20	To help you learn.	0----1----2----3----4----5----6----7----8----9----10
21	To stop you buying things such as in game purchases.	0----1----2----3----4----5----6----7----8----9----10

About you

What is your age?	☐ 10 years ☐ 11 years ☐ 12 years ☐ 13 years ☐ 14 years ☐ 15 years ☐ 16 years ☐ Other (Please indicate below)
What is your gender?	☐ Male ☐ Female

Figure A.2: Child Survey Page 2

Adult Survey Questions

Please read the following questions and for each indicate the extent to which you would take an action (intervene) in your child's use of digital technologies. Examples of an intervention could include:

- Preventing your child from making a financial in game transaction
- Helping your child find online educational materials if they were stuck with their homework.

Digital technologies describe computers, games consoles and mobile devices such as smart phones and tablets.

If you have more than one child, please focus on a single child and indicate their age in the first box when you are asked to indicate the age and gender of your children at the end of this survey.

		Not at all -------------------------- Very much
1	How much does your child use digital technologies?	0----1----2----3----4----5----6----7----8----9----10
2	How often do you intervene in your child's use of digital technologies such as computers or mobile devices?	0----1----2----3----4----5----6----7----8----9----10
3	How much would concern for your child's safety or wellbeing influence whether you would intervene in their use of a digital technology?	0----1----2----3----4----5----6----7----8----9----10

Questions 4 – 12 list reasons you may choose to take action (intervene) in your child's use of a digital technology. For each action please indicate how highly motivated you would be to intervene on the scale. For example, question 4 is asking:

> How much would concern for your child's security influence whether you would intervene in their use of a digital technology?

		Not at all -------------------------- Very much
4	Your child's security.	0----1----2----3----4----5----6----7----8----9----10
5	Your curiosity as to what your child is doing.	0----1----2----3----4----5----6----7----8----9----10
6	Your desire to control your child's digital activity.	0----1----2----3----4----5----6----7----8----9----10
7	Your desire to help your child.	0----1----2----3----4----5----6----7----8----9----10
8	Your desire to ensure that any materials viewed are appropriate.	0----1----2----3----4----5----6----7----8----9----10
9	Your desire to ensure that your child is enjoying themselves.	0----1----2----3----4----5----6----7----8----9----10
10	Your desire to ensure that your child is making effective use of their time.	0----1----2----3----4----5----6----7----8----9----10
11	Your desire to ensure that your child is learning effectively.	0----1----2----3----4----5----6----7----8----9----10

Figure A.3: Adult Survey Page 1

12	Your desire to ensure that appropriate financial safeguards are in place.	0----1----2----3----4----5----6----7----8----9----10
13	Please describe any other factors which influence your decision to monitor and intervene in your child's use of digital technologies.	
14	If applicable please describe any occasion when you have intervened in your child's use of a digital technology and the impact of that action.	

If a technology existed that could monitor your child's actions and make appropriate interventions on your behalf when your child was using a digital technology. To what extent would you **trust the technology**? For each of the questions below indicate the level of trust you would have in each of the circumstances described. For example, question 17 is asking:

To what extent would you trust the technology to monitor your child's safety?

Question 18 is asking:

To what extent would you trust the technology to take appropriate action when monitoring your child's safety?

		Not at all --------------------------- Very much
15	To what extent would you trust the technology to monitor your child's everyday use of a digital technology?	0----1----2----3----4----5----6----7----8----9----10
16	To what extent would you trust the technology to take appropriate action when monitoring your child's everyday use of a digital technology?	0----1----2----3----4----5----6----7----8----9----10
17	Monitor your child's safety.	0----1----2----3----4----5----6----7----8----9----10
18	Take appropriate action when monitoring your child's safety.	0----1----2----3----4----5----6----7----8----9----10
19	Monitor your child's security.	0----1----2----3----4----5----6----7----8----9----10
20	Take appropriate action when monitoring your child's security.	0----1----2----3----4----5----6----7----8----9----10
21	Monitor the appropriateness of the content your child is viewing.	0----1----2----3----4----5----6----7----8----9----10
22	take appropriate action if it categorises content as inappropriate.	0----1----2----3----4----5----6----7----8----9----10

Figure A.4: Adult Survey Page 2

23	Monitor the enjoyment your child experiences.	0----1----2----3----4----5----6----7----8----9----10
24	Take appropriate action to improve your child's enjoyment of a digital technology.	0----1----2----3----4----5----6----7----8----9----10
25	Monitor your child's productivity.	0----1----2----3----4----5----6----7----8----9----10
26	Take appropriate action to improve your child's productivity.	0----1----2----3----4----5----6----7----8----9----10
27	Monitor how effectively your child is learning.	0----1----2----3----4----5----6----7----8----9----10
28	Take appropriate action to improve your child's learning.	0----1----2----3----4----5----6----7----8----9----10
29	Monitor financial transactions your child is making.	0----1----2----3----4----5----6----7----8----9----10
30	Take appropriate action if your child was making a financial transaction.	0----1----2----3----4----5----6----7----8----9----10

About you and your children

What is your age?	☐ 18-24 years old ☐ 25-34 years old ☐ 35-44 years old ☐ 45-54 years old ☐ 55-64 years old ☐ 65-74 years old ☐ 75 years or older ☐ Prefer not to say
What is your gender?	☐ Male ☐ Female ☐ Prefer not to say
What are the ages and gender of your children?	

Age	Gender
	☐ Male ☐ Female ☐ Prefer not to say
	☐ Male ☐ Female Prefer not to say
	☐ Male ☐ Female ☐ Prefer not to say
	☐ Male ☐ Female ☐ Prefer not to say
	☐ Male ☐ Female ☐ Prefer not to say
	☐ Male ☐ Female ☐ Prefer not to say

Figure A.5: Adult Survey Page 3

What digital technologies do your children have access to? (Please check all that apply)	☐ PC or laptop ☐ Smart Phone ☐ Tablet ☐ Games Console ☐ Internet Access ☐ Other (Please indicate below)

If you wish to withdraw

After submitting your survey, you will be given a unique reference number. If you decide at any point that you don't want your data to be used as part of the research project, please email Graham Parsonage gparsonage1@uclan.ac.uk quoting your reference number. As we don't collect any data that identifies you personally, we can only delete your survey if you provide the reference number so please make a careful note of it.

Further details are available on the project webpage at www.chici.org/apps-with-attitude

Figure A.6: Adult Survey Page 4

Appendix B

Robot Qualitative Data

B.1 Use Cases

B.1.1 Safety

You are carrying out some action that could potentially cause you harm eg Giving your personal details to someone you do not know.

Motivation	Safety	
Actor	Acceptance Numbers	Notes
Computer	6	Computer can stop
Watching Robot	2	The robot knows more than the computer
Parent	9	
Friend	7	

B.1.2 Security

You are carrying out an action that could potentially result in damage to you or the system you are using. Eg Installing unsafe software or disabling security features on the software.

Motivation	Security	
Actor	Acceptance Numbers	Notes
Computer	10	Computer knows how to protect itself
Watching Robot	1	
Parent	8	
Friend	4	Friends may be false

B.1.3 Curiosity

The technology or person is curious about what you are doing. Eg You take an action in a game and the technology or person wants to know why

Motivation	Curiosity (Group 1)	
Actor	Acceptance Numbers	Notes
Computer	0	Computers can't be curious they just predict.
Watching Robot	5	Robots can learn
Parent	10	
Friend	10	
Motivation	Curiosity (Group 2)	
Actor	Acceptance Numbers	Notes
Computer	2	Terrifying idea!
Watching Robot	2	Robots can learn
Parent	4	General theme that the group were concerned about privacy and motivation
Friend	10	

B.1.4 Helpfulness

The person or technology wants to help you complete a task.

Eg The person or technology makes suggestions about actions you could take in order to complete a game or finish some work.

Motivation	Helpfulness	
Actor	Acceptance Numbers	Notes
Computer	9	
Watching Robot	9	
Parent	7	
Friend	10	

B.1.5 Appropriateness

The person or technology believes materials you are accessing are inappropriate.

Eg You are accessing material which has an 18+ age restriction.

Motivation	Appropriateness	
Actor	Acceptance Numbers	Notes
Computer	9	
Watching Robot	1	
Parent	6	Depends on context. Parents may be judgmental
Friend	7	

B.1.6 Enjoyment

The person or technology wants to make you happy

Eg The person or technology suggests an action that they think you will enjoy.

Motivation	Enjoyment	
Actor	Acceptance Numbers	Notes

Computer	6	
Watching Robot	5	
Parent	10	My parents want me to have fun
Friend	8	

Appendix C

Designing and Conducting Experiments with Children

Hi I’m Poppy the newest member of the team

I need to learn to do things. Can you help?

Please write down 3 things you think I can learn to do.

On the other side of the paper draw a picture of me doing one of the things you wrote down.

Draw a picture of me doing one of the things you thought of.

Meet Poppy the Robot

Poppy the robot needs programming so it can do things.

In the box below write down 3 thing you think Poppy can be programmed to do.

On the other side of the paper draw a picture of Poppy doing one of the things you wrote down.

Draw a picture of Poppy doing one of the things you thought of.

Appendix D

Adult Free Text Survey Responses

Table D.1: Adult Free Text Responses

User ID	Question 13	Question 14
	Please describe any other factors which influence your decision to monitor and intervene in your child's use of digital technologies.	If applicable please describe any occasion when you have intervened in your child's use of a digital technology and the impact of that action
211		
212	The age and maturity of my children would greatly influence my level of monitoring and intervention	

User ID	Question 13	Question 14
213	None, safety and security are my main influences on intervening. I suppose it can sometimes be a punishment restricting use.	I have made him block he friends when they have been nasty to him when playing online games together I have talked a lot to Alex about security when playing but never needed to take any action on this. I have restricted his use due to him getting stressed/upset when playing - and also as a punishment.
214	Social networking and campaigns that question the safety of the application Over tiredness linked with disabilities, frustration and concentration.	Frustration has caused them to become angry at the device. This was removed for a time and correct behaviour was encouraged Playing a particular game and not want to come off so I limited the evenings he could play. He was upset at first until it was recognised as the norm.

User ID	Question 13	Question 14
215		
216	Parental controls also we have full access to what they search look at and who they talk to	none
217		I have intervened on: - viewing inappropriate content - occasional upset on social media - in-app purchases - getting hold of my passwords and using without permission - too much time spent on-line
218		
219		

User ID	Question 13	Question 14
220	When he gets stuck (can't make progress in a game, for example), he whines a lot...which would cause me to intervene He is still very young...so I often help him get to the place he wants to go.	helped him complete part of a game. For example, in lego games...the age is 7 but my son started playing these when he was 4...and they have instructions in writing on the screen, and I would need to be available to help him understand.
221	Detection of Grooming activities. Suitable online content.	Due to online bullying.
222	Socialising Studying Online Videos Mini Transaction games Home Launcher or Restricted Apps are help full but not the best.	Watching inappropriate videos online. Downloaded games with permission. Some games have inappropriate adds for age range. Some ratings for games don`t seem right.

User ID	Question 13	Question 14
223	I had to intervene when my daughter was bullied online (on Instagram). I also worry that the kids spend to much time on online games like fortnite and what effect this is having in them	I stopped my daughter using Instagram and this had an effect on her social circle and ultimately she ended up moving school because if it
224	To check that inappropriate sites were not being accessed. Worries that children gain access acting as parents! change settings and as in numerous cases gain control. It is not uncommon for children to be more comfortable and capeable using a computer than the parent, which could switch the control roles.	When people, life beyond a screen or important task are being ignored, the equipment is confiscated, internet cut off/restricted. Result- Bad tempers, moods and battle of wills. Child seeks digital technology from another source- libraries and friends house.
225		

User ID	Question 13	Question 14
226	Violent content Chat features - who they can / cannot communicate to	Downloading apps - they ask before they can download PS4 on my account -- shown things he cannot do as my spend money Deleted apps in the past and as they made he very grumpy
227	Message from school saying something not done ie prompting work or admin. Also asking child to install anti virus software or suchlike.	Just sending links mainly
258		
259	I don't monitor my children's use of digital technologies much but that is because they don't use them much (less than one a month) and they only use them for specific things e.g. homework in the case of my 10 year old and the 5 year old only uses the CBeebies website so far.	I think I've only intervened to find better games for my 5 year old and to find the most useful links for the questions my 10 year old needs to answer.
260		

User ID	Question 13	Question 14
261	Triggers of fear, causing nightmares or viewing material that might introduce mature content that is not appropriate	Child is still young enough that we install apps and monitor all viewing. Time restrictions are also in place.
262		Limiting screen time around bedtime
263		
264	Time spent. Try to steer her towards the educational side	
265		I have a parental filter for the CBBC buzz app. This allows me to check my child's uploads before they go through to the app. They also have staff who filter content as well. This gives me confidence that my child is staying safe.

User ID	Question 13	Question 14
266	Age appropriateness, ensuring it does not influence other areas of life eg eating habits, mood or time outside	They are very young (3 and 6) so I often help them choose games, download new games (impact additional engagement, enjoyment, learning) and mediate turn taking or remove the technology (impact usually tantrums)
267		

Appendix E

Literature Classifications

Table E.1: CHI 2019 Subject Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
AI Ethics and Philosophy	AI Ethics	4	Wang, Yang, Abdul and Lim (2019) Hohman et al. (2019) Mirnig and Meschtscherjakov (2019) Yin et al. (2019)
Communication	NLP	2	Liang et al. (2019) Jo et al. (2019)
Learning	Machine Learning	1	Williams et al. (2019)
Perception	Computer Vision	2	Arakawa and Yakura (2019) Perusquia-Hernández et al. (2019)
Services	AI Services	11	Chen et al. (2019) Guzdial et al. (2019) McCormack et al. (2019) Choi et al. (2019) Constant and Levieux (2019) Jiang et al. (2019) Samson and Sumi (2019) Kocielnik et al. (2019) Koch et al. (2019) Hu et al. (2019) Wang, Ming, Jin, Shen, Liu, Smith, Veeramachaneni and Qu (2019)
None AI		2	Phelan et al. (2019) Türkay and Adinolf (2019)

Table E.2: CHI 2019 Implementation Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
AI Ethics and Philosophy	AI Ethics	1	Mirnig and Meschtscherjakov (2019)
Integration and Interaction	Robotics and Automation	1	Williams et al. (2019)
Learning	Machine Learning	15	Liang et al. (2019) Wang, Yang, Abdul and Lim (2019) Chen et al. (2019) Jo et al. (2019) Guzdial et al. (2019) McCormack et al. (2019) Arakawa and Yakura (2019) Choi et al. (2019) Constant and Levieux (2019) Kocielnik et al. (2019) Yin et al. (2019) Koch et al. (2019) Perusquia-Hernández et al. (2019) Hu et al. (2019) Wang, Ming, Jin, Shen, Liu, Smith, Veeramachaneni and Qu (2019)
Services	AI Services	2	Hohman et al. (2019) Jiang et al. (2019)
None AI		1	Samson and Sumi (2019)
Not Considered, None AI Subject		2	Phelan et al. (2019) Türkay and Adinolf (2019)

Table E.3: CHI 2020 Subject Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
AI Ethics and Philosophy	AI Ethics	13	Kaur et al. (2020) Mallari et al. (2020) Kuzminykh et al. (2020) Lindley et al. (2020) Long and Magerko (2020) Yang et al. (2020) Kontogiorgos et al. (2020) Gero et al. (2020) Troiano et al. (2020) Völkel et al. (2020) Madaio et al. (2020) Wang et al. (2020) Shi et al. (2020)
AI Ethics and Philosophy	Philosophy of AI	1	Dove and Fayard (2020) Sun, Feng, Chen, Wang, Zeng, Yuan, Pong and Qu (2020)
Integration and Interaction	Robotics and Automation	1	Noguchi and Tanaka (2020)
Services	AI Services	11	Yan et al. (2020) Xu and Warschauer (2020*c*) Hohman et al. (2020) Wentzel et al. (2020) Agarwal and Sivakumar (2020) Cheema et al. (2020) Mayer et al. (2020) Bachynskyi and Müller (2020) Asai et al. (2020) Sun, Li, Chen, Lee, Liu, Zhang, Huang, Shi and Xu (2020) Sun, Feng, Chen, Wang, Zeng, Yuan, Pong and Qu (2020)
None AI		1	Okuya et al. (2020)

Table E.4: CHI 2020 Implementation Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
Integration and Interaction	Robotics and Automation	2	Kontogiorgos et al. (2020) Noguchi and Tanaka (2020)
Learning	Machine Learning	8	Sun, Li, Chen, Lee, Liu, Zhang, Huang, Shi and Xu (2020) Kaur et al. (2020) Hohman et al. (2020) Sun, Feng, Chen, Wang, Zeng, Yuan, Pong and Qu (2020) Cheema et al. (2020) Shi et al. (2020)
Perception	Computer Vision	1	Yan et al. (2020)
Services	AI Services	8	Wentzel et al. (2020) Agarwal and Sivakumar (2020) Völkel et al. (2020) Mayer et al. (2020) Bachynskyi and Müller (2020) Asai et al. (2020) Xu and Warschauer (2020*c*) Kuzminykh et al. (2020)
None AI		7	Mallari et al. (2020) Lindley et al. (2020) Long and Magerko (2020) Yang et al. (2020) Dove and Fayard (2020) Madaio et al. (2020) Wang et al. (2020)
Not Considered, None AI Subject		1	Okuya et al. (2020)

Table E.5: CHI 2021 Subject Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
AI Ethics and Philosophy	AI Ethics	15	Poursabzi-Sangdeh et al. (2021) Tsai et al. (2021) Jiang et al. (2021) Liao and Sundar (2021) Richardson et al. (2021) Anik and Bunt (2021) Ehsan et al. (2021) Bansal et al. (2021) Lima et al. (2021) Park et al. (2021) Cheng et al. (2021) Suresh et al. (2021) Benjamin et al. (2021) Lu and Yin (2021) Ross et al. (2021)
AI Ethics and Philosophy	Philosophy of AI	1	Zhu et al. (2021)
Communication	Natural Language Processing	2	Zaheer et al. (2021) Molina et al. (2021)
Integration and Interaction	Multi-agent Systems	1	Ashktorab et al. (2021)
Learning	Machine Learning	3	Zhang and Banovic (2021) Wang et al. (2021) Xin et al. (2021)

AI Domain	AI Sub Domain	Count	Papers
Services	AI Services	15	Hughes and Roy (2021) Hong et al. (2021) Chen, Takashima, Fujita and Kitamura (2021) Guo et al. (2021) Gordon et al. (2021) Han et al. (2021) Bunian et al. (2021) Bennett et al. (2021) Kang et al. (2021) Lee et al. (2021) Saquib et al. (2021) Lambton-Howard et al. (2021) Prange et al. (2021) Lemmer et al. (2021) AlOmar et al. (2021)

AI Domain	AI Sub Domain	Count	Papers
None AI		21	Miyatake et al. (2021) Key et al. (2021) Ettehadi et al. (2021) Corbett and Dantec (2021) Shinohara et al. (2021) Mack et al. (2021) Li et al. (2021) Lakshmi et al. (2021) Yoo et al. (2021) Chen, Vitale and McGrenere (2021) Bruns et al. (2021) Venkatasubramanian et al. (2021) Dai and Moffatt (2021) Lewis and Venkatasubramanian (2021) Song and Paulos (2021) Lin and Brummelen (2021) Uzor et al. (2021) Tigwell (2021) Sin et al. (2021) Kirabo et al. (2021) Gray et al. (2021)

Table E.6: CHI 2021 Implementation Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
Communication	Natural Language Processing	2	Guo et al. (2021) Han et al. (2021)
Integration and Interaction	Multi-agent Systems	1	Ashktorab et al. (2021)
Learning	Machine Learning	11	Poursabzi-Sangdeh et al. (2021) Gordon et al. (2021) Bansal et al. (2021) Bunian et al. (2021) Zaheer et al. (2021) Molina et al. (2021) Lee et al. (2021) Lu and Yin (2021) Wang et al. (2021) Ross et al. (2021) AlOmar et al. (2021)
Services	AI Services	11	Tsai et al. (2021) Jiang et al. (2021) Chen, Takashima, Fujita and Kitamura (2021) Liao and Sundar (2021) Richardson et al. (2021) Zhang and Banovic (2021) Kang et al. (2021) Saquib et al. (2021) Lambton-Howard et al. (2021) Prange et al. (2021) Lemmer et al. (2021)

AI Domain	AI Sub Domain	Count	Papers
None AI		12	Hughes and Roy (2021) Hong et al. (2021) Anik and Bunt (2021) Ehsan et al. (2021) Lima et al. (2021) Zhu et al. (2021) Park et al. (2021) Bennett et al. (2021) Cheng et al. (2021) Suresh et al. (2021) Benjamin et al. (2021) Xin et al. (2021)
Not Considered, None AI Subject		21	See (Table E.5) None AI

Table E.7: CHI 2022 Subject Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
AI Ethics and Philosophy	AI Ethics	3	Park et al. (2022) Lyons et al. (2022) Panigutti et al. (2022)
Communication	Natural Language Processing	8	Yeh et al. (2022) Kim et al. (2022*b*) Chung et al. (2022) Zheng, Tang, Liu, Liu and Huang (2022) Arakawa et al. (2022) Cai et al. (2022) Jasim et al. (2022*b*) Hope et al. (2022)
Integration and Interaction	Multi-agent Systems	3	Mahmood et al. (2022) Zhu et al. (2022) Cila (2022)
Learning	Machine Learning	6	Lai et al. (2022) Cambo and Gergle (2022) Boggust et al. (2022) Dang et al. (2022) Hadash et al. (2022) Wang et al. (2022)

AI Domain	AI Sub Domain	Count	Papers
Services	AI Services	10	Grace et al. (2022) Windl et al. (2022) Langer et al. (2022) Yan et al. (2022) Pan et al. (2022) Liao et al. (2022) Cimolino and Graham (2022) Zheng, Wang, Wang and Ma (2022) Bäuerle et al. (2022) Louie et al. (2022)
None AI		2	Renom et al. (2022) Pang et al. (2022)

Table E.8: CHI 2022 Implementation Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
Communication	Natural Language Processing	1	Yeh et al. (2022)
Integration and Interaction	Multi-agent Systems	1	Mahmood et al. (2022)
Learning	Machine Learning	3	Lai et al. (2022) Cambo and Gergle (2022) Boggust et al. (2022)
Services	AI Services	14	Grace et al. (2022) Yan et al. (2022) Pan et al. (2022) Kim et al. (2022*b*) Liao et al. (2022) Chung et al. (2022) Arakawa et al. (2022) Cai et al. (2022) Dang et al. (2022) Jasim et al. (2022*b*) Zheng, Wang, Wang and Ma (2022) Hope et al. (2022) Bäuerle et al. (2022) Louie et al. (2022)

AI Domain	AI Sub Domain	Count	Papers
None AI		11	Windl et al. (2022) Langer et al. (2022) Park et al. (2022) Lyons et al. (2022) Zheng, Tang, Liu, Liu and Huang (2022) Panigutti et al. (2022) Cimolino and Graham (2022) Zhu et al. (2022) Hadash et al. (2022) Wang et al. (2022) Cila (2022)
Not Considered, None AI Subject		2	Renom et al. (2022) Pang et al. (2022) [1]

[1] Pang et al. (2022)'s prime contribution is to research methods albeit using AI techniques in the implementation. Hence their contribution is not considered for further analysis.

Table E.9: IDC 2019 Subject Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
Communication	NLP	1	Lovato et al. (2019)
Integration and Interaction	Robotics and Automation	1	Michaelis and Mutlu (2019)
Learning	Machine Learning	2	Fails et al. (2019) Zimmermann-Niefield et al. (2019)
Services	AI Services	3	Kang et al. (2019) Sharma, Papavlasopoulou and Giannakos (2019) Badillo-Urquiola et al. (2019)
None AI		6	Cabrera et al. (2019) Maldonado and Zekelman (2019) Soni et al. (2019) Seraj et al. (2019) Jones et al. (2019) Sharma, Kallioniemi, Hakulinen, Keskinen and Turunen (2019)

Table E.10: IDC 2019 Implementation Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
Communication	NLP	1	Lovato et al. (2019)
Integration and Interaction	Robotics and Automation	1	Michaelis and Mutlu (2019)
Learning	Machine Learning	1	Zimmermann-Niefield et al. (2019)
Perception	Computer Vision	1	Sharma, Papavlasopoulou and Giannakos (2019)
Services	AI Services	1	Kang et al. (2019)
None AI		2	Badillo-Urquiola et al. (2019) Fails et al. (2019)
Not Considered, None AI Subject		6	Cabrera et al. (2019) Maldonado and Zekelman (2019) Soni et al. (2019) Seraj et al. (2019) Jones et al. (2019) Sharma, Kallioniemi, Hakulinen, Keskinen and Turunen (2019)

Table E.11: IDC 2020 Subject Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
Communication	NLP	4	Santos et al. (2020) Xu and Warschauer (2020*b*) Xu and Warschauer (2020*a*) Spitale et al. (2020)
Integration and Interaction	Robotics and Automation	4	Malinverni and Valero (2020) Cagiltay et al. (2020) Boffi (2020) van Ewijk et al. (2020)
Learning	Machine Learning	1	Wan et al. (2020)
Services	AI Services	7	Ruan et al. (2020) Cheung et al. (2020) Shin and Holtz (2020) Zimmermann-Niefield et al. (2020) Lee-Cultura et al. (2020) Silva et al. (2020) DiPaola et al. (2020)
None AI		7	McEwan et al. (2020) Potapov and Marshall (2020) Lechelt et al. (2020) Cumbo and Iversen (2020) Almjally et al. (2020) Van Mechelen et al. (2020) Long et al. (2020)

Table E.12: IDC 2020 Implementation Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
Communication	NLP	4	Santos et al. (2020) Xu and Warschauer (2020*b*) Xu and Warschauer (2020*a*) Spitale et al. (2020)
Learning	Machine Learning	1	Zimmermann-Niefield et al. (2020)
Services	AI Services	3	Wan et al. (2020) Cheung et al. (2020) Lee-Cultura et al. (2020)
None AI		8	Malinverni and Valero (2020) Ruan et al. (2020) Cagiltay et al. (2020) Boffi (2020) Shin and Holtz (2020) van Ewijk et al. (2020) Silva et al. (2020) DiPaola et al. (2020)
Not Considered, None AI Subject		7	McEwan et al. (2020) Potapov and Marshall (2020) Lechelt et al. (2020) Cumbo and Iversen (2020) Almjally et al. (2020) Van Mechelen et al. (2020) Long et al. (2020)

Table E.13: IDC 2021 Subject Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
AI Ethics and Philosophy	AI Ethics	2	Charisi et al. (2021) Melsión et al. (2021)
Communication	Natural Language Processing	3	Petousi et al. (2021) Hiniker et al. (2021) Motozawa et al. (2021)
Integration and Interaction	Multi-agent Systems	8	White et al. (2021) Chiou et al. (2021) Stower and Kappas (2021) Tolksdorf et al. (2021) Ho et al. (2021) Elbeleidy et al. (2021) Sanoubari et al. (2021) Fuhrmann et al. (2021)
Learning	Machine Learning	6	Tseng et al. (2021) Voulgari et al. (2021) Aki Tamashiro et al. (2021) Huan and Brewster (2021) Agostinelli et al. (2021) Zhou et al. (2021)
Services	AI Services	9	Zhang, Zhou, Wu, Hu, Shao, Liu, Hu, Ying and Yao (2021) Druga and Ko (2021) Hope Currin et al. (2021) Zhang, Liu, Ying, Huang, Yao and Ying (2021) Stefanidi et al. (2021) Im and Rogers (2021) Sharma et al. (2021) Schloss et al. (2021) Van Brummelen et al. (2021)

AI Domain	AI Sub Domain	Count	Papers
None AI		10	Read et al. (2021) Marconi et al. (2021) Simko et al. (2021) Lee-Cultura et al. (2021) Bae et al. (2021) Yim et al. (2021) Bhaduri et al. (2021) Dao-Kroeker et al. (2021) Nouwen and Duflos (2021) Zarei et al. (2021)

Table E.14: IDC 2021 Implementation Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
Communication	Natural Language Processing	3	Hope Currin et al. (2021) Petousi et al. (2021) Hiniker et al. (2021)
Integration and Interaction	Multi-agent Systems	8	White et al. (2021) Chiou et al. (2021) Stower and Kappas (2021) Tolksdorf et al. (2021) Ho et al. (2021) Charisi et al. (2021) Elbeleidy et al. (2021) Sanoubari et al. (2021)
Learning	Machine Learning	6	Druga and Ko (2021) Aki Tamashiro et al. (2021) Huan and Brewster (2021) Sharma et al. (2021) Motozawa et al. (2021) Melsión et al. (2021)
Services	AI Services	10	Tseng et al. (2021) Voulgari et al. (2021) Zhang, Zhou, Wu, Hu, Shao, Liu, Hu, Ying and Yao (2021) Zhang, Liu, Ying, Huang, Yao and Ying (2021) Stefanidi et al. (2021) Agostinelli et al. (2021) Im and Rogers (2021) Zhou et al. (2021) Schloss et al. (2021) Van Brummelen et al. (2021)
None AI		1	Fuhrmann et al. (2021)

AI Domain	AI Sub Domain	Count	Papers
Not Considered, None AI Subject		3	Read et al. (2021) Marconi et al. (2021) Simko et al. (2021) Lee-Cultura et al. (2021) Bae et al. (2021) Yim et al. (2021) Bhaduri et al. (2021) Dao-Kroeker et al. (2021) Nouwen and Duflos (2021) Zarei et al. (2021)

Table E.15: IDC 2022 Subject Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
AI Ethics and Philosophy	AI Ethics	3	Zhao et al. (2022) Escobar-Planas (2022) Antle et al. (2022)
Integration and Interaction	Multi-agent Systems	6	Cagiltay, White, Ibtasar, Mutlu and Michaelis (2022) Du and Breazeal (2022) Guneysu Ozgur et al. (2022) Cagiltay, Michaelis, Sebo and Mutlu (2022) Rubegni et al. (2022*b*) Yadollahi et al. (2022)
Learning	Machine Learning	2	Ruan et al. (2022) Dietz et al. (2022*b*)
Services	AI Services	15	Gagan et al. (2022) Zarei et al. (2022) Mansi et al. (2022) Yu and Roque (2022) Gürbüzsel et al. (2022) Andrade et al. (2022) Nguyen (2022*a*) El Shemy (2022) Ho et al. (2022) Thomas et al. (2022) Lin et al. (2022) Sasaki Otani (2022) Chatain et al. (2022) Câmara Olim et al. (2022) Tisza et al. (2022)

AI Domain	AI Sub Domain	Count	Papers
None AI		11	Read et al. (2022) Eriksson et al. (2022) Worsley (2022) Rodrigues et al. (2022) McDermott et al. (2022) Druga et al. (2022) Morales-Navarro et al. (2022) Vacca et al. (2022) Bilstrup et al. (2022) Ferreira (2022) Li et al. (2022)

Table E.16: IDC 2022 Implementation Classifications by Domain

AI Domain	AI Sub Domain	Count	Papers
Integration and Interaction	Multi-agent Systems	4	Cagiltay, White, Ibtasar, Mutlu and Michaelis (2022) Du and Breazeal (2022) Cagiltay, Michaelis, Sebo and Mutlu (2022) Yadollahi et al. (2022)
Learning	Machine Learning	3	Ruan et al. (2022) Thomas et al. (2022) Tisza et al. (2022)
Services	AI Services	13	Mansi et al. (2022) Gürbüzsel et al. (2022) Andrade et al. (2022) Nguyen (2022*a*) Escobar-Planas (2022) Antle et al. (2022) Ho et al. (2022) Guneysu Ozgur et al. (2022) Lin et al. (2022) Dietz et al. (2022*b*) Sasaki Otani (2022) Chatain et al. (2022) Câmara Olim et al. (2022)
None AI		6	Gagan et al. (2022) Zarei et al. (2022) Zhao et al. (2022) Yu and Roque (2022) El Shemy (2022) Rubegni et al. (2022*b*)

AI Domain	AI Sub Domain	Count	Papers
Not Considered, None AI Subject		11	Read et al. (2022) Eriksson et al. (2022) Worsley (2022) Rodrigues et al. (2022) McDermott et al. (2022) Druga et al. (2022) Morales-Navarro et al. (2022) Vacca et al. (2022) Bilstrup et al. (2022) Ferreira (2022) Li et al. (2022)

Appendix F

Literature Themes

Table F.1: CHI 2019 Semantic Classification

implicit communication.pdf	game ai	implicitly communicate	gameplay ai	ai games	implicature games
user centric.pdf	explainable ai	explaining ai	explanations ai	characterize ai	ai based
Gamut.pdf	model interpretability	datasets interpretability	interpretability models	interpretability designing	operationalizing interpretability
trolled by the trolley.pdf	trolley dilemmas	trolley dilemma	dilemmas trolley	using trolley	trolley problems
messageontap.pdf	suggest messageontap	messageontap suggestive	messageontap useful	messageontap intelligent	conversation apps
web blog.pdf	credibility weblogs	weblog credibility	assessment weblogs	credibility classification	credibility evaluations
friend.pdf	games ai	ai games	game designer	design ai	ai creative
in a silent way.pdf	musicians improvising	musicians improvise	improvising musicians	musical improvisers	musical improvisation
rescue.pdf	cues coaching	cues coaches	coaches detection	nonverbal cues	coaches rescue
Aila.pdf	attention labeling	attention labeler	labeling assistant	labeler attention	attention module
dynamic difficulty.pdf	game study	games psychology	games difficulty	confidence games	games overconfidence
orc.pdf	adaptive layout	orc layouts	gui layouts	layouts orc	layouts flow
exploring factors.pdf	driving navigating	driving routes	explored drivers	driving navigation	drivers desire
ATMseer.pdf	models automated	automated model	model searching	models search	automl algorithms
imperfect ai.pdf	emails ai	ai meeting	meeting ai	expectations ai	accept ai
Accuracy on trust.pdf	accuracy trust	trust predictions	interpretability trust	trust automated	trust observed
may ai.pdf	interactive ideation	ideation designers	ideation material	ideation tool	ideation design
A is for.pdf	robots children	children robots	robot children	robots child	intelligence robots
invisible potential.pdf	facial emg	emg facial	facial electromyography	emg usability	emg signals
Vizml.pdf	datasets vizml	visualizations corpus	visualizations training	visualization choices	chart corpus

Table F.2: CHI 2020 Semantic Classification

fdhelper.pdf	fraud feature	detecting frauds	selecting fraud	unsupervised fraud	feature fraud
interpreting.pdf	interpretability tools	interpretability ml	interpretability tool	interpretability advances	interpretability machine
recidivism.pdf	race recidivism	racial information	recidivism race	race information	judgements racial
frown on error.pdf	conversation interruptions	conversation interruption	interrupting responses	intentional interruptions	interruption mind
what are you talking to.pdf	children technological	child conversations	children conceptualize	children understanding	children interaction
genie.pdf	conversational agent	conversational agents	anthropomorphization conversational	perceptions agents	perceptions conversational
iteration.pdf	iteration visualizations	extending visualization	visualizations data	visualizations examine	visualizations integrated
AI legibility.pdf	legibility designing	designing legibility	legibility ai	ai legibility	legibility design
reach bound.pdf	virtual body	body virtual	movements vr	movement vr	reality ergonomics
ai literacy.pdf	ai literacy	literacy ai	ai education	learner ai	literacy hci
bci.pdf	bci battery	bci wearable	bci wearables	wake bci	wearable eeg
human-ai .pdf	design ai	ai designing	designed ai	ai designers	ai design
embodiment.pdf	robot embodiments	robots failure	failing robots	robot failures	perceptions robots
mental models.pdf	ai games	ai agents	game ai	ai agent	ai users
Dfseer.pdf	models demand	model demand	demand visualization	model forecasting	models forecasting
how i met your mother.pdf	robots sex	robots sexual	robot sex	robotic sex	sex robots
monsters.pdf	generative metaphors	metaphor monster	metaphor monsters	generative metaphor	metaphors machine
how to trick ai.pdf	chatbot personality	personality chatbot	chatbot personalities	chatbot assesses	chatbot assessment
co-designing checklist.pdf	ai ethics	fairness ai	ai fairness	ethics checklists	ai fair
OMOY.pdf	robotic emotions	weight robotic	robotic emotional	robotic gadget	weight embodiment
fatigue.pdf	interaction fatigue	simulated fatigue	realism fatigue	human simulation	predict fatigue
deictic.pdf	gestures virtual	deictic gestures	gestures considered	gestures collaborative	gestures conducted
mid-air.pdf	air movements	air movement	modeling movements	movements models	predict movement
scatter plot.pdf	program visualizes	analytics programming	interactive scatter	data scatter	plot editor
fairness.pdf	fairness algorithmic	algorithmic fairness	algorithms fairness	fairness algorithms	fairness research
bot identities.pdf	identities chatbots	persuasive chatbots	humanizing chatbots	chatbots involve	chatbots increasingly

Table F.3: CHI 2021 Semantic Classification

keeper.pdf	conversation facilitation	conversations facilita	conversation participation	online facilitation	conversation participants
manipulating.pdf	models interpretable	model interpretability	interpretable models	model interpretable	interpretable model
symptom checker.pdf	ai informing	diseasesdiagnostic transparency	diagnosis ai	information explanations	trustworthy symptom
spectroscopy.pdf	near infrared	infrared spectroscopy	infrared nir	infrared	infrared quantitative
playbook.pdf	prototyping ai	ai prototyping	nl prototyping	requirements ai	prototypes ai
pinpointfly.pdf	pointing drone	interactive drone	drone controller	position drone	controls drone
shing.pdf	communication fraud	payment conversation	alerting customers	alert customers	conversational agent
deconvolution.pdf	accuracy tasks	metric tasks	metrics accuracy	task accuracy	performance metrics
ai agent.pdf	communication ai	communicated ai	ai interaction	perceive ai	agents ai
role framing.pdf	information users	personal information	information use	helping users	assisting users
rubric.pdf	fairness toolkits	toolkits fairness	fair toolkits	fairness toolkit	fairness tools
data centric.pdf	trust explanations	explanations data	explanations training	explanations trust	data explanations
chi21.bib	trust explanations	explanations data	explanations training	explanations trust	data explanations
social transparency.pdf	ai social	ai explanations	explanations ai	explaining ai	ai explainability
gans.pdf	gans sample	images gan	gans generate	gans evaluation	gan evaluation
dementia.pdf	dementia qualitative	dementia exploring	insights dementia	dementia researchers	dementia agenda
team performance.pdf	ai explanations	explanations ai	adaptive explanations	explaining ai	human explanations
bail.pdf	ai moral	ai ethics	ai ethical	responsibility ai	ethical ai
chatbot.pdf	chatbot research	design chatbots	designing chatbot	chatbots interview	chatbot researchers
vins.pdf	ui retrieval	ui searching	visual searching	visual search	image retrieval
network games.pdf	ai games	games ai	playing ai	neuroevolution games	game ai
semour.pdf	speech database	utterances semour	emotional corpus	emotion speech	speech databases
clickbait.pdf	clickbait perceptions	clickbait characteristics	clickbaity characteristics	characteristics clickbait	classifying clickbait
HR.pdf	employees ai	ai employees	ai hr	work ai	hrm ai
image descriptions.pdf	racial identity	descriptions race	communicating appearance	nonvisual accessibility	describing race
metamap.pdf	visual metaphors	visual metaphor	metaphors visual	metaphor design	metaphors design
rehabilitation.pdf	therapist ai	assessment ai	personalized rehabilitation	rehabilitation assessment	assessment rehabilitation
child maltreatment.pdf	fairness learning	learning fairness	fairness algorithmic	eliciting fairness	fairness elicitation
beyond expertise.pdf	expertise interpretability	interpretability stakeholders	stakeholders interpretability	expertise interpretable	stakeholders interpretable
algebra.pdf	embodied algebra	embodied math	sketching abstraction	sketching embodied	math embodied
phenomenological.pdf	uncertainty designerly	technologies uncertainty	designers ml	ml design	design ai
peer support.pdf	peer support	support peer	support social	health peer	social support
heuristics.pdf	relying ai	models ai	feedback heuristics	ai driven	human ai
autods.pdf	ml automation	automation data	automated data	use automation	automating data
dementia screening.pdf	test cognitive	cognitive testing	cognitive test	test dementia	cognitive tests
generative models.pdf	interpretability generative	models interpretable	interpretable models	interpretability deep	interpretable representation
trackers.pdf	tracker initialization	tracker initialized	crowdsourced initialization	tracker crowdsourced	tracker robustness
automl.pdf	automation ml	ml automated	ml automating	automation tasks	increasing automation
haystack.pdf	accessibility reviews	review accessibility	reviews accessibility	users accessibility	accessibility review

Table F.4: CHI 2022 Semantic Classification

q-chef.pdf	recipes perceived	recipes selection	evaluating recipes	selecting recipes	recipe recommender
discovery time.pdf	ai designers	designers ai	ai designing	design ai	ai designer
how to.pdf	chatbot tasks	chatbot research	chatbot participants	chatbot complex	chatbot guidance
look.pdf	algorithmic decisions	algorithmic decision	evaluations algorithmic	algorithms perceived	algorithms human
ai errors.pdf	ai apology	apology ai	agent apology	erroneous ai	errors agents
tensions.pdf	hrm stakeholders	stakeholders hrm	stakeholders ai	ai stakeholder	ai stakeholders
emoglass.pdf	wearable emotion	wearables emotion	wearable emotional	emotions emoglass	monitor emotions
whats the appeal.pdf	permitted copy	algorithms reviewers	algorithmic review	algorithmic reviewers	review procesacm
smartphones.pdf	voice commands	voice command	voice interface	automatic voice	voice interfaces
stylette.pdf	predicting css	style web	familiarize css	styling web	stylete styling
user trust.pdf	recommender perceived	recommendation systems	trust recommender	recommender compared	recommender systems
talebrush.pdf	generating stories	generating story	generate story	generate stories	generative story
conditional delegation.pdf	delegation ai	ai collaborative	ai collaborate	distrust ai	ai collaboration
reflexivity.pdf	data annotators	data annotations	annotators comment	annotation behaviors	annotators knowledge
ux.pdf	conversation ai	conversational interfaces	conversation autonomous	conversational ai	interactions conversational
vocab encounter.pdf	vocabulary vocabencounter	vocabencounter vocabulary	vocabulary browsing	learning vocabencounter	computers vocabulary
impacts.pdf	trust conversational	personality trust	trust interpersonal	trust perceived	perceived trust
advice taking.pdf	ai explanations	explanations ai	ai explainability	advice ai	ai informing
two heads.pdf	design shared	designing shared	collaborative game	designs shared	collaboratively control
shared interest.pdf	reasoning saliency	models saliency	model saliency	model saliencies	model salient
robots.pdf	robots gaze	robot perceived	robot gaze	robot look	robots eyes
GANslider.pdf	interactive generative	generative visual	generative image	exploring generative	visualization ganslider
bias.pdf	analyzing reviews	reviews exploration	analyze reviews	exploring reviews	explore reviews
improving.pdf	explained features	understandability feature	features readability	features unexplained	feature contributions
chi22.bib	explained features	understandability feature	features readability	features unexplained	feature contributions
telling stories.pdf	notebooks presentation	slides notebooks	slides notebook	notebooks slides	notebook slides
ideas.pdf	finding ideas	idea descriptions	ideation design	ideas products	idea representations
whose ai.pdf	annotation career	annotation jobs	annotation industry	annotators interview	work annotators
symphony.pdf	interactive ml	machine learn	training analyzing	machine learning	framework ml
affinder.pdf	activity contexts	experiences context	context features	context awareness	context programming
human agent.pdf	collaboration hci	collaboration agent	collaboration agents	collaborate agents	agent collaboration

Table F.5: IDC 2019 Semantic Classification

papercraft.pdf	prototyping children	prototyping tools	interactive prototyping	prototyping ideas	prototyping
joint emotional state.pdf	emotions collaborative	collaborative emotional	emotions coding	collaboration emotions	sharing emotions
science learning.pdf	robots social	robots socially	social robot	robot encourage	social robots
do unicorns exist.pdf	conversational agents	conversational agent	children asking	children requests	discussion children
stranger danger.pdf	mobile threats	children privacy	privacy children	apps children	apps cyberbullying
query formulation.pdf	children queries	child queries	queries children	functionality kidzsearch	kids search
youth learning machine.pdf	modeling youth	youth athletics	models youth	youth athletes	modeling athletics

Table F.6: IDC 2020 Semantic Classification

about robots.pdf	robot artistic	robot creativity	robótica educativa	robots imaginaries	arts robotics
therapist vibe.pdf	emotions chatbots	emotions chatbot	storytelling chatbot	chatbot emotion	storytelling chatbots
conversational agent.pdf	children conversations	children conversation	child conversations	conversation reading	reading conversational
maths learning.pdf	story interface	interactive narrative	narratives tutoring	augmented narrative	narratives interactive
in-home robot.pdf	robot parents	robot families	robot home	family robot	robot participatory
apps on the market.pdf	literacy apps	apps preschoolers	preschoolers apps	apps preschool	preschool apps
smiley cluster.pdf	skills ml	ml knowledge	ml activities	ml technologies	learning elicit
TAR.pdf	augmented tangibles	tangible learning	tangible tabletop	tangible augmented	tangibles augmented
ding dong.pdf	remote readers	child readers	children reader	embodied remote	child reader
linguistic.pdf	conversational impairment	impairment linguistic	children conversational	impairment language	children linguistic
managing diabetes.pdf	technology diabetes	smartphones children	mobile health	health app	apps interventions
teachers perspective.pdf	robots educational	robots moral	robots classroom	robots teachers	robots teaching
gesture controlled.pdf	creatively gestures	computing youth	programming youth	projects gestures	youth modeling
sensing technologies.pdf	interaction children	children interaction	avatar educational	games children	avatars educational
blue whale.pdf	children cognitivemaps	children maps	maps children	maps cognitive	children mapping
design agendas.pdf	technology ethics	technology ethical	youtube stakeholders	design agendas	design agenda

Table F.7: IDC 2021 Semantic Classification

toys.pdf	interactive toy	interactive toys	toys interactive	toys plushpal	interactive plush
game based.pdf	educational game	ai literacy	games ai	games artificial	digital game
bio sketchbook.pdf	sketching nature	biodiversity learning	observational sketching	children biodiversity	learn biodiversity
child-robot.pdf	robot emotional	robot emotions	robot emotionally	robot commentary	robot affective
machine intelligence.pdf	children smart	intelligence kids	intelligence children	perceive smart	child smart
shy.pdf	executive functions	executive function	executive functioning	children executive	preschool executive
machine learning.pdf	design fiction	design children	fiction prototyping	students design	learning design
cybersecurity.pdf	classroom cybersecurity	students cybersecurity	children cybersecurity	teaching cybersecurity	robot cybersecurity
ModHera.pdf	baby wearable	infant monitoring	monitor newborns	monitor babies	parent wearable
AR.pdf	programming children	children programming	program intelligent	programming intelligent	children computers
cozmo.pdf	educational robot	educational robots	robots educational	robot education	robots classroom
proxemics.pdf	robot shy	shyness children	children shyness	shy children	children shy
MCAST.pdf	automated engagement	investigating storytelling	investigating children	investigate storytelling	engaging children
rubiks cube.pdf	skills children	ai children	children challenging	task children	designing children
draw2code.pdf	playful tangible	tangible computational	animation children	tangible programming	animations children
robomath.pdf	robots educational	kids robot	robot game	robot children	numbers robot
bridge.pdf	stem learning	scaffolding ml	discovery scaffolding	learning explored	facilitate learners
empathy.pdf	chatbots education	educational chatbots	potential chatbots	facilitate conversation	literature chatbots
japan.pdf	fairness storytelling	fairness children	fairness robot	fairness psychological	fair robot
talk.pdf	children conversations	children conversational	child conversation	children communication	conversational technologies
causal analysis.pdf	emotions coding	gaze emotions	children coding	coding causal	coding affect
intercultural.pdf	translation children	translation intercultural	intercultural collaboration	children intercultural	translation collaborative
autism.pdf	robots autism	robotics autism	robots teleoperation	therapists teleoperation	robot teleoperation
gender bias in ai.pdf	ai sexist	bias ai	gender bias	gender classifier	ai interpretability
in his belly.pdf	vr children	vr characters	characters vr	character vr	virtual characters
alexa.pdf	students alexa	conversation alexa	ai conversational	artificial conversational	conversational artificial
bullies.pdf	bullying robots	robots bullying	bullying robot	robots bullied	robot bullied
middle school.pdf	robot students	computational thinking	thinking computational	science classrooms	experiments curriculum

Table F.8: IDC 2022 Semantic Classification

asd.pdf	conversational artificial	children conversations	conversational agent	conversational agents	conversational ability
cues.pdf	classroom writing	children writing	interactive writing	video writing	child writing
koala.pdf	privacy children	children privacy	apps children	app children	online privacy
k-2.pdf	art educa	art classroom	museum education	art education	museums classrooms
coding.pdf	coding kids	coding educators	coding experiences	teaching coding	children coding
eliciting.pdf	developmental motor	development toys	motor developmental	designing toys	skill toys
in-home.pdf	robot children	educational robots	children robots	robot child	robots children
seastory.pdf	seastory interactive	design seastory	seastory design	interactive storytelling	novel interactive
exploring.pdf	teachers design	design teachers	pedagogy design	design educators	classrooms design
learning theory.pdf	interact learn	child learning	children learning	childhood learning	develop learning
teenagers.pdf	conversational agents	conversational technologies	conversational agent	competence conversational	technologies teenagers
language.pdf	technology augmented	experiences augmented	learning autism	mobile augmented	disabilities autism
towards trust.pdf	trustworthy conversational	children interaction	trustworthy child	chatbots	conversational agents
real time.pdf	metacognitive monitoring	emotion metacognition	metacognition children	metacognition asd	emotion metacognitive
bio.pdf	literacy biowearables	biowearables teaching	design children	ideation biowearable	technological literacy
kid connect.pdf	vr children	kidconnect vr	vr kidconnect	connect children	kids connect
identifying features.pdf	children sketches	sketches children	drawings children	child sketches	children drawings
robot.pdf	multiplayer tangible	game tangible	tangible robots	robots tangibles	tangible robotics
fish.pdf	interaction children	child interaction	children interaction	child interactions	facilitate children
ARtonomous.pdf	learn robotics	learning robotics	educational robotics	robotics learning	robotics educa
track track.pdf	puppet scenography	scenography puppet	privacy puppets	children biometric	spectator biometric
grasping deri.pdf	mathematics embodied	embodied mathematical	interaction embodied	embodied interaction	embodied interactions
social robot.pdf	robot caretaking	robot care	robot children	children robot	robot child
Grasping.pdf	mathematics embodied	embodied mathematical	interaction embodied	embodied interaction	embodied interactions
dogs.pdf	perceptions robots	robots perceived	robots children	fears robots	children robots
fable.pdf	children chemistry	chemistry children	concepts chemistry	education chemistry	students chemistry
adapt.pdf	robot perspective	perspective child	perspective spatial	children perspectives	children spatial
modal.pdf	learning fun	understanding fun	measuring fun	playful coding	investigating fun
theory.pdf	interact learn	child learning	children learning	childhood learning	develop learning
topic.pdf	topic correlated	topic cci	topics research	topics cci	integrated topics

Appendix G

Literature Keywords

Table G.1: CHI Papers By Keyword - AI Ethics and Philosophy

Paper	Author	AI Domain	AI Sub Domain	Keyword
Designing Theory-Driven User-Centric Explainable AI	Wang, Yang, Abdul and Lim (2019)	AI Ethics and Philosophy	AI Ethics	explainability
Gamut: A Design Probe to Understand How Data Scientists Understand Machine Learning Models	Hohman et al. (2019)	AI Ethics and Philosophy	AI Ethics	explainability (Interpretability)
Trolled by the Trolley Problem On What Maters for Ethical Decision Making in Automated Vehicles	Mirnig and Meschtscherjakov (2019)	AI Ethics and Philosophy	AI Ethics	Multiple Keywords
Understanding the Effect of Accuracy on Trust in Machine Learning Models	Yin et al. (2019)	AI Ethics and Philosophy	AI Ethics	Accuracy and Trust
Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning	Kaur et al. (2020)	AI Ethics and Philosophy	AI Ethics	explainability/interpretability
Do I Look Like a Criminal? Examining how Race Presentation Impacts Human Judgement of Recidivism	Mallari et al. (2020)	AI Ethics and Philosophy	AI Ethics	fairness
Genie in the Bottle: Anthropomorphized Perceptions of Conversational Agents	Kuzminykh et al. (2020)	AI Ethics and Philosophy	AI Ethics	Anthropmorphism Embodiment
Researching AI Legibility through Design	Lindley et al. (2020)	AI Ethics and Philosophy	AI Ethics	explainability/transparency
What is AI Literacy? Competencies and Design Considerations	Long and Magerko (2020)	AI Ethics and Philosophy	AI Ethics	explainability/transparency
Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design	Yang et al. (2020)	AI Ethics and Philosophy	AI Ethics	explainability/transparency
Embodiment Effects in Interactions with Failing Robots	Kontogiorgos et al. (2020)	AI Ethics and Philosophy	AI Ethics	Embodiment
Mental Models of AI Agents in a Cooperative Game Setting	Gero et al. (2020)	AI Ethics and Philosophy	AI Ethics	Mental Models
And This, Kids, Is How I Met Your Mother: Consumerist, Mundane, and Uncanny Futures with Sex Robots	Troiano et al. (2020)	AI Ethics and Philosophy	AI Ethics	Embodiment
Monsters, Metaphors, and Machine Learning	Dove and Fayard (2020)	AI Ethics and Philosophy	Philosophy of AI	artificial general intelligence
How to Trick AI: Users' Strategies for Protecting Themselves from Automatic Personality Assessment	Völkel et al. (2020)	AI Ethics and Philosophy	AI Ethics	privacy
Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI	Madaio et al. (2020)	AI Ethics and Philosophy	AI Ethics	General Design
Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences	Wang et al. (2020)	AI Ethics and Philosophy	AI Ethics	fairness
Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies	Shi et al. (2020)	AI Ethics and Philosophy	AI Ethics	Embodiment
Manipulating and Measuring Model Interpretability	Poursabzi-Sangdeh et al. (2021)	AI Ethics and Philosophy	AI Ethics	Interpretability
Exploring and Promoting Diagnostic Transparency and Explainability in Online Symptom Checkers	Tsai et al. (2021)	AI Ethics and Philosophy	AI Ethics	Explainability
User Trust in Assisted Decision-Making Using Miniaturized Near-Infrared Spectroscopy	Jiang et al. (2021)	AI Ethics and Philosophy	AI Ethics	Trust
How Should AI Systems Talk to Users when Collecting their Personal Information? Efects of Role Framing and Self-Referencing on Human-AI Interaction	Liao and Sundar (2021)	AI Ethics and Philosophy	AI Ethics	privacy
Towards Fairness in Practice: A Practitioner-Oriented Rubric for Evaluating Fair ML Toolkits	Richardson et al. (2021)	AI Ethics and Philosophy	AI Ethics	fairness
Data-Centric Explanations: Explaining Training Data of Machine Learning Systems to Promote Transparency	Anik and Bunt (2021)	AI Ethics and Philosophy	AI Ethics	transparency

Paper	Author	AI Domain	AI Sub Domain	Keyword
Expanding Explainability: Towards Social Transparency in AI systems	Ehsan et al. (2021)	AI Ethics and Philosophy	AI Ethics	transparency
Does the Whole Exceed its Parts? The Efect of AI Explanations on Complementary Team Performance	Bansal et al. (2021)	AI Ethics and Philosophy	AI Ethics	explainability
Human Perceptions on Moral Responsibility of AI: A Case Study in AI-Assisted Bail Decision-Making	Lima et al. (2021)	AI Ethics and Philosophy	AI Ethics	fairness morality
Player-AI Interaction: What Neural Network Games Reveal About AI as Play	Zhu et al. (2021)	AI Ethics and Philosophy	Philosophy of AI	Human AI Interaction
Human-AI Interaction in Human Resource Management: Understanding Why Employees Resist Algorithmic Evaluation at Workplaces and How to Mitigate Burdens	Park et al. (2021)	AI Ethics and Philosophy	AI Ethics	fairness
Soliciting Stakeholders' Fairness Notions in Child Maltreatment Predictive Systems	Cheng et al. (2021)	AI Ethics and Philosophy	AI Ethics	fairness
Beyond Expertise and Roles: A Framework to Characterize the Stakeholders of Interpretable Machine Learning and their Needs	Suresh et al. (2021)	AI Ethics and Philosophy	AI Ethics	Interpretability
Machine Learning Uncertainty as a Design Material: A Post-Phenomenological Inquiry	Benjamin et al. (2021)	AI Ethics and Philosophy	AI Ethics	Uncertainty
Human Reliance on Machine Learning Models When Performance Feedback is Limited: Heuristics and Risks	Lu and Yin (2021)	AI Ethics and Philosophy	AI Ethics	reliance accuracy metrics
Evaluating the Interpretability of Generative Models by Interactive Reconstruction	Ross et al. (2021)	AI Ethics and Philosophy	AI Ethics	Interpretability
Designing Fair AI in Human Resource Management: Understanding Tensions Surrounding Algorithmic Evaluation and Envisioning Stakeholder-Centered Solutions	Park et al. (2022)	AI Ethics and Philosophy	AI Ethics	fairness/transparency/interpretability
What's the Appeal? Perceptions of Review Processes for Algorithmic Decisions	Lyons et al. (2022)	AI Ethics and Philosophy	AI Ethics	fairness/transparency/interpretability
Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems	Panigutti et al. (2022)	AI Ethics and Philosophy	AI Ethics	explainability

Table G.2: CHI Papers By Keyword - Communication

Paper	Author	AI Domain	AI Sub Domain	Keyword
Implicit Communication of Actionable Information in Human-AI teams	Liang et al. (2019)	Communication	Natural Language Processing	Natural Languaue Understanding
How Do Humans Access the Credibility ofWeblogs: Qualifying and Verifying Human Factors with Machine Learning	Jo et al. (2019)	Communication	Natural language processing	text classification
SEMOUR: A Scripted Emotional Speech Repository for Urdu	Zaheer et al. (2021)	Communication	Natural Language Processing	sentiment analysis
Does Clickbait Actually Atract More Clicks? Three Clickbait Studies You Must Read	Molina et al. (2021)	Communication	Natural Language Processing	text classification
How to Guide Task-oriented Chatbot Users, and When: A Mixed-methods Study of Combinations of Chatbot Guidance Types and Timings	Yeh et al. (2022)	Communication	Natural Language Processing	chatbot
Stylete: Styling theWeb with Natural Language	Kim et al. (2022*b*)	Communication	Natural language processing	natural language understanding
TaleBrush: Sketching Stories with Generative Pretrained Language Models	Chung et al. (2022)	Communication	Natural language processing	natural language generation
UX Research on Conversational Human-AI Interaction: A Literature Review of the ACM Digital Library	Zheng, Tang, Liu, Liu and Huang (2022)	Communication	Natural language processing	Various
VocabEncounter: NMT-powered Vocabulary Learning by Presenting Computer-Generated Usages of ForeignWords into Users' Daily Lives	Arakawa et al. (2022)	Communication	Natural language processing	computational linguistics

Paper	Author	AI Domain	AI Sub Domain	Keyword
Impacts of Personal Characteristics on User Trust in Conversational Recommender Systems	Cai et al. (2022)	Communication	Natural language processing	trust
Supporting Serendipitous Discovery and Balanced Analysis of Online Product Reviews with Interaction-Driven Metrics and Bias-Mitigating Suggestions	Jasim et al. (2022*b*)	Communication	Natural language processing	natural language understanding
Scaling Creative Inspiration with Fine-Grained Functional Aspects of Ideas	Hope et al. (2022)	Communication	Natural language processing	natural language understanding

Table G.3: CHI Papers By Keyword - Integration and Interaction

Paper	Author	AI Domain	AI Sub Domain	Keyword
OMOY: A Handheld Robotic Gadget that Shifts its Weight to Express Emotions and Intentions	Noguchi and Tanaka (2020)	Integration and Interaction	Robotics and Automation	human-ai interaction
Efects of Communication Directionality and AI Agent Diferences in Human-AI Interaction	Ashktorab et al. (2021)	Integration and Interaction	Multi-agent systems	intelligent agent
Owning Mistakes Sincerely: Strategies for Mitigating AI Errors	Mahmood et al. (2022)	Integration and Interaction	Integration and Interaction	intelligent agent
The Trusted Listener: The Influence of Anthropomorphic Eye Design of Social Robots on User's Perception of Trustworthiness	Zhu et al. (2022)	Integration and Interaction	Robotics and Automation	social robot
Designing Human-Agent Collaborations: Commitment, responsiveness, and support	Cila (2022)	Integration and Interaction	Multi-agent systems	intelligent agent

Table G.4: CHI Papers By Keyword - Learning

Paper	Author	AI Domain	AI Sub Domain	Keyword
A is for Artificial Intelligence The Impact of Artificial Intelligence Activities on Young Children's Perceptions of Robots	Williams et al. (2019)	Learning	Machine Learning	supervised learning/generative model
Method for Exploring Generative Adversarial Networks (GANs) via Automatically Generated Image Galleries	Zhang and Banovic (2021)	Learning	Machine Learning	generative adversarial network
AutoDS: Towards Human-Centered Automation of Data Science	Wang et al. (2021)	Learning	Machine Learning	automated machine learning
Whither AutoML? Understanding the Role of Automation in Machine Learning Workflows	Xin et al. (2021)	Learning	Machine Learning	automated machine learning
Human-AI Collaboration via Conditional Delegation: A Case Study of Content Moderation	Lai et al. (2022)	Learning	Machine Learning	
Model Positionality and Computational Reflexivity: Promoting Reflexivity in Data Science	Cambo and Gergle (2022)	Learning	Machine Learning	Modelling
Shared Interest: Measuring Human-AI Alignment to Identify Recurring Paterns in Model Behavior	Boggust et al. (2022)	Learning	Machine Learning	neural networks
GANSlider: How Users Control Generative Models for Images using Multiple Sliders with and without Feedforward Information	Dang et al. (2022)	Learning	Machine Learning	generative adversarial network
Improving understandability of feature contributions in model-agnostic explainable AI tools	Hadash et al. (2022)	Learning	Machine Learning	supervised learning
Whose AI Dream? In search of the aspiration in data annotation.	Wang et al. (2022)	Learning	Machine Learning	supervised learning

Table G.5: CHI Papers By Keyword - Perception

Paper	Author	AI Domain	AI Sub Domain	Keyword
REsCUE: A framework for REal-time feedback on behavioral CUEs using multimodal anomaly detection	Arakawa and Yakura (2019)	Perception	Computer Vision	Multiple Keywords

Paper	Author	AI Domain	AI Sub Domain	Keyword
The Invisible Potential of Facial Electromyography A Comparison of EMG and Computer Vision when Distinguishing Posed from Spontaneous Smiles	Perusquia-Hernández et al. (2019)	Perception	Computer Vision	Action Recognition (smile)

Table G.6: CHI Papers By Keyword - Services

Paper	Author	AI Domain	AI Sub Domain	Keyword
MessageOnTap: A Suggestive Interface to Facilitate Messaging-related Tasks	Chen et al. (2019) 2019	Services	AI Services	intelligent user interface
Friend, Collaborator, Student, Manager: How Design of an AI-Driven Game Level Editor Affects Creators	Guzdial et al. (2019) 2019	Services	AI Services	computational creativity
In a Silent Way Communication Between AI and Improvising Musicians Beyond Sound	McCormack et al. (2019) 2019	Services	AI Services	computational creativity
REsCUE: A framework for REal-time feedback on behavioral CUEs using multimodal anomaly detection	Arakawa and Yakura (2019) 2019	Perception	Computer Vision	Multiple Keywords
AILA: Attentive Interactive Labeling Assistant for Document Classification through Attention-based Deep Neural Networks	Choi et al. (2019) 2019	Services	AI Services	decision support
Dynamic Difficulty Adjustment Impact on Players' Confidence	Constant and Levieux (2019) 2019	Services	AI Services	intelligent control
ORC Layout: Adaptive GUI Layout with OR-Constraints	Jiang et al. (2019) 2019	Services	AI Services	intelligent user interface
Exploring Factors that Influence Connected Drivers to (Not) Use or Follow Recommended Optimal Routes	Samson and Sumi (2019) 2019	Services	AI Services	decision support
ATMSeer: Increasing Transparency and Controllability in Automated Machine Learning	Wang, Yang, Abdul and Lim (2019) 2019	Services	AI Services	ai application
Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems	Kocielnik et al. (2019) 2019	Services	AI Services	ai application
May AI? Design Ideation with Cooperative Contextual Bandits	Koch et al. (2019) 2019	Services	AI Services	computational creativity
The Invisible Potential of Facial Electromyography A Comparison of EMG and Computer Vision when Distinguishing Posed from Spontaneous Smiles	Perusquia-Hernández et al. (2019)	Perception	Computer Vision	Action Reconition (smile)
VizML: A Machine Learning Approach to Visualization Recommendation	Hu et al. (2019) 2019	Services	AI Services	data analytics
FDHelper: Assist Unsupervised Fraud Detection Experts with Interactive Feature Selection and Evaluation	Sun, Li, Chen, Lee, Liu, Zhang, Huang, Shi and Xu (2020)	Services	AI Services	data analytics
FrownOnError: Interrupting Responses from Smart Speakers by Facial Expressions	Yan et al. (2020)	Services	AI Services	intelligent user interface
What Are You Talking to?: Understanding Children's Perceptions of Conversational Agents	Xu and Warschauer (2020*c*)	Services	AI Services	intelligent user interface
Understanding and Visualizing Data Iteration in Machine Learning	Hohman et al. (2020)	Services	AI Services	data analytics
Improving Virtual Reality Ergonomics Through Reach-Bounded Non-Linear Input Amplifcation	Wentzel et al. (2020)	Services	AI Services	virtual reality
Charge for a whole day: Extending Battery Life for BCI	Agarwal and Sivakumar (2020)	Services	AI Services	intelligent software development
DFSeer: A Visual Analytics Approach to Facilitate Model Selection for Demand Forecasting	Sun, Feng, Chen, Wang, Zeng, Yuan, Pong and Qu (2020)	Services	AI Services	ai software toolkit

Paper	Author	AI Domain	AI Sub Domain	Keyword
Predicting Mid-Air Interaction Movements and Fatigue Using Deep Reinforcement Learning	Cheema et al. (2020)	Services	AI Services	intelligent software development
Improving Humans' Ability to Interpret Deictic Gestures in Virtual Reality	Mayer et al. (2020)	Services	AI Services	intelligent software development
Dynamics of Aimed Mid-air Movements	Bachynskyi and Müller (2020)	Services	AI Services	intelligent user interface
Integrated Development Environment with Interactive Scatter Plot for Examining Statistical Modeling	Asai et al. (2020)	Services	AI Services	data analytics
Keeper: A Synchronous Online Conversation Environment Informed by In-Person Facilitation Practices	Hughes and Roy (2021)s	Services	AI Services	Intelligent Software
Planning for Natural Language Failures with the AI Playbook	Hong et al. (2021)	Services	AI Services	Decision Support
PinpointFly: An Egocentric Position-control Drone Interface using Mobile AR	Chen, Takashima, Fujita and Kitamura (2021)	Services	AI Services	Intelligent Software
Shing: A Conversational Agent to Alert Customers of Suspected Online-payment Fraud with Empathetical Communication Skills	Guo et al. (2021)	Services	AI Services	Intelligent user interface
The Disagreement Deconvolution: Bringing Machine Learning Performance Metrics In Line With Reality	Gordon et al. (2021)	Services	AI Services	data analytics
Designing Efective Interview Chatbots: Automatic Chatbot Profiling and Design Suggestion Generation for Chatbot Debugging	Han et al. (2021)	Services	AI Services	ai application
VINS: Visual Search for Mobile User Interface Design	Bunian et al. (2021)	Services	AI Services	ai application
"It's Complicated": Negotiating Accessibility and (Mis)Representation in Image Descriptions of Race, Gender, and Disability	Bennett et al. (2021)	Services	AI Services	ai application
MetaMap: Supporting Visual Metaphor Ideation through Multi-dimensional Example-based Exploration	Kang et al. (2021)	Services	AI Services	ai application
A Human-AI Collaborative Approach for Clinical Decision Making on Rehabilitation Assessment	Lee et al. (2021)	Services	AI Services	ai application
Constructing Embodied Algebra by Sketching	Saquib et al. (2021)	Services	AI Services	ai application
Blending into Everyday Life: Designing a Social Media-Based Peer Support System	Lambton-Howard et al. (2021)	Services	AI Services	ai application (tentative)
Explainable Automatic Evaluation of the Trail Making Test for Dementia Screening	Prange et al. (2021)	Services	AI Services	ai application
Crowdsourcing More Effective Initializations for Single-Target Trackers Through Automatic Re-querying	Lemmer et al. (2021)	Services	AI Services	ai application
Finding the Needle in a Haystack: On the Automatic Identification of Accessibility User Reviews	AlOmar et al. (2021)	Services	AI Services	ai application
Q-Chef: The impact of surprise-eliciting systems on food-related decision-making	Grace et al. (2022)	Services	AI Services	ai application
It Is Not Always Discovery Time': Four Pragmatic Approaches in Designing AI Systems	Windl et al. (2022)	Services	AI Services	ai application
"Look! It's a Computer Program! It's an Algorithm! It's AI!": Does Terminology Afect Human Perceptions and Evaluations of Algorithmic Decision-Making Systems?	Langer et al. (2022)	Services	AI Services	decision support
EmoGlass: an End-to-End AI-EnabledWearable Platform for Enhancing Self-Awareness of Emotional Health	Yan et al. (2022)	Services	AI Services	ai application
Automatically Generating and Improving Voice Command Interface from Operation Sequences on Smartphones	Pan et al. (2022)	Services	AI Services	intelligent user interface

Paper	Author	AI Domain	AI Sub Domain	Keyword
User Trust in Recommendation Systems: A comparison of Content-Based, Collaborative and Demographic Filtering	Liao et al. (2022)	Services	AI Services	decision support
Two Heads Are Beter Than One: A Dimension Space for Unifying Human and Artificial Intelligence in Shared Control	Cimolino and Graham (2022)	Services	AI Services	decision support
Telling Stories from Computational Notebooks: AI-Assisted Presentation Slides Creation for Presenting Data Science Work	Zheng, Wang, Wang and Ma (2022)	Services	AI Services	ai application
Symphony: Composing Interactive Interfaces for Machine Learning	Bäuerle et al. (2022)	Services	AI Services	ai application
Afinder: Expressing Concepts of Situations that Aford Activities using Context-Detectors	Louie et al. (2022)	Services	AI Services	ai application

Table G.7: IDC Papers By Keyword - AI Ethics and Philosophy

Paper	Author	AI Domain	AI Sub Domain	Keyword
Exploring the Concept of Fairness in Everyday, Imaginary and Robot Scenarios: A Cross-Cultural Study With Children in Japan and Uganda	Charisi et al. (2021)	AI Ethics and Philosophy	AI Ethics	fairness
Using Explainability to Help Children Understand Gender Bias in AI	Melsión et al. (2021)	AI Ethics and Philosophy	AI Ethics	explainability
KOALA Hero: Inform Children of Privacy Risks of Mobile Apps	Zhao et al. (2022)	AI Ethics and Philosophy	AI Ethics	privacy
Towards Trustworthy Conversational Agents for Children	Escobar-Planas (2022)	AI Ethics and Philosophy	AI Ethics	trust
There are a LOT of moral issues with biowearables" ... Teaching Design Ethics through a Critical Making Biowearable Workshop	Antle et al. (2022)	AI Ethics and Philosophy	AI Ethics	Design Ethics

Table G.8: IDC Papers By Keyword - Communication

Paper	Author	AI Domain	AI Sub Domain	Keyword
"Hey Google, Do Unicorns Exist?": Conversational Agents as a Path to Answers to Children's Questions	Lovato et al. (2019)	Communication	Natural language processing	chatbot
Therapist Vibe: Children's Expressions of their Emotions through Storytelling with a Chatbot	Santos et al. (2020)	Communication	Natural language processing	chatbot
Exploring Young Children's Engagement in	Xu and Warschauer (2020*b*)	Communication	Natural language processing	chatbot
A Content Analysis of Voice-based Apps on the Market for Early Literacy Development	Xu and Warschauer (2020*a*)	Communication	Natural language processing	chatbot
Whom would you like to talk with? Exploring Conversational Agents for Children's Linguistic Assessment	Spitale et al. (2020)	Communication	Natural language processing	chatbot
Social bots of conviction as dialogue facilitators for history education: Promoting historical empathy in teens through dialogue	Petousi et al. (2021)	Communication	Natural language processing	Conversational agent
Can Conversational Agents Change the Way Children Talk to People?	Hiniker et al. (2021)	Communication	Natural language processing	Conversational agent
Conversation Analysis for Facilitation in Children's Intercultural Collaboration	Motozawa et al. (2021)	Communication	Natural language processing	machine translation

Table G.9: IDC Papers By Keyword - Integration and Interaction

Paper	Author	AI Domain	AI Sub Domain	Keyword
Supporting Interest in Science Learning with a Social Robot	Michaelis and Mutlu (2019)	Integration and Interaction	Robotics and Automation	social robot
What is a robot? An artistic approach to understand children's imaginaries about robots	Malinverni and Valero (2020)	Integration and Interaction	Robotics and Automation	social robot
Investigating Family Perceptions and Design Preferences for an In-Home Robot	Cagiltay et al. (2020)	Integration and Interaction	Robotics and Automation	social robot
Ding- Dong: The Storybell and Its Wizard	Boffi (2020)	Integration and Interaction	Robotics and Automation	social robot
Teachers' Perspectives on Social Robots in Education: An Exploratory Case Study	van Ewijk et al. (2020)	Integration and Interaction	Robotics and Automation	social robot
Designing Emotionally Expressive Social Commentary to Facilitate Child-Robot Interaction	White et al. (2021)	Integration and Interaction	Robotics and Automation	social robot
Teacher Views on Storytelling-based Cybersecurity Education with Social Robots	Chiou et al. (2021)	Integration and Interaction	Robotics and Automation	social robot
CozmoNAOts: Designing an Autonomous Learning Task with Social and Educational Robots	Stower and Kappas (2021)	Integration and Interaction	Robotics and Automation	social robot

Paper	Author	AI Domain	AI Sub Domain	Keyword
Do Shy Children Keep more Distance from a Social Robot? Exploring Shy Children's Proxemics with a Social Robot or a Human	Tolksdorf et al. (2021)	Integration and Interaction	Robotics and Automation	social robot
RoboMath: Designing a Learning Companion Robot to Support Children's Numerical Skills	Ho et al. (2021)	Integration and Interaction	Robotics and Automation	social robot
Analyzing Teleoperation Interface Usage of Robots in Therapy for Children with Autism	Elbeleidy et al. (2021)	Integration and Interaction	Robotics and Automation	social robot
Robots, Bullies and Stories: A Remote Codesign Study with Children	Sanoubari et al. (2021)	Integration and Interaction	Robotics and Automation	social robot
Scientific Inquiry in Middle Schools by combining Computational Thinking,Wet Lab Experiments, and Liquid Handling Robots	Fuhrmann et al. (2021)	Integration and Interaction	Robotics and Automation	human-ai interaction
Understanding Factors that Shape Children's Long Term Engagement with an In-Home Learning Companion Robot	Cagiltay, White, Ibtasar, Mutlu and Michaelis (2022)	Integration and Interaction	Robotics and Automation	social robot
Exploring changes in special education teachers' attitudes and design belief towards pedagogical agents in co-designing with children	Du and Breazeal (2022)	Integration and Interaction	Multi-agent systems	agent-based modelling
Designing Tangible Robot Mediated Colocated Games to Enhance Social Inclusion for Neurodivergent Children	Guneysu Ozgur et al. (2022)	Integration and Interaction	Robotics and Automation	social robot
Exploring Children's Preferences for Taking Care of a Social Robot	Cagiltay, Michaelis, Sebo and Mutlu (2022)	Integration and Interaction	Robotics and Automation	social robot
"Don't let the robots walk our dogs, but it's ok for them to do our homework": children's perceptions, fears, and hopes in social robots.	Rubegni et al. (2022*b*)	Integration and Interaction	Robotics and Automation	social robot
Do Children Adapt Their Perspective to a Robot When They Fail to Complete a Task?	Yadollahi et al. (2022)	Integration and Interaction	Robotics and Automation	social robot

Table G.10: IDC Papers By Keyword - Learning

Paper	Author	AI Domain	AI Sub Domain	Keyword
Query Formulation Assistance for Kids: What is Available, When to Help & What Kids Want	Fails et al. (2019)	Learning	Machine Learning	recommender system
Youth Learning Machine Learning through Building Models of Athletic Moves	Zimmermann-Niefield et al. (2019)	Learning	Machine Learning	classification
SmileyCluster: Supporting Accessible Machine Learning in K-12 Scientific Discovery	Wan et al. (2020)	Learning	Machine Learning	clustering
PlushPal: Storytelling with Interactive Plush Toys and Machine Learning	Tseng et al. (2021)	Learning	AI Learning	supervised learning
Learn to Machine Learn: Designing a Game Based Approach for Teaching Machine Learning to Primary and Secondary Education Students	Voulgari et al. (2021)	Learning	AI Learning	supervised learning
Introducing Teenagers to Machine Learning through Design Fiction: An Exploratory Case Study	Aki Tamashiro et al. (2021)	Learning	Machine Learning	supervised learning
Designing an Engaging Story-stem taken from the MCAST test	Huan and Brewster (2021)	Learning	Machine Learning	classification
Designing Children's New Learning Partner: Collaborative Artificial Intelligence for Learning to Solve the Rubik's Cube	Agostinelli et al. (2021)	Learning	Machine Learning	supervised learning
Scaffolding Design to Bridge the Gaps between Machine Learning and Scientific Discovery for K-12 STEM Education	Zhou et al. (2021)	Learning	Machine Learning	clustering
Real-time Feedback based on Emotion Recognition for Improving Children's Metacognitive Monitoring Skill	Ruan et al. (2022)	Learning	Machine Learning	neural networks
ARtonomous: Introducing Middle School Students to Reinforcement Learning Through Virtual Robotics	Dietz et al. (2022*b*)	Learning	Machine Learning	reinforcement learning

Table G.11: IDC Papers By Keyword - Services

Paper	Author	AI Domain	AI Sub Domain	Keyword
PrototypAR: Prototyping and Simulating Complex Systems with Paper Craft and Augmented Reality	Kang et al. (2019)	Services	AI Services	ai application
Joint Emotional State of Children and Perceived Collaborative Experience in Coding Activities	Sharma, Papavlasopoulou and Giannakos (2019)	Services	AI Services	ai application
Stranger Danger! Social Media App Features Co-designed with Children to Keep Them Safe Online	Badillo-Urquiola et al. (2019)	Services	AI Services	ai application
Supporting Children's Math Learning with Feedback-Augmented Narrative Technology	Ruan et al. (2020)	Services	AI Services	ai application
Techniques for Augmented-Tangibles on Mobile Devices for Early Childhood Learning	Cheung et al. (2020)	Services	AI Services	ai application
Identifying Opportunities and Challenges: How Children Use Technologies for Managing Diabetes	Shin and Holtz (2020)	Services	AI Services	ai application
Youth Making Machine Learning Models for Gesture-Controlled Interactive Media	Zimmermann-Niefield et al. (2020)	Services	AI Services	ai application
Using Sensing Technologies to Explain Children's Self-Representation in Motion-Based Educational Games	Lee-Cultura et al. (2020)	Services	AI Services	ai application
Blue Whale Street Art as a Landmark: Extracting Landmarks from Children's Cognitive Maps for the Design of Locative Systems	Silva et al. (2020)	Services	AI Services	ai application
Decoding Design Agendas: An Ethical Design Activity for Middle School Students	DiPaola et al. (2020)	Services	AI Services	ai application
Bio Sketchbook an AI-assisted Sketching Partner for Children's Biodiversity Observational Learning	Zhang, Zhou, Wu, Hu, Shao, Liu, Hu, Ying and Yao (2021)	Services	AI Services	ai application
How do children's perceptions of machine intelligence change when training and coding smart programs?	Druga and Ko (2021)	Services	AI Services	ai application
Supporting Shy Preschool Children in Joining Social Play	Hope Currin et al. (2021)	Services	AI Services	ai application
ModHera: A modular kit for parents to take care babies	Zhang, Liu, Ying, Huang, Yao and Ying (2021)	Services	AI Services	ai application
When Children Program Intelligent Environments: Lessons Learned from a Serious AR Game	Stefanidi et al. (2021)	Services	AI Services	ai application
Draw2Code: Low-Cost Tangible Programming for Creating AR Animations	Im and Rogers (2021)	Services	AI Services	ai application
Information flow and children's emotions during collaborative coding: A causal analysis	Sharma et al. (2021)	Services	AI Services	ai application
"I'm in his belly!": Children's Responses to Different Types of Characters in Virtual Reality	Schloss et al. (2021)	Services	AI Services	virtual reality
"Alexa, Can I Program You?": Student Perceptions of Conversational Artificial Intelligence Before and After Programming Alexa	Van Brummelen et al. (2021)	Services	AI Services	intelligent user interface
Designing A Virtual Talking Companion to Support the Social-Emotional Learning of Children with ASD	Gagan et al. (2022)	Services	AI Services	intelligent user interface
Designing Interactive Contextual Cues for Children's Video-Stimulated Writing	Zarei et al. (2022)	Services	AI Services	ai application
Ready, Set, Art: Technology Needs and Tools for Remote K-2 Art Education	Mansi et al. (2022)	Services	AI Services	ai application
Young Children's Perceptions of Coding and Implications	Yu and Roque (2022)	Services	AI Services	ai application
Eliciting parents' insights into products for supporting and tracking children's fine motor development	Gürbüzsel et al. (2022)	Services	AI Services	ai application

Paper	Author	AI Domain	AI Sub Domain	Keyword
SeaStory: An interactive narrative using collaborative features	Andrade et al. (2022)	Services	AI Services	ai application
Examining Teenagers' Perceptions of Conversational Agents in Learning Settings	Nguyen (2022*a*)	Services	AI Services	intelligent user interface
Language Learning with Mobile Augmented Reality and Artificial Intelligence for Children with Autism Spectrum Disorder	El Shemy (2022)	Services	AI Services	ai application
KidConnect VR: Technology to Stay Connected	Ho et al. (2022)	Services	AI Services	virtual environment
Identifying Features that Characterize Children's Free-Hand Sketches using Machine Learning	Thomas et al. (2022)	Services	AI Services	ai application
What color are the fish's scales? Exploring parents' and children's natural interactions with a child-friendly virtual agent during storybook reading	Lin et al. (2022)	Services	AI Services	intelligent user interface
"Track-track: Let's follow the cat!" Reflecting on children's biometric data processing through a micro puppet show	Sasaki Otani (2022)	Services	AI Services	ai application
Grasping Derivatives: Teaching Mathematics through Embodied Interactions using Tablets and Virtual Reality	Chatain et al. (2022)	Services	AI Services	ai application
Periodic Fable Augmenting Chemistry with Technology, Characters and Storytelling	Câmara Olim et al. (2022)	Services	AI Services	ai application
Understanding Fun in Learning to Code: A Multi-Modal Data approach	Tisza et al. (2022)	Services	AI Services	ai application

Appendix H

PDLS Survey Data

Table H.1: Children's Responses to Survey Questions (scale 1 - 10)

Classmate's Judgment	Own Judgment	Acceptance of System	Trust in System
9	9	7	10
8	9	7	10
9	10	8	7
10	10	10	10
9	9	5	5
6	9	8	7
3	8	7	3
5	5	3	3
8	9	4	5
5	8	4	4
8	10	5	4
8	6	4	4
8	10	9	9
8	7	4	6
9	9	9	9
10	10	8	6
9	9	9	9
9	10	9	9
6	5	4	4
10	10	6	8
10	-	-	-
10	10	7	8

Appendix I

PDLS Validation

Table I.1: PDLS Logged Study Statuses

Start	Duration	Status
Study 166		
Validator 1		
4	16	ENGAGED
20	6	DISENGAGED
26	56	ENGAGED
82	3	DISENGAGED
85	40	ENGAGED
125	2	DISENGAGED
127	125	ENGAGED
252	3	DISENGAGED
255	75	ENGAGED
330	2	DISENGAGED
332	74	ENGAGED
406	3	DISENGAGED
409	19	ENGAGED
Validator 2		
26	400	ENGAGED

Start	Duration	Status
Pupil Observer		
92	331	ENGAGED
Study 171		
Validator 1		
5	140	ENGAGED
145	5	DISENGAGED
150	201	ENGAGED
351	2	DISENGAGED
353	77	ENGAGED
430	3	DISENGAGED
433	92	ENGAGED
525	13	DISENGAGED
538	105	ENGAGED
643	2	DISENGAGED
645	58	ENGAGED
703	2	DISENGAGED
705	50	ENGAGED
755	3	DISENGAGED
758	50	ENGAGED
808	3	DISENGAGED
811	29	ENGAGED
840	5	DISENGAGED
845	123	ENGAGED
968	3	DISENGAGED
971	15	ENGAGED
986	3	DISENGAGED
989	11	ENGAGED
1000	2	DISENGAGED
1002	94	ENGAGED
Validator 2		
2	527	ENGAGED

Start	Duration	Status
529	8	DISENGAGED
537	561	ENGAGED
Pupil Observer		
5	521	ENGAGED
526	4	DISENGAGED
530	564	ENGAGED
Study 172		
Validator 1		
6	15	ENGAGED
21	3	DISENGAGED
24	23	ENGAGED
47	9	DISENGAGED
56	40	ENGAGED
96	8	DISENGAGED
104	46	ENGAGED
150	9	DISENGAGED
159	93	ENGAGED
252	6	DISENGAGED
258	54	ENGAGED
312	4	DISENGAGED
316	30	ENGAGED
346	2	DISENGAGED
348	52	ENGAGED
400	4	DISENGAGED
404	290	ENGAGED
694	3	DISENGAGED
697	70	ENGAGED
767	4	DISENGAGED
771	4	ENGAGED
Validator 2		

Start	Duration	Status
2	44	ENGAGED
46	9	DISENGAGED
55	30	ENGAGED
85	17	DISENGAGED
102	46	ENGAGED
148	8	DISENGAGED
156	79	ENGAGED
235	14	DISENGAGED
249	148	ENGAGED
397	11	DISENGAGED
408	284	ENGAGED
692	13	DISENGAGED
705	62	ENGAGED
767	8	DISENGAGED
Pupil Observer		
0	42	ENGAGED
42	3	DISENGAGED
45	8	ENGAGED
53	3	DISENGAGED
56	65	ENGAGED
121	3	DISENGAGED
124	26	ENGAGED
150	4	DISENGAGED
154	159	ENGAGED
313	6	DISENGAGED
319	26	ENGAGED
345	5	DISENGAGED
350	55	ENGAGED
405	4	DISENGAGED
409	125	ENGAGED
534	3	DISENGAGED

Start	Duration	Status
537	162	ENGAGED
699	3	DISENGAGED
702	70	ENGAGED
Study 173		
Validator 1		
11	27	ENGAGED
38	10	DISENGAGED
48	88	ENGAGED
136	6	DISENGAGED
142	190	ENGAGED
332	3	DISENGAGED
335	130	ENGAGED
465	8	DISENGAGED
473	71	ENGAGED
544	3	DISENGAGED
547	69	ENGAGED
616	3	DISENGAGED
619	43	ENGAGED
662	4	DISENGAGED
666	76	ENGAGED
742	7	DISENGAGED
749	110	ENGAGED
859	3	DISENGAGED
862	79	ENGAGED
941	3	DISENGAGED
944	273	ENGAGED
Validator 2		
2	225	ENGAGED
227	8	DISENGAGED
235	307	ENGAGED
542	3	DISENGAGED

Start	Duration	Status
545	164	ENGAGED
709	7	DISENGAGED
716	76	ENGAGED
792	2	DISENGAGED
794	148	ENGAGED
942	3	DISENGAGED
945	136	ENGAGED
1081	2	DISENGAGED
1083	133	ENGAGED
Pupil Observer		
17	452	ENGAGED
469	334	DISENGAGED
803	93	ENGAGED
896	198	DISENGAGED
1094	119	ENGAGED
Study 196		
Validator 1		
17	17	ENGAGED
34	2	DISENGAGED
36	23	ENGAGED
59	6	DISENGAGED
65	9	ENGAGED
74	6	DISENGAGED
80	35	ENGAGED
115	6	DISENGAGED
121	28	ENGAGED
149	7	DISENGAGED
156	31	ENGAGED
187	3	DISENGAGED
190	41	ENGAGED
231	8	DISENGAGED

Start	Duration	Status
239	58	ENGAGED
Validator 2		
3	51	ENGAGED
54	4	DISENGAGED
58	11	ENGAGED
69	10	DISENGAGED
79	156	ENGAGED
235	2	DISENGAGED
237	59	ENGAGED
Pupil Observer		
0	75	ENGAGED
75	4	DISENGAGED
79	35	ENGAGED
114	4	DISENGAGED
118	175	ENGAGED
Study 212		
Validator 1		
5	6	ENGAGED
11	3	DISENGAGED
14	6	ENGAGED
20	15	DISENGAGED
35	125	ENGAGED
160	6	DISENGAGED
166	163	ENGAGED
329	3	DISENGAGED
332	23	ENGAGED
355	8	DISENGAGED
363	35	ENGAGED
398	5	DISENGAGED
403	329	ENGAGED
Validator 2		

Start	Duration	Status
2	154	ENGAGED
156	3	DISENGAGED
159	194	ENGAGED
353	8	DISENGAGED
361	369	ENGAGED
Pupil Observer		
-403	427	ENGAGED
24	7	DISENGAGED
31	6	ENGAGED
37	9	DISENGAGED
46	26	ENGAGED
72	5	DISENGAGED
77	8	ENGAGED
85	5	DISENGAGED
90	43	ENGAGED
133	14	DISENGAGED
147	11	ENGAGED
158	5	DISENGAGED
163	22	ENGAGED
185	7	DISENGAGED
192	19	ENGAGED
211	6	DISENGAGED
217	14	ENGAGED
231	4	DISENGAGED
235	66	ENGAGED
301	10	DISENGAGED
311	46	ENGAGED
357	16	DISENGAGED
373	26	ENGAGED
399	4	DISENGAGED
403	54	ENGAGED

Start	Duration	Status
457	11	DISENGAGED
468	259	ENGAGED
Study 213		
Validator 1		
8	209	ENGAGED
217	3	DISENGAGED
220	124	ENGAGED
344	4	DISENGAGED
348	37	ENGAGED
385	7	DISENGAGED
392	74	ENGAGED
466	3	DISENGAGED
469	75	ENGAGED
Validator 2		
4	463	ENGAGED
467	4	DISENGAGED
471	73	ENGAGED
Pupil Observer		
2	139	ENGAGED
141	11	DISENGAGED
152	52	ENGAGED
204	26	DISENGAGED
230	21	ENGAGED
251	9	DISENGAGED
260	90	ENGAGED
350	7	DISENGAGED
357	24	ENGAGED
381	40	DISENGAGED
421	32	ENGAGED
453	26	DISENGAGED
479	23	ENGAGED

Start	Duration	Status
502	12	DISENGAGED
514	74	ENGAGED
588	11	DISENGAGED
599	119	ENGAGED
718	24	DISENGAGED
742	182	ENGAGED
924	7	DISENGAGED
931	10	ENGAGED
Study 219		
Validator 1		
8	13	ENGAGED
21	16	DISENGAGED
37	25	ENGAGED
62	7	DISENGAGED
69	36	ENGAGED
105	7	DISENGAGED
112	25	ENGAGED
137	2	DISENGAGED
139	45	ENGAGED
184	4	DISENGAGED
188	44	ENGAGED
232	4	DISENGAGED
236	7	ENGAGED
Validator 2		
4	15	ENGAGED
19	19	DISENGAGED
38	25	ENGAGED
63	6	DISENGAGED
69	35	ENGAGED
104	8	DISENGAGED
112	4	ENGAGED

Start	Duration	Status
116	2	DISENGAGED
118	112	ENGAGED
230	4	DISENGAGED
234	10	ENGAGED
Pupil Observer		
24	7	DISENGAGED
31	62	ENGAGED
93	12	DISENGAGED
105	28	ENGAGED
133	5	DISENGAGED
138	25	ENGAGED
163	3	DISENGAGED
166	6	ENGAGED
172	8	DISENGAGED
180	61	ENGAGED
Study 231		
Validator 1		
12	166	ENGAGED
Validator 2		
5	167	ENGAGED
172	6	DISENGAGED
Pupil Observer		
6	169	ENGAGED
Study 237		
Validator 1		
13	200	ENGAGED
213	3	DISENGAGED
216	114	ENGAGED
330	4	DISENGAGED
334	84	ENGAGED
Validator 2		

Start	Duration	Status
3	78	ENGAGED
81	2	DISENGAGED
83	335	ENGAGED
Pupil Observer		
12	274	ENGAGED
286	39	DISENGAGED
325	91	ENGAGED
Study 238		
Validator 1		
6	429	ENGAGED
435	3	DISENGAGED
438	23	ENGAGED
Validator 2		
2	333	ENGAGED
Pupil Observer		
10	450	ENGAGED
Study 239		
Validator 1		
10	88	ENGAGED
98	8	DISENGAGED
106	456	ENGAGED
562	3	DISENGAGED
565	448	ENGAGED
Validator 2		
1	1013	ENGAGED
Pupil Observer		
9	158	ENGAGED
167	31	DISENGAGED
198	815	ENGAGED
Study 242		
Validator 1		

Start	Duration	Status
6	130	ENGAGED
136	4	DISENGAGED
140	69	ENGAGED
209	2	DISENGAGED
211	73	ENGAGED
284	10	DISENGAGED
Validator 2		
3	88	ENGAGED
91	3	DISENGAGED
94	40	ENGAGED
134	3	DISENGAGED
137	145	ENGAGED
282	12	DISENGAGED
Pupil Observer		
22	269	ENGAGED
Study 243		
Validator 1		
11	431	ENGAGED
Validator 2		
2	312	ENGAGED
314	3	DISENGAGED
317	121	ENGAGED
438	4	DISENGAGED
Pupil Observer		
12	427	ENGAGED
Study 244		
Validator 1		
6	203	ENGAGED
209	6	DISENGAGED
215	84	ENGAGED
299	2	DISENGAGED

Start	Duration	Status
301	395	ENGAGED
Validator 2		
3	695	ENGAGED
Pupil Observer		
9	618	ENGAGED
627	4	DISENGAGED
631	50	ENGAGED
681	13	DISENGAGED
Study 245		
Validator 1		
22	8	DISENGAGED
30	9	ENGAGED
39	2	DISENGAGED
41	24	ENGAGED
65	2	DISENGAGED
67	62	ENGAGED
129	3	DISENGAGED
132	15	ENGAGED
Validator 2		
3	32	ENGAGED
35	2	DISENGAGED
37	112	ENGAGED
Pupil Observer		
16	129	ENGAGED
Study 246		
Validator 1		
10	93	ENGAGED
103	2	DISENGAGED
105	49	ENGAGED
154	3	DISENGAGED
157	63	ENGAGED

Start	Duration	Status
220	2	DISENGAGED
222	64	ENGAGED
286	3	DISENGAGED
289	20	ENGAGED
Validator 2		
4	17	ENGAGED
21	1	DISENGAGED
22	272	ENGAGED
294	13	DISENGAGED
Pupil Observer		
32	265	ENGAGED
297	8	DISENGAGED