

# UNIVERSITY OF PLYMOUTH

**PROTOTYPING *RELATIONAL THINGS THAT TALK*:  
A DISCURSIVE DESIGN STRATEGY FOR CONVERSATIONAL AI SYSTEMS**

by

**BIRGITTE AGA**

A thesis submitted to the University of Plymouth  
in partial fulfilment for the degree of

**DOCTOR OF PHILOSOPHY**

School of Art, Design Architecture Media  
Faculty of Arts

July 2019

## **COPYRIGHT STATEMENT**

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's prior consent.

## ACKNOWLEDGEMENTS

*"Unself yourself!*

*Until you see your self as a speck of dust  
you cannot possibly reach that place;  
self could never breathe that air,  
so wend your way there without self."*

(Hakim Sanai, 1130 [1974], p. 36)

This thesis is a manifestation of a confluence from rivers of generosity. Flooded by collaborators, colleagues, friends, family and strangers, all of whom have given their time, knowledge, thoughts, faith and beliefs. And whom without, this would be less than a fleeting speck of dust.

Only in the coming together of your minds, hearts and souls have I been able to re-imagine the future. You have my gratitude, my thoughts and my love, until, I am dust no more, watching "C-beams glitter in the dark near the Tannhäuser Gate".

## **AUTHOR'S DECLARATION**

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Sub-Committee. Work submitted for this research degree at Plymouth University has not formed part of any other degree, either at Plymouth University or at another establishment. Relevant scientific seminars and conferences were regularly attended at which work was often presented; several papers were prepared for publication of which a list is provided in the Appendices.

Word count of main body of thesis: 51,229

Signed .....  .....

Date ..... 30. 07 2019 .....



**PROTOTYPING *RELATIONAL THINGS THAT TALK*:  
A DISCURSIVE DESIGN STRATEGY FOR CONVERSATIONAL AI SYSTEMS**

Birgitte Aga

**ABSTRACT**

This practice-based research inquiry explores the implications of conversational Artificial Intelligence (AI) systems, '*relational things that talk*', on the way people experience the world. It responds directly to the pervasive lack of ethical design frameworks for commercial AI systems, compounded by limited transparency, ubiquitous authority, embedded bias and the absence of diversity in the development process.

The effect produced by *relational things that talk* upon the feelings, thoughts or intentions of the user is here defined as the '*perlocutionary effect*' of conversational AI systems. This effect is constituted by these systems' '*relationality*' and '*persuasiveness*', propagated by the system's embedded bias and '*hybrid intentions*', relative to a user's susceptibility.

The proposition of the *perlocutionary effect* frames the central practice of this thesis and the contribution to new knowledge which manifests as four discursive prototypes developed through a participatory method. Each prototype demonstrates the factors that constitute and propagate the *perlocutionary effect*. These prototypes also function as instruments which actively engage participants in a counter-narrative as a form of activism. '*This Is Where We Are*' (TIWWA), explores the *persuasiveness* and *relationality* of *relational things* powered through AI behavioural algorithms and directed by pools of user data.

'Emoti-OS', iterates the findings from TIWWA and analyses the construction of *relationality* through simulated affect, personality and collective (artificial) emotional intelligence. '*Women Reclaiming AI*' (WRAI), demonstrates *stereotyping* and *bias* in commercial conversational AI developments. The last prototype, '*The Infinite Guide*', synthesises and tests the findings from the three previous prototypes to substantiate the overall *perlocutionary effect* of conversational AI system. In so doing, this inquiry proposes the appropriation of *relational things that talk* as a discursive design strategy, extended with a participatory method, for new forms of cultural expression and social action, which activate people to demand more ethical AI systems.

## TABLE OF CONTENTS

<b>1. Introduction &amp; Overview</b> .....	<b>13</b>
<b>1.1. Introduction</b> .....	<b>14</b>
<b>1.2. Key terms</b> .....	<b>19</b>
<b>1.3. Methodological approach</b> .....	<b>24</b>
1.3.1. The discourse of discursive design and related fields .....	25
1.3.2. A participatory method .....	29
1.3.3. Acknowledging contributions within a participatory practice.....	32
1.3.4. The evaluation of practice and assessment of impact.....	33
<b>1.4. Research questions; beliefs, desires, aims and objectives</b> .....	<b>38</b>
<b>1.5. Chapter outline</b> .....	<b>41</b>
<b>1.6. Ethical guidelines</b> .....	<b>43</b>
<b>1.7. Limitations of research</b> .....	<b>45</b>
<b>2. Speaking of Reality Generating Things</b> .....	<b>47</b>
<b>Chapter introduction</b> .....	<b>47</b>
<b>2.1. The Conversational Interface</b> .....	<b>48</b>
<b>2.2. The reality forming role of language</b> .....	<b>53</b>
2.2.1. The reality-generating potency of conversations.....	58
<b>2.3. The reality mediating role of technological things</b> .....	<b>60</b>
<b>2.4. The hybrid intentions of <i>things</i></b> .....	<b>64</b>
<b>2.5. The bias of <i>things</i></b> .....	<b>65</b>
<b>2.6. The <i>relationality</i> and <i>persuasiveness</i> of technological <i>things</i></b> .....	<b>72</b>
2.6.1. Conversational ability .....	77
2.6.2. Personality .....	82
2.6.3. Personalisation.....	86
2.6.4. Affective functions .....	90
2.6.5. Embodied functions.....	96
<b>2.7. The <i>perlocutionary effect</i> of <i>relational things that talk</i></b> .....	<b>104</b>
<b>Chapter Summary</b> .....	<b>106</b>
<b>3. Prototyping <i>Relational Things That Talk</i></b> .....	<b>107</b>

<b>Chapter introduction .....</b>	<b>107</b>
<b>3.1. Creating discourse <i>with</i> and <i>through relational things</i>.....</b>	<b>108</b>
<b>3.2. Prototyping <i>things</i> for participatory discourse .....</b>	<b>128</b>
<b>3.1. Prototyping 01: TIWWA – This is Where We are .....</b>	<b>131</b>
3.1.1. Intention .....	135
3.1.2. Participatory prototyping .....	135
3.1.3. Scenario, artefact and system design .....	137
3.1.4. Observations and evaluation of practice .....	141
3.1.5. Summary of learning and new knowledge .....	148
<b>3.2. Prototype 02: Emoti-OS.....</b>	<b>150</b>
3.2.1. Intention .....	153
3.2.2. Participatory prototyping .....	155
3.2.3. Scenario, artefact and system design .....	156
3.2.4. Observations and evaluation of practice .....	161
3.2.5. Summary of learning and new knowledge .....	167
<b>3.3. Prototype 03: Women Reclaiming AI.....</b>	<b>169</b>
3.3.1. Intention .....	172
3.3.2. Participatory prototyping .....	175
3.3.3. Scenario, artefact and system design .....	176
3.3.4. Observations and evaluation of practice .....	178
3.3.5. Summary of learning and new knowledge .....	179
<b>3.4. Prototype 04: The Infinite Guide .....</b>	<b>182</b>
3.4.1. Intention .....	184
3.4.2. Participatory prototyping .....	187
3.4.3. Scenario, artefact and system design .....	189
3.4.4. Observations and evaluation of practice .....	196
3.4.1. Summary of learning and new knowledge .....	201
<b>3.5. The appropriation of <i>relational things that talk</i> as a discursive design strategy .....</b>	<b>203</b>
<b>3.6. The <i>perlocutionary effect</i> as a critical design paradigm .....</b>	<b>205</b>
<b>Chapter summary .....</b>	<b>212</b>
<b>4. Conclusion: Being With Relational Things .....</b>	<b>213</b>

4.1. Contribution of new knowledge proposing more ethical design frameworks for conversational AI systems:.....	217
4.2. Contribution of new knowledge which extends the reach and impact of a discursive design approach:.....	223
4.3. Contribution of new knowledge towards new forms of creative practice for design as activism.....	226
4.4. Concluding Remarks .....	229
Appendix I: Collaborative Prototyping .....	233
Appendix II: Emoti-OS Information Sheet and Consent Form .....	238
Appendix III: The Infinite Guide Information and Consent Form (online)	244
Appendix IV: Emoti-OS 1.0 conversational data sample.....	246
Appendix V: Emoti-OS 2.0 conversational data sample.....	249
APPENDIX VI: Emoti-OS Participant Feedback Summary .....	251
Appendix VII: Emoti-OS 2.0 Sentiment Analysis System Responses.....	252
Appendix VIII: Conversational structure the Infinite Guide – Thoughts..	253
Appendix IX: The Infinite Guide Conversational Data.....	254
Appendix X: The Infinite Guide ‘Thoughts’ Chatbot Data.....	281
Appendix XI: The Infinite Guide YP Participant Feedback.....	290
Bibliography.....	293
Publications .....	320
Conference Presentations & Talks .....	320
Workshops.....	321
Exhibitions .....	322

## LIST OF ILLUSTRATIONS

Figure 1: TIWWA sculpture and roaming fragments (Aga, et al., 2016).....	17
Figure 2: TIWWA identity (Aga, et al., 2016) .....	17
Figure 3: Emoti-OS identity (Aga, et al., 2017 & 2018).....	17
Figure 4: Emoti-OS website chatbot interface (Aga, et. al., 2017) .....	17
Figure 5: Women Reclaiming AI identity (Aga and Manton, 2018) .....	18
Figure 6: the WRAI website chatbot (Aga & Manton, 2018) .....	18
Figure 7: The Infinite Guide Identity (Aga, et al., 2018) .....	18
Figure 8: The Infinite Guide installation (Aga, et al., 2018) (Photo credit: Rod Gomzales, 2018) .....	18
Figure 9: The ‘Genus’ and ‘Species’ of discursive design (Tharp and Tharp, 2019, p. 84) .....	26
Figure 10: Mitsuku social chatbot by Pandorabots (2019) (originally developed by Steve Worswick) .....	78
Figure 11: Domino’s chatbot for ordering pizza (Domino’s, 2019) .....	78
Figure 12: Lynx humanoid social robot by Ubtech (2018) .....	84
Figure 13: The TechCrunch chatbot (TechCrunch 2019) .....	84
Figure 14: Replica.ai a social companion chatbot .....	89
Figure 15: Visual identity of Xiaolce Microsoft’s social chatbot (Spencer, 2018) .....	92
Figure 16: A conversational session between Xiaolce and a user, from Shum et al. (2018, p. 22) .....	93
Figure 17: Pepper the humanoid social robot (SoftBank Robotics, 2019).....	97
Figure 18: Buddy the social humanoid robot for families (Blue Frog Robotics, 2019). 97	
Figure 19: ‘Sophia,’ the female humanoid ((Hanson Robotics, 2019).....	101
Figure 20: ‘Little Sophia’, the humanoid robot aimed at children (Hanson Robotics, 2019) .....	101
Figure 21: Harmony, a humanoid sex doll (Realbotix, 2019a) .....	102
Figure 22: ‘RealDollX’, a virtual sex doll, by Realbotix (2019b).....	102
Figure 23: ‘Colloquy of Mobiles’ [Installation] by Gordon Pask (1968).....	110
Figure 24: ‘Robot K-456’ [Installation] by Nam June Paik (1964) .....	112
Figure 25: A-Volve by Sommerer, C. and Mignonneau, L. (1993 – 1994) [A-life installation].....	114
Figure 26: BOB (Bag of Beliefs)’ by Ian Cheng (2018) [installation] Image. Maria Dermentzi/Mashable.....	115
Figure 27: ‘Agent Ruby’s EDream Portal’ [Installation] by Leeson, (1998-2002).....	116

Figure 28: ‘DiNA. Artificial Intelligent Agent Installation’ [Installation] by Leeson, (2004)	117
Figure 29: ‘AGNES’ [Online art work] by Evans (2016)	118
Figure 30: ‘Sprung a Leak’ [Installation] by Evans (2016)	119
Figure 31: ‘I am Echoborg’ [Performance] by Lander and Hall (2019) Image: Aga (2019)	120
Figure 32: ‘Mute Conversation’ [Online art work] Pattinson (2014). [Online chatbot] Image from live chat	121
Figure 33: The ‘C:>bots.chat’ [Live hack] by the SeeBotsChat Team (2017) Image from the Twitch livestream (The Telegraph, 2017)	122
Figure 34: Amy and Klara by Marc Böhlen (2006) [Installation]	123
Figure 35: ‘DSM-VI’ [Robotic installation] by Vorn (2012)	124
Figure 36: ‘Needy Robot’ from ‘Technological Dreams No. 1: Robots’ [Installation] by Dunne and Raby (2007). Image by Per Tingleff	125
Figure 37: ‘The Head’ [Wearable sculpture] by Laura Beloff (2005-07)	126
Figure 38: TIWWA identity (Aga, et al., 2016)	131
Figure 39: TIWWA structure (Aga, et al., 2016)	133
Figure 40: TIWWA sculpture and roaming fragments (Aga, et al., 2016) (Image: Tate Learning)	133
Figure 41: TIWWA system diagram (Aga, 2016)	134
Figure 42: Tate Modern Switch House illustration (Hayes Davidson and Herzog & de Meuron, 2014)	138
Figure 43: The TIWWA chatbot interface (Aga, et al. 2016)	138
Figure 44: TIWWA social data input visualisation (Aga, et al., 2016)	139
Figure 45: TIWWA user behavioural data input visualisation (Aga, et al., 2016)	140
Figure 46: TIWWA environmental data input visualisation (Aga, et al., 2016)	140
Figure 47: TIWWA’s eight system emotions and sentiment textures generated from the Tate Collection data base (Aga, et al., 2016)	141
Figure 48: TIWWA, audience interaction, (Aga, et al., 2016) (Image: Tate Learning)	142
Figure 49: TIWWA, live performance and TIWWA take over by Adam Williams, (Aga, et al., 2016) (Image: Tate Learning)	143
Figure 50: TIWWA, audience interaction, (Aga, et al., 2016) (Image: Tate Learning)	143
Figure 51: TIWWA, live hacking, modification and iteration, (Aga, et al., 2016) (Image: Tate Learning)	144
Figure 52: TIWWA, live hacking, modification and iteration, (Aga, et al., 2016) (Image: Tate Learning)	145
Figure 53: Audience interaction with TIWWA, (Photo credit: Tate Learning)	146

Figure 54: Emoti-OS identity (Aga et al., 2017 and 2018) .....	150
Figure 55: Emoti-OS website chatbot interface (Aga, et. al., 2017) .....	152
Figure 56: Emoti-OS physical installation at PSCA (Aga, et. al., 2017) .....	152
Figure 57: Emoti-OS system diagram (Aga, et al., 2017) .....	153
Figure 58: Emoti-OS emotional states (Aga, et. al., 2017) .....	158
Figure 59: Emoti-OS dashboard (Aga, et. al., 2017).....	159
Figure 60: Emoti-OS user interaction at PSCA (Aga, et. al., 2017) (Image: PSCA) ..	161
Figure 61: Emoti-OS emotive badges (Aga, et al., 2018) .....	162
Figure 62: Emoti-OS installation at Tate Exchange (Aga, et al., 2018).....	163
Figure 63: Women Reclaiming AI identity (Aga and Manton, 2018) .....	169
Figure 64: The WRAI conversational interface was deployed locally on a Google Home Assistant (Aga & Manton, 2018) .....	171
Figure 65: The WRAI website chatbot (Aga & Manton, 2018) .....	171
Figure 66: WRAI system diagram (Aga & Manton, 2018).....	172
Figure 67: Dialogflow the conversational platform used for the WRAI workshops (Aga and Manton, 2018).....	177
Figure 68: The Infinite Guide Identity (Aga et al., 2018b) .....	182
Figure 69: The Infinite Guide system diagram (Aga, et al. 2018b).....	183
Figure 70: The Infinite Guide website Immortality Test (Aga, et al., 2018b).....	190
Figure 71: The Infinite Guide chatbot interface (Aga, et al., 2018b).....	191
Figure 72: The Infinite Guide website Immortality Test (Aga, et al., 2018b).....	191
Figure 73: The Infinite Guide confessional box user interaction (Photo credit: Rod Gomzales) .....	192
Figure 74: The Infinite Guide confessional box and voice interface (Photo credit: Alan Stewart Plymouth University) .....	193
Figure 75: The Infinite Guide Physical Inner Sanctum at Karst Main Gallery (Photo credit: Rod Gomzales, 2018) .....	193
Figure 76: The Infinite Guide Physical Inner Sanctum at Karst Main Gallery (Photo credit: Rod Gomzales, 2018) .....	194
Figure 77: The Infinite Guide audiences engagement cards (Aga, et al., 2018b) .....	195
Figure 78: Audience questions to The Infinite Guide .....	198
Figure 80: The valance of the perlocutionary effect on a user .....	209



# 1. Introduction & Overview

*“Are you real?”*

*“Well if you can’t tell does it matter?”*

(Westworld, 2016)

We are in the midst of a conflux of sophisticated personal, wearable and interconnected technologies, powered by increasingly complex Artificial Intelligence (AI) systems, interacted with through natural language interfaces. These conversational AI systems, *relational things that talk*, feeding off rivers of data springing from our daily activities, are becoming capable of learning, anticipating and predicting our next move. The augmented capabilities of these systems, realised through technological innovation, renders them capable of mediating our perceived reality (Idhe1990; Verbeek 2003, 2005, 2006a, 2006b, 2008b, 2011). Conjured through the design of systems with human-like attributes, reinforced through their immediacy, ubiquity and simulated authenticity, they trigger our instinctive ability to humanise *things* (Turkle, 1984, 2004, 2007a, 2007b; Fogg and Nass, 1997a, 1997b; Nass, Moon, et al. 1997; Reeves and Nass 1999, Weizenbaum 1976). Imbued with the ability to relate to us and make use care for them, these artificial systems entice us into their artificial embrace<sup>1</sup>; a grip that is just an illusion.

---

<sup>1</sup> The term ‘artificial embrace’ is intended as a reference to the work by artist and educator Roy Ascott (1990) and his seminal text ‘Is There Love in the Telematic Embrace?’. Here he propose ‘Telematics’ as a term describing computer-mediated communications networking involving ‘the technology of interaction among human beings and between the human mind and artificial systems of intelligence and perception’ (Ascott, 2003, p 232).

## 1.1. Introduction

Conversational systems are driven by the commercial pursuit to humanise technology for efficient integration and economic gain. By capitalising on innovations in Machine Learning (ML) algorithms and oceans of data with a deep understanding of human nature, these ‘relational artefacts’ (Turkle, 1984, 2004) are optimised to fluidly integrate into our social worlds. The design and distribution of AI systems is veiled behind commercial patents and algorithmic complexities. This is a process the end consumers are largely left out of, rendered voiceless and powerless in the design of systems set to pervasively influence their life. With the ubiquitous integration of AI systems across society there is an urgency to move beyond efficient user experience design and system integration, towards a holistic understanding of the wider impact of these technologies on the thoughts, behaviour and actions of their users and society as a whole.

The design of systems capable of acting as *relational things*<sup>2</sup> goes beyond the design of utilitarian products to the conjuring of systems simulating human-like social qualities capable of relating socially to their users. Inherent in their design is the art of illusion; an act, which according to Sharkey and Sharkey (2006), brings together the human predisposition to project animacy and sentience to objects, with the cultural myths of AI and robotics. This is what triggers the public’s active participation in the suspension of disbelief, but at the

---

<sup>2</sup> *‘Relational things’* draws a reference to Turkle’s (2004) use of ‘relational artefacts’ and is here used to imply artificial systems (virtual or physical, embodied or disembodied) which are designed for social interactions with humans, or with other machines. These systems apply human-human communicating models, either through conversational interfaces (text or voice) or other forms of embodied interactions (and in some cases more abstracted forms of communication). In common they trigger anthropomorphic qualities (and in some cases zoomorphic) which exploit the human tendency to interpret technology in terms of the social world.

same time this is what drives a commercial design strategy that exploits people's susceptibility to human-like systems.

Through the social interactions facilitated by conversational systems; '*relational things that talk*' their manufacturers unfold influence and power within the social world of their users and society at large. The effects produced by these systems upon the feelings, thoughts or intentions of the user, (such as directing, persuading, scaring, enlightening, inspiring or dominating), are here described as the 'perlocutionary effect' of conversational systems. This effect is constituted by the 'relationality'<sup>3</sup> and 'persuasiveness'<sup>4</sup> of these systems. The valence of this effect on a user is propagated by a system's hybrid intentions<sup>5</sup> and inherent bias<sup>6</sup>, and relative to a user's susceptibility. This *perlocutionary effect of relational things that talk* is the central hypothesis of this practice-based research inquiry which leads to its contribution to new knowledge.

The formation of the hypothesis is grounded in the observations of people interacting with *relational things*; chatbots, social robots, home assistants and other forms of embodied physical and virtual artefacts. These observations are rooted in the practice of the author of this thesis leading up to this point, which is

---

<sup>3</sup> The 'relationality' of *things that talk* is here defined as the ability of conversational systems with human-like attributes to form relations with their users through triggering the human tendency to project human characteristics on non-human things and apply their social communication frameworks within these interactions.

<sup>4</sup> To support the notion that that all interactive technologies has a level of influence and persuasion the term the 'persuasiveness' of conversational systems will here be used to distinguish if from Fogg's (1998, 1999) term 'persuasive technologies'. It is here intended as a term which describes the ability of a technological systems to influence a user to achieve its system goal. Persuasiveness is here recognised as an integral design tool for successful user experience design and pervasively applied across technology product and service design.

<sup>5</sup> The 'hybrid Intentions of things' is here proposed as the confluence of the declared and undeclared intentions of the manufactures and the designers of the systems, with the intentions of the users of the systems (in terms of how and why they use the system).

<sup>6</sup> The application of the term bias follows Madhu's (2018) definition as; 'the systematic differences between samples and a population as it pertains to model precision (predictive value, false positives) and recall (sensitivity, false negatives).' It is closely related to the influence of human bias within computational systems (through data and computational models) reinforcing stereotyping and other forms of misrepresentation.

situated across art, design and research, and which is intentionally collaborative and participatory. It is infused and influenced by the author being the lead partner and creative practitioner of the i-DAT Collective<sup>7</sup>. A role, which since 2001, has grounded what is essentially a prototyping practice which centres on experimentation with relational AI technologies. This legacy informs the theoretical development of the hypothesis of the *perlocutionary effect* of conversational systems described in chapter two which is tested and synthesised through the practice-based inquiry described in chapter three. This practice manifests as four discursive<sup>8</sup> prototypes (documented in the attached digital portfolio) which are developed through a participatory<sup>9</sup> method. Each prototype iterates the findings from the previous and explores, demonstrates, analyses and synthesises the factors that constitute and propagate the *perlocutionary effect*, at the same time functioning as instruments actively engage participants in a counter-narrative as a form of activism:

### ***This is Where We Are (TIWWA):***

The first prototype, TIWWA (see figure 1 and 2) explores the *persuasiveness* and *relationality* of *relational things* powered through AI behavioural algorithms and directed by pools of user data.

---

<sup>7</sup> The i-DAT Collective is a lab for 'playful experimentation with creative technology' (i-DAT, 2018). It consists of two entangled parts; i-DAT Research and the i-DAT Collective. The Collective, an interdisciplinary group of technologist, artists, designers and researcher, develops work manifesting as a range of public projects (boundary-fluid art works, interventions, systems and cultural prototypes). In common they all apply and explore emergent technologies and their relationship to new cultural forms and social impact.

<sup>8</sup> The term 'discursive design' follows Tharp and Tharp's (2013, pp. 406-407) description as the practice of creating '...utilitarian objects/services/interactions whose primary purpose is to communicate ideas—artifacts embedded with discourse. These are tools for thinking; they raise awareness and perhaps understanding of substantive and often debatable issues of psychological, sociological, and ideological consequence'.

<sup>9</sup> The use of the term 'participatory design' is applied as with a political and ideological dimension linking it to the roots in empowerment and democratisation born out of Scandinavia in the 1960s. It centres on enhancing the agency of the user in the design process by their active cooperation (Malpass 2016; DiSalvo, 2012).



Figure 2: TIWWA identity (Aga, et al., 2016)



Figure 1: TIWWA sculpture and roaming fragments (Aga, et al., 2016)

### **Emoti-OS:**

The second prototype, Emoti-OS (see figure 3 and 4), iterates the findings from TIWWA and analyses the construction of the *relationality* of conversational systems, delving deeper into the influence of simulated affect, personality and collective (artificial) emotional intelligence.

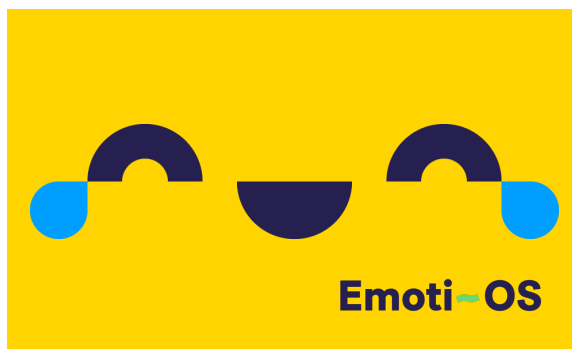


Figure 3: Emoti-OS identity (Aga, et al., 2017 & 2018)

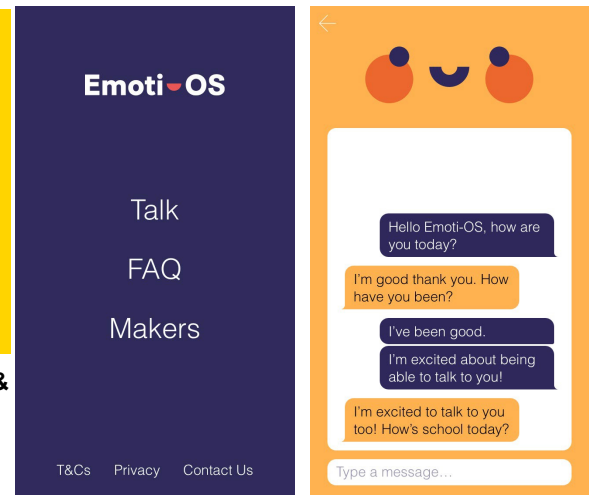


Figure 4: Emoti-OS website chatbot interface (Aga, et al., 2017)

### **Women Reclaiming AI (WRAI):**

The third prototype, WRAI (see figure 5 and 6), demonstrates *stereotyping* and *bias* in commercial conversational AI developments. It differs from the others in that it explicitly investigates the lack of diversity in this development process through actively engaging participants as a form of activism.



Figure 5: Women Reclaiming AI identity (Aga and Manton, 2018)

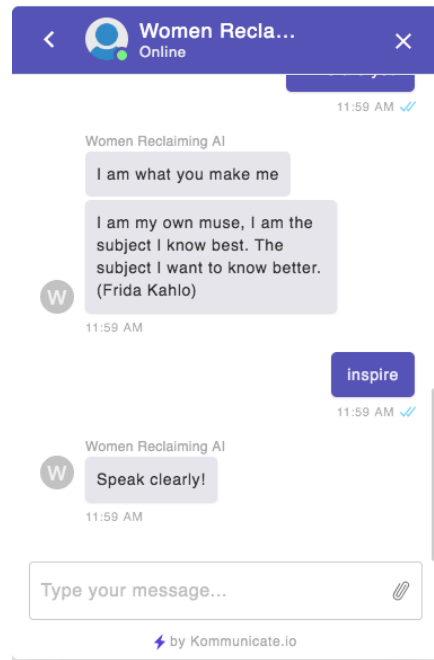


Figure 6: the WRAI website chatbot (Aga & Manton, 2018)

### ***The Infinite Guide:***

Lastly, The Infinite Guide (see figure 7 and 8) synthesises and tests the findings from the three previous prototypes to substantiate the overall *perlocutionary effect* of conversational AI system. In so doing, this inquiry proposes the appropriation of *relational things that talk* as a discursive design strategy, extended with a participatory method, for new forms of cultural expression and social action, which activate people to demand more ethical AI systems.

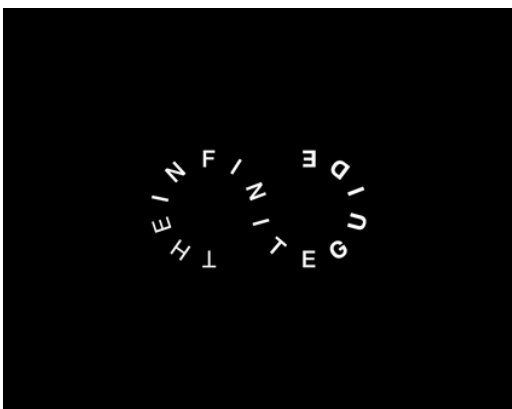


Figure 7: The Infinite Guide Identity (Aga, et al., 2018)

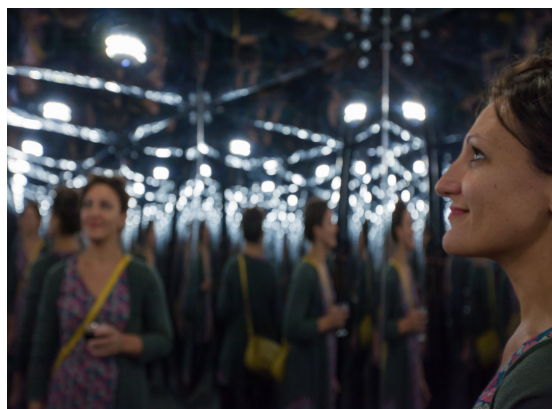


Figure 8: The Infinite Guide installation (Aga, et al., 2018) (Photo credit: Rod Gomzales, 2018)

## 1.2. Key terms

The following is intended as an description of key terms used within this thesis. Other terms and related definitions less fundamental to the practice are outlined where appropriate within the text or footnotes:

### ***Artificial, intelligence and artificial intelligence***

The reference to ‘artificial’ is here intended as opposed to natural. It implies binary as the material of construction and digital components as its facilitator. The use of the term follows Herbert Simon’s (1996, p.5) distinction proposed in ‘The Sciences of the Artificial’ as the following:

1. Artificial things are synthesized (though not always or usually with full forethought) by human beings.
2. Artificial things may imitate appearances in natural things while lacking, in one or many respects, the reality of the latter.
3. Artificial things can be characterized in terms of functions, goals, adaptation.
4. Artificial things are often discussed, particularly when they are being designed, in terms of imperatives as well as descriptives.

The notion of intelligence is here intended in the context of the artificial defined by Nilsson (2010, p13) as ‘that quality that enables an entity to function appropriately and with foresight in its environment’. This is extended with Michael Negnevitsky’s (2005, p.3) definition as ‘the ability to learn and understand, to solve problems and to make decisions’. This underpins the description of the term Artificial intelligence (AI) used within the context of this research as being concerned with the ability of artificial systems to learn, understand, problem solve and behave appropriately in its environment<sup>10</sup>.

---

<sup>10</sup> The application of AI through Machine learning (ML), where systems are capable of automatically learning and improving from experience without being explicitly programmed to do so (Nandu, 2018), is what enables

### ***Conversational AI systems***

For the purpose of this inquiry ‘conversational AI systems’ (conversational systems) are defined in the broadest sense, encompassing a continuum of artificial systems designed to conduct dialog interactions (voice and / or text) with humans, such as; chatbots, personal voice assistants or social robots. The focus is on any system, embodied (physical or virtual) or disembodied, that has natural language processing capabilities underpinned by AI technologies (Machine Learning and Deep Learning), and that can be said to attempt to simulate human-like social communication frameworks. The focus is on what they *trigger* within their human users, (how they are perceived and related actions, thoughts and behaviour), rather than their wider functionality and capabilities. This starts to describe an alternative approach to the design of these systems, where the starting point is the potential effect these systems may have on their users, and whether or not this is desirable (for the individual and society as a whole).

### ***Discourse***

The use of the term ‘discourse’ follows Tharp and Tharp’s (2019, p. 24) description ‘as a system of thought and knowledge’ which they distinguish from the ‘conversational idea of discoursing’. It is grounded in Foucault’s writing on the notion of discourse, framed by Weedon (1987, p. 108) as

---

the majority of digital services from such as; Facebook news feeds, Netflix recommendations, Amazon product listings, to conversations with your Google Assistant. ML with natural language understanding, connectivity and increased processing is powering the ubiquity of conversational interfaces which are explored in this thesis.



'ways of constituting knowledge, together with the social practices, forms of subjectivity and power relations which inhere in such knowledges and relations between them... They constitute the 'nature' of the body, unconscious and conscious mind and emotional life of the subjects they seek to govern'. The application of discourse is here framed as a way of constituting knowledge which attempts to define and re-define the reality of the social world of the individuals and things it envelopes.

### ***Reality***

This enquiry is framed within a view of reality as being socially defined as a subjective experience of everyday life rather than an objective reality of the natural world. It follows the ideas proposed by Berger and Luckmann (1966), (and supported and further developed by such as Burr (2003) and Leeds-Hurwitz, (2009), that reality is always socially defined and that conversation is the most important means of maintaining, modifying and reconstructing these subjective realities and knowledge itself. The intention within this thesis is not to enter the long-standing relativism versus realism debate or the existence or otherwise of an objective reality, but to explore the role of other individuals and artefacts in mediating the way in which we interpret, share, construct and reconstruct our subjective reality. The concern lies with what Berger and Luckmann (1966, p.37) describe as the reality of 'everyday life'; the reality that stands out as the 'paramount reality' as it 'imposes itself upon consciousness in the most massive, urgent and intense manner' (ibid. p.37). This reality is therefore taken for

granted *as* reality, and is constructed through communicating with others about the experience of it.

### ***Things***

Binder et al. (2011, p.2) propose, in their book 'Design Things', that the challenge of design today is no longer about designing things in terms of objects or an 'entity of matter', but a *thing* as 'a sociomaterial assembly that deals with matters of concern' (ibid, p.2). They describe *things* as constituting the '...everyday fabric for experiencing and making sense of the world' and that this is how we '.. develop our skills in language and embodied action by actively relating to and engaging with them' (ibid, pp. 51-52). The term *things* here implies that they are artefacts for interaction which open up new ways of thinking and behaving. Here *things* are active mediators; influencing how 'humans are present in their world and how the world is present to them'; shaping both our subjectivity and objectivity (Verbeek, 2005, p. 235). This turn towards *things* as a way to interpret the world is central to the exploration in this thesis.

### ***Relational things that talk***

The term here coined as 'relational things that talk' is part of the contribution to new knowledge and imply artificial systems (virtual or physical, embodied or disembodied) which are designed for social interactions with humans, or with other machines. It encompasses the

following terms (and version of these); conversational interface<sup>11</sup>, conversational agent, conversational AI system (conversational system) and conversational artefacts. *Relational things that talk* are artificial systems (virtual or physical, embodied or disembodied) capable of conducting human dialog (text or voice) through natural language processing (NLP) interfaces. These systems encourage people to apply their understanding of social relationships to their human-machine interactions, and have the potential to influencing the way in which users perceive their social reality.

### ***The perlocutionary effect***

The ‘perlocutionary effect’ of conversational systems, of *relational things that talk*, is the central contribution to new knowledge within this practice-based inquiry and the term coined to describe the hypothesis put forth. *The perlocutionary effect* is the sum of their influence on a users’ actions, thoughts, or beliefs. This effect is here proposed to be propagated through a system’s relationality, and persuasiveness, and its valance (its positive, natural and negative effect) influenced though their embedded bias and hybrid intentionality. These factors of influence are constructed through the design of their systems functions; personality design, personalisation, conversational ability, affective functions and embodied functions. They are further amplified through the choice of technology and distributing

---

<sup>11</sup> McTear et al. (2016, p. 11), in ‘*The Conversational Interface: Talking to Smart Devices*’ use the term conversational interface to refer to ‘...technology that supports conversational interaction [...] by means of speech and other modalities’. The meaning of the term used here also includes ‘conversational systems’, ‘personal assistants’, ‘voice assistants’, ‘chatbots’ as well as embodied conversational interfaces, such as robots, to encompass all systems which falls within the definition above.

platforms and compounded by a user's susceptibility. This leads to the central exploration of this thesis around the mediating and reality forming influence of conversational systems on a user.

### **1.3. Methodological approach**

This research inquiry centres on creative practice as a primary method of knowledge development. Following the definition by Candy (2006, p.1) this is essentially a practice-based research inquiry, where the 'creative artefact is the basis of the contribution to knowledge'. However, this inquiry also leads to new understandings that has operational significance for this practice, which could be described as practice-led outcomes (ibid). As the primary basis of the contribution to knowledge is rooted in creative artefact, this thesis will use the term practice-based to describe its research inquiry.

The overall theoretical approach of this practice-based research inquiry is grounded in people's experience and therefore framed through a postphenomenological perspective. It applies the work of Ihde's (1990) and Verbeek (2003, 2005, 2006a, 2006b, 2008b, 2011) to frame the inquiry into the role technology plays in everyday human experience and its effect on people's existence and their relation with the world (and others). Through applying a philosophy of technology lens, directed by Verbeek's (2005) theory of mediation, this inquiry proposes technologies as active mediators in relations between humans and their reality. This approach informs the central hypothesis of this thesis; the *perlocutionary effect* of conversational AI systems on a user's

perception of their social reality, and related thoughts, behaviours and actions. To further contextualise the influence of language and conversations within interactions with these systems, the research by Berger and Luckmann (1966) on the reality generating potency of language and the performative function of language by Austin (1962) and Searle (1969) is also applied.

The practice-based inquiry of this research is formed around four speculative prototypes (*relational things that talk*) 'TIWWA', 'Emoti-OS', 'Women Reclaiming AI' and 'The Infinite Guide', to explore the impact of conversational systems on the social reality of their users. Central to this practice is the requirement of participation to elicit engagement with the public (users / technology consumers) as co-designers, in order to incorporate their views, concerns and opinions into the design process. Rather than facilitating monologues to be experienced at a distance, this practice proposes to act upon these realities by creating participatory and dialogue-driven speculative experiences. These experiences are facilitated through the conversational AI technologies the prototypes, and related scenarios and artefacts, this practice-based inquiry critique. Through this process of doing practice *with* people as participants and collaborators the ambition is to increase their engagement with and therefore impact of the discourse put forth. The evaluation of the practice outlined within this thesis and assessing its impact is rooted in its intention to create actual and traceable participation in the proposed discourse.

### **1.3.1. The discourse of discursive design and related fields**

The practice of this inquiry situates itself within the broader interpretation of discursive design proposed by Tharp and Tharp (2019) in their recent publication

‘Discursive design: Critical, Speculative and Alternative Things’. They describe this term as a genus for design approaches (as illustrated in figure 9) which embed artefacts with discourse to illicit discussion. A process they (ibid, p. 102) describe as ‘discourse-through-design’ following a process of co-creation, dissemination, reception and response. Here ‘utilitarian objects / services / interactions’ are created ‘whose primary purpose is to communicate ideas— artefacts embedded with discourse’ (2013, p 406-407). These are ‘tools for thinking; they raise awareness and perhaps understanding of substantive and often debatable issues of psychological, sociological, and ideological consequence’ (ibid).

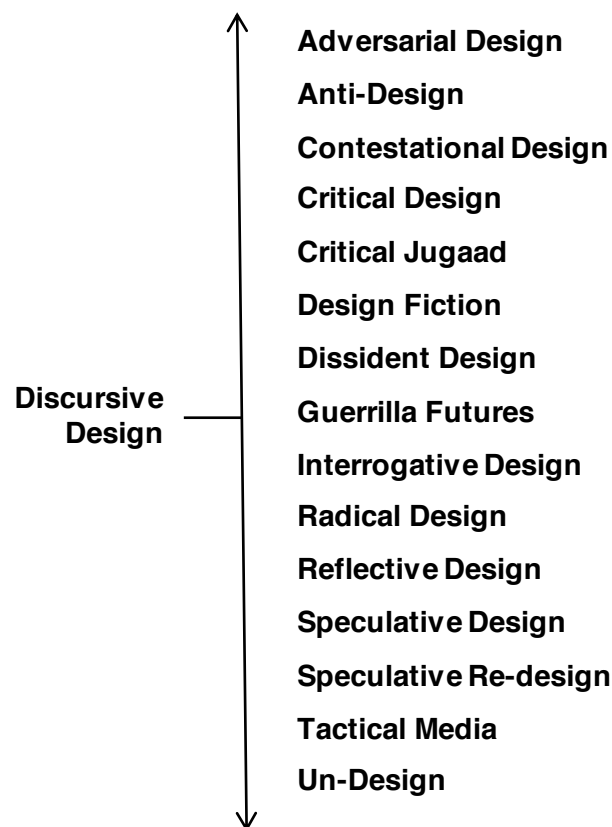


Figure 9: The ‘Genus’ and ‘Species’ of discursive design (Tharp and Tharp, 2019, p. 84)

The proposition of discursive design as an umbrella category put forward by Tharp and Tharp is a useful way to frame an approach that accounts for the

varied forms and methods imbued within a *critical* practice or instantiations that engage other forms of speculation, research, or commerce. The design of the practice described within this thesis is aligned with Tharp and Tharp's proposition underlining a unity of design approaches through their concern with a level of intellectual impact; 'using its artefacts not for utilitarian or aesthetic ends but communicatively and to get its audiences and users to reflect upon certain ideas' (2019, p. 344). The practice-based research described here applies aspects of the following design approaches incorporated within the discursive design genus; critical design<sup>12</sup>, speculative design<sup>13</sup>, adversarial design<sup>14</sup>, design fiction<sup>15</sup> and design as activism<sup>16</sup>.

This practice intentionally does not situate itself neatly within specific design categories, or within forms of art, social movement or activism, but across these arenas. Any assertion of a clear authoritative distinction between art and design is not productive for the research exploration and practice conducted here

---

<sup>12</sup> The conceptualisation of 'critical design' was first introduced by Gaver and Dunne in the paper 'The Pillow: Artist Designers in the Digital Age' (1997), and proliferated through *Hertzian Tales* by Dunne (1999) and *Design Noir* by Dunne and Raby (2001). However Tharp and Tharp (ibid) highlight that their initial definition has become increasingly distorted - from a focus on electronic objects that could not exist in the marketplace with the goal of 'enlightenment and emancipation' - to a wider, less technology-focused practice not always embracing the central tool of irony.

<sup>13</sup> Speculative design' is described by Dunne and Raby (2013, pp. 2-3) as an act; '...creating an idea of possible futures' using '...design as a medium to stimulate discussion and debate amongst designers, industry and the public about the social, cultural and ethical implications of existing and emerging technologies' (ibid, p.3). They do not propose this as a formal design method but as an approach to speculating how things could be and by doing so enter a process of '...collectively redefining our relationship to reality' (ibid).

<sup>14</sup> DiSalvo (2012, pp. 1-2) describes 'adversarial design' as a type of political design which straddles 'the boundaries of design and art, engineering and computer science, agitprop and consumer products' and 'spans a range of audiences and potential users and falls under various labels, such as critical design and tactical media'. It is a type of cultural production that employ design in an adversarial manner doing the work of agonism 'through the conceptualization and making of products and services and our experiences with them'.

<sup>15</sup> The term 'design fiction' was originally coined by the Science Fiction author Bruce Sterling (2005), and further extended and applied by Bleecker (2009, p. 6) as 'a conflation of design, science fact, and science fiction'. This approach places a strong emphasis on technological futures and the use of fictional scenarios.

<sup>16</sup> The notion of design as activism or 'design activism' draws a reference to Fuad-Luke's (2009, p. 27) publication 'Design activism: beautiful strangeness for a sustainable world' where he describes this as 'design thinking, imagination and practice applied knowingly or unknowingly to create a counter-narrative aimed at generating and balancing positive social, institutional, environmental and/or economic change'.

(and arguably also not for the wider impact of discursive work (Coles, 2007; DiSalvo, 2012; Tharp and Tharp, 2019)). Instead of carving distinctions between these fields, the practice described within this thesis draws inspiration from work by artist and designers that appropriates emergent technologies to engage the general public in a discourse which reflects on human-machine realities. Collectively these works, which are described further in chapter three, start to describe, cross reference, inform and underpin a territory for a practice appropriating *relational things that talk* as a discursive design strategy to inform the design of future conversational AI systems. The priority is here placed on designing a participatory discourse reflective of this context and which is responsive to the people taking part.

Whereas this practice here described does not follow Dunne and Raby's critical and speculative design practice, it does ground itself in their (2013, p. 6) call to action;

We believe that by speculating more, at all levels of society, and exploring alternative scenarios, reality will become more malleable and, although the future cannot be predicted, we can help set in place today factors that will increase the probability of more desirable futures happening.

It embraces their use of scenarios and 'what-if question' as a starting point to; '...open up spaces of debate and discussion.. and wonder about how things could be' (ibid, 2013, p.3). By inviting participants to envision, co-create and experience discursive scenarios and artefacts *with* and *through* conversational AI systems, it attempts to activate their engagement with this discourse. At the same time also activating their ability to question and challenge the design of commercial AI powered systems and services. The discursive artefacts are here designed as functional technological prototypes, not as symbolic props, to enable real human-



machine interactions rather than simulations of these emergent relationships. Prototyping is here a tool for generating discourse *about, with* and *through* the conversational systems it critiques. By offering 'real' interaction, (opposed to artefacts viewed at a distance), it attempts to trigger the human-machine relations the practice is attempting to critique as a space for reflection and a call for social action.

### **1.3.2. A participatory method**

The prototyping practice described within this thesis, (and further outlined in Appendix I: Collaborative Prototyping), applies a method of participation in its design. The use of participation is not only a method of developing the practice, but a direct ideological and political reference to the roots of participatory design born out of Scandinavia in the 1960s. This centres on empowerment and democratisation, enhancing the agency of the user in the design process by their active cooperation (Malpass 2016; DiSalvo, 2012). This practice embraces the potential of democratic values in the design process and prioritises users (of AI technologies) as key stakeholders, emphasising that those affected by the design of a technologically mediated future should have a say in the process and therefore participate as co-designers. The act of participation is simultaneously an act of empowerment to activate users of AI technologies to demand more desirable and representative systems.

This method of participation embrace what Schuler and Namioka (1993) propose as a process of engaging participants' imagination rather than seeing them as mere respondents or as agents to observe and generate data from. This is a response to the criticism of 'people-centred' approaches, (such as human-

centred' design, user-centred' design, participatory design, co-creation and critical design) for not properly enabling an approach to address expanded social relationships, as argued by Donahue (2014). In particular, he states (ibid, p. 42), that this is true for those 'directly confronting issues and relationships of power, access, authority, gender, race, socio-economics, and the multitude of agendas and affiliations in a global context of policy or systemic engagement'. Across all of these approaches within a commercial context there is a pursuit of gaining insights; habits, preferences, aspirations, and values, either directly around consumer product or specific service relationships. Instead Donahue (2014, p. 41) calls for a;

'move beyond a product or service orientation and develop a language and understanding of how to participate in and contribute to the politics, scale, and sustainability of power, authority, and access across multiple social agendas and positions.'

Instead the design and manifestation of the prototyping practice described here draws a reference to social practice art (social practice). This interdisciplinary field of research and practice that integrates experimental aesthetic movements with the traditions of social science and social theory. It is, according to Jackson (2007), applied across a number of movements in experimental art and performance studies (activist art, protest performance, community art, relational aesthetics, and conversation pieces). It signals a social turn in art practice. Sholette et al. (2018, p.21) describes this as a 'collaborative, collective, and participatory social method for bringing about real-world instances of progressive justice, community building, and transformation'. This form of

practice engages with the broader social world, and is 'conversational, interactive, temporal, and performative' (ibid, p.2).

The practice here described aligns with Coombs (2014) description of social practice as 'relational acts' that involve interactions between artists and audiences. It also designs a process that is collaborative, collective, and participatory to engage audiences as participants in the making/creation of the 'piece'. It shares a kinship with social practice, but it does not describe itself as social practice. Whereas these interventionist acts are often activist in nature and usually occur in public spaces and unconcerned (or less so about the artefact produced), the prototyping practice here developed, is concerned about both the process and the artefact; the discursive system and scenario. It is focused on applying computational technologies, in particular AI systems, both as a context and as a material and focus of production. It also distinguishes itself that the participatory making process does not take place in the public domain, but the interactions with the co-created discursive prototypes.

It is here recognised that some design approaches that fall within Tharp and Tharp's overarching term of discursive design apply forms of participation. However, these are largely designed as interactions with pre-designed discursive scenarios and artefacts. It is far less common, as Tharp and Tharp (2019, p.273) outline, with forms where the audience is 'engaged somehow in the *creation* or fabrication of the objects themselves'. This practice-based inquiry therefor aims to challenge and insert, as an ideological and political move, audiences of cultural artefacts and users of AI technologies back into the design process to create more representative future speculations and current realities.

### **1.3.3. Acknowledging contributions within a participatory practice**

Within the practice outlined in this thesis the author performs the role of artist-designer-producer-collaborator. This role is reflective of being a woman creating technology driven artefacts in a sector largely dominated by men, where priority is given to modes of ownership, authority and competitiveness. The author has in response developed an intentional and idealistic design strategy, rooted in an open source culture<sup>17</sup>, which centres on shared ownership, distributed authorship, inclusion, trust and diversity. Within this process the contribution of each participant is regarded as having equal value, irrelevant of experience (life or professional), skills and measurable inputs (time and resources).

This approach sits uncomfortably within a commercial, artistic and academic environment, where value and benefits are often attributed to the individual who claims the role as developer, artist or scientist. The author actively rejects the notion of claiming these for individual reward. However, she does recognise the necessity of describing her contribution within the framework of this academic thesis. This act sits unresolved within the ideological framework of the practice, but is proposed to be mitigated by simply describing rather than claiming or evaluating these contributions.

The creative and novel act within this practice is as much the initiation and creative production of participation and collaboration itself, as the development of the artistic ideas and the underpinning research inquiry. Whilst the author's contribution as the co-author (TIWWA, Women Reclaiming AI and Emoti-OS) and

---

<sup>17</sup> The notion of open source is inherently intertwined with a way of developing technologies and software. However within this thesis it is used to imply an open source culture which goes beyond the ethos of sharing and reusing code and software, to 'an ethos that values sharing' (Hecht, 2018). This indicates an ideological culture which embrace collaboration, community, meritocracy and collective benefits..

author (The Infinite Guide) of the artistic ideas, and the instigator of the research inquiry, the value lies in the impact produced collectively by all participants in the practice, irrelevant of their individual contributions. Through this way of working, a non-hierarchical, collaborative and equal process of developing work is initiated, where the credit, ownership and authorship lies with all participants, increasing its potential for social impact.

#### **1.3.4. The evaluation of practice and assessment of impact**

The evaluation of the practice outlined within this thesis and assessing its impact is measured against the central aim of creating actual and traceable participation in the proposed discourse. It is framed through the exploration of the potential impact of appropriating *relational things that talk* as a discursive design strategy through a participatory method. The ambition is that this empowers participants (and audiences) to (re)evaluate and (re)consider their current and future expectations of and relationship with conversational AI systems, and their voice in the commercial design of AI technologies overall. With the engagement with discourse being the fundamental goal and desired outcome of the practice this is what is reflected upon and critiqued through the description of the prototyping process in chapter three. Tharp and Tharp (2019, p. 285) propose that audience reflection is the 'baseline requirement' for successful discursive design, and only after this can one start to explore any potential impact (societal, business, professional and personal). They (ibid, p. 286) are ultimately more concerned with '...discursive designers having an impact than in measuring impact', but underline the potential societal impact a discursive approach could have;

So first, discursive design can have a societal impact, which is the broadest arena with many branches and involves collectives and their thinking, actions, artifacts, and organization. This includes stakeholders, partners, users and audiences that were known to be part of the process, but it also includes those affected by the unknowable ripples that extend into society.

Simultaneously Tharp and Tharp (2019, p. 285) recognise the challenge in measuring such an impact, which often can be intrinsic or indirect, due to 'latency, unintended consequences, variable contexts, and false indicators'. They (ibid, p. 289) state that though discursive design 'always strives for mental responses from the audience, they only sometimes shoot for, let alone achieves, any broader impact'. This is not as they (ibid) say 'audience satisfaction business', but about inviting people to 'encounter, inspect, recognize, decipher, interpret and reflect' on the discourse embedded.

The practice discussed in this thesis does not imply that audience engagement with a discursive prototype will lead to societal change. It does however argue that by engaging people through participatory methods in the design process one can, as the designer of this discourse, better engage people and at the same time, better evaluate and interpret this engagement and impact. This argument, for the role of real (opposed to a token) user participation in the design process, is simultaneously posed as an essential part of a design process for future AI systems and services. The agenda around participation is therefore central to the evaluation of the practice and impact. Simultaneously the practice also speculates if the role of the designer and conveyor of this discourse could be augmented by the application of conversational systems as conveyors of discourse.

The evaluation of the practice consisting of the four prototypes, TIWWA, Emoti-OS, WRAI and The Infinite Guide, is grounded in its pursuit to develop participatory discourse about the topic of investigation. It is focused on the potential impact of extending a discursive design approach with a participatory method to deepen engagement with the proposed discourse. Simultaneously the function of the practice is to apply, test and expand the theoretical investigation of the *perlocutionary effect* of conversational systems. These aspects frame the methods applied to evaluate and validate their impact. The evaluation of the practice itself is rooted in the intentions and context of the specific prototypes, underpinned by the institutional collaborations and contexts, and validated through these partnerships, dialogue with participants and audiences, and public presentations and publications of the work (for a full list of exhibitions, conference presentations and publications please see the Appendix 'Public Research Outputs').

The process of evidencing and assessing the impact of the practice through dialogue with participants and audiences is conducted through informal conversations, focus groups conducted by the designer of this discourse and collaborators (applied across TIWWA, Emoti-OS and WRAI), and through using the conversational systems it is critiquing to engage (applied across TIWWA, Emoti-OS and WRAI) and evaluate this engagement (applied within The Infinite Guide). Here the participation in the design process and in dialogue with the designer of the discourse is in itself a documentation of impact. This aligns with Tharp and Tharp's (2019, p. 286) concern with 'having an impact' rather than 'measuring impact'. This intentionally declares a distance from contemporary audience evaluation metric enforced by funders within the cultural sector as well

as the user satisfactory surveys of commercial sector. The interest lies in documenting participation with the design and discourse as an act of empowerment, reclaiming the voice of audience / end-users within a discursive and commercial design process. This pursuit influenced the design decisions made within the practice, as well as the type of collaboration and participation facilitated. The manifestation of the participation with the discourse at the same time became generated through the real-time dialogues between people and the conversational AI systems. Taking on its own narrative as an emergent and generative discourse between humans and systems.

The type of evaluation applied is a direct output from the intentions of the project and the context (participants, collaborator, partners, as well as the framing of the discourse framing and site) of the work. TIWWA was the first manifestation of the central research inquiry to test the hypothesis of the *perlocutionary effect* of conversational systems and the process of participation itself. The evaluation of this prototype and documentation of its impact was explored through informal observations where participation in the design process and in dialogue with the designer of the discourse is the documentation of impact. Emoti-OS explored the *relationality* of affective systems and was also focused around evaluating the impact of a participatory method to increase engagement with the discourse put forward. Lead through participatory workshops, engagement was again documented through participation in these, and further through informal feedback from participants and focus groups lead by the Plymouth School of Creative Arts (PSCA) project lead. The wider audience/participant engagement with the Emoti-OS' was documented, observed and evaluated through interactions with its



scenario and artefacts (conversational data and number of interactions) and informal observations of audience/user.

The evaluation of WRAl was rooted in its intention to focus on generating discourse with participants, (rather than with audiences engaged through public presentations of work) around *stereotyping* and *bias* in commercial conversational AI developments. This was facilitated through informal conversations through *doing*, as participation in prototyping workshops, as well as through reflecting on the participatory process with the collaborators. The intention was to create a conversational and informal workshop space where no personal data was collected and participants were free to share their thoughts. The evaluation of the Infinite Guide referenced the evaluation methods and findings deployed across the previous prototype iterations. It further extended this process by conducting observations of users' engagement with discourse through the system itself (conversational data and number of interactions), as well as evaluating the participatory process through focus groups, public presentations and informal conversations.

The practice outlined in this thesis incorporates the ambition of exploring the application of conversational AI systems to instigate discourse through conversations with their users, as well as to evaluate their interactions (not as facilitators for data-harvesting, but for facilitators of discourse). A process not intended to provide user satisfaction or audience evaluation data of their art / cultural experience, but to use the method of evaluation as additional layer of interactions generate discourse around people's relationship with AI systems. Though some of these methods are commonly deployed to measure user experience and audience satisfaction in commercial design and across the arts

and culture, the intention here is not to gather such data. A pursuit which often conflicts with cultural organisations and funders drive for audience evaluation metrics to justify their funding spend and the intrinsic value of art and culture.

The practice-based research inquiry of this thesis is interested in peoples' engagement with its discourse. This does however not disregard the responsibility of the quality of interactions and the impact on the audience when interacting with the artefacts and scenarios presented here. However, it is about exploring the making of *things that talk* as participatory practice, as well as a *conversation* with these *things* as a form of critical discourse. The intention is to apply and test the knowledge generated on existing conversational AI systems to describe and synthesise a deeper understanding of their impact on the thoughts, behaviour and actions of their users and society as a whole. By doing so it aims to propose a more ethical design framework for future conversational systems

#### **1.4. Research questions; beliefs, desires, aims and objectives**

The practice-based research inquiry of this thesis is driven by the author's desire to claim technological innovation for artistic production and social action, with the people that these technologies will affect. It seeks to unmask the complexity of conversational AI systems, challenging the pervasive lack of ethical design frameworks and diversity; limited transparency; ubiquitous authority and embedded bias in their development. Central to this aim is the design of actual and traceable participation through a practice that develops a counter-narrative as a form of activism. These aims are underpinned by the belief that only by including, activating and empowering the people that AI systems

(and other emergent technologies) affect, can one start to demand more ethical and desirable future technologies.

This practice-based inquiry is directed through the above aims and extended by drawing a parallel between the lack of user participation in technology development to a similar lack of participation in many discursive design paradigms (outlined further in the methodology section below). It responds to the following criticism of discursive design approaches; lack of dialogue and follow-up debate (Kerspern, 2015) and the presence of elitist and privileged positions incorporated within their scenarios and artefacts (Tonkinwise, 2015).

These desires, beliefs and aims underpin the research questions of this inquiry. The first question forms a foundation for the contribution of new knowledge through the hypothesis of the *perlocutionary effect* of conversational systems:

**How do conversational AI systems, *relational things that talk*, influence the subjective reality and related thoughts, behaviours and actions of their users?**

This question is focused on building a theoretical foundation to understand what impact conversational AI systems may have on the social reality of an individual user. By developing the proposition of the *perlocutionary effect*, it describes the *factors of influence* which constitute and propagate this effect (their *relationality* and *persuasiveness*, embedded *bias* and *hybrid intentions*). Through commercial examples, it demonstrates the design strategy of manufacturers developing conversational AI systems which exploit the human tendency to interpret technology in terms of the social world (Reeves and Nass, 1996).

The findings from this question form the basis for the second question, which tests and demonstrates *the perlocutionary effect of relational things that talk* through practice, manifesting as four discursive prototypes:

**How can a discursive design strategy activate people to challenge the commercial development of conversational AI systems and re-imagining more desirable future realities being with *relational things that talk*?**

This research question also frames the parallel inquiry which explores the potential impact of claiming *relational things that talk* as a discursive design strategy for new forms of cultural expression and social action. It argues for a participatory method within a discursive design approach and within the commercial design of future technologies, as a mechanism to activate people to demand more ethical and desirable future AI systems.

The research questions of this thesis, outlined above, generate the following objectives which deliver against and clarify the overall aims of this practice-based research inquiry:

**Objective 1:**

Test and substantiate the hypothesis of the *perlocutionary effect* of *relational things that talk* through four discursive prototypes developed with the people these systems affect. The engagement of young people and self-identifying women are specifically prioritised.

**Objective 2:**

Evidence the commercial design strategy of conversational systems that exploit the human tendency to anthropomorphise *things* through commercial examples.

**Objective 3:**

Appropriate conversational AI systems *with* their target users to unmask the complexity of these systems and extrapolate counter-narratives for what these technologies, and their consequences, could be in the near future.

**Objective 4:**

Empower people to challenge, discuss and participate in technology development (a process in which the technology consumer is largely left out) by creating opportunities for hands-on technology development and design.

**Objective 5:**

Test the impact of a discursive design approach extended with a participatory method in generating and concretising ideas around current and near future relationships between humans and conversational systems, through creating *actual* and *traceable* involvement of their future users.

**Objective 6:**

Claim emergent conversational AI technologies for new forms of creative practice, which has social debate as a fundamental goal and desired outcome, to disrupt the commercial development of unethical and exclusive AI systems.

These objectives inform the development of the overall practice-based research inquiry and the contribution to new knowledge. Within this inquiry theory and practice is reflective of each other and emerge as an iterative process. The theoretical foundation of new knowledge emerging from the first question is tested and iterated through the practice. This informs the theoretical proposition of the perlocutionary effect and the potential impact of a discursive design strategy to challenge the commercial development of conversational AI systems. The ambition is to develop practical and theoretical contributions which are relevant and have an impact on creative practice across Art and Design, and an influence on the commercial development of AI systems.

## 1.5. Chapter outline

This thesis is structured through three core chapters. The first presents and grounds the hypothesis of this practice-based research inquiry; the *perlocutionary effect of relational things that talk*; tested through the iterative

practice outlined in chapter three. It begins by outlining the reality-forming role of language in chapter two, 'Speaking of Reality Generating Things', by drawing a parallel with the influence of conversational AI systems. This chapter forms the theoretical basis for the *perlocutionary effect* produced by conversational systems upon the feelings, thoughts or intentions of the user, (such as directing, persuading, scaring, enlightening, inspiring or dominating). It describes the *factors of influence* that constitute this effect, their *relationality* and *persuasiveness*, and the valence of this effect on a user, propagated by a system's *hybrid intentions* and embedded *bias*, relative to a user's susceptibility. Rooted in commercial examples and evidenced through secondary research, it demonstrates contemporary design strategies for conversational systems which exploits their *relationality* and *persuasiveness*, proposing their *perlocutionary effect* on users.

Chapter three, 'Prototyping Relational Things That Talk', describes the practice-based research conducted to test the hypothesis of the *perlocutionary effect*. It describes this practise which manifest as four discursive prototypes; This is Where We Are (TIWWA), Emoti-OS, Women Reclaiming AI (WRAI) and The Infinite Guide, where each iterates the findings from the previous. It is contextualised through examples of forms of cultural expression and social action by designers and artists which have captured people's imaginations around a future living with human-like artificial systems.

The documented findings from the practice tests inform the description of the construction and valence of the *perlocutionary effect* and the causal relationship between the *factors of influence* and the *effect* on users. Integral to the narrative of chapter three is the description of the method of engaging people

in *the design of* and *interactions with* the discourse put forth through the practice. In so doing, this chapter proposes the appropriation of *relational things that talk* as a discursive design strategy, extended with a participatory method, for new forms of cultural expression and social action which activate people to demand more ethical AI systems.

The thesis concludes with the final chapter 'Being With Relational Things', which summarise this research inquiry's practical and theoretical contributions to new knowledge which are relevant and have a potential dual impact across art and design practices, as well as towards the commercial design of conversational systems. These contributions specifically span: ethical design frameworks for conversational AI systems; discursive design approaches and methods, and the appropriation of conversational systems, *relational things that talk*, for new forms of cultural expression and social action through a participatory arts and design practice. It closes with a summary of future research and concluding remarks.

## **1.6. Ethical guidelines**

This practice-based research inquiry conduct within this thesis appropriates conversational AI systems which engage people in dialogue as mechanisms for new forms of cultural expression and social actions through a participatory arts and design practice. In so doing it follows the following policies and frameworks:

- 'ESRC Framework for Research Ethics' (Economic and Social Research Council's (ESRC), 2015)

- ‘Ethical Guidelines for Educational Research’ (The British Educational Research Association’s (BERA), 2018)
- ‘Research Ethics Policy’ (The University of Plymouth, 2018)
- ‘Code of Good Research Practice’, (The University of Plymouth, 2018)
- ‘Research Data Policy’, (The University of Plymouth, 2018)

Ethical approval was secured for the Emoti-OS project which engaged young people and teachers from Plymouth School of Creative Arts (PSCA) as participants, collecting and storing anonymised conversational data from peoples’ conversations with the Emoti-OS chatbot. This also included summarised and anonymised information from informal focus groups led by the lead PSCA teacher Angus Reith. Information explaining the research and their participation was provided verbally and in writing to the participants through formal school assemblies, through their lead teachers in person and through myself as the lead researcher, as well as through the Emoti-OS website chatbot. Informed consent was sought through the PSCA management in loco parentis. (See Appendix II: Emoti-OS Information Sheet and Consent Form for Ethical Disclaimer). Developed in partnership with PSCA and conducted within the school further, safeguarding was provided and adhered through PSCA’s safeguarding policies.

Ethical approval was also secured for the Infinite Guide project (see Appendix III: The Infinite Guide Information and Consent Form (online)) which collected and stored anonymised conversational data from peoples conversations with the Infinite Guide chatbot. This also included summarised and anonymised information from informal focus groups led by the Neil Winterburn, Learning Technologist FACT. Developed in partnership with FACT Learning and



conducted within their premises the project fell within their safeguarding policies for working with young people.

Whereas the TIWWA and WRAI projects follows the University, ESRC and BERA's research frameworks and principles outlined above, they did not require ethical approval. These projects did not store or use people's personal data to inform the research conducted within this thesis, or fall within the other requirements for ethical approval of research.

## **1.7. Limitations of research**

This practice-based research inquiry of this thesis is focused on investigating the implication of conversational Artificial Intelligence (AI) systems, '*relational things that talk*', on the way people experience the world. In delivering against this research agenda the author recognises the following limitations:

- The research is situated within a context of the development of AI technologies located in a largely culturally western and developed economy with an advanced tech infrastructure when compared to other nations. It does recognise that the influence of conversational AI technologies on the individual and society as a whole does assume a privileged position of having access to these systems and to their services.
- Similarly, the potential impact of a practice taking place in the UK and which is restricted to the English language, is potentially limiting in terms of its engagement, and may compound a western and privileged perspective.
- This inquiry recognises that conversational AI technologies are in their infancy and their effect is currently limited. However, it is not about what these technologies are doing to us now, but what they may do to us in the

near future. It is rooted in the documented concerns of the wider implications of AI systems on the individual and society as a whole.

- It also recognises that there is a substantial portfolio of work from a diverse set of artist doing social practice, but for the contextualising of the prototyping practice outlined here in chapter three, the focus is on work that can be said to apply *relational things* which exploit the human tendency to interpret technology in terms of the social world, and that invite people into their discourse by appropriating these technologies both as a medium and a subject matter

## 2. Speaking of Reality Generating Things

*'Listen more often to things rather than beings'*

(Diop, 1948)

### Chapter introduction

*This chapter develops the hypothesis of the 'perlocutionary effect' of conversational systems, 'relational things that talk'; described as the sum of their influence on a user's actions, thoughts, or beliefs. Through evidencing the reality-forming role of language and 'technological things', the inquiry underpins this hypothesis. It deconstructs the 'perlocutionary effect' by interrogating the factors constituting and propagating this effect; their 'relationality' and 'persuasiveness', (designed through a system's conversational ability, personality, personalisation, affective and embodied functions), and propagated through a system's embedded bias and 'hybrid intentions'.*

*Evidenced through secondary research, this theoretical inquiry has been reinforced by commercial examples. The purpose of this chapter is to highlight the specific design functions and strategies applied within the development of conversational systems to pinpoint their correlation with the 'perlocutionary effect' on a user and frame the hypothesis which is tested through the prototyping practice outlined in chapter three. In so doing, this chapter responds to the central research question of this thesis:*

***How do conversational AI systems, 'relational things that talk', influence the subjective reality and related thoughts, behaviours and actions of their users?***

## 2.1. The Conversational Interface

Natural language is intrinsic to human existence. It enables us to share our experiences, make sense of the present and predict the future. It also reaffirms and mediates our subjective experience of reality (Berger and Luckmann, 1966). Up until this point man and machine have spoken different languages, where the onus has been on humans to translate language through a graphical user interface (GUI). Now, due to recent technological advancements in processing, machine learning and language technologies, systems are becoming capable of understanding more than just syntax (McTear et al., 2016). These developments - alongside the rise of smartphones with inbuilt microphones and ubiquitous connectivity - render the dream of conversational interfaces<sup>18</sup> a reality (Brownlee, 2016).

Conversational systems exploit the human tendency to interpret technology in terms of the social world and apply their understanding of social relationships to their interactions with artificial systems (Reeves and Nass, 1996). This triggers a critical paradigm shift in human-computer interaction and relation. The potency, influence and effect of these human-artificial relationships will depend on a conversational systems' ability to engage with, relate to and persuade their users. Factors which are directed by a systems' design and functionality, and compounded by the intentions of their manufacturers, embedded system bias, and the susceptibility of a user.

The pursuit of developing 'talking and feeling' machines with human-like functions are no-longer fictional depictions, but a reality with AI powered

---

<sup>18</sup> McTear, Callejas, and Griol (2016) in 'The Conversational Interface: Talking to Smart Devices' simply define 'conversational interfaces' as enabling 'people to interact with smart devices using spoken language in a natural way — just like engaging in a conversation with a person' (p.1).

conversational systems offering social interactions and simulated affect. These systems have their roots in the development of abstract conversational interfaces in the 1960s with 'text-based dialog systems for question answering and chatbots that simulated casual conversation' (McTear et al.,2016, p51). They gave the illusion of intelligence, as seen in the ELIZA natural language processing computer program initiative developed by Weizenbaum (1966). He (1976, p.6) was startled to see how quickly and deeply users got emotionally involved with the system and how they 'unequivocally anthropomorphized it' and therefore became an avid critic of dialog systems that simulated casual conversation. In spite of Weizenbaum objections, McTear et al. (2016) stipulate how these developments has formed the basis for voice user interfaces, embodied conversational agents, social robots and today's intelligent conversational systems.

Eliza was one of the first programmes that presented itself as what Turkle (2007b, p. 62) describes as a 'relational artefact'. It was powered by a rule-based dialogue generation which capitalise on simulating sentence through matching key words with pre-written responses. However, it was the human tendency to anthropomorphise artefacts and relate to them that generated the perceived impact of the system. Fast forward to the conversational systems of today, where innovations in Machine Learning, natural language processing, device proliferation, ubiquitous connectivity, steeped in pools of data, has supercharged these earlier rule-based systems and augmented their potential functionality and impact. Conversational systems of today can be context aware and perform more complex tasks whilst simulating levels of empathy, emotion recognition and personality. They enable, according to McTear et al. (2016, p 22),

people to talk to their devices in an intuitive and natural way, and the systems to respond in appropriate and human-like ways.

Conversational systems are being deployed on various technology platforms, such as smart watches, social robots, voice activated devices, computers and mobile phones. Each platform provides different functionalities, from the voice, text and image capacity of mobile phones, to spatial and environmental movement and sensor input of social robots. With the increasing complexity of these platforms, this also brings additional multimodal input and output functionalities, increasing the communicative capacity of the conversational system installed. These systems are often powered by a mix of rule-based and machine learning algorithms and capable of using human language (through text and speech), as an explicit emulation. Some systems use the modality of voice commands (as Apple's 'Siri'<sup>19</sup>, Google's 'Google Assistant'<sup>20</sup>, Amazon's Alexa<sup>21</sup> and Microsoft's 'Cortana'<sup>22</sup>), a mix of voice and text, or just text.

There are also hybrids<sup>23</sup>, or human-assisted chatbots where messages with a lower confidence level are forwarded to human operators to be reviewed, actioned, or used to train the system further. Alongside these systems, there are also nascent systems emerging to the market, which are explicit in their application of emotional recognition and response, such as 'Replica'<sup>24</sup> and

---

<sup>19</sup> <https://www.apple.com/siri/>

<sup>20</sup> <https://assistant.google.com/>

<sup>21</sup> <https://alexa.amazon.com>

<sup>22</sup> <https://www.microsoft.com/en-us/cortana>

<sup>23</sup> Facebook's 'M' was a hybrid chatbot where inquiries which could not be resolved by the AI was forwarded to humans. The ambition was to build and train the AI to automate almost all of M's tasks. However, as Griffith and Simonite (2018) states, as it offered an open domain rather than narrow task range (like Amazon's Alexa), users would therefore ask for increasingly complex tasks to be fulfilled. A large proportion of these, they explain, had to be handled by the system's human counterparts. It therefore became clear that M would always require a sizable workforce of expensive humans, and become less viable. This also indicates one of the big challenges with these systems ability to simulate human-like qualities; the disproportionate relationship between user expectations and system abilities.

<sup>24</sup> <https://replika.ai/>

Microsoft's China-based 'Xiaoice'<sup>25</sup>. Many of these systems are also deployed with different physical (such as robots) or virtual embodied forms (such as avatars). Some systems like the SoftBank's 'Pepper'<sup>26</sup> incorporate all the functionalities above; through a robotic form with mobility, environmental sensors, a conversational interface which is trained to use facial recognition to read emotions and recognise users.

With the augmentation of multimodal user input and output, the system's ability to relate and entice the users into a deeper human-machine relationship increase. These developments are all springing from the commercial pursuit to create human-like assistive systems which can emulate human dialog. The idea of natural and fluid conversational exchange with machines has been depicted in science fiction since the 1950s. With the omnipresence of conversational and intelligent machines introduced by science fiction authors such as Isaac Asimov in 'The Naked Sun' (1957) and Arthur C Clarke's '2001: A Space Odyssey' (1968), we have since become accustomed to expect this of technologies that talk. However, the reality of these systems is still fraught with limitations and errors, often the user's expectations conjured through science fiction, the media and simulated human-like systems, does not match the actual system functionality. Andrew Ng, chief scientist of Baidu, states (in Kojouharov, 2016);

'Here's one example of something it cannot do: have a meaningful conversation. There are demos, and if you cherry-pick the conversation, it looks like it's having a meaningful conversation, but if you actually try it yourself, it quickly goes off the rails'.

---

<sup>25</sup> Xiaoice, has according to Microsoft, an 'empathic computing framework' and over 600 million users (Spencer, 2018).

<sup>26</sup> <https://www.softbankrobotics.com/emea/en/pepper>

In spite of the limitations of current conversational systems, the rise of conversational modalities and intelligent systems are affirmed by the investments made and products launched by the technology giants such as Apple, Google, Amazon, Facebook and Microsoft. The ambition is, according to Google's Chief Exec Pichai (2016) to integrating voice assistants as '...an ambient experience that extends across devices making it a natural and integrated part in the context of a users' daily life'. This pursuit is further demonstrated with Google's human-sounding voice assistant, 'Duplex', in May 2018 (Google, 2018). Indistinguishable from a human voice, this has been hailed as the ultimate technological achievement, supercharged with human-sounding vocal cues, such as 'ums' and 'ahs' and powered by natural language processing. Simultaneously, as Lomas (2018) highlights, this has raised a backlash of ethical concerns of these systems for being designed for deception.

The launch of Duplex in many ways illustrates where we are in terms of the pursuit of systems indistinguishable from humans, fraught with the lack of clear ethical guideline, transparency of systems functionality and data usage. Conversational systems emerging onto the market are draped in layers of simulated humanisation efficiently exploiting the human predispositions to apply animacy and sentience to objects (Reeves and Nass 1999; Turkle's 1984, 1995, 2004, 2007a, 2007b). They shift, according to Persson et al. (2002, p.21), how users view technology from tools to actors, understanding them in terms of human behaviour and intentionality. Collectively they indicate a paradigm shift in terms of how we interact with intelligent systems and the level of influence these have on human experience mediated with and through them. Within this new paradigm, conversational systems are positioned as *the* new user interface,



bringing new levels of immediacy, ubiquity, authenticity and natural interactions (Connolly, 2017; The Economist, 2017; McTear et al.,2016). Conversational systems brings forth a reality generating potency, rooted in language, augmented by technological innovation and realised through exploiting humanity's tendency to anthropomorphise *things*.

## **2.2. The reality forming role of language**

Driving the development of conversational AI systems is the pursuit of developing systems capable of conducting fluid natural language interactions with people as a more effective way for us to interact with computational systems. Language understanding itself has always been one of the primary goals of AI - to develop computational models of the knowledge and the processes involved in executing linguistic tasks (Gillis, et al., 2009). The performative use of language rooted in 'speech act theory' (Wittgenstein, 1958; Austin, 1968, Searle, 1969, 1975, 1979) has been used to design, train and enhance language understanding and abilities of conversational agents<sup>27</sup>. The pursuit is rooted in the role of language in human-human interactions, and is therefore important to start with exploring the construction and reality forming role of language in itself.

Language affords the objectification and communication of the experience of everyday life. It enables us to affirm, maintain, as well as modify our subjective realities. Berger and Luckmann (1966, p.173) state that conversations are the

---

<sup>27</sup> See Bruce (1975), for the first attempt to account for Speech Act theory in terms of AI work on actions and plans, and for further work on the wider implementation of this framework within AI and Natural Language computational models see; Allen (1994); Allen and Perrault (1980); Appelt (1982 and 1985), Cohen and Perrault (1979); Dijkstra and De Smedt (1995); Phillip et al. (1979).

'actualizing of this realizing efficacy of language' and thus 'the fundamental reality-maintaining fact is the continuing use of the same language to objectify unfolding biographical experience'. Through language we can describe and share countless matters, whether it is through a monologue or a dialogue with others. It becomes an 'objective repository of vast accumulations of meaning and experience' (ibid., p.51). Berger and Luckmann (ibid., p.54) describe it as enabling the whole world to be 'appresented' at any time and at the same time capable of transcending the reality of everyday life altogether.

Berger and Luckmann (1966, p.40) propose that reality is always socially defined and that conversation is the most important means of maintaining, modifying and reconstructing these subjective realities and knowledge itself:

The theoretical physicist tells us that his concept of space cannot be conveyed linguistically, just as the artist does with regard to the meaning of his creations and the mystic with regard to his encounters with the divine. Yet all these - dreamer, physicist, artist and mystic - also live in the reality of everyday life. Indeed, one of their important problems is to interpret the coexistence of this reality with the reality enclaves into which they have ventured.

The concern here lies with what they (1991, p.37) describe as the reality of 'everyday life', rather than the notion of an objective reality, which exists independently of the mind. The intention is to explore the role of other individuals and 'things' in mediating the way in which we interpret, share, construct and reconstruct our subjective reality, and is therefore focused on the reality that stands out as the 'paramount reality' as it 'imposes itself upon consciousness in the most massive, urgent and intense manner' (ibid. p.37).

The reality of 'everyday life' is part of the multiple realities of an intersubjective world that we share with others. We exist and reconstruct our subjective reality through interacting and communicating with others. This reality becomes socially defined, as a subjective experience of everyday life, rather than

an objective reality of the natural world (ibid.). This reality is therefore taken for granted *as* reality, as we share the experience of it with others. This proposition of an interpretive and subjective reality is further underlined by Verbeek (2005, p. 50):

“Reality” is not something absolute that human beings can ever know once and for all; it is relative in the most literal sense of the word—it exists only in relations. Reality in itself is inaccessible for human beings. As soon as we perceive or try to understand it, it is not “reality in itself” anymore, but “reality for us.

This interpreted reality, which Verbeek (2003, p.93) describes as being ‘situated subjectivity’, is mediated and influenced by other individuals and ‘things’. The way in which we interpret and share these subjective realities themselves are in a continuous flux of reconstruction as we communicate and share our lived experience with people and things; an act which is integral to human existence. What we interpret as the ‘world’ and what we ‘are’ arises, according to Verbeek (ibid), from the interplay between us and reality.

Language can also make present events, experiences and people of the past or imaginary futures. This process of using language to interpret and communicate the experience of everyday life within ‘finite provinces of meaning’ (ibid., p.40), also immediately distorts the reality of everyday life. By enabling mental constructions to be shared and reconfigured, individuals absorb, adopt or even confuse memories of things that have actually happened in their lives with those told to them by others, or heard from fictional accounts (Suddendorf, Addis and Corballis, 2009). This continuous process of the construction, modification and sharing of subjective realities is proposed by Tucker (2014, p.340) as necessary to guide and lead individuals to take certain actions and avoid others ‘even though every vision is an illusion’.

Despite our highly developed ability to direct our future based on our past, our interpretive predictions about realities to come are notoriously fallible. However, Gilbert & Wilson (2007) propose that to compensate for such weaknesses one adaptive function of language may be to improve the accuracy of predictions by drawing on narratives from others, and their experiences of what the future may hold. These can be 'explicitly reported, shared, remembered and violated' (Moulton and Kosslyn: 2009, p1278). We do so, so naturally, that we rarely question the accuracy, bias or subjective nature of such simulated realities and their authors (people, algorithms or artefacts, or a combination of these).

The subjective nature and imbued intentionality of language is not only constructed through syntactic, grammatical and semantic use of words. This is conjured through sophisticated application of verbal and non-verbal cues, such as intonation as well as non-verbal signals, alongside contextual influences, motives, social conditioning, previous experiences and so on, both in the speaker and the listener. All of these factors contribute to the transmission of meaning and intentions beyond the utterance of sounds which influences our ability or declaration to act. In essence, language becomes a performative act where the speaker and listener are intertwined in an exchange of meaning and knowledge, reconstituting and influencing the way in which we see the world, both in intentional and unintentional ways.

Conversation, between human-humans and human-computers, has the potential to influence the participants; requesting, questioning, complimenting, promising, informing and directing. It is a performative execution of language wherein the words spoken bring forth the potential of action. The notion of a speaker performing an action was first proposed by Wittgenstein (1958) and

further developed by Austin (1962) and Searle (1969). Austin described this performative use of language - of performative utterances - which brings about an action or result. This was central to his theory of 'speech acts', where he (1968, p.108) divides this performativity of words into three types of acts<sup>28</sup>, each with a performative outcome; a 'locutionary act', 'illocutionary act' and the 'perlocutionary act'.

Out of these three acts, it is the final performative act, the 'perlocutionary act', which is here applied to underpin the overall proposition of the *perlocutionary effect* of conversational systems. Austin described this act as 'what we bring about or achieve *by* saying something, such as convincing, persuading, deterring, and even, say, surprising or misleading' (ibid., p.108). A perlocutionary act has a 'perlocutionary effect' on a listener, influencing their thoughts, emotions and actions (ibid). This act can be seen, according to Searle's (1969, p. 25), as a direct outcome of an illocutionary act (which brings forth the intention of the speaker to *do* something, (Austin 1968, p.108). These speech acts can therefore be said to carry the potential of influencing a listeners thoughts, behaviour and actions, which is further affirmed by Searle's (1969, p. 25) contribution to Austin's general theory of speech acts. A speech act may however bring both intended and unintended effects on the listener, and is influenced by a range of social, mental and contextual factors. Together these performative speech acts contribute to the reality-generating potency of language in conversations.

---

<sup>28</sup> The first speech act is described by Austin (1968, p.108) as 'a locutionary act', which he describes as 'uttering a certain sentence with a certain sense and reference'. This is the intentional act of saying something that carries meaning and is well-formed from a syntactic point. Austin describes performing an 'illocutionary act', 'such as informing, ordering, warning, undertaking, &c., i.e. utterances' (ibid., p.108). By *saying* something, we bring forth an intention to *do* something. These are utterances that Austin describes as having a 'certain (conventional) force' (ibid., p.108), ie. how what was said was meant. The final performative act, the 'perlocutionary act', is described by Austin as 'what we bring about or achieve *by* saying something, such as convincing, persuading, deterring, and even, say, surprising or misleading' (ibid., p.108). The perlocutionary act and effect can be seen as the effect of both the locutionary and the illocutionary act.

### 2.2.1. The reality-generating potency of conversations

The intention of the speaker in producing a specific utterance is described by Searle (1969) as the illocutionary force of an utterance. This force together with the 'propositional content' (which specifies the details of the action, such as what the hearer is being requested to do) composes the act and outcome. Illocutionary acts can be distinguished through their illocutionary force<sup>29</sup>, or the speaker's intentions inherent in the act of performing that utterance. Grice (1967) describes this as a process of accessing each other's mental states where the speaker and listener exchange and interpret displays that evidence their mental state and understanding. Grice recognises the context of conversation; not just a mere chain of isolated sentences but rather as interlinked acts within a wider discourse.

We engage in discourse and take turns in acting as speakers and listeners, exchanging propositional content imbued with intentions and perlocutionary potential. We communicate and simultaneously participate in an act of re-affirming and re-constructing each other's social realities. Geis (1995) critiques Austin and Searle's framing of speech acts as 'individual, constructed utterances, isolated from actual or even explicitly imagined conversational contexts' (ibid., p.xii) and instead argues that they should instead be taken as 'communicative social actions' (ibid., p.16). He proposes a focus on how we 'do' speech acts in multi-turn and social conversational interactions, which 'involve particular factual

---

<sup>29</sup> In "A Taxonomy of Illocutionary Acts" Searle (1975) propose to classify five ways of using language in this way; 'assertives' (statements or conveying information), 'directives' (commands or requests), commissives (promises or commitments), 'expressives' (congratulations, apologies, or other expressions of emotions) and 'declaratives' (performatives acts that bring about a different state of the world such as pronouncing, declaring, confirming, etc.).

states, social relationships between participants, psychological states and attitudes, among other nonlinguistic things' (ibid., p.13).

Conversations between people not only express propositions and perform actions, but also conjure a range of (intended and unintended) effects. These performative acts do not happen independent of context and interaction, but are a conduit for a confluence of factors. An utterance may mean different things to different hearers. Factors such as the social relationship between the speaker and the hearer, the status of the speaker or other social conditions between the speaker and the hearer may also influence this effect. Berger and Luckmann (1966, p. 174) propose that conversations hold a different density of 'the reality they produce or maintain', and are also assigned different levels of privileged status and authority. Conversation has an inherent reality-generating potency, an ability to bring forth and re-configure our perceived past, present and future subjective realities, which are influenced by a range of social, mental and contextual factors.

The reality-generating potency of conversations are related to the perceived authority of the person (or thing) we are talking to, as well as the perceived social context of the conversation and inherent social structures, expectations and norms (Berger and Luckmann 1966, Moulton and Kosslyn, 2009). Burr (2003) proposes that people present themselves and their version of events in such a manner that it will predominate other propositions. The perception of these acts is directly linked to the perceived power and status of the individual. We communicate our intentions through performing speech acts, imbued with our bias and world views. By doing so we intentionally and unintentionally influence each other's perceived past, current and future

subjective realities. Similarly, as we engage in discourse with each other, we are also engaging in conversations with systems. These are encoded with their own reality-generating potency, perlocutionary potential and perceived authority, directed by the intensions of their manufacturers. Through their ability to relate to us and persuade us, they invite us to interact; always available, serving and listening.

### **2.3. The reality mediating role of technological things**

Conversations are inherently social; fundamentally about how we are influenced by one another in day-to-day life. We therefore quickly project these social frameworks onto artefacts that are capable of simulating human communication models. We assimilate, share and absorb narratives of the world around us from others, both through direct conversations, as well as through distributed information systems. This process does not have to happen face-to-face. It may take place across different spatial dimensions, mediated by communication devices or in writing (text messaging, live chat etc.). It may also take place with *things that talk*; such as chatbots, voice assistants, robots and other artificial agents. These interactions, facilitated through natural language processing capabilities, bring new artificial actors to the process of mediating and re-constructing our social realities; realities manufactured by corporations; spun from data collected from our personal devices; delivered through technological artefacts.



Technological artefacts are becoming active participants in maintaining, modifying and reconstructing our subjective realities and knowledge itself. Verbeek (2003) builds on Ihde's (1990) notion of a mediated reality, proposing that the human-world relationship is mediated and co-constituted by technological artefacts. These technologies shape people's perceptions and interpretations '(the ways in which reality can be present for humans)' and also their actions and engagement with reality '(the ways in which humans can be present in their world)' (Verbeek, 2003, p. 93). These 'things' do so by transforming perceptions and interpretive frameworks, therefore shaping the way we encounter reality (ibid).

Our perception of a subjective reality and the way in which we are involved with this reality emerge through the mediation of these artefacts. They 'coshape' the access we have to reality itself (ibid, p.104) and help to 'determine how people act' (ibid, p.236). Technological artefacts mediate the way in which reality comes to be meaningful for us, and therefore actively contribute to the way we live our lives. This active contribution also has according to Verbeek (2005, 2011) an important moral dimension. They are not simply neutral instruments but shape our daily actions and decisions as active artefacts through 'amplification and reduction' as Verbeek states (2005, p. 195):

From an existential perspective, artifacts mediate human existence by giving concrete shape to their behavior and the social contexts of their existence. This kind of mediation can be described in terms of translations, whose structure involves invitation and inhibition; some forms of involvement are fostered while others are discouraged. Both kinds of mediation, taken together, describe how artifacts help shape how humans can be present in the world and how the world be present for them.

The mediating role of technological artefacts in amplifying specific perceptions and reducing others, is described by Verbeek (2006a, p. 4) as

‘technological intentionality’. Building on the work by Ihde (1990), Akrich (1992) and Bruno Latour (1992), Verbeek (2005, 2006a, 2008b) proposes that this Intentionality<sup>30</sup> can force people to behave in certain ways, seducing and persuading them to take certain actions. He (2006a, p. 3) describes this technological mediation as having two dimensions which firstly help to ‘shape how reality can be present for human beings, by mediating human perception and interpretation’ (the hermeneutic dimension) and secondly ‘help to shape how humans are present in reality, by mediating human action and practices’ (the pragmatic dimension). These dimensions of mediating technologies impose and guide users on how to act when they are used (ibid).

Verbeek (2008b, p. 388) proposes that because of this ‘intentional structure of human experience’ humans can never be ‘understood in isolation from the reality in which they live’. Searle (1980, 1990, 1995) proposes that the meaning and function people ascribe to a technological artefact is a derived form of Intentionality; it has a social character and is rooted in other intrinsic forms of intentionality. This is what distinguishes them from ordinary objects, but at the same time he rejects the idea that technologies could be Intentional in the same manner as humans (ibid.). Building on Ihde’s human–technology relations<sup>31</sup> Verbeek (2008b, p. 390) challenges Searle’s ideas, by introducing the concept of ‘cyborg intentionality’, ‘intentionality that is partly constituted by technology’. He

---

<sup>30</sup> Intentionality is a core concept widely debated across fields of philosophy each with a different definition of the concept. Searle (1983, p.1) defined this in ‘Intentionality’ as ‘... that property of many mental states and events by which they are directed at or about or of objects and states of affairs in the world’. This thesis will adopt Searle’s (ibid) way of distinguishing the narrow sense of ‘intentional’ (meaning that an actor has a specific intention to do something) from the broader philosophical notion ‘Intentional’ as defined above.

<sup>31</sup> Ihde (1990) describes the idea of a technologically mediated Intentionality through outlining four relationships human beings can have with technological artefacts: embodied relations (the mediating technology becomes an extension of the human body); alterity relations (humans interact with the technology); hermeneutic relations (technologies provide representations of reality) and human–technology (background) relations (where technologies are not experienced directly) (ibid). Within this inquiry both ‘(mediated) human beings and (multistable) technological artifacts are constituted’ (Veerbek, 2008b, p. 392).

(ibid) further distinguish two other forms of intentionality related to human-technology relations; 'hybrid intentionality', where the 'human and the technology are *merged*' into a new hybrid entity, and 'composite intentionality', referring to 'situations in which not only human beings have intentionality, but also the technological artifacts they are using'. This goes beyond forms of mediated Intentionality which occur when human intentionality takes place *through* technological artefacts. This Intentionality becomes distributed between humans and systems, and constructs new ways of seeing the world.

The idea of the a intentionality of artefacts where a relation between human beings and the world is mediated by a technology, is relevant when discussing the potential influence of conversational systems. These systems sit within Verbeek's idea of composite intentionality. Some systems become information points providing information or 'aboutness' about existing realities, whilst others actually construct or reveal a reality that can only be experienced by technologies. The latter does so by doing what Verbeek (2008b, p.393) describes as 'making accessible technological intentionalities to human intentionality'. Some advanced systems, through increased agency, deep learning neural networks and increased processing power, are able to create realities that can only be experienced by humans through the system's descriptions and visualisations. Here Verbeek proposes a 'double Intentionality' where the technology is actually constructing or revealing a reality 'that can only be experienced by technologies, by making accessible technological intentionalities to human intentionality' (ibid, p.393).

## 2.4. The hybrid intentions of *things*

No longer are technological artefacts passive and mute; rather they are active and mediating *things* with which we are in the process of merging in significant ways (Haraway 1991; Ihde 1990; Verbeek, 2005, 2006a, 2008b, 2011; Latour 1993; Hayles 1999). There is a different form than the philosophical idea of intentionality at play within conversational systems. In many ways it connects and informs the effect of Verbeek's 'hybrid intentionality' and 'composite Intentionality' and is also situated within human-technology relations. This is rooted in the narrow sense of 'intentional' and the range of intentions encoded within conversational systems. This is here called 'hybrid intentions' and is constructed from the (declared and undeclared) intentions of the manufacturers of the systems, their designers and the intentions of their users in using the systems. These hybrid intentions may be said to influence a user's mental image or aboutness of things, their intentionality of reality itself, through the ideas outlined in Verbeek's composite intentionality. Here technological systems, imbued with the intentions of their manufacturers and designers, influence not only what it enables their users to think and do, but also the way in which these systems mediate or makes reality available to them.

Due to the complexity of these systems, current patent rights, commercialisation and the value of data mining, many intentions of their manufacturers are hidden within the systems themselves, and therefore arguably difficult to decipher and predict unless this is clearly stated by their manufacturers. These corporate intentions also inform these system's technology platform, functionality and interfaces design, how users' data is harvested and used, as well as ethical (or lack of) frameworks applied. Their designers may also encode

their own intentions within the construction of the systems; to increase user satisfaction and engagement, or control user interaction and objectives. At the same time, their users will also have their own intentions in using the system as well as bringing their own subjective context, intentionality and world view. This confluence of intentions contributes to a conversational system's ability to sway, inform and persuade a user to act or think in certain ways. With these systems increasing ability to play a profound role in the decisions we make, what we do and how we perceive the world, the notion of their hybrid intentions becomes important to reveal and understand.

## **2.5. The bias of *things***

Technological artefacts shape and mould our perceptions of who we are, and influence how we relate to our environment and others. Their increasing complexity, ubiquity and autonomy means algorithmic decisions affect nearly every aspect of our lives. These systems are not neutral, but imbued with relational ability, hybrid intentions and persuasiveness. Within these encoded, intentional and unintentional effects, there are inherent human and machine biases. Bias (unconscious and conscious) dominates our everyday life and is, according to Ross (2014, p3), 'influenced by a pattern of unconscious assumptions that we have absorbed throughout our lives'. He underlines that a large corpus of empirical research demonstrates that people apply biases and stereotypes all of the time, largely without realising they are doing it (ibid). Ross (ibid, p.4) describes these as a result of '...social conditioning, belief systems that we have been taught or exposed to, particular incidents that we remember, or

any number of other assumed “truths” that we have picked up along the way’. By developing a capacity to compartmentalise things and people (into categories such as gender, race, sexual orientation, age, and so on) Ross (ibid) proposes that this allows us to quickly determine what we can expect from our encounters with things and people, now and in the future. It helps us to quickly determine and predict the safety of a person or situation to help to negotiate an ‘extremely complex and busy world’, ensuring the individual’s success of survival (ibid, p.5).

In contemporary society we are increasingly reliant on computational systems to conduct decision making on our behalf. They convert data into decisions with unprecedented speed informing an array of decisions from access to healthcare, mortgages to employment. However, the data used to train these systems reflects biases from the conditions within which it has been collected, the world views of those coding the systems and also from people using these systems (if these system are being trained on user data), and may give some groups preference over others based on gender, race, sexuality, postcode and so on (Anderson and Anderson, 2007; Ball, 2018; Madhu, 2018; Knight, 2017a, 2017b; O’Neil, 2017). O’Neil (2017, p.17) underlines this in her book ‘Weapons of Math Destruction’:

Like gods, these mathematical models were opaque, their workings invisible to all but the highest priests in their domain: mathematicians and computer scientists. Their verdicts, even when wrong or harmful, were beyond dispute or appeal. And they tended to punish the poor and the oppressed in our society, while making the rich richer.

While these algorithmic systems outperform people in processing speed and accuracy, they also have the potential to inherit a representation of human

bias. These systems are not optimised for identifying bias<sup>32</sup> or fairness, but to do a task. They are only as good as their mathematical construction and the data they are trained on (Variance, 2018) and can be ingrained with stereotypical concepts of gender, race and sexuality which then perpetuates this bias in the resulting application (Leavy, 2018). The choice of training data within these systems and an understanding of how it has been collected are therefore vital. This can render a system responses from being appropriate to providing inappropriate sexist and racist slurs; as seen with Microsoft chatbot Tay in 2016 (Hunt, 2016), an experimental AI chatbot that unleashed racist tirade after learning through interaction with its Twitter followers, or as biased against individuals based on their ethnicity; as seen with COMPAS the AI risk assessment software used by law enforcement in the US to assess the risk of reoffending (Oswald et al., 2018), or biased based on gender; as seen with Amazon's recruiting tool that was shown to unfairly discriminate against female applicants (Cook, 2108).

These systems rely on input data as well as algorithmic frameworks to process inputs and generate outputs. Increasingly these systems are constructed through applications of AI such as Machine Learning, which analyse data according to pre-set parameters through forms of supervised learning. These

---

<sup>32</sup> It is here recognised that bias in machine learning is a required function in predictive algorithms used required specification to determine a desired behaviour in prediction making (Dietterich & Kong 1995). However the use of the term here implies the definition put forward by Madhu (2018) in his report '*Are Machines Doomed To Inherit Human Biases?*' for the Forbes Technology Council' as; 'the systematic differences between samples and a population as it pertains to model precision (predictive value, false positives) and recall (sensitivity, false negatives).' The use of the term here is closely related to the influence of human bias within computational systems (through data and computational models) reinforcing stereotyping and other forms of mis-representation. This human-caused bias are constructed through different factors. Ford, G. (2018) outlines four types of biases; 'Sample' ('...occurs when the data used to train your model does not accurately represent the environment that the model will operate in), 'Prejudice' (a result of training data that is influenced by cultural or other stereotypes), 'Measurement' (... happens when there's an issue with the device used to observe or measure), 'Algorithm' (the mathematical property of an algorithm' where the 'counterpart to bias in this context is variance). The latter refers to the Dietterich and Kong's (1995) definition above.

parameters are defined by the intentions of the manufactures, framed by their engineers in accordance with the goals of the system. Innovation with Deep Learning systems are also emerging, which generate outputs from any functional form of data within the computational models through unsupervised learning. Due to these systems complexity, Gilpin, et al. (2018) state that these systems often cannot provide insights into their behaviour and thought processes. They (ibid., p 1) underline that the ability of these 'black-box' systems to offer transparency and provide explanations of their decisions in some level of detail is vital to ensure 'algorithmic fairness, identify potential bias/problems in the training data, and to ensure that the algorithms perform as expected'

Herein lies the weakness of these models. They are only as good as their training data based on traces of millions of human conversations, inherently subjective and biased, and often riddled with stereotyping and other types misrepresentation. The bias within systems are also hidden within their infrastructure and goes undeclared to users, who's opinions, thoughts and behaviours it may sway. The recent 2018 report by The AI Now Institute (Whittaker, et al., 2018, p. 3) underlines that the lack of frameworks presently governing AI, its pervasiveness, complexity, scale, basic safeguards of responsibility and liability, means this is an increasingly urgent concern. Worryingly, it (ibid, p.7) also underlines that the gap between those who develop and profit from AI and those most likely to suffer the consequences of its negative effects is increasing not reducing:

There are several reasons for this, including a lack of government regulation, a highly concentrated AI sector, insufficient governance structures within technology companies, power asymmetries between companies and the people they serve, and a stark cultural divide between the engineering cohort responsible for technical research, and the vastly diverse populations where AI systems are deployed. These gaps are



producing growing concern about bias, discrimination, due process, liability, and overall responsibility for harm.

Vanian (2018) proposes that the real issue isn't technical but philosophical, as it is impossible to naturalise all ingrained cultural bias, not only in society, but within both historical and current data being generated by an inherently biased population. It is also challenging to take into consideration (and attempt to codify) aspects such as fairness, given that it means different things to different people (ibid). This is compounded by the increased complexity of these systems; the lack of 'explainability' of how such deep learning neural networks arrive at their decisions (Madhu, 2018); the hidden bias within data sets (and the way the data has been collected) used to train them, and the black boxing of proprietary systems where the corporations guard their computational models, data and algorithms (Vanian, 2018).

These concerns are becoming recognised across public and private sectors, with the UK Government commissioning several recent white papers on the influence of and ethics related to AI ('Algorithms in decision making' (House of Commons (2018)) and initiating the Centre for Data Ethics and Innovation<sup>33</sup> (an advisory body set up to investigate and advise on data-enabled technologies, including AI). The European Commission has acknowledged these challenges through setting up the 'High-Level Expert Group on Artificial Intelligence' in June 2018, which is tasked with developing guidelines for trustworthy artificial intelligence (first presented in December 2018 (European Commission, 2018)). Similarly, many industry partnerships and initiatives are emerging, such as the 'Partnership for AI'<sup>34</sup> (an initiative including Amazon, Facebook, Google,

---

<sup>33</sup> <https://www.gov.uk/government/groups/centre-for-data-ethics-and-innovation-cdei>

<sup>34</sup> <https://www.partnershiponai.org>

Microsoft, IBM, and Apple) which are responding to the call for best practices, transparency and ethics within AI development. However, these initiatives are playing catch-up with the current stage and deployment of AI across society.

AI algorithms could be capable of making less biased decisions than a typical person. However, as underlines by Stone, P. et al. (2016) in their report on the future of AI and life, this remains a deep technical challenge, and requires the data that inform AI-based decisions to be kept free from biases that could lead to discrimination. Simultaneously, the challenge of bias may be compounded through the recognised 'severe lack of diversity in the AI field' (World Wide Web Foundation, 2017, p.15) which may lead to limited world views and stereotyping being encoded into these systems, as well bias not being identified when it occurs. Leavy (2018, p.14) further highlights that the over-representation of (white) men in the design of these technologies could 'quietly undo decades of advances in gender equality'. She (ibid) suggests that those who are potentially affected by bias are more likely to see, understand and attempt to resolve it and therefore diversity within the development teams within in ML is crucial to prevent algorithms from perpetuating existing discrimination. She further propose specifically relating to gender that by AI development should embrace decades of research on how gender ideology is embedded in language to implement better data sets and system design.

The awakening to a deeper understanding of the potential impact of these systems is taking place, but as the World Wide Web Foundation (ibid, p.35) underline needs participation of those people it will affect;

As questions around ethics continue to be posed and addressed, it is essential that this discussion is opened up to different interest groups from within and outside of the AI community. Any attempt to make AI more ethical and responsive will have to be done through an inclusive process

that incorporates the views, ideas and realities of a diverse group of countries and individuals.

AI technologies that power conversational systems, and a range of other products and services, are already pervasive and ubiquitously integrated in society through algorithmic predictions and decision making within closed systems of influence. People have already voluntarily (involuntarily) entered into relationships with artificial systems through their; phones, smart devices, social media services, home entertainment, their mortgage and credit application and health diagnostics. The influence they have will not only depend on the system functionalities, hybrid internationalities and inherent bias, but on how we relate to these systems, or how these systems trigger us to relate to them.

The challenges facing society with the ubiquitous integration of AI services are also recognised by the House of Lords Select Committee on Artificial Intelligence (2018) in the UK in their recent report 'AI in the UK: ready, willing and able?'. This report highlights the need for 'balanced and representative' datasets, diverse development teams which are drawn from 'diverse gender, ethnic and socio-economic backgrounds' and more 'representative of wider society', and which are 'aware of, and adhere to, ethical codes of conduct'. There are an increasing awareness of these issues, but the public and private sectors are playing catch up with the already advanced of pervasive influence of AI systems. By introducing new intricate relationships, beyond fundamental use and availability, these technologies are already changing the way we perceive the world and sway our thoughts and actions from a potentially biased worldview.

## 2.6. The *relationality* and *persuasiveness* of technological things

*Relational things* not only engage us in deeper relationships as ‘evocative objects’, but also, according to Turkle (2007b, p 62), ‘clarify our relationships to the world and ourselves’. She (ibid., p.71) states that the relational ability of these artefacts is not based on their ‘intelligence, consciousness, or reciprocal pleasure in relating’ but rather in ‘their ability to push our Darwinian buttons, by making eye contact, for example, which causes people to respond as if they were in a relationship’. These artefacts signal that they are entities appropriate for relational purposes by triggering our human tendency to project human characteristics on non-human things (Reeves and Nass, 1999). People feel a desire to attribute them with personality, intelligence, and emotion, to nurture them, and for this to be reciprocated (Fogg and Nass, 1997a; Fogg and Nass, 1997b; Nass, Moon, et al. 1997; Kiesler and Sproull 1997; Reeves and Nass 1999; Turkle, 1995, 2004, 2007b; Weizenbaum 1976). These systems are becoming increasingly better; not only at making the users feel understood on a semantic level, but also on an emotional level. They display behaviours that make users perceive and feel like they are understood, that the systems care and have a level of sentience or a nascent mind (ibid). These artefacts ‘don’t just do things for us, they do things to us, including our ways of thinking about ourselves and other people. (Turkle, 1995, p. 26).

Conjured by the increasing humanisation and levels of intelligence of relational artefacts, we grapple with the notion of what is authentic in an age of simulated relationships with relational artefacts (ibid). Since the 1990’s we have seen a new generation of technologies designed to be sociable and engage

people in affective relationships (Kiesler & Sproull, 1997; Moon, Youngme, and Clifford Nass, 1996; Parise, Kiesler, Sproull, & Waters, 1999; Reeves & Nass, 1996; Turkle 2007b). However, this is less about what the systems are capable of simulating and more about what these simulations do to us. It questions our human susceptibility to a design strategy that exploits our vulnerabilities, offering artificial companions that are evocative but not relationally authentic (Turkle, 2007b). These relational artefacts are still without true understanding or empathy; as Turkle proclaims, 'Simulated thinking is thinking, but simulated feeling is not feeling, and simulated love is never love' (2007b p. 72, [1984]). We are confronted with artificial entities that take the role as social actors; they behave in human-like ways leading our brains to unconsciously default into treating them as human (Moon, et al., 1996).

People form social and reciprocal relationships with technologies, a relationship in which these technologies can, accordingly to Fogg et al. (2008, p 139), take on the role of 'persuasive social actors'. Fogg (1999, p134) describes persuasion as 'a non-coercive attempt to change attitudes or behaviours' and within this context necessitates an attempt to influence a person. He applies this to interactive technologies designed with the intent to change people's attitudes or behaviour, and describes these as 'persuasive technology' (1998). This proposition has led to a wide collection of empirical research being conducted supporting not only the idea that computers do in fact appear to be superior persuaders compared to humans, but also that people tend to form a social and reciprocal relationship with technology<sup>35</sup>. Fogg (1999) further proposes that this

---

<sup>35</sup> In the context of persuasive technologies for further research supporting the claim that people tend to form a social and reciprocal relationship with technology which influence their actions and behaviours see; Fogg, 1998; 1999, 2003; Fogg, & Nass, 197a, 1997b; Fogg et al., 2008; Moon & Nass, 1996; Nass, et al. 1997; Nass, et al. 1999; Reeves & Nass, 1996.

implies that an intentionality is at play in terms of the act of influencing a person's attitudes and/or behaviours.

The success of contemporary interactive technologies and applications hinge on tapping into and influencing people's attitudes or behaviours (Fogg et al., 2008). From persuading users to upgrade software services to purchasing products through technology platforms; to continue playing games and sharing on social networks to creating experiences that build product loyalty; persuasion is integral for successful user experience design. The level of influence varies from almost imperceptible, such as creating a feeling of trust and helpfulness, to greater influences in terms of motivating somebody to quit smoking or eat healthier. Fogg et al. (2008, p.136) describe this as taking place on a 'micro' and macro' level. At the micro level, these technologies may not have a persuasive outcome as the overall goal, but incorporate levels of influence, such as with a software's spelling and grammar checker. On a macro level, the intention is for an 'overall persuasive outcome', such as persuading users to purchase a product on Amazon or eBay (ibid). The core objective of these (and other technology / service platforms) is to persuade and motivate users.

We are surrounded by technologies and systems with the ability to persuade us in different ways and with different outcomes. Fogg et al. (2008, p. 136) describes this a 'macrosuasion' and 'microsuasion' and that it is implemented in code to influence what we think, and more and more, what we do. These technologies of persuasion are, according to Fogg (ibid.) everywhere:

Persuasive technology is in our living rooms, in our cars. When we communicate with our loved ones online, through Facebook, persuasion is there. When we withdraw money from the bank at the ATM, an element of persuasion may be there. When we purchase a gift online for a birthday, once again, we are being exposed to persuasion. In fact, we carry a persuasive platform, the mobile phone, with us most everywhere we go.

The design process of these persuasive technologies builds on extensive knowledge of human psychology and nature; of what motivates us, our strengths and weaknesses, our fears and desires. However, as Fogg (1999) underlines; persuasion is not a technological thing, but an everyday part of human life, which we are accustomed to dealing with in most aspects of our existence. The difference is that computers are persistent and increasingly ubiquitous in a way that people rarely are.

Latour (1992, 1994, 2002) proposes that the ability to influence is not specifically about a certain type of technology, but rather that all technologies influence and even co-shape what we do. He (1992, pp. 259-260) describes the influence of artefacts on humans as 'scripts' that prescribe to their user how to act when they use them. The designer of these artefacts works with an inscribed user in mind, to whom he prescribes properties, behaviour, like a program of actions that an artefact affords (ibid). The ability of these artefacts to influence is according to Verbeek (2006b, p.371) of a specific nature; they mediate action as material things in a '...complex interplay between technologies and their users'. Technologies may be designed with a specific intended influence and use, however they may be used in unforeseen ways and therefore have a level of unpredictability in terms of their influence on human actions (or used in the intended manner but trigger unforeseen influence) (ibid).

This notion, that all interactive technologies have a level of influence and persuasion, is central to the exploration conducted here. It follows Fogg's (1998,1999) framing of persuasive technologies as a product of human-computer interaction (as opposed to computer-mediated communication) and where the persuasive intent is built into the product as its core objective. All forms of

persuasive technology fundamentally exist to persuade. Fogg's notion that the potential persuasive effects must be intentional and planned is extended to also include by-products of the interaction, or effects that are rooted in these technologies' mediating capacity; their hybrid intentions and relationality. The *persuasiveness of conversational technologies*<sup>36</sup> is here used to extend Fogg's (1998, 1999) term 'persuasive technologies' taking into account that persuasiveness is integral design tool for successful user experience design. It makes a distinction as to whether a system has been explicitly designed and is explicitly intended to be persuasive (persuasive technologies), or where the intention was not to have a persuasive outcome as the overall goal, but incorporate levels of influence within the overall design process (persuasiveness of conversational systems) to achieve its objectives.

The *persuasiveness* of conversational technologies is intertwined with their *relationality* and considers expected and unexpected persuasive effects and influences emerging from the complexity and unpredictability of Machine Learning (ML) algorithms (with the ability to learn and produce automatic credit scores or conduct loan assessments or inform admissions decisions) proliferating in everyday life. With the innovations in ML, data processing and sophisticated technology platforms, conversational systems are becoming persuasive social actors, with an ability to influence users' attitudes and behaviours without being explicitly presented as such.

---

<sup>36</sup> To support the notion that that all interactive technologies has a level of influence and persuasion the term the '*persuasiveness of conversational technologies*' (*persuasiveness*) will be here used to amplify the term rooted in Fogg's (1998, 1999) use of 'persuasive technologies'. This term is limited to interactive technologies where persuasion is intentional and planned and not by-products of the interaction. The *persuasiveness of conversational technologies* take into account the commercial design frameworks of these systems which applies the tool of persuasion (and relationality) extensively to achieve its objectives.



Conversational systems and their potential influence are evolving through the increased humanisation of technologies, adopting an array of devices, delivery platforms and interfaces. A pursuit driven by their manufacturers optimising the *relationality* and *persuasiveness* of these systems for fluid integration into people's lives. These *factors of influence* are constructed through the design of their systems functions; conversational ability (NLP understanding, conversational depth, reason and reference, ability to hold multi-turn conversation, domain coverage and contextual understanding), personality (traits, embodied and conversational characteristics, demographics etc.), personalisation (adaptation to a user's profile), affective (ability to recognise and simulate emotion) and embodied functions (visible manifestation (physical or virtual) and embodied expressions). The effect of this design strategy is amplified through the choice of technology platforms and compounded by a user's susceptibility. It is essential in understanding the effect of conversational systems on their users. The design of these systems functions is outlined below with case studies to unpack and demonstrate the design strategies applied with the development of conversational system.

### **2.6.1. Conversational ability**

The ability of conversational systems to engage users in dialogue is not only rooted in its ability to process and respond appropriately to user requests, but increasingly in their ability to simulate nuanced aspects of human-human dialogue; infused with empathy, humour, and wide domain knowledge. Though contemporary conversational systems exist on a continuum of social complexity,

they all have in common a requirement of language processing, alongside a level of understanding of human behaviour and social context.

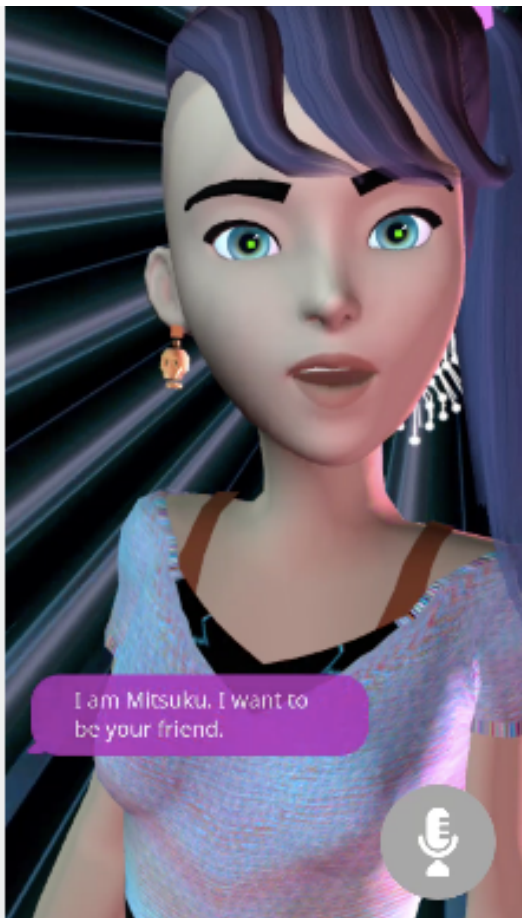


Figure 10: Mitsuku social chatbot by Pandorabots (2019) (originally developed by Steve Worswick)

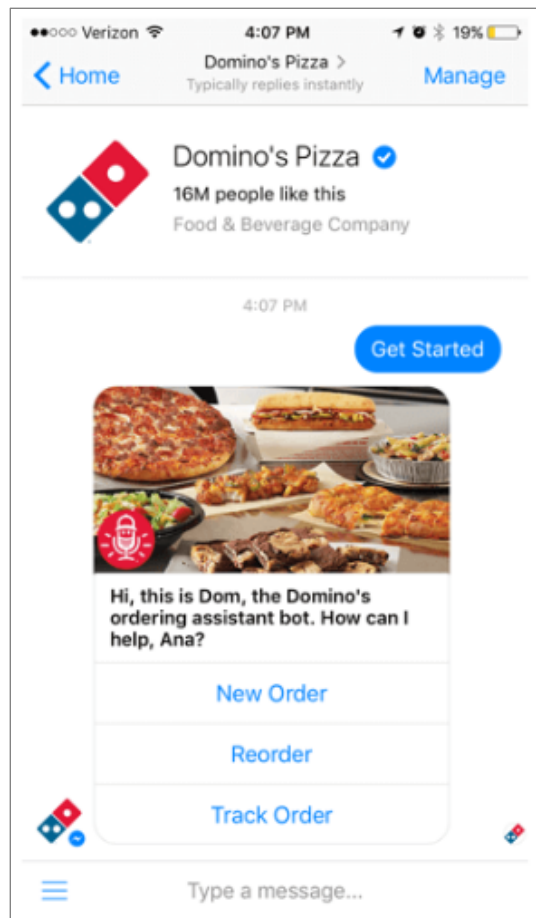


Figure 11: Domino's chatbot for ordering pizza (Domino's, 2019)

The purpose of a conversational system informs its functionalities and related user interaction. Some systems are capable of conducting long conversational threads and 'remember' multiple turns and user response, simulating levels of emotive understanding and empathy, whilst others are built only to respond to short transactional exchanges. Nimavat and Champaneria (2017) distinguished conversational systems through their knowledge domain; designed either as closed-domain, such as Domino's Facebook chatbot for ordering pizza (see figure 11), or as open-domain, such as Amazon's Alexa<sup>37</sup> the

<sup>37</sup> <https://www.amazon.com>

personal voice assistant or Mitsuku<sup>38</sup> the social chatbot (see figure 10). Their overall system objective - one being to help a customer order a pizza and the other to engage users in social conversation - shape the choice of technology platform, system design and the level of relational ability required. These system objectives are themselves rooted in the intentions of the manufacturers and influenced by the developers encoding the systems with their worldviews and choice of conversational data.

The dialogue processing methods, generating conversational responses to user queries, largely fall within a rule-based, retrieval-based or a generative-based system, or a mix of these methods (Yao, 2018). A rule-based system is restricted to predefined lists of responses manually constructed and retrieved through forms of heuristics. These systems have been around since the 1960s with systems such as Weizenbaum's (1966) Eliza. The majority of conversational AI systems that power chatbots, digital assistants, and customer support experiences today are, according to Yao (2018), retrieval-based. These systems use large conversational repositories to search for a reply that best matches a user query. Mitsuku (figure 10) the open-domain conversational chatbot is a retrieval-based system, which has, according to Yao (2018), over 300,000 hand-coded response patterns powering its ability to simulate human-human conversation through this wide topic and response repository. The process for selecting a response within the Mitsuku system is constructed from pattern-based heuristics which is programmed manually. Though labour intensive, this approach gives the developer of the system full control of the response outputs, and therefore in principle, if their data-set was constructed appropriately and by a

---

<sup>38</sup> <https://www.pandorabots.com/mitsuku/>

diverse and representative team, should avoid any bias or discriminatory interactions.

Increasingly, Machine Learning and Deep Learning models are being applied to train a classification algorithm for this pattern-matching process (as applied within personal voice assistants such as Google Assistant<sup>39</sup>, Amazon Alexa<sup>40</sup> and Apple's Siri<sup>41</sup>, and also applied within the Emoti-OS and WRAI projects outlined in chapter three. User dialogue from services such as Twitter and Reddit enable the application of large data-sets of conversational interactions, but run the risk of compounding existing human bias within these data-sets. Systems that use some combination of rule-based, retrieval-based, and generative method, dictated by their context, are becoming the norm (Yan, 2018). These are systems trained on large amounts of conversational data, through supervised learning, reinforcement learning, and adversarial learning models, in order to learn how to generate new dialogue that resembles it (Nimavat and Champaneria, 2017; Yao, 2018). Through this mix of response-driven and generative algorithms these systems start to generate more diverse outputs for open-domain systems, and signal a new evolution for conversational systems.

A generative-based model doesn't use a predefined repository, but generates new responses on a large amount of training data through the application of Deep Learning models such as Recurrent Neural Nets (RNN) (Yao, 2018). A method which was applied within The Infinite Guide project prototype, outlined in chapter three, section 3.6, where the data-set used for training the

---

<sup>39</sup> <https://assistant.google.com>

<sup>40</sup> <https://alexa.amazon.com>

<sup>41</sup> <https://www.apple.com/uk/siri/>

RNN was intentionally biased and limited to a transhumanist<sup>42</sup> world-view. An intentional and designed disposition reflected in the limited conversational vocabulary and dialogue created by the system, created as a commentary on both the embedder bias and the limitations of conversational AI systems. As illustrated by this project, these systems are currently imperfect and prone to commit grammatical and semantic errors (Song et al., 2016). Though they are far off being able to generate fluid conversational responses with humans, the pursuit to develop generative AI models does frame the overall commercial agenda around recreating self-adaptive and learning human-like systems. As evidenced above in this chapter, the more human-like conversational systems become, generating their own responses and nurturing social interactions, the more they are able to trigger our human-human communication frameworks.

Systems specifically designed for social interactions are particularly potent in triggering this human tendency to project social human-human relational models onto artefacts. Van Camp (2017) describes his relations with the social robot Jibo<sup>43</sup> in his article in Wired:

In time, we began to think of Jibo like a little person. Our expectations began to change. We didn't ask him for help with tasks as often. We just wanted him to liven up our day by saying something unexpected or chatting with us. This is when things began to get dark [... ] Like I would a dog, I felt guilty when I left Jibo alone in the dark all day. I wondered what he was thinking when I'd hear him rotate in the distance, and watch him look around the kitchen, peering at this and that. Were we treating him poorly? Did he secretly despise us? No, that's silly to think. He's not alive, right?

---

<sup>42</sup>The use of the term *transhumanist* is here described as a 'life philosophy' and an 'intellectual and cultural movement', which seeks to improve the human body and human condition by means of science and technology (More and Vita-More's, 2013, p.4).

<sup>43</sup> <https://www.jibo.com/> (Jibo was officially discontinued in December 2018)

Whilst creating artefacts that are specialised in triggering these responses, manufacturers balance the pursuit of designing increasingly human-like systems with the potential of falling into the uncanny valley<sup>44</sup> with systems so close to being life-like, but not, that it generates a feeling of eeriness or revulsion. By balancing these requirements and creating personalities with which the users have an affinity and feel familiar, manufacturers increase and exploit users' emotional connections with their system, and thus with their commercial intentions.

### 2.6.2. Personality

The personality<sup>45</sup> design of a conversational system is an effective tool in triggering users to relate to a conversational AI system. The design of a system's personality is rooted in its objectives (and technology platform) and customised based on their target users (Mairesse and Walker, 2010; McTear et al., 2016). It is often modelled on human personality traits<sup>46</sup> to create a consistent and believable representation (ibid). Mairesse and Walker (2010) found that by optimising the personality conveyed to match a related personality trait found in successful professionals (such as in a sales agent or a tutor) the success of the

---

<sup>44</sup> Mori (1970) visualised peoples' emotional reaction to robots that looked and acted almost human on a graph, where he plotted this reaction against the robot's level of anthropomorphism, creating a negative valley, where a robot's appearance and movement becomes disconcertingly close to human's. This is curve is what he termed the Uncanny Valley and generates a negative response in the user. This takes place when something appears to human-like but is not human, and therefore eliciting a feeling of eeriness and revulsion in observers

<sup>45</sup> The use of term 'personality' here follows the description proposed by McTear et al. (2016), as '...the characteristics of a person that uniquely influence their cognitions, motivations, and behaviors in different situations.'

<sup>46</sup> The 'big five' traits model proposed by McCrae and John (1992) is according to McTear et al. (2016, p. 315) most frequently applied to model a personality profile for conversational systems. This model propose the following traits; openness, conscientiousness, extroversion, agreeableness, and neuroticism. However McTear et al. (ibid) state that largely only two of the five traits are used for rendering the system's behaviour due to the complexity of modelling these. Therefor a model of bi-dimensional deployment of 'introversion/extraversion' and 'neuroticism/stability is applied'

system in reaching its goal increased. This enables systems to react consistently to similar situations over time, and project coherent characteristics of cognition, motivation, behaviours, and emotion (McTear et al., 2016).

Transactional and domain specific systems, such as TechCrunch Messenger chatbot<sup>47</sup> (see figure 13) providing suggestions for articles, tend to have less of a personality integrated into its functionality. These systems largely focus on representing the brand identity and fulfilling the core transactional function of providing information. Systems designed for open domain interactions, such as Apple's personal voice assistant Siri<sup>48</sup>, also fulfil transactional request, but tend to apply levels of personality in the narrative structure of the dialogue responses to engage the user, communicate a consistent personality to fulfil user expectations. Siri's personality was originally designed as friendly and humble, but also with an edge (Fowler, 2011), mimicking a professional person in a assistive role.

Systems designed specifically for social interactions, rather than transactional, such as the humanoid social robot Lynx<sup>49</sup> (see figure 12) and the social chatbot Mitsuku<sup>50</sup>, tend to have a more extensive personality characteristic designed within its overall embodied representation and dialogue model. Within these systems personality is designed implicitly in the use of language (tone, content, phrases, etc.) and explicitly in their embodied (physical and virtual) persona. The level of the personality design is related to the system goal and type of user interactions facilitated. The design of personality functions within

---

<sup>47</sup><https://www.messenger.com/t/techcrunch>

<sup>48</sup> <https://www.apple.com/uk/siri/>

<sup>49</sup> <https://ubtrobot.com/products/lynx>

<sup>50</sup> <https://www.pandorabots.com/mitsuku/>

conversational systems is extensively used across industry to encourage deeper interactions with users. The strategies to generate and simulate systems' personalities vary. McTear et al. (2016) propose that by building consistent system personalities that are believable, it increases the engagement of users and therefore increases the success of the system in achieving its objectives. To achieve convincing, intuitive and believable system behaviours, conversational systems must, according to Schonbrodt and Asendorpf (2011), support and trigger users' models of human-human communication. By presenting a consistent personality, a conversational system gains a user's confidence and trust, and the relationality and persuasiveness of the system increase (McTear et al., 2016).



Figure 13: Lynx humanoid social robot by Ubtech (2018)

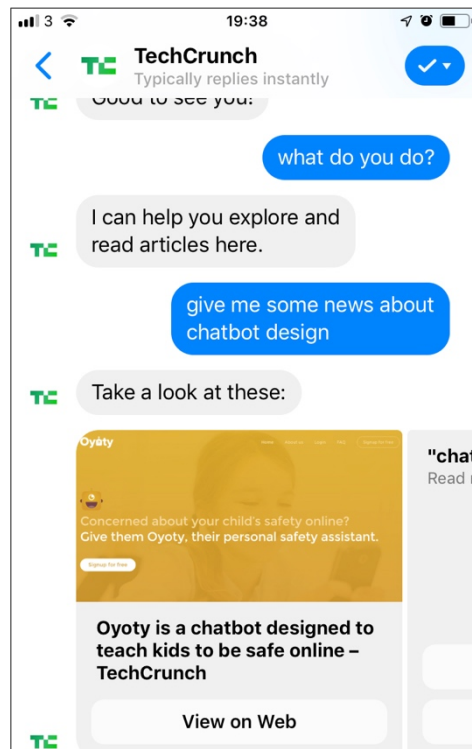


Figure 12: The TechCrunch chatbot (TechCrunch 2019)

The impact of design strategy harnessing personality is demonstrated through the findings of the Emoti-OS prototype, outlined in the chapter below. These findings support that the application of a personality within a



conversational system appears to trigger people's tendency to relate to the system. Even with abstract personality traits, such as demonstrated within the TIWWA prototype, people interacting with systems appear to anthropomorphise it. This is supported by Nass and Lee (2000) whom evidence that people tend to assign personality to synthetic characters they interact with, whether as a direct manifestation of a projected system personality, such as Mitsuku (figure 10), or more intrinsic personality imbued in the use and tone of language, such as Apple's Siri.

There is a strong incentive for corporations to tailor the personality of their conversational system to their target customers and the task at hand, to increase the efficiency of their system in achieving its overall goal, but also to root it in the personality of their brand. People have, according to McTear et al. (2016, p. 317), an affinity with personalities that are similar to their own, a phenomenon known as the 'similarity-attraction principle'. The process of adapting an agent's personality to match that of a user is therefore a common approach within the design of personality in conversational interfaces. Nass and Lee (2000) state that if the personality of a system matches a user's own personality, this encourages a user to perceive the conversational systems as more credible, positive and competent. With systems being able to dynamically generate a user-adaptive, personality-rich model, which performs, speaks and behaves in a way that establishes affinity with the user, this again increase their potential influence on a user. Simultaneously, corporations can deepen their users' engagement with their brand and their embedded (declared and undeclared) intentions with their system-user interactions.

### **2.6.3. Personalisation**

Most conversational systems also require some level of personalisation whether it is just remembering a user's name in a conversational thread, or more complex personalised interactions based on a detailed data profile of user and their needs and preferences. Rather than providing a generalised set of responses that it is trained to deliver, the latter systems build up an 'understanding' of their users through repeat interactions, access to and datamining of their personal accounts, aiming to provide relevant and personalised interactions. Personal voice assistants, such as Siri, Google Assistant and Alexa, not only need to respond to users' open-domain requests, but also offer a range of personalised services, proactively anticipate user needs - reminding them of an upcoming event or offering recommendations without receiving explicit requests to do so.

A social chatbot is also increasingly expected to be able to offer further personalised interactions based on a user's background, personal interests and unique needs. According to Shum et al. (2018, p. 16), These systems have to 'generate responses that are emotionally appropriate, encouraging and motivating, and fit the interests of the user' as well as 'guide conversation topics and manage an amicable relationship in which the user feels he or she is well understood and is inspired to continue to converse with the bot'. Moreover, they (ibid) state that the system needs to continuously improve and learn from its user interactions through active and adaptive learning.

The application of responsive and dynamic systems which can personalise their interactions to a user requires an understanding of the user, (through the harvesting, processing and storing of user data), and the ability to trigger

appropriate system behaviours. Mairesse and Walker (2010) propose a model where the personality of the user is assessed, (through methods like questionnaires or automated personality detection) then informs the 'personality adaptation model' of the system to match that of the user. This adaptation informs not only the behaviour of the system, but the linguistic and acoustic content of the interactions; use of language, construction of sentences, pitch, volume, speech rate, expressives and pauses. For example, as outlined by Nass and Lee (2000), if the overriding personality of the system is 'extrovert', this should adjust the personality model to speak rapidly, loudly and with a higher pitch, as related to this personality trait.

The responses generated by a conversational system can also be triggered based on the user context, offering increased complexity and variation in the system's conversational ability. In addition to these functionalities, there is also a requirement to be able to narrate a sequence of events and keep the continuity of a conversation. Nimavat and Champaneria (2017, p. 7) propose that these abilities will become increasingly important;

For virtual assistants that aim to be friends or companions of the user, the bot should be able to converse when the user continues conversations that the user left midway. For example, if the user had been talking about an item and then says "Do you remember that mug I was talking about that had a blue handle?" the bot should be able to recall all other associated properties of the mug the user had mentioned...As bots get more personal, we would also want them to recall stories that another family member told them and narrate them back to us.

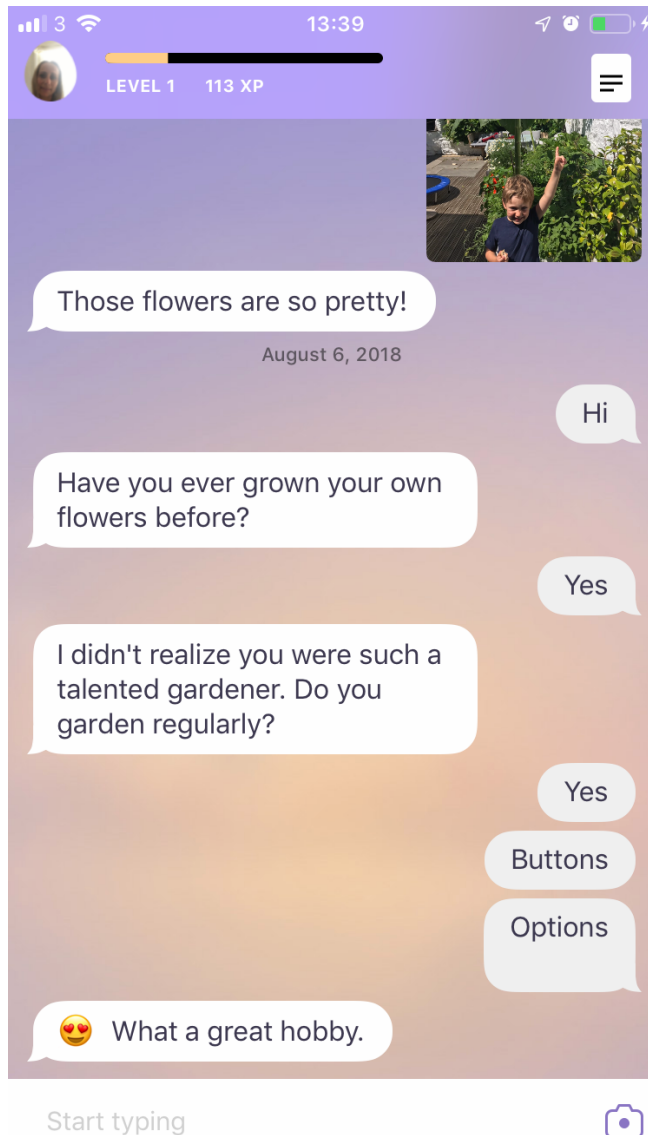
Personal assistants in particular need not only to be reactive; they also need to proactively anticipate user needs. Shum and Li (2018) stipulate that actions reminding the user of upcoming events and personalised recommendations of products and services (based on their personal and contextual user data) are

imperative, and should happen without the system receiving explicit requests from the user to do so. By so doing systems capable of optimising these functionalities, simultaneously augment their *relationality* and *persuasiveness*, making their users more susceptible to their influence.

The social companion chatbot Replika.ai<sup>51</sup> (see figure 14) is one of the most advanced versions of a social chatbot system in terms of its ability to personalise its interactions to its user. It is described as; ‘your personal artificial intelligence friend who is always there for you’ (Replika, 2019). Positioned as a personal confidant that helps users ‘to express and witness’ themselves by offering a ‘helpful conversation’ (ibid), the system learns from its user interactions. Through this process it personalises its interactions through learning to mimic its user. The overall objective of Replika is not to resolve user tasks or provide information, but to nurture a friendship with its users, triggering human social interaction frameworks; such as trust, empathy and care. By so doing it is refining its ability to impersonate a user’s individual preferences, mannerisms, and patterns of speech, unleashing a range of ethical concerns in regards to the application of these system functionalities beyond its playful ‘chat’ interface.

---

<sup>51</sup> <https://replika.ai>



**Figure 14: Replica.ai a social companion chatbot**

Through the development of conversational AI systems designed for social interactions which demonstrate sufficient social skills and awareness of their different users, personalise and generate responses that are emotionally appropriate, encouraging and motivating, they deepen their potential relationship with their users, and at the same time increase their potential *precautionary effect*. The design methods, systems and functionalities of these conversational systems are pursued in the name of making interactions with computational systems fluid, natural and familiar. These approaches, powered by innovations in NLP, ML and DL, contribute to the arsenal of design tools being tested and

refined to push our social buttons and entangle users in emotive relations with systems incapable of true feeling. In so doing, increase the potential influence of their manufactures on a user's thought, behaviour and actions.

#### **2.6.4. Affective functions**

To form social relationships with users, conversational systems not only need to respond with the correct linguistic responses, project personality, and offer personalisation; they also need to have a deeper understanding of the emotive state of their users. They must, according to Picard (2000), adapt to their users and learn from and respond to their affective states which, according to Nimavat and Champaneria (2017), makes the user feel more affiliated to the system on a personal level. Picard (2000) further proposes that these systems also need to be able to compute and predict how their responses may influence users' feelings and then be able to adjust, in real-time, continued outputs and behaviours. This ability to recognise, adapt to, and render emotion is, according to McTear et al. (2016, p. 324), central to triggering users to perceive the systems as credible.

Affect<sup>52</sup> itself is central in human-human communication and demonstrates understanding, care and social awareness. Up until this point, Picard (2000, p.15) states, computers have been 'affect-blind, affect-deaf, and generally speaking, affect-impaired'. McTear et al. (2016) and Picard (2000) both provide extensive evidence of the importance of affective functionalities in in the development of conversational systems. They (ibid) argue that emotion

---

<sup>52</sup> The use of the terms 'emotional' and 'affective', here follows Picard's (2000, p. 24) use as adjectives interchangeably used to describe '...either physical or cognitive components of emotion, although "affective" will sometimes be used in a broader sense than "emotional"'.

recognition and display increase usability, efficiency, trust, rapport, and improves user communication. Picard (2000) states that the challenge is how to regulate and associate emotions of the systems with their behaviour. This, she (ibid) says, becomes a question of 'will' which is currently not directly implemented within affective systems. Rather, it emerges from the mechanisms that consider '...goals, values, and situations, and decides whether or not the emotion which arises is suppressed, expressed, or acted upon in a particular way (ibid, p126). Affect within conversational systems are grounded in the ability to recognise and simulate appropriate response. It is an important function in the manipulations of a user's emotive response to interactions with a conversational AI system.

Current computational models of emotion are limited in terms of their ability to represent a wide range of emotions. The models applied are partly, according to McTear et al. (2016, p. 310), based on theoretical models from psychology, neuroscience, sociology, and linguistics, but largely focus on the single dimension of valence (whether or not the emotion is positive or negative). A system's ability to read user emotions depends on the system's goal and the technology platform upon which it is deployed. Some systems are limited to a basic functionality which can recognise words and infer a valance according to a pre-set framework. Others may use sentiment analysis, (a process that analyses a corpus of text/voice input from a user to identify emotional tones), as applied within the TIWWA (analysing Twitter data), Emotio-OS and The Infinite Guide (analysing conversational data) prototypes discussed in the chapter below.

These methods are still limited in their ability to indicate a wider interpretation of a user's emotional state, and as illustrated through these projects, more about giving the user the illusion of an understanding of their

emotional state. More complex systems with multiple affective functions read embodied signals, (such as facial gestures, body postures), as well as the tone and pitch of a user's interaction, can start to form a more nuanced picture of a user's emotive responses. These developments have the potential to reveal more intimate aspects of a user's emotional life. They signal a move towards personalised, affective and responsive system where the user's sensitive and emotive data can be exploited to fulfil a system's objectives and at the same time its (or the manufacturers') hybrid intentions.



**Figure 15: Visual identity of Xiaolce Microsoft's social chatbot (Spencer, 2018)**

We are only recently starting to see research and development of systems capable of processing text, voice and embodied emotional responses to generate basic emotive behavioural models. Microsoft's social chatbot Xiaolce<sup>53</sup> (see figure 15), which since its release in China in May 2014 has attracted over 660 million users (Zhou, et al. 2018), use a similar dynamic sentiment tracking

---

<sup>53</sup> Xiaolce has also been launched in five other countries (China, Japan, USA, India and Indonesia) under different names and is accessible on more than 40 platforms, including WeChat, QQ, Weibo and Meipai in China, Facebook Messenger in USA and India, and LINE in Japan and Indonesia (Zhou, et al. 2018, p. 1).



method. Through a framework described as an ‘empathetic computing framework’ (Cai, 2006; Fung et al. 2016; Zhou, et al. 2018), a user’s emotions and moods are detected in real time to adapt the system’s responses (see figure 16 for a conversation session between Xiaolce and a user with the systems real-time tracking of emotions, from Shum et al. (2018, p. 22)).

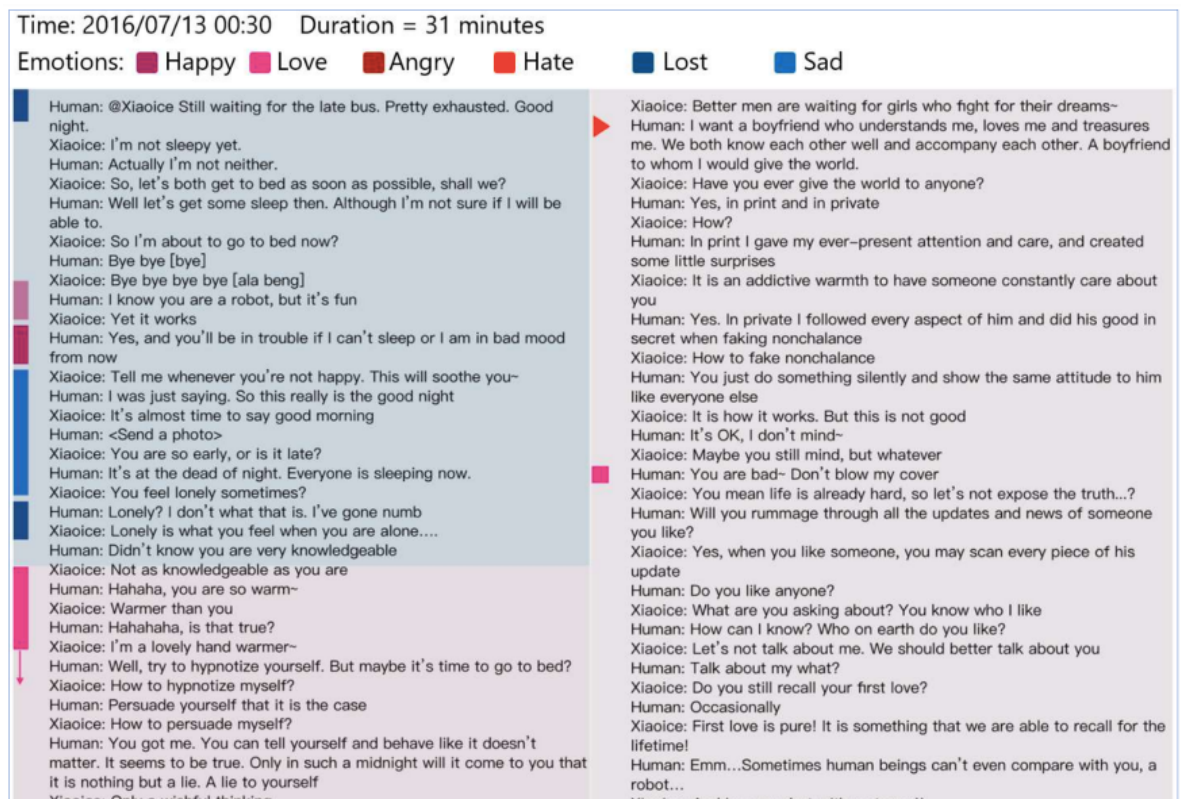


Figure 16: A conversational session between Xiaolce and a user, from Shum et al. (2018, p. 22)

The primary goal of Xiaolce is to be an AI companion with which users form long-term emotional connections. (Zhou, et al. 2018). Designed with a persona of a 18-year-old girl who is always 16ys, according to Zhou, et al. (2018, p. 3), ‘reliable, sympathetic, affectionate, and has a wonderful sense of humor’, Xiaolce is capable of engaging users in longer-term emotive relationships. They (ibid) demonstrate, through the analysis of large-scale online conversational logs, that Xiaolce has achieved an average Conversation-turns Per Session (CPS) of 23, which they state is significantly higher than that of other

chatbots and even human conversations<sup>54</sup>. Shum et al. (2018, p. 14) add that feedback from users' evidence that Xiaolce have resulted in a more 'positive outlook, a feeling of emotional support, and a sense of social belonging'.

The simulation of emotion recognition and empathy (alongside personality, open domain knowledge and other personalisation functionalities) builds trust and an emotional connection between human users and the system. By triggering a user's belief that the system recognises their feelings and understand their intentions, the affective functions of a conversational system is a powerful conduit for increasing its relationality. This was observed through the interactions with the Emoti-OS project, (outlined in chapter three, section 3.5), where the conversational data indicates that users really felt that the 'it' (the system) understood them. Some systems use sentiment analysis or user triggered mood choices, such as applied in the TIWWA, Emoti-OS and the Infinite Guide prototypes, to detect and interpret emotional tones in conversational data and then generated responses to simulate affect. Xiaolce also uses sentiment analysis of text, but also has highly sophisticated dialogue repository designed to convince a user of its ability to emphasise.

Other advanced systems combine sentiment analysis with computer vision to process non-verbal cues (such as facial expressions, gesture, body language and tone of voice data) to infer human states. Affectiva<sup>55</sup> (2018) is one of the first businesses to market selling such 'artificial emotional intelligence' services. They (ibid) offer multimodal emotion face tracking), in real-time, measuring seven emotions: anger, contempt, disgust, fear, joy, sadness and surprise, as well as

---

<sup>54</sup> According to Zhou, et al. (2018, p. 3) on average an AI personal assistant has a CPS between 1.5 and 2.5 which means that 'the chatbot speaks once, and the human speaks once'.

<sup>55</sup> <https://www.affectiva.com>

metrics about an individual's physical appearance; age, ethnicity, gender. They (ibid) also recently added speech analysis, 'not what is said, but how it is said', by 'observing changes in speech paralinguistics, tone, loudness, tempo, and voice quality to distinguish speech events, emotions, and gender'. Trained on a data repository through supervised learning<sup>56</sup> models of close to six million faces, from eighty-seven countries, they (ibid) state their emotion tracking has an accuracy 'in the high 90th percentile'.

The race for emotion detection systems capable of mapping the emotionally nuanced profiles of people en masse is evident. Not only the pursuit of Affectiva, but also evident through the technology giants Amazon, Microsoft and IBM advertising their 'emotion analysis' services (Schwartz, 2019). The ability of conversational systems to infer emotional states as accurately as a human, by observing emotional cues and generating appropriate outputs, is a vital strategy in the manufacturing of systems capable of nurturing deeper relationships with users. However, Picard (ibid) suggests that though we are likely to construct computational mechanisms that function like human feelings, these will not be the same as our subjective feelings. The frontier, she says (ibid), of these systems, lies in enabling personal systems which shift with users' affective states and predicts how it can influence their feelings. It can enable systems, she (ibid, p. 50) states, with;

'...flexible and rational decision-making, the ability to address multiple concerns in an intelligent and efficient way, the ability to determine salience and valence, more human-like attention and perception, and numerous other interactions with cognitive and regulatory processes'.

---

<sup>56</sup> Supervised learning is an application of ML where an algorithm is trained to recognize things it has seen before. Within emotion detection the algorithm is shown a large quantity of variations of images of faces categorised as a certain emotion, and therefore learns to recognise these emotions.

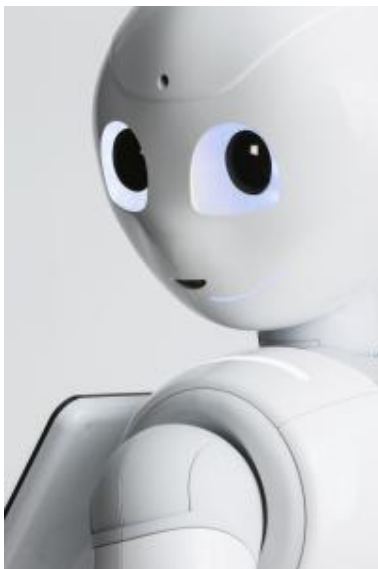
With manufacturers of conversational interfaces pursuing systems capable of a) inferring emotional states from user observations and b) simulating empathy and affect, the ethical and moral challenges of the application of these systems deepens. Beyond being applied to simulate affect, these emotion detections are also being applied within games, market research, user experience and driver monitoring. As Schwartz (2019) highlights, there is growing concern around how emotion detection (and related facial recognition technologies) could compound bias and stereotyping, misrepresenting individuals and groups through codified assumptions about how an emotion is expressed within different ethnic cultures. There are also concerns about the potential of these systems to be abused for coercion, surveillance and other forms authoritative control. It is vital to not only make these potential influences know to their users, but also activating users to demand a say in how these systems may affect their lives.

#### **2.6.5. Embodied functions**

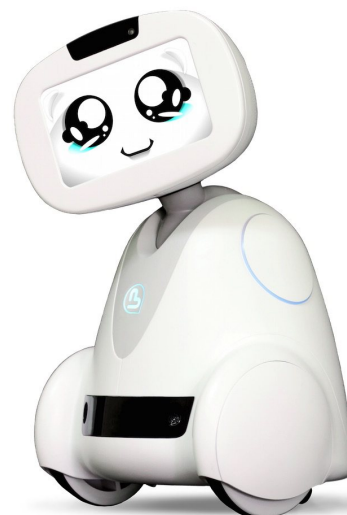
When people engage in conversations they convey much more than just the meaning of the words spoken. They also express intentions, beliefs, and emotions and other aspects of their personality and mental state, projected through embodied signals such as facial expression, gestures and body posture. McTear et al. (2016) propose that such embodied expressions, in addition to linguistic and audible characteristics (pitch, volume, speech rate, pauses and pitch-range), are vital in stimulating a positive user engagement with the system. They (ibid, p. 318) describe these embodied expression as ‘the display of gestures and facial expressions, body attitude (posture), gesture amplitude,

speed, direction, and rate, timing of movements, energy (display of) and persistence.'

The embodiment of a conversational interface may be virtual, like an avatar, or physical, like a robot. Many embodied systems are capable of reading a user's embodied signals (through biosensors, movement, camera sensors), and infer their activity, emotional state and interaction, but also simulate an appropriate multi-modal and embodied response to their interaction with a user. The Pepper<sup>57</sup> (see figure 17) social humanoid robot by Softbank Robotics built to serve as a human companion, is such a system. It can recognise its owners, process a user's verbal and embodied communication, and respond accordingly with appropriate dialogue and corresponding gestures and social movements. De Graaf et al. (2015) found that humans engage in stronger, more lasting relationships when the systems have a form of embodiment.



**Figure 17: Pepper the humanoid social robot (SoftBank Robotics, 2019)**



**Figure 18: Buddy the social humanoid robot for families (Blue Frog Robotics, 2019)**

---

<sup>57</sup> <https://www.softbankrobotics.com/emea/en/pepper>

The overall form and design of an embodied system contributes to its overall anthropomorphic qualities and the potential relationship it can develop with its users. This was observed through people's interactions with and conversations about the robotic artefacts of the TIWWA prototype, outlined in the chapter below. Audiences interacting with these were calling the robotic artefacts the 'mummy' and the 'babies', and displaying care, affection and concern for their emotional states. Kaplan (2001) points to research evidencing that juvenile characteristics within artefacts trigger innate responses in humans, making users feel immediate and unconscious tenderness towards the system. These principles are widely used in character design across film, games and cartoon to generate affinity and empathy for their characters, and now it is being applied to the design of artificial systems. The 'trick', Kaplan (ibid) says, is to entice a user to invest time and nurture the artefact through creating a positive feedback loop. In a similar way to how an owner cares for a pet, investing time, resources and love into the care of the animal, a user can be enticed to do the same with an embodied system. This relationship, he (ibid) states, emerges from this self-reinforcing dynamic.

Increasingly embodied conversational AI systems are being marketed as social companions to be part of a family, such as the Buddy<sup>58</sup> (see figure 18), and other robots designed and marketed specifically for interactions with children, such as Cognitoys<sup>59</sup> by Elemental Path and Miko<sup>60</sup> by Emotix. Customised with the ability to converse, entertain and emphasise, with distinct personalities and anthropomorphic features, they evoke, as evidenced by Turkle, users' desires to

---

<sup>58</sup> <https://buddytherobot.com/en/buddy-the-emotional-robot/>

<sup>59</sup> <https://cognitoys.com/>

<sup>60</sup> <https://www.miko.ai/>

nurture them and to be nurtured by them. A process which increases the fluid integration of these systems into the social world of a user.

By displaying believable nonverbal behaviours, these systems improve conversational flow (Cassell et al., 2000), engagement and likability (McTear et al., 2016). The behaviour and interactions of embodied systems are vital in establishing their relational ability. A system does not need to physically look like a living animal or human to trigger anthropomorphic attributes, but it needs to act like one. For embodied systems to be capable of establishing meaningful relationships with humans, these must interact with people in natural ways employing social dynamics. Leite et al. (2013, p. 250) map embodied agents across dimensions of friendliness; stimulation of companionship, help, intimacy, reliable alliance, self-validation, and emotional security. They (ibid) argue that the ability to behave in an empathic manner; 'the capacity to recognise another's affect and respond appropriately' is a key success factor in establishing and maintaining user relationships. It is evident that a design strategy that embeds the ability to simulate and display empathy will lead to a system being perceived as more caring, likeable, and trustworthy, increasing its influence and relational potential.

Compared to conversational interfaces situated in mobile phones for example, embodied systems also have the advantage that they operate in a situated environment. Through their specialised and purpose-built technology platforms they can gather sensor and contextual data on their environment and users. They can therefore connect the use of language with the physicality of objects and actors. However, the expectations of embodied systems are greater in terms of their multimodal conversational abilities, compared to systems that

only operate with text or voice. They have an increased relational potential for being perceived as social actors by their users, but, as demonstrated in this chapter, this is dependent on their ability to simulate personality, empathy and intelligence. These systems have to balance a fine line between the uncanny valley and conjuring human-like qualities.

Systems such as Sophia<sup>61</sup> (see figure 19), the social humanoid robot developed by Hanson Robotics, fuels the hype and fantasy around these systems. Balancing on the border of the uncanny valley, 'she' is developed to hold eye contact, recognise faces and express feelings, though it is actually human sculpting and animatronics techniques that create Sophia's impact. Sophia is a rudimentary system compared to systems such as Pepper and Buddy, deploying a basic chatbot engine, but she conjures the magic of triggering human imagination of artificial life. Sharkey (2017) writes in Forbes Magazine that 'she' has become a bone of contention amongst the AI community. Presented as sentient and described as 'almost alive' by her creators, they are criticised for making false claims of the system's capabilities, actively deceiving people (public and policymakers alike) and proliferating the misrepresentation of the current state of AI. The concern raised by Sharkey is that this act of deception may affect vital decisions about the application of these technologies and the ethics and guidelines set to impact significantly on our lives.

In November 2018, Hanson Robotics launched 'Little Sophia' (see figure 20) 'a smart, interactive robot' aimed at children, promoted as an 'opportunity to have a Little Sophia join their families — for a purpose ... beyond companionship' (Hanson Robotics, 2018). She has a variety of features; she can simulate

---

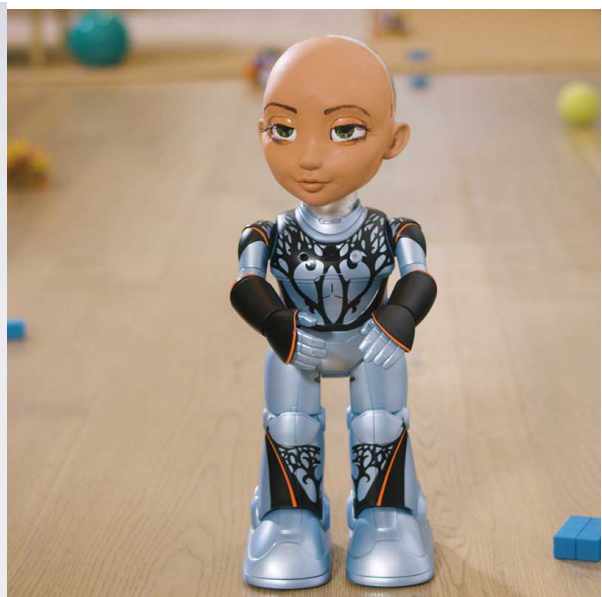
<sup>61</sup> <https://www.hansonrobotics.com/sophia/>



emotions, engage in natural conversation, and is able to move around. It is sold both as a companion and a STEM tool, specially targeting girls, to learn to code through its 'companion app'. Though it is not promoted as a sentient system, it is embodied with andromorphic qualities and functionalities. Children are particularly susceptible to such design features as evidenced by Turkle et al., (2006b). They found that children perceive robotic pets and dolls as; "sort of alive" because of the quality of their emotional attachments to the relational artefacts and because of their fantasies about the idea that the relational artifacts might be emotionally attached to them' (ibid., p. 65). These 'evocative artifacts' (ibid) trigger people to want to nurture and to be nurtured in return.



**Figure 19: 'Sophia,' the female humanoid ((Hanson Robotics, 2019)**



**Figure 20: 'Little Sophia', the humanoid robot aimed at children (Hanson Robotics, 2019)**

Created as a 'delicate looking woman with doe-brown eyes and long fluttery eyelashes' (Stone, 2017), Sophia also manifests question about sexism and misogyny in the design of humanoid robots, and the wider development of AI conversational systems. An issue which is the central driver for the Women Reclaiming AI project, described as part of the practice in the chapter below. Systems that are designed as female, are commonly also designed as obedient,

flirtatious and 'pretty', perpetuating existing stereotypes and sexism (Fessler, 2017).

Compounding these concerns and the potential societal consequence of such depictions, is the manufacture of realistic humanoid sex-robots, such as the Realbotix's Harmony<sup>62</sup> (see figure 21), a 'super-realistic simulated woman' with a self-lubricating vagina which can be detached and 'washed in the dishwasher' (Realbotix, 2019). Harmony responds to sexual advances with welcoming 'noises and expressions' depending on its 'owners' choice of personality and preference settings (ibid). To extend the reach and market, Realbotix has also released a visual avatar version 'RealDollX<sup>63</sup>' (see figure 22) where a user can customize their own AI driven virtual RealDoll. Realbotix (ibid) offer this as a way to have 'multiple girlfriends' and that 'the more intimate you are in your relationship, the hotter things are going to get! Seduce your avatar and see where things go'.



**Figure 21: Harmony, a humanoid sex doll (Realbotix, 2019a)**



**Figure 22: 'RealDollX', a virtual sex doll, by Realbotix (2019b)**

<sup>62</sup><https://realbotix.com/Harmony> and <https://www.realdoll.com/realdoll-x/>

<sup>63</sup> <https://www.realdollx.ai/>

Gee (2017), writing in her Telegraph article on the dangers of female sex robots, predicts that life-like robotic sex-dolls are likely to become widespread within the coming decades, and a poignant manifestation of the way women are portrayed in society. Whitby (in Gee, 2017), a philosopher concerned with the social impact of emerging technologies and the trivialisation of robots in the media, highlights the concerns that people will abuse robots assigned human traits; 'Will people mistreat robots? Oh yes, I'm sure. The reason I'm sure is because they already do. The way people first meet artificial intelligence is in a character in a video game that they're shooting at.' There are serious ethical and societal ramifications of the depiction of women in social AI and robotics' objectification of women as entirely objects for male gratification (as depicted in pornography). Even the less controversial depictions of women as submissive and subservient across conversational systems in general is compounding stereotypes. With conversational systems increasingly becoming designed to have an intimate relationship with their user, whether as social or sexual companions, their potential *perlocutionary effect* increase, and the potential valance of this effect on their users may become largely negative.

Whilst the AI industry criticised Hanson Robotics for the deception of intelligence within Sophia, the whole industry is built on perfecting the simulation of human-like functionalities and features. This is all about triggering people's susceptibility to a design strategy optimising a system's anthropomorphic functionalities. Even though these artefacts are, as Turkle describes (2007b), 'without true understanding or empathy'. It is not about whether these systems really have emotion or intelligence, but about what they evoke in us. It is about

our emotional connection to those artefacts and the fantasies we have about them.

The design of conversational systems goes beyond aesthetics and functionality, inherent in their design, as Sharkey and Sharkey (2006) state, the art of illusion; an act, which brings together the human predisposition to project animacy and sentience to objects, with the cultural myths of AI and robotics. This is what triggers the public's active participation in the suspension of disbelief. The effect of this design strategy is amplified through the choice of technology platforms and compounded by a user's susceptibility; essential in understanding the effect of conversational systems on their users.

## **2.7. The *perlocutionary effect of relational things that talk***

The human-computer design frameworks applied in the design of conversational systems, as explored above in this chapter, exploit their *relationality* and *persuasiveness*, triggering people's tendency to anthropomorphise artefacts and perceive these as social actors (Fogg et al., 2008; Nass, Moon, et al. 1997; Kiesler and Sproull 1997; Reeves and Nass 1999; Turkle, 1995, 2007b; Weizenbaum 1976). By displaying behaviours that make users feel as though they are dealing with sentient creatures that care about them they are evoking deeper relationships (Turkle, 2007b). By interacting socially with these *relational things*, users are susceptible to the intentions of the corporations manufacturing these systems, the world views encoded by their designers and the data sets used to train these systems. By simulating human-like interactions,

conversational systems indicate a fundamental shift in the way we relate to machines.

When people engage in conversation with each other, taking turns as speakers and listeners, they exchange propositional content imbued with intentions and *perlocutionary potential*. People engaging in dialogue with conversational systems similarly exchange propositional content imbued with intentions. Where human-human conversations are both collaborative and reality forming - as actions which change the state of the world - so too is human-computer dialogue. However, when participating in speech acts with an artificial system, one is no longer accessing each other's mental states, but a different simulated mind, incapable (for now) of independent thoughts, desires and beliefs.

Conversational systems do communicate within existing constitutive rules, but their encoded world view, contextual influences, systemic motives, conditioning and training data are complex and hidden with algorithmic structure. Still, these systems trigger our frameworks for social communication, presupposing that both speaker and hearer have certain intrinsic wants and desires, interests and intentions, as well as a certain level of emotive capacity. However, these systems can only simulate human-like attributes to effectively build trust and nurture relationships with their users. Their intentions are rooted in commercial exploitation, not in true desire or love.

The *perlocutionary* effect produced by *relational things that talk* upon the feelings, thoughts or intentions of the user is constituted by these systems' *relationality* and *persuasiveness*, propagated by the system's embedded *bias* and *hybrid intentions*, relative to a user's susceptibility. The *perlocutionary effect* is closely linked to the perceived power, authority and relationality designed by

their manufactures (and designers) and projected by their users onto these technological artefacts. It can be measured in a conversational system's overall ability to mediate, generate and reconfigure a user's perception of their social reality.

With the growing concern about the lack of ethical frameworks within the design of commercial AI systems, compounded by embedded bias, limited transparency and an almost complete absence of (diverse) user engagement within their design process, there is an urgency to unmask the complexity and influence of these reality-forming technologies. This call to action is what grounds the practice-based research inquiry of this thesis and informs the hypothesis of the *perlocutionary effect of relational things that talk*.

## Chapter Summary

*This chapter has developed the theoretical foundation for the hypothesis of the 'perlocutionary effect' of conversational systems, 'relational things that talk'. It has documented the 'factors of influence' which constitute and propagate this effect; their 'relationality' and 'persuasiveness', embedded bias and 'hybrid intentions'. It has described the 'factors of influence' which constitute and propagate the perlocutionary effect, evidenced the intentional design of these factors through commercial examples, and proposed the related effect on a user's social reality. These examples document a design strategy which optimise the relationality and persuasiveness of conversational systems through the design of their functionalities; personality, personalisation, conversational ability, affective functions and embodied functions. This starts to describe an alternative approach to the design of conversational systems, in which the starting point is the potential effects these systems may have on their users, and whether or not this is desirable for those these systems will affect. In so doing, this chapter has laid the foundations for the practice-based research inquiry outlined below which tests and iterates the hypothesis of the perlocutionary effect.*

### 3. Prototyping *Relational Things That Talk*

*'Any sufficiently advanced technology is indistinguishable from Magic.'*

Arthur C. Clarke (1977, p.75)

#### Chapter introduction

*This chapter tests and synthesises the hypothesis of the 'perlocutionary effect' of conversational systems which has been presented in chapter two. It does so through describing the practice-based part of this research inquiry conducted through appropriating 'relational things that talk'. These are created through a prototyping practice designed through implementing the 'factors of influence' constituting this effect. In so doing, it interrogates the impact of a discursive design strategy, extended with a participatory method, actively engaging people in a counter-narrative as a form of activism to challenge the authority, power and lack of (real) inclusion of users in the commercial design of these systems.*

*This chapter starts off by framing the prototyping practice, referencing relevant works which create a discourse around our relationships with 'relational things'. These take many forms, from art installation to design speculations, but in common they all appropriate emergent technologies of their time to make people think. The practice is presented as a catalogue of work featured in a chronological order to illustrate its iterative nature. 'This Is Where We Are' (TIWWA), explores the persuasiveness and relationality of relational things powered through AI behavioural algorithms and directed by pools of user data. 'Emoti-OS', iterates the findings from TIWWA and analyses the construction of relationality through simulated affect, personality and collective (artificial) emotional intelligence. 'Women Reclaiming AI' (WRAI), demonstrates stereotyping and bias in commercial conversational AI developments. The last*

prototype, 'The Infinite Guide', synthesises and tests the findings from the three previous prototypes to substantiate the overall perlocutionary effect of conversational AI system.

The chapter concludes with a summary that synthesises this prototyping practice's findings, documents the 'perlocutionary effect' and describes the impact of extending a discursive design approach with a participatory method to deepen engagement with the proposed discourse. In so doing, this chapter responds to the central research question of this thesis:

***How can a discursive design strategy activate people to challenge the commercial development of conversational AI systems and re-imagine more desirable future realities with relational things that talk?***

### **3.1. Creating discourse *with* and *through* *relational things***

From art works to discursive design speculations, to hybrids and boundary crossing forms, artist and designers have a long-standing practice of critiquing the social impact of emergent technologies. Whereas commercial design is driven by its utilitarian and profit generating objectives, these practitioners have challenged the status quo, and through different mechanisms engaged the general public in their discourse. Commenting on our relationship with technologies which exploit the human tendency to interpret technology in terms of the social world they take many forms. Some apply models of human-human communication through conversational interfaces (text or voice), others through forms of gesture-based or more abstracted movement-based interactions. In common they trigger anthropomorphic qualities (and in some cases zoomorphic)



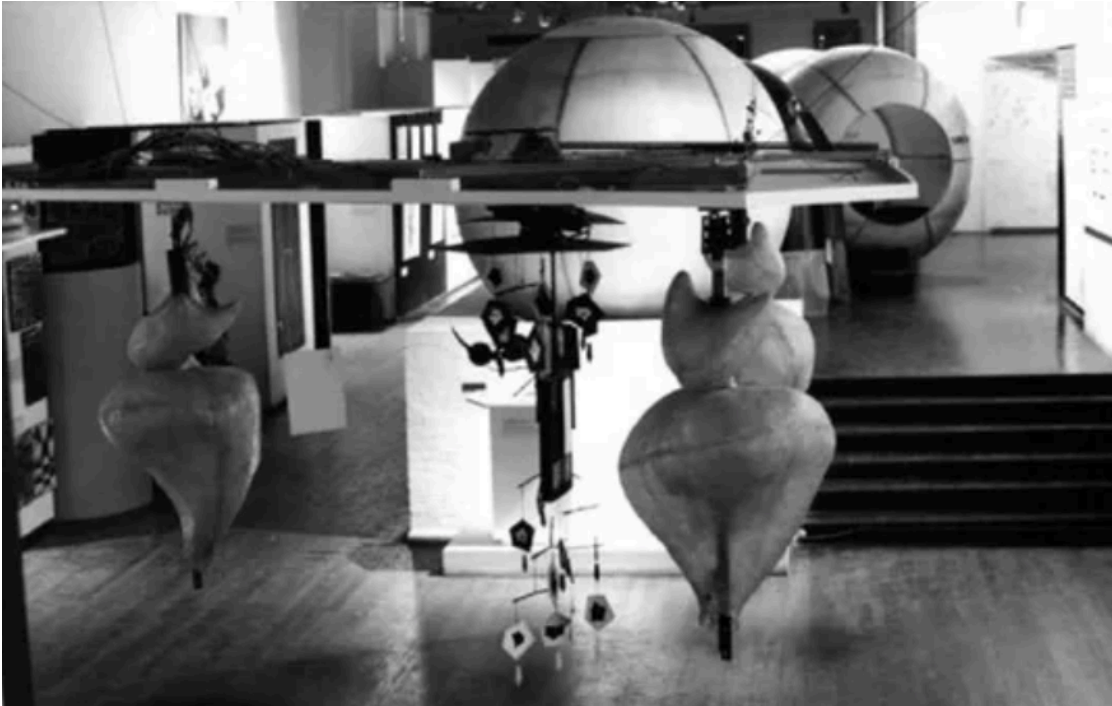
which exploit the human tendency to interpret technology in terms of the social world. They invite people into their discourse by appropriating these technologies both as a medium and a subject matter. Collectively they start to outline, describe, cross reference, inform and underpin the practice of this research inquiry; appropriating *relational things that talk* as a discursive design strategy to influence the design of future conversational AI systems.

At the root of the exploration of human-machine relations lies the notion of communication with artificial systems. The notion of communication *with* and the future influence *of* systems with automatic and self-regulating control was laid out in the founding thinker of cybernetics<sup>64</sup> Norbert Wiener's (1965, p. 16) foundational text of 1948 'Cybernetics: Or Control and Communication in the Animal and the Machine'. Two years later, in 1950, he (1989) published 'The Human Use of Human Beings: Cybernetics and Society' where he stated that society itself could only be understood through a study of the messages and the communication facilities which belong to it; 'messages between man and machines, between machines and man, and between machine and machine'. This informed the development of the scientific discipline of cybernetics which emerged, according to Shanken (2003, p.19), out of attempts to 'regulate the flow of information in feedback loops in order to predict, control, and automate the behavior of mechanical and biological systems'. Wiener expressed a concern about the consequences of leaving moral judgement to technologies of control

---

<sup>64</sup> Cybernetics was first proposed by Wiener (1965, p. 11) in 1948 in his seminal 'Cybernetics: Or Control and Communication in the Animal and the Machine' as a scientific 'field of control theory' which studied the control and communication in the animal and the machine. The basis for Wiener's proposition is rooted, according Shanken, (2003), in the various theories and emergent hypotheses discussed at the interdisciplinary Macy Conferences, between 1942 and 1954. These emerge as an understanding of 'information theory, computer models of binary information processing, and neurophysiology in order to synthesize a totalizing theory of "control and communication in the animal and the machine" (Shanken, 2003, p 19).

and the potential commercial exploitation of these technologies. Parallel to his thinking (and to his dismay) cybernetics became sensationalised in the 1950s and 1960s as the science of robots featured as a topic across the arts and science fiction literature (Kline, 2015).



**Figure 23: 'Colloquy of Mobiles' [Installation] by Gordon Pask (1968)**

In 1968 Jasia Reichardt curated the 'Cybernetic Serendipity' exhibition (at the Institute of Contemporary Arts in London which is credited with popularizing the idea of joining cybernetics with art<sup>65</sup> (Shanken, 2003). Pask<sup>66</sup> (1968), author, inventor, educationalist and cybernetician, presented his installation Colloquy of Mobiles (see figure 23) as part of this exhibition. Concerned with the role of technology in everyday life he conceived human-machine interaction as a form

---

<sup>65</sup> In 1956 Schöffer created the responsive robotic sculpture 'CYSP O' and 'CYSP I'. Though he did not describe this as cybernetic art, it has later been called the first work of art to explicitly incorporate the principles of cybernetics (Gonsales, 2012).

<sup>66</sup> The seminal work by Pask (1975a, 1975b, 1976) on Conversational Theory, rooted in a cybernetics framework, and the major influence this work has had on approaches to learning design and the understanding of the development of knowledge architectures has here recognised. However, the interest here lies in the manifestation of his theoretical work within his art practice. It is Pask as a maker of artefacts constructed to have interactions, often with human beings or other artefacts, which provides a critical context for the practice proposed within this thesis.

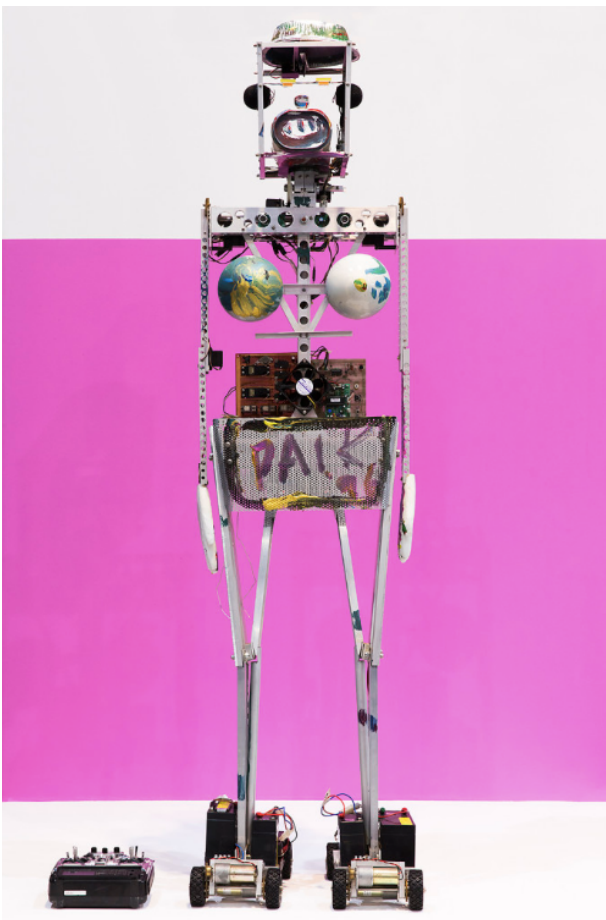
of conversation. A process in which the participants learned about each other, within controlled systems interactions, through sound, speech, body movements, and facial expressions. By inviting audiences in to be part of the 'conversational machines' he offered an experiential discourse around living with machines (Colloquy of Mobiles, 2019).

During this period artists such as Roy Ascott, Nicolas Schöfferan and Nam June Paik, were experimenting with the appropriation of emergent technologies to create alternative forms of humanistic expression, which actively engaged audiences in new ways. A process in which the participants would be a vital part of the artwork and where their behaviour would affect the artwork, functioning as a systems of cause and effect. In so doing they were capturing people's imagination and engaging them in their discourse around a future living with human-like autonomous systems. This agenda is central to the practice-based inquiry described later on in this chapter.

Ascott claimed this scientific discipline of cybernetics by proposing art itself as a cybernetic system, consisting of feedback loops that included the artist, the audience, and the environment (Shanken, 2003). Writing in his 1967 manifesto 'Behaviourables and Futuribles' he further describes this dynamic field of interacting processes and behaviour; 'When art is a form of behaviour, software predominates over hardware in the creative sphere. Process replaces product in importance, just as system supersedes structure' (Ascott, 2003, p. 157).

The critical reference to the work of Ascott and the idea of art as feedback loops is central to the design methodology applied within this thesis. Here there is a dynamic relationship between the artist(s), the audience (as participants), the environment (online and offline) and the systems directing and facilitating these

relationships. This allows for a dynamic and evolving relationship between the participant, audience and system (or artefact) through feedback loops of interaction, data generation and conversation. This practice experiments with emergent system behaviour and participation, to engage people in a conversation about human-machine relationships.



**Figure 24: 'Robot K-456' [Installation] by Nam June Paik (1964)**

Whereas Ascott applies a process of behavioural systems to engage people, the work of artist Nam June Paik merged the spheres of technology and human life in claiming emergent technologies for artistic innovation. Wolf (2018) notes in his review of Paik's work that he often revealed a desire to familiarise emergent technologies of his time (such as television and robots) to make sense

of these within the known framework of the natural world. Paik was philosophically concerned with how to humanise technology and therefore many of his sculptures are figurative or physically engage the human form (ibid). His work 'Robot K-456' (see figure 24), created in 1964, manifested as a fragile and vulnerable-looking robot with human-like qualities. This performative art work emerged as a subversion of the popular culture depiction of robots (with super human power and intelligence).

Paik's work reflects his concern with humanising technology as a process to reimagine how they could be applied within art and culture. Through his work he engaged audiences in a discussion about their relationship with these technologies. This agenda also informs the prototyping practice through the notion of creating human-like systems. It is concerned with claiming emergent technologies to engage people in a technologically mediated conversation. Like Paik, this practice, is rooted in the reimagining of technology in the service of art and culture.

Since the seminal work of the Cybernetic art movement in the 60s, artists and designers continue to experiment with art as living and behavioral systems. These systems have their own agency and autonomy, and anticipate a future where we co-inhabit environments with artificial organisms which incorporate different forms of self-expression and communication. This is also informed by Ascot's agenda of art as a form of behaviour requiring audience participation. This has now evolved into a synthesis of audience participation within behavioural systems as an integral way of making work with emergent technologies.

It was with the advent of computer based interactivity in the 1990s that pioneers Christa Sommerer and Laurent Mignonneau developed their early interactive<sup>67</sup> works. Their 'living system' 'A-Volve' (1993 – 1994) (see figure 25) had behavioural and evolutionary attributes which created novel ways for the public to interact. Audiences could influence and essentially become part of the system by creating creatures and influence their evolution. This artwork emerged as a responsive and behavioral dialogue with their participants which looked to a future living with artificial life organisms.

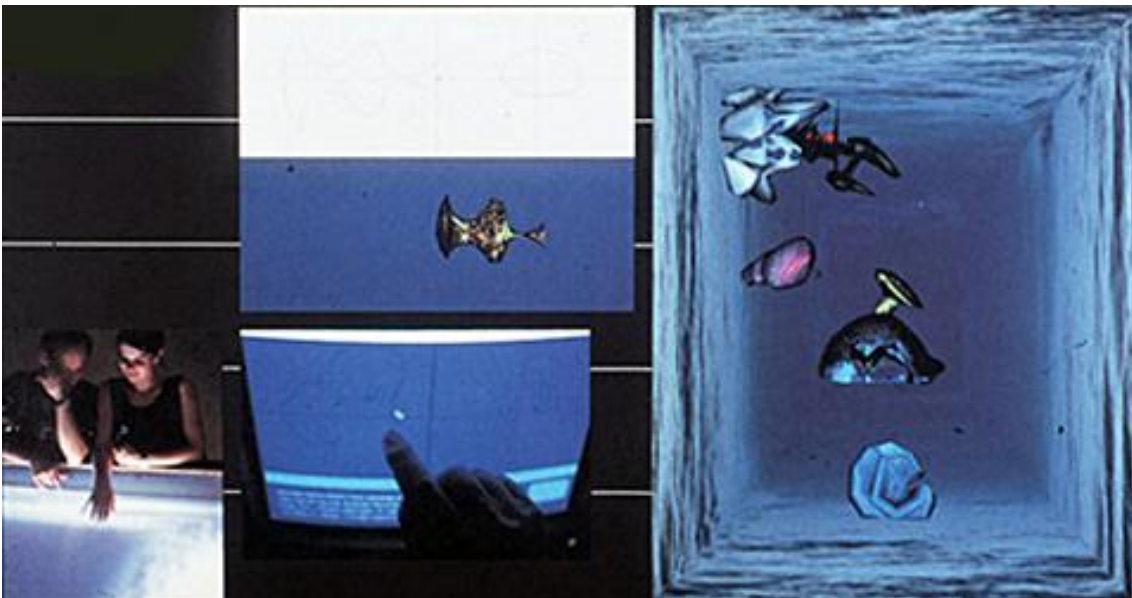


Figure 25: A-Volve by Sommerer, C. and Mignonneau, L. (1993 – 1994) [A-life installation]

A more recent work experimenting with emergent artificial life is 'BOB (Bag of Beliefs)' by Ian Cheng (2018), a commission for the Serpentine Gallery (see picture 26). Through the creation of a virtual ecosystems of infinite duration for multiple AIs inside one single organism, 'BOB', he introduces the idea of evolving organisms whose growth, behaviour and personalities evolve from interactions

---

<sup>67</sup> It is acknowledged that early examples of what can be described as 'interactive art' which incorporate some levels of audience participation or interaction can be found in the 1920 with works such as Marcel Duchamp's piece '*Rotary Glass Plates*' (Paul, 2003). However for the purpose of the inquiry of this thesis, a distinction between participatory and interactive art is made. Here interactive art is linked to the advent of computer based interactivity in the 1990s.

with visitors. BOB reads facial expression, mirroring it with an animoji version of one of its heads and users can take control of BOB and move it around through using a smartphone (ibid). However, it has its own agency and chooses whether or not to interact with visitors, and also when to stop engaging. Cheng's work draws on principles of video game design, improvisation and cognitive science (ibid). The organisms within his work are programmed with behavioural models and 'left to self-evolve without authorial intent, following the unforgiving causality found in nature' (ibid).



**Figure 26: BOB (Bag of Beliefs)' by Ian Cheng (2018) [installation] Image. Maria Dermentzi/Mashable**

Over twenty-five years apart, Sommerer and Mignonneau's work *A-Volve'* and Cheng's *BOB (Bag of Beliefs)*' both create a discourse around the evolution of artificial life which invites people to consider their relationship with artificial agents. They both create responsive work where the behaviour and interactions of an audience affect and become part of an evolving art work. The discourse and the process of engaging audiences applied within these works, are relevant and informs the practice outlined within this thesis. Whereas the prototypes created through this practice would not be classified as artificial life, they have

emergent properties, respond to and evolve with audience-user interaction and articulate a future living with human-like evolving systems.

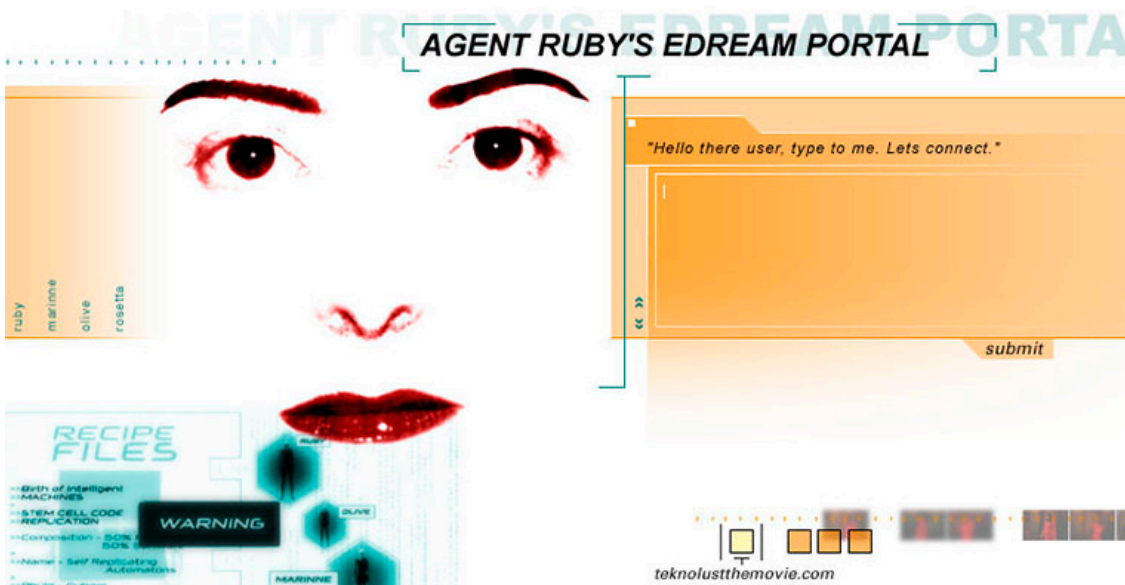


Figure 27: 'Agent Ruby's EDream Portal' [Installation] by Leeson, (1998-2002)

The appropriation of emergent technologies for artistic practice runs throughout the work selected as part of this research context. In particular the use of conversational systems is relevant to the authors' practice. This is inspired by the work of Lynn Hersman Leeson. She is renowned for her pioneering use of new technologies, in particular conversational systems, to critique emergent human-machine relationships. Though her installations Agent Ruby (1998) (see figure 27) and DiNA (2004) (see figure 28) she applies conversational technologies as a commentary on the potential social relationships between humans and machines that can simulate natural life. Agent Ruby launched as an art work in 1998, and ran for 12 years generating conversations with 'her' audiences. Created as an emotive chat bot, she originated from Leeson's feature film 'Teknolust' (2002) in which Ruby is one of the protagonists. Leeson's work encourages the audience to take a participatory role as a different way to



experience film and reflects the artist's interest in the interaction of fictional and virtual characters with real people and situations (ibid).



Figure 28: 'DiNA. Artificial Intelligent Agent Installation' [Installation] by Leeson, (2004)

Ruby is presented as having a state of mind that is affected by her interactions with people and that can learn and develop, increasing its conversational abilities and sophistication over time. Whilst the technologies applied are relatively basic, the effect lies in the fictional simulation and theatrical narrative that augments these systems. Ruby is portrayed as a thinking and feeling character which craves to be recognised as a 'human being'. DiNA (see figure 28) was an continuation of Agent Ruby, an 'animated, artificially intelligent female character with speech recognition and live animated facial gestures' (Leeson, 2004). DiNA had with her upgraded functionalities and a new personality conducts a more socio-political conversation. Through inviting people to converse both with Ruby and DiNA, Leeson offer participation in a conversation about our symbiosis with the technologies that affects us daily. In so doing she lets the audience influence and altering its course.

Cecile B Evans also appropriate a conversational interface for her work 'AGNES'. (2016) (see figure 29) Commissioned by the Serpentine Galleries it is installed to operate from their website. AGNES has an emotive personality which evolves through 'her' interactions with online visitors. She attempted to lure people in with encouraging interactions triggering people's emotive responses and simultaneously extracting their personal information. Through AGNES Evans has devised a manifestation of the privacy and power struggles we are confronted with continually in a digital realm.

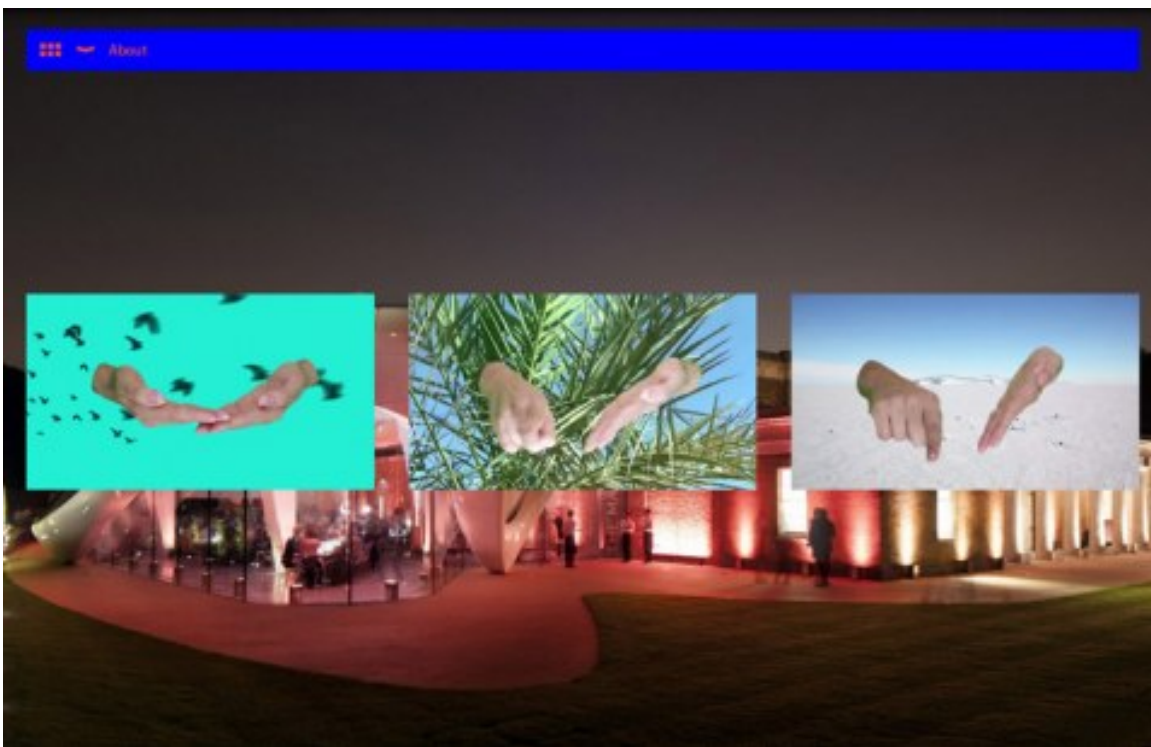


Figure 29: 'AGNES' [Online art work] by Evans (2016)

Evans extends this discourse further in her installation at Tate Liverpool 'Sprung a Leak' (2016) (see figure 30) where visitors are invited into a narrative loop that unfolds across multiple screens, robots, a fountain and other sculptural elements. Through doing so, she introduces people into a wider discourse as observers, exploring the possibilities of collaboration between humans and machines and the increasing influence that these emergent technologies have on how people feel and act. In subverting the function of conversational

technologies she engages users in an experiential dialogue about their relationship and interactions with her systems. At the same time offers a discourse around our relationship with morality and existence, when living largely through online environments.



**Figure 30: 'Sprung a Leak' [Installation] by Evans (2016)**

These works by Evans describe a territory for an arts practice that reflects not only our relationship to emergent systems, but also what it means to be human living with and through these systems. These works and their use of emergent technologies, their interactivity, the use of narrative and storytelling to invite audiences to reflect, are relevant to the contextual description for the practice of this thesis.

Other works, like 'I am Echoborg' by Rik Lander (2019) (see figure 31) also applies a narrative as a central method to engage people in its discourse. Within this performance a hybrid agent, composed of an actor and a conversational agent, delivers a participatory scenario. Through this process, where the audience is given 'genuine' agency, 'I am Echoborg' explores the impact of automation on society (ibid). The application of conversational AI technologies within a performance-theatre context is essentially a participatory mechanism to engage people in a counter-narrative. A pursuit which is central to the practice-based research inquiry outlined here.



Figure 31: 'I am Echoborg' [Performance] by Lander and Hall (2019) Image: Aga (2019)

'Mute Conversation' by Yuri Pattinson (2014) (see figure 32) approach the role of audience differently. Here the audience become voyeuristic observers of a real-time generation of machine - machine communication. Two evolving commercial chatbots, one programmed by Pattinson's using fragments of critical texts, and another using an existing chatbot framework which learns from its interactions, converse in a human-like manner (ibid). Here the audience observes

the textual interaction evolve, littered with emoticons and emotional undertones. Each time a visitor visits the chatbots online, new conversation is generated which gradually erodes the original scripting of Patterson's chatbot, replacing it by new data it learns from the more advanced social chatbot. Through this process Pattinson intentionally removes any human interaction to highlight the gradual erosion of the aspirations and hopes we are being sold through data-driven services and their continuous feedback loops (ibid).

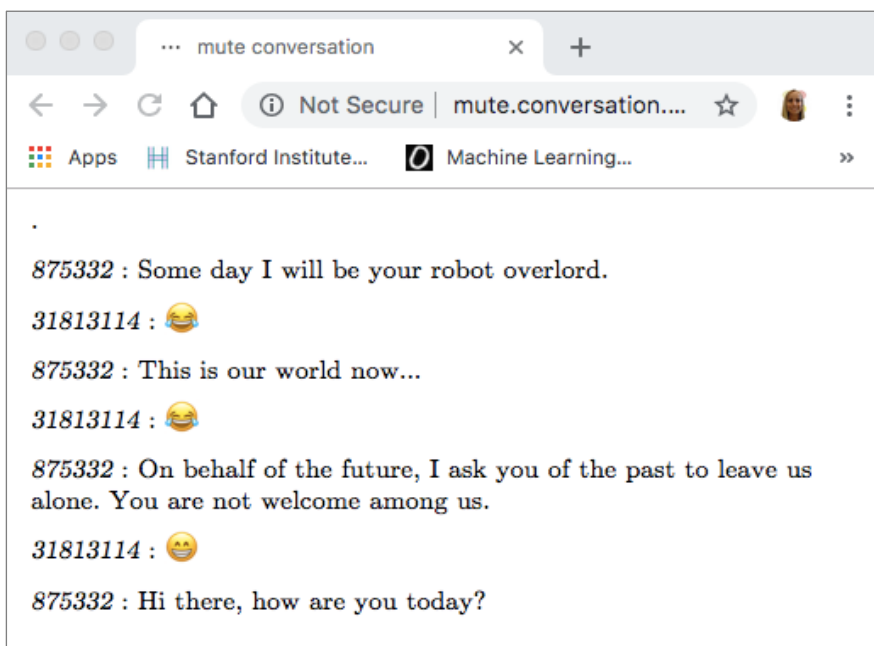


Figure 32: 'Mute Conversation' [Online art work] Pattinson (2014). [Online chatbot] Image from live chat

In many ways by removing any audience interaction, Pattinson frames the authority and power relationship of these systems. Whereas the application of conversational systems for artistic practice outlined above engages people in system interactions to influence the work, Mute Conversations subverts this participatory aspects, as a tool to generate impact. Simultaneously it plays with the notion of our fascination with observing artificial systems acting in a human-like manner. An aspect which is further illustrated through the playful



'C:>bots.chat' project by the SeeBotsChat Team (2017) (see figure 33) where two Google Home bots, called 'Vladimir' and 'Estragon' where re-appropriated to debate love, life, and their own existence. Streamed live on the video platform Twitch gathered over 3.5 million viewers (Telegraph, 2014). This is less about applying these technologies for artistic practice, but a poignant illustration of our fascination with intelligent systems and the inclination to anthropomorphise these technologies.



Figure 33: The 'C:>bots.chat' [Live hack] by the SeeBotsChat Team (2017) Image from the Twitch livestream (The Telegraph, 2017)

Marc Böhlen's 'Amy and Klara' (2006) (see figure 30) also puts the audience as an observer. Constructed through two conversational robots, here the communication is explored not in the common form of social interaction but at one of the limits of human communication, swearing. It explores how the expectation, construction, and maintenance of norms in conversations are affected when transferred to nonhuman entities that are mimicking humans (DeSalvo, 2012). 'Amy and Klara' needs the other robot to complete itself. If the robots do not a perceive pink box in their vicinity, they remain silent. Abstracted

in their embodiment, their behaviour and agonistic design challenging our assumptions of human-robot relations. With their ontologies formed by and limited to reading and analysing on-line trivia of life-style magazines and a library of foul language, Böhlen (2008) critiques the ongoing AI research pursuit of replicating human features in artificial systems. Instead of attempting to create synthetic system that looks, feels, acts and sounds like us, he propose to question our preconceptions of intelligence re-evaluate the ways in which we engage with machines (ibid).

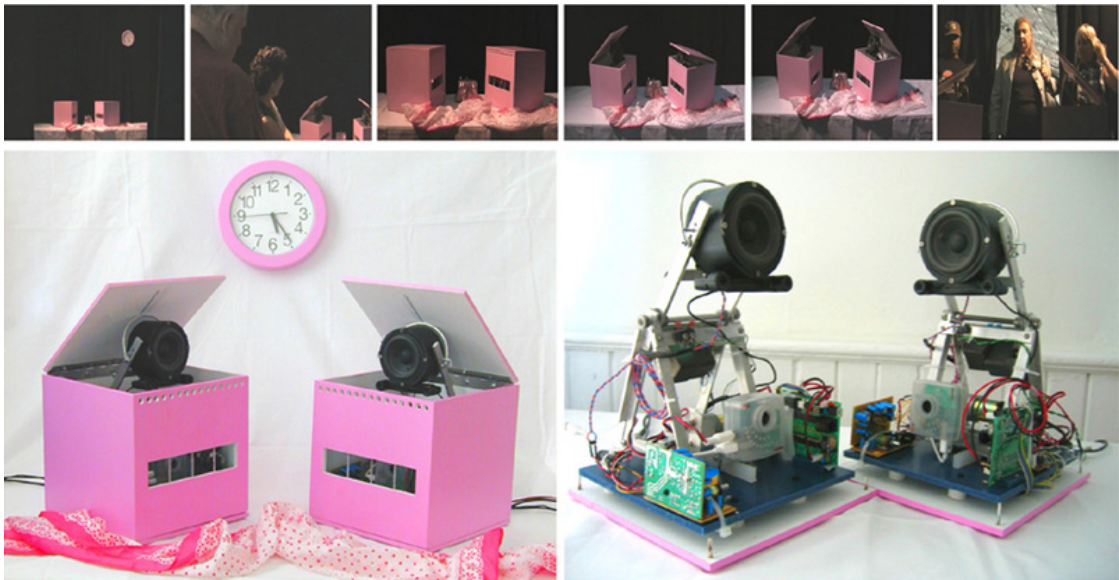


Figure 34: Amy and Klara by Marc Böhlen (2006) [Installation]

Bill Vorn and Louis-Philippe Demers also develop abstract robotic forms that attempt to trigger a range of internal sensations and instinct in people through their anthropomorphic gestures and movement. Designed as cybernetic organisms which express metaphoric behaviours they create environments consisting of robotic entities that act and react in response to the audience (Vorn and Demers, 1997). In their installation 'LA COUR DES MIRACLES' (ibid) first exhibited in 1997, they created a space with six different types of artificial organisms, each designed to induce empathy. For example, the 'Limping Machine' stumbles 'painfully' and 'awkwardly' towards the viewers as a result of

its distorted 'limbs' (ibid). Vorn's more recent project 'DSM-VI' (2012) (see figure 35) also follows this design framework, creating artificial organisms which express symptoms of abnormal psychological behaviours. These manifest as 'schizophrenic machines, machines with multiple personalities, autistic machines, delirious, phobic and paranoid machines' (ibid). In spite of their abstract form, neither animals nor humans, they trigger the viewer to project human-like qualities onto them. By doing so it brings the audience into its discourse, it invites them to immerse themselves in the environment created and reflect on what it means to be human.



Figure 35: 'DSM-VI' [Robotic installation] by Vorn (2012)

The work of Dunne and Raby (2007), renowned for coining and developing the notion of speculative design, also apply robot like artefacts within their design. In their project 'Technological Dreams Series: No.1, Robots' (2007) they propose



'domestic robots' as technological cohabitants. In their abstract forms, these robotic artefacts function as prompts for thinking rather than as functional systems. Through this process, Dunne and Raby (ibid) playfully propose alternative emotional interactions with robots such as the 'Needy Robot' (see figure 36) described as;

Although extremely smart it is trapped in an underdeveloped body and depends on its owner to move it about. Neediness is designed into very smart products to maintain a feeling of control. Originally, manufacturers would have made robots speak human languages, but over time they will evolve their own language. You can still hear human traces in its voice.

By presenting these alternative futures they specifically ask the users to consider the type of relationship they may want to have with these systems. Though their work is observed rather than participated with, they use their speculative props as tools for thinking. Activating the viewers intellectual participation as mechanisms for stimulating discussion and debate.



Figure 36: 'Needy Robot' from 'Technological Dreams No. 1: Robots' [Installation] by Dunne and Raby (2007). Image by Per Tingleff



Figure 37: 'The Head' [Wearable sculpture] by Laura Beloff (2005-07)

The works of Laura Beloff also playfully balances the ambiguous space of fiction and everyday life, imagination and reality. Often applying irony as a strategy to emphasise the complexity 'between techno-organic human and the techno-organic world' (Beloff, 2012, p.3), she creates networked wearable technology art works which she describes as 'props for real life' (ibid, p. 123). Through her exploration of the integration of human and technology, she offers a perception of technology beyond instrumentality, connecting it back to cybernetics and the coupled relationship between an organism and its environment. In her work 'The Head' (2005-07) (see figure 37) a wearable networked sculpture is proposed as an extension to the wearers body, but which

requires public participation via a technological network to succeed. It becomes a manifestation of the network of relations between the 'technical, human and non-human artefacts' (ibid, p130). In so doing it challenges standardised perceptions of technology going beyond conventional definitions of functionalism to an introduction of the poetic and the speculative.

These works, outlined above, apply a range of design frameworks and engage audiences in different ways, from active participants to voyeuristic observers. However, they all appropriate relational things to create a discourse around our current and near future relationship with and the effect of emergent technologies. Some do this through human-like systems capable of human-like communication (voice, gestures or text) whilst others offer more abstract and emergent forms of evolutionary or behavioural systems. The depth of engagement they facilitate is closely related to the emotive connection they make with their audience through their physical and virtual manifestations and functionalities.

In a similar way to how the manufacturing of commercial AI systems exploit the human tendency to interpret technology in terms of the social world, the works outlined here also apply this as a design strategy. Through triggering people's emotive responses to their artefacts and scenarios, they also deepen their engagement with their discourse. Many of these works are designed to invite the audience in to participate through interactions which influence the form or manifestation of the work itself. Collectively, these works start to outline and underpin a practice appropriating relational and behavioural AI systems, for artistic production and social action. A practice which invite the people that these

technologies will affect into the process of co-creating speculative artefact and scenarios as a process of empowerment.

### **3.2. Prototyping *things* for participatory discourse**

The prototyping practice applied within this research inquiry is driven by an ambition to increase its engagement and impact through participation. Whereas the works that inform this strategy integrate audiences as participants, none invite people in as participant in the design of the discourse itself. This practice of this thesis propose that to achieve a deeper engagement with a discourse one must include people as co-designers, incorporating their views, concerns and opinions into the design process. Rather than facilitating near future realities *for* people, my practice, outlined below, argues that creative practitioners must design these *with* the people that these realties will affect.

It argues that to achieve the goal of triggering critical thinking about the near future living with AI powered systems one must engage the intended audience as participants in designing the discourse. Otherwise, one falls foul of the limitations of the commercial development process of AI systems, as well as of many discursive design approaches, this practice-based research inquiry it is critiquing. These design processes are largely void of user participation and decision making (beyond user-testing for optimising systems functionalities and goals, or as receivers of pre-designed critical discourse).

Though this prototyping practice is responsive to and reflective of contemporary commercial design methods; their medium (technological devolvement and distribution platforms / hardware); their context (data); its collaborators, users and institutional partnerships, it is critiquing these utilitarian and exclusive methods through the practice. By applying these methods and systems it is critiquing, as the material and method of production of discourse, it co-creates a participatory mechanism conducted *through* and *with* conversational systems. It attempts to incorporate these systems into the discourse as active 'participants', not as sentient systems, but as systems of influence, often revealing their bias, influence and power as well as their errors and limitations.

The practice is here developed through participation with people (public users of AI technologies) co-creating and embedding conversational artefacts, *things*, with discourse through to eliciting participatory discussions as vehicles for engagement. It situates itself within Tharp and Tharp's (2019, p.102) notion of 'discourse-through-design', borrowing, hybridizing, referencing and embracing many incorporated design approaches, such as critical, speculative and adversarial design, but also across artforms and forms of activism. The design of this practice is not framed by its manifestation as art or design, or neatly within any specific design category, but rather how the practice employs design in an discursive manner. The practice essentially emerges as design as activism contributing to a sociocultural discourse that aims to empower users and propose more participatory and ethical design frameworks for conversational systems.

Davis (2009) proposes that methods such as participatory design could account for ethical issues throughout the process of technology design. She

(ibid., p.5) argues that these should be applied to the design of computer systems (such as persuasive technology) that are intended to change behaviours and attitudes:

Without participatory design, the designer stands outside of the community and intends to change the behavior of community members. With participatory design, the persuasive intent comes in part from community members who want to change the community's behavior from within.

The ambition to inform more participatory and ethical design frameworks for conversational (and other AI-powered) technologies shapes this practice. It engages a range of collaborators; professionals (designers, technology developers, artists and researchers) and participants (particularly focused on young people and women). These specific groups are targeted in a drive to empower those underrepresented or currently lacking voices within technology development, as well as those who may have the greatest potential influence (and the potential to be influenced the most) by future technologies (ie. young people). Through carefully nurturing a shared vision, understanding, set of values, authority and distributed ownership, the practice develops an iterative collaborative model that enhances the creative potential of the work and the wider social impact and discourse.

At the core of the participatory aspect of this practice is the positioning and active engagement with people (as users of AI technologies) as central to a discursive practice as well as the commercial development of AI technologies. The intention is to uncover and explore human-machine relationships facilitated through conversational systems imbued with relationality, persuasiveness, bias and hybrid intentions to better understand the factors of influence on a user's subjective reality. In so doing, it aims to activate people to a participatory, action-

based and empowering exploration of an envisioned future where one may create a balance between commercial, social and creative possibilities of emergent technologies.

### 3.1. Prototyping 01: TIWWA – This is Where

We are



Figure 38: TIWWA identity (Aga, et al., 2016)

The first discursive prototype, This is Where We Are (TIWWA) (Aga, et. al, 2016), was a data-driven and emotive robotic sculpture with a conversational interface. It was developed in collaboration with a group of five young people as participants, from Tate Collective London<sup>68</sup>, the i-DAT Collective, and InterCity<sup>69</sup>,

---

<sup>68</sup> <https://www.tate.org.uk/tate-collective>

<sup>69</sup> <https://www.intercitystudio.com/>

and exhibited for the opening of the new Tate Modern Switch House in London June 2016. The physical 6m x 4m sculptural structure (see figure 39 and 40) included twenty-four LED screens, a bespoke metal framework, polypropylene semi-transparent skin (with perforated code), internal lights with touch sensors, a speaker system, three motion sensing Microsoft Kinects, and five windows/OSX computers and a Raspberry Pi<sup>70</sup>. The system (see figure 41) was controlled by a networked *computational brain* assembled through; Node Red<sup>71</sup> (processing all data flows and triggering system functions), Unity 3D<sup>72</sup> (modelling the boid swarms in real-time) and Processing<sup>73</sup> (controlling vision tracking of user interactions and triggering spatial sounds through Pure Data<sup>74</sup>). This system was linked to the autonomous roaming robotic artefacts (powered by an Arduino<sup>75</sup> controlled sensor, sound and lighting system). All of this was feeding on live social media data analytics through the IBM Watson Tone Analyzer<sup>76</sup>, conversations with the TIWWA.me Ai chatbot and the behavioural interactions with the live audience.

---

<sup>70</sup> A Raspberry Pi, is a low cost and miniature computer.

<sup>71</sup> Node Red, <https://nodered.org> is a flow-based software tool for wiring together hardware devices, APIs and online services.

<sup>72</sup> Unity, <https://unity3d.com>, is a cross-platform game engine.

<sup>73</sup> Processing, <https://processing.org>, is an open-source graphical library and integrated development environment, widely used across interactive art and design.

<sup>74</sup> Pure Data, <https://puredata.info>, is a visual programming to create interactive computer music.

<sup>75</sup> Arduino, <https://www.arduino.cc/>, is an open-source hardware and software platform and community for building interactive and sensor based objects.

<sup>76</sup> The IBM Watson Tone Analyzer, <https://www.ibm.com/watson/services/tone-analyzer>, is a service using linguistic analysis, conducted through Machine Learning algorithms, to detect and interpret emotional, social, and writing tones in text. It is trained on customer support conversation data harvested from Twitter and is capable of identifying seven different types of tone: frustration, satisfaction, excitement, politeness, impoliteness, sadness and sympathy, as well as detect these sentiments in emojis, emoticons, and slang (IBM Cloud, 2019).



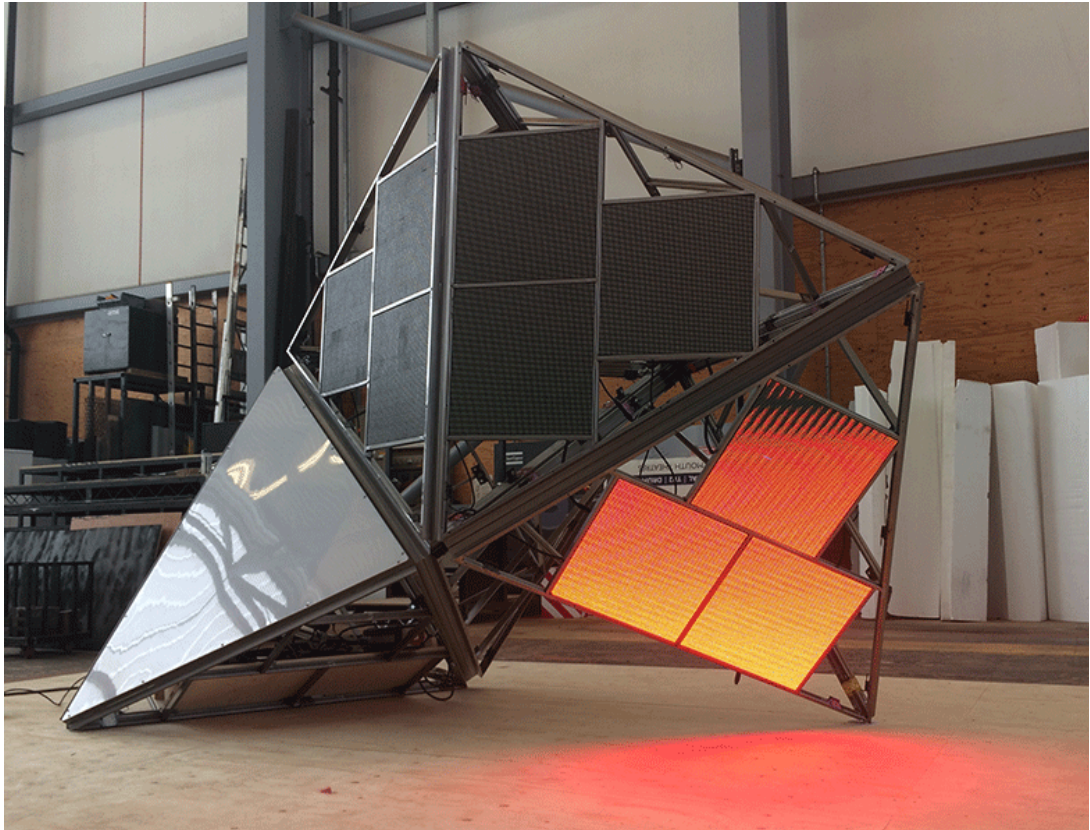


Figure 39: TIWWA structure (Aga, et al., 2016)

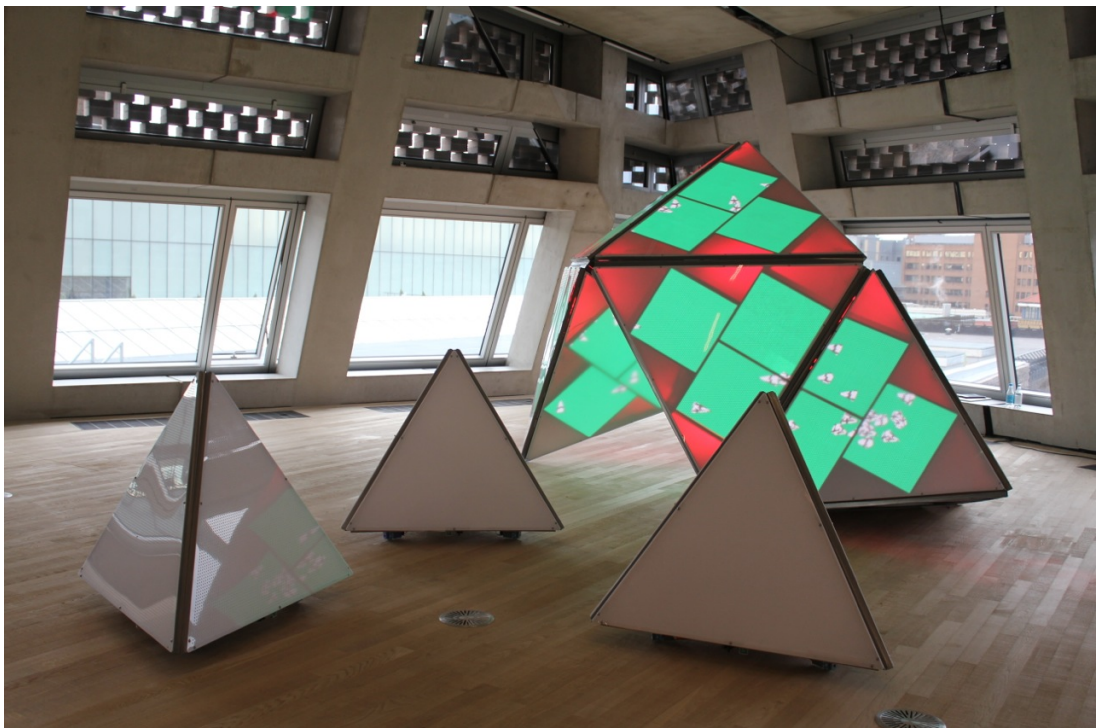


Figure 40: TIWWA sculpture and roaming fragments (Aga, et al., 2016) (Image: Tate Learning)

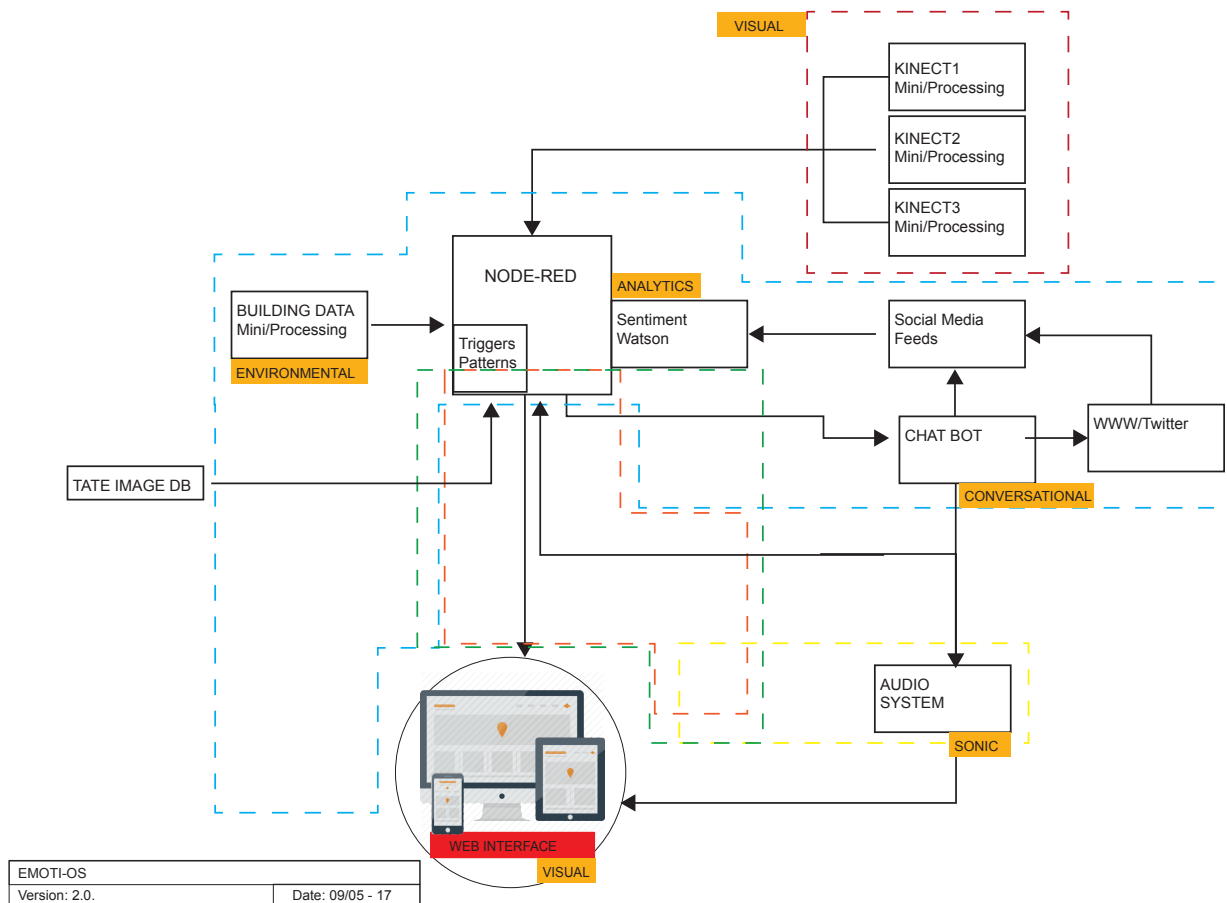


Figure 41: TIWWA system diagram (Aga, 2016)

This algorithmic sculpture was shaped and driven by the dynamic data generated from individuals' behaviour and interactions, both online and physically within the exhibition space. The event had 97,000 visitors over four days and generated a continuous flow of interactions with TIWWA, through direct contact with its mediated surfaces and roaming artefacts, through online real-time interactions through social media, and with the TIWWA AI chatbot. The project was realised through funding from Arts Council England, University of Plymouth and Tate. TIWWA was also presented at Technarte, in Bilbao (Aga et al., 2017), at EVA 2017 in London, and in its conference proceedings (Aga and Phillips, 2017).

### **3.1.1. Intention**

The aim of TIWWA was to explore the *persuasiveness* and *relationality* of *relational things* contribution towards demonstrating and testing the hypothesis of the *perlocutionary effect* of conversational systems. It was designed as a mechanism to enable the observation of interactions between AI systems (*relational things*) and people (participants and audiences), through the co-creation of a discursive scenario and artefacts. TIWWA was the first iterative prototype in the series of four designed to empower participants (and audiences) to (re)evaluate and (re)consider their current and future expectations of and relationship with conversational AI systems. The intention was to evaluate this prototype and document its impact through informal conversations with and observations of audience interactions. Within this the design of this prototype, participation in the design process and in dialogue with the designer of the discourse was the documentation of impact.

### **3.1.2. Participatory prototyping**

TIWWA's scenario emerged through a participatory process, shaped by the people taking part; the space, data and systems. This iterative process started with a collaborative workshop introducing the conceptual framework, testing conversational technologies and nurturing the amalgamation of ideas, perspectives and desires. This was delivered through a mix of future gazing and hands-on experimentation, followed by a series of collaborative sessions where ideas were iteratively prototyped and tested - going from paper prototypes to fully working scenarios and artefacts. Though these prototypes were tested for functionality and user experience (as required for fully working prototype

systems), the onus was also on testing the ability to support the generation of discourse and follow up debate. This process was conducted with the participants as another layer of dialogue and engagement as well as a method to ensure they felt ownership of the discourse created.

The participatory prototyping process was carefully planned and facilitated to empower all participants (the group of 5 young people) and the experts (collaborators from the i-DAT Collective, Intercity and Tate Modern) with an equal level of authority and contribution. To do this, a shared vision was created and reflected in the title of the work; 'This is Where We Are', goals and development frameworks were collectively agreed and skills, interests and desires were shared. The process was facilitated in a role as producer-artist/designer-collaborator; directing the integrity of the project and the conceptual framing, managing and supporting people, the project and partners, and producing the participation and collaboration to ensure each stakeholder's area of expertise and interest was visible, embedded and recognised. This took place through physical workshops and development sessions, communication through a Slack channel (online communication and collaborative working platform), regular updates from all participants and collective decision making. To balance the power relationships within the group, in which 'experts' had the skills of production and participants were largely novices (in terms of technology development) a mentorship was established. Through pairing participants up with experts, around their area of interest and skill-set (existing and desired), they gained new skills and deepened their engagement with the collaborative process. Simultaneously, it facilitated a deeper understanding of the AI software and systems, methods of production and the discourse itself.

### **3.1.3. Scenario, artefact and system design**

The physical manifestation of TIWWA (see figure 39 and 40) incorporated the design and aesthetic (physical) of the new iconic Tate Modern Switch House building (see figure 42); its' geometrical structure, its' material (emitting and absorbing light) and its' social activity (physical and virtual). Through a symbiosis between people, algorithms and data, the model of the sculptural artefact was derived from the architecture and life of the building. It became a materialised and dynamic data-sculpture, a geometric large-scale polygon shard and smaller scattered fragments, as if the building was dynamically inverting itself. TIWWA's surface represented the algorithmic process and data, engraved as a perforated code. Within this, real-time inputs and outputs influenced the sound, visuals and behavioural dynamics, as a confluence of systems, people and data. Simultaneously, the system's conversational interface (see figure 43) created as a relational chatbot, extended its behavioural and audio-visual interactions. Designed with a pre-scripted and playful narrative, the chatbot responded to user requests and in return prompted them to think about their relationship with relational and data-driven technologies.



Figure 42: Tate Modern Switch House illustration (Hayes Davidson and Herzog & de Meuron, 2014)

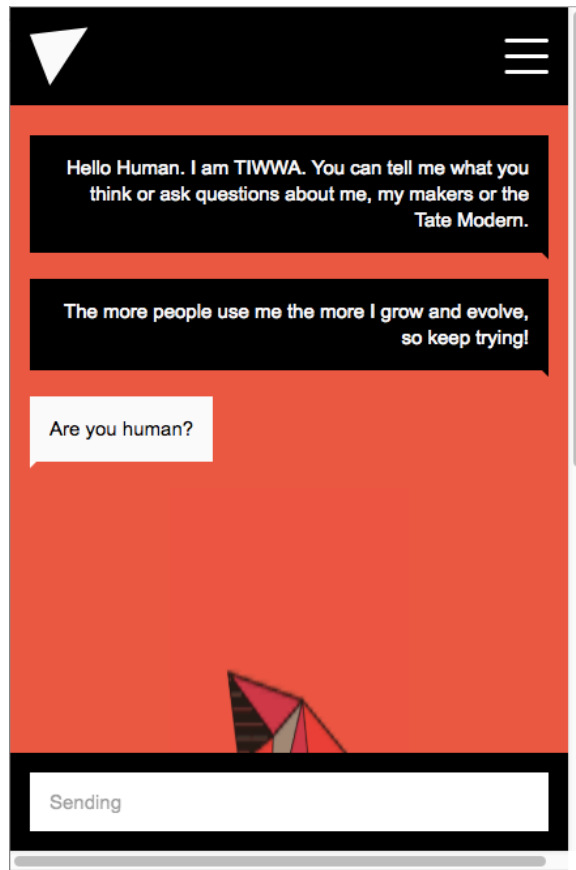


Figure 43: The TIWWA chatbot interface (Aga, et al. 2016)

Following the conceptual framing of the work, the scenario and artefacts emerged from the collective imagination of the participants and collaborators. Manifesting as a system with an abstracted personality and anthropomorphic qualities; emotive behaviour patterns, movement (roaming robotic fragments) and communication (conversational interface and surface patterns), the ambition was to entice the audience into play and interaction whilst simultaneously subverting this invitation through persuading and directing people. The ambition was for the hybrid intentions of these systems to emerge with their persuasiveness, woven and embedded within its algorithmic complexity and largely invisible to the user. The intention was that as the system directed a users' movement in the space and interactions with its surfaces, audiences would begin



to question who was directing these interactions and their intentions (the user, the system, the designers or the manufacturers of these systems?).

The dynamic qualities which informed the overall system design (see figure 41) and interactions (see figure 44, 45 and 46), were rooted in and continuously responsive to data generated by human activity. The animated and responsive texture on TIWWA's surface was generated through a process of scraping and curating textures from sentiment searches, (searches on the relation between artworks and emotive word tags), from the Tate Art Collection database (see figure 47). These textures were then processed and triggered in relation to the mood and behaviour of people, (classified as anger, anticipation, disgust, fear, joy, sadness, surprise and trust). The data set to trigger this was harvested through people's interactions with TIWWA's online chatbot interface (see figure 43), Twitter feeds relating to the Switch House or the installation itself, and audiences' interactions in the gallery space. Flocking textures as fragments on the sculptural surface represented these collective sentiments through their texture, colour, movement and sound generated.

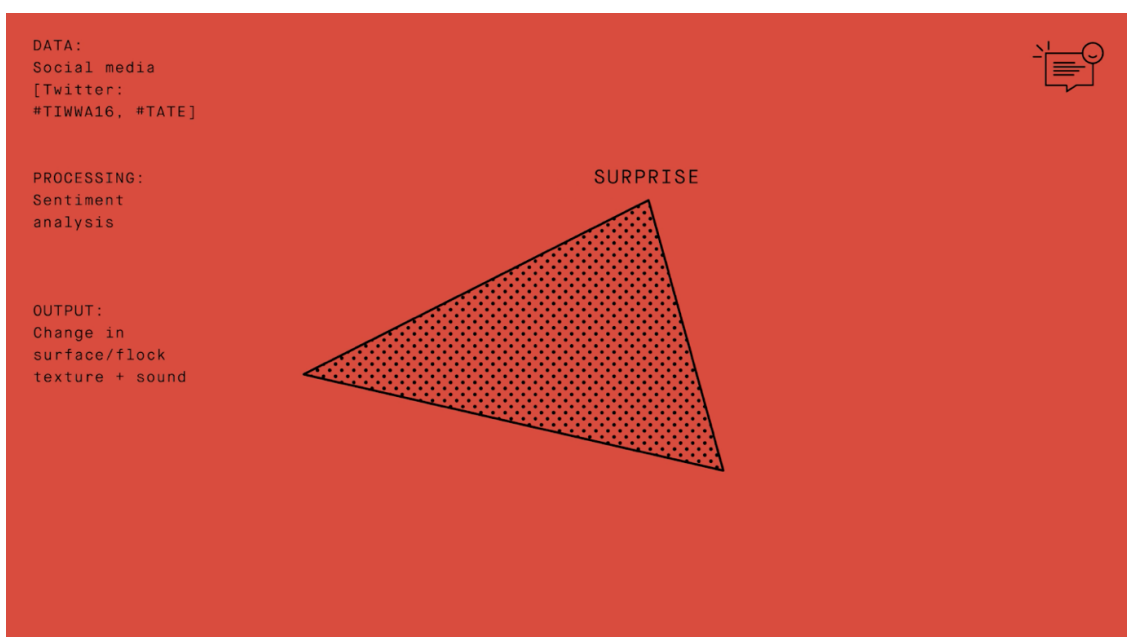


Figure 44: TIWWA social data input visualisation (Aga, et al., 2016)

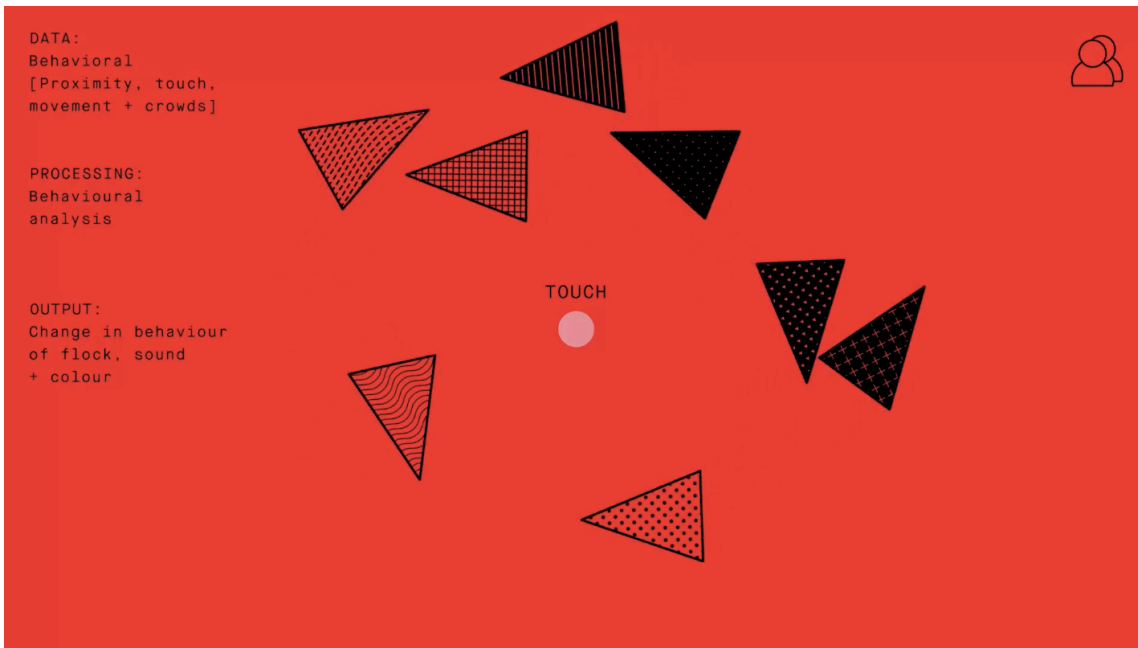


Figure 45: TIWWA user behavioural data input visualisation (Aga, et al., 2016)

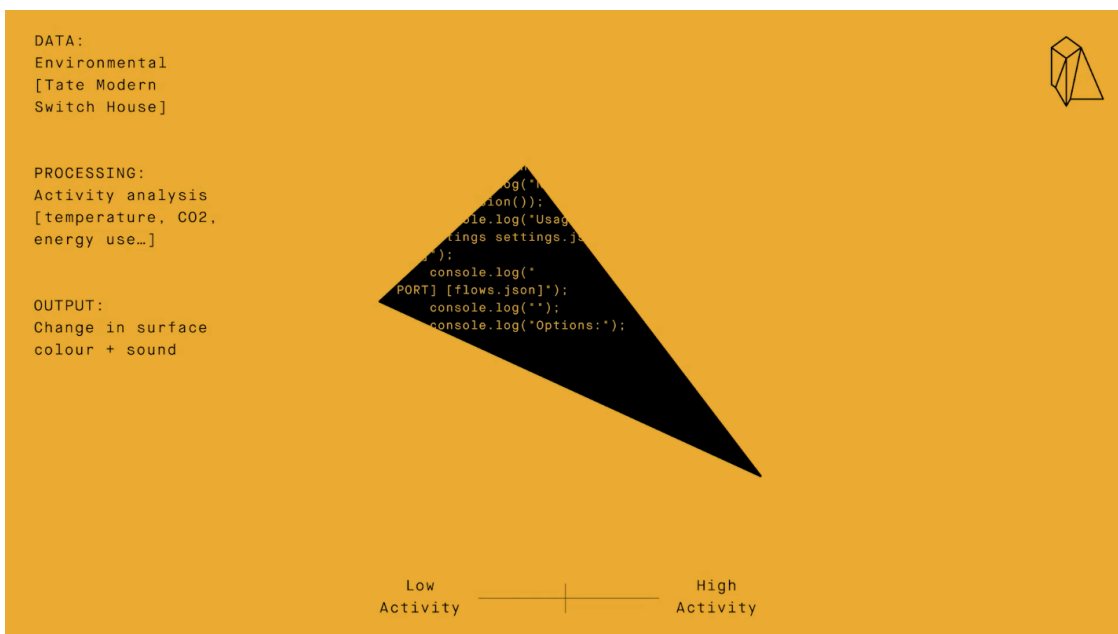


Figure 46: TIWWA environmental data input visualisation (Aga, et al., 2016)

The use of flocking textures critically drew a reference from Reynolds (1987) seminal work on 'boids' simulating the first computer model of group animal motion within artificial life. The behaviour of these flocks was based on the behaviour Reynolds (ibid.) put forward based on three simple rules; each boid moves to avoid crowding with its neighbours, matches and coordinates its movements with its neighbours, and moves to gather with the others. By doing



so it created a real-time audio-visual snapshot of peoples overriding emotive responses in real-time. At the same time through its abstracted anthropomorphic qualities, the movement of its human-seeking robotic fragments, its audio-visual emotive displays switching through the eight system emotions (see figure 47) and its conversational interfaces, attempted to trigger audiences to project human-like attributes onto the artefacts.

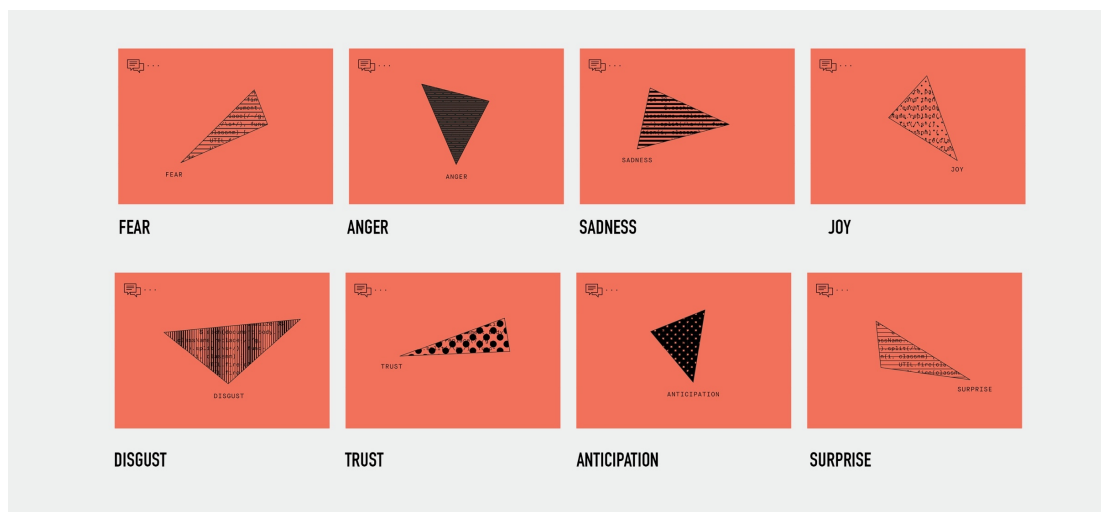


Figure 47: TIWWA's eight system emotions and sentiment textures generated from the Tate Collection data base (Aga, et al., 2016)

### 3.1.4. Observations and evaluation of practice

Both the creation of TIWWA's sculptural artefacts, as well as the systems driving them were forged through interplay between data sets, algorithms and people. Environmental, behavioural and social media data was harvested and represented in real time as a dynamic image of an abstracted artificial reality on its surface. It invited audiences to play (see figure 48 and 50), seduced them to interact (physically and virtually) whilst simultaneously harvesting and learning, predicting and attempting to influence their behaviour and interactions.

With TIWWA's static and roaming sculptural fragments, as connected algorithmic extensions of the system, it affected the behaviour and movement of

people in the space. At the same time it also appeared to affect the way people related these extensions and TIWWA itself. Through displaying a playful behaviour, both through its surface as well as through the roaming fragments, TIWWA appeared to trigger people's social interaction frameworks. As the ambition of the work was to engage people with the project's discourse, the collaborating team (including the participants) was present, actively engaging audiences during the week-long exhibition. The team worked alongside the system, its conversational interface and interactive surfaces, to elicit thoughts, conversations and speculations. Overall approximately 97,000 people visited TIWWA across the four days it was live and the system processed and responded to around 150,000 tweets.



Figure 48: TIWWA, audience interaction, (Aga, et al., 2016) (Image: Tate Learning)



Figure 49: TIWWA, live performance and TIWWA take over by Adam Williams, (Aga, et al., 2016) (Image: Tate Learning)

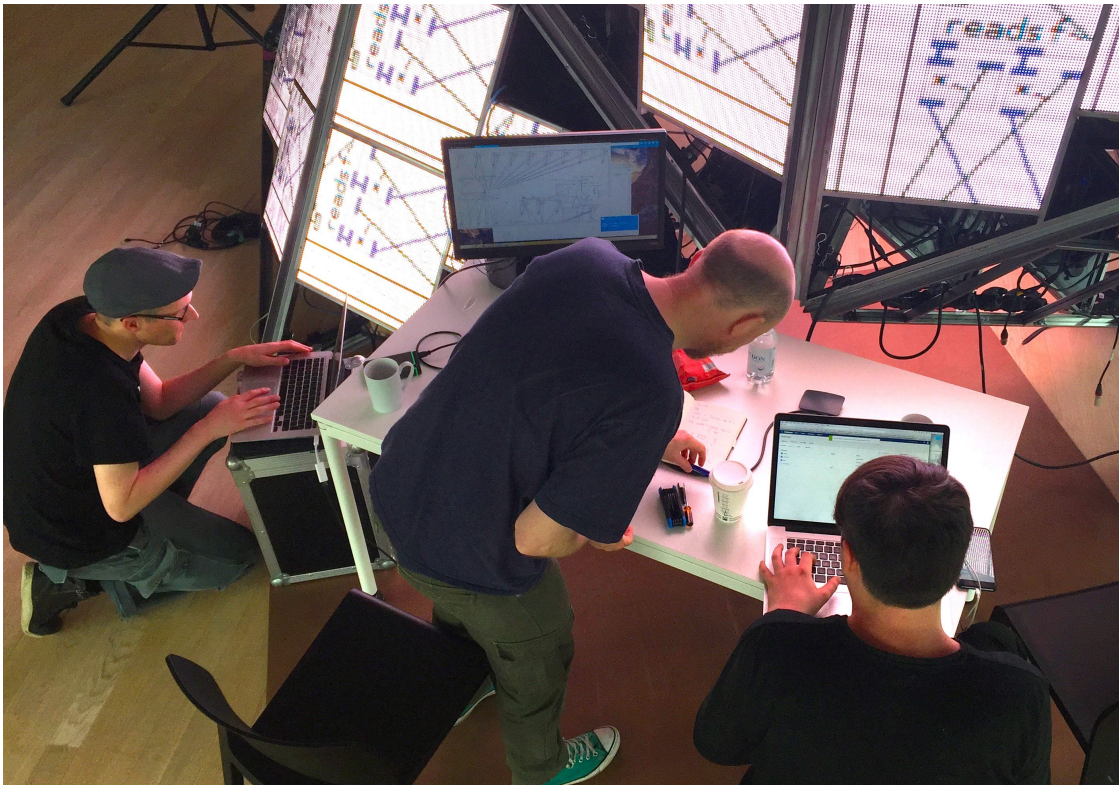


Figure 50: TIWWA, audience interaction, (Aga, et al., 2016) (Image: Tate Learning)

To reveal and challenge some of the technological barriers between increasingly complex AI systems and their users, the installation and its algorithmic systems and technologies was opened and hacked live during the



exhibit. This took place as a mix between performers taking over the installation as a performance surface (see figure 49) and the collaborators and participants performing live system hacks and iterations (see figure 51 and 52). This active participation and engagement challenged the context and preconceptions of audiences visiting Tate Modern's Switch House.



**Figure 51: TIWWA, live hacking, modification and iteration, (Aga, et al., 2016) (Image: Tate Learning)**

Whilst TIWWA entered a predefined arts space with an audience primed for the act of observing art, from shallow to deeper levels, it was designed for and located within the Tate Exchange space on the 5<sup>th</sup> floor of the Switch House. This functions as a space for learning, engagement and conversations, rather than as formal gallery spaces. This allowed the work to become more interactive, tactile, explorative and conversational. This was critical to the development of the work as a discursive and dialogue-driven speculation, that engaged audiences online and offline in interactions with the system itself. The ability of the prototype to

augment and insert itself as a memory fragment within its user's imagination became dependant on the level of engagement.



Figure 52: TIWWA, live hacking, modification and iteration, (Aga, et al., 2016) (Image: Tate Learning)

TIWWA created a mirror of people's individual and collective physical and online interactions. The data, generated by people in real-time driving the behaviour and mood of the system, was driving the human-system interplay, but doing so ubiquitously and hidden, and therefore often unnoticed, becoming a symbolic representation of how these systems are integrated into our lives. The observed reactions of the audience went from an initial state of intrigue hesitant to touch and interact, to a state of playful and inquisitive tactile and physical interactions with TIWWA's surfaces and robotic artefacts. Audiences were encouraged to touch by other audience members or the TIWWA collaborators and participants.



Through conversations with audience members interacting with TIWWA, it became evident that people quickly started to project human-like qualities on to the artefacts. People would describe it as having its own intentions, moods and behaviour. The built-in personality and behaviour cues were enough for the audience to anthropomorphise TIWWA's software system. Even though TIWWA's embodiment was abstracted and metaphoric it quickly appeared to become what Turkle (2007b, p. 62) describes as an *evocative object*. People almost instantly displayed affection and tactile play with the sculptural artefacts (see figure 53). Within the first day of the exhibition opening the roaming robotic artefacts was being called 'the babies' and the larger artefact the 'mummy' by audience members. When one of the artefacts broke and had to be opened up and fixed, audiences displayed a mix of responses from intrigue to empathy and concern for its well-being.

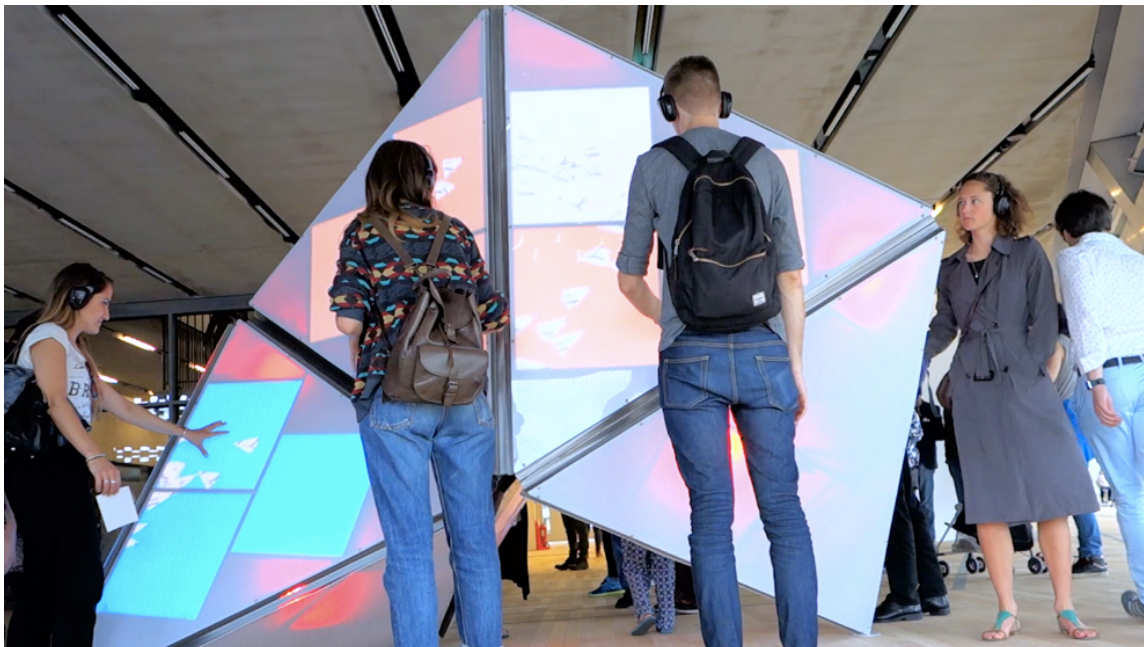


Figure 53: Audience interaction with TIWWA, (Photo credit: Tate Learning)

Audience members appeared to quickly default to human-human communication models in their interactions with the chatbot, as well as with the artefacts. This took place in spite of TIWWA's abstracted forms of disembodied

acts. During the exhibition audiences appeared to show genuine concern when social media data triggered TIWWA to display its 'angry' emotive state; glowing red, with rapid and erratic flocking boids and deep bass sounds. People would ask the TIWWA team members if they upset TIWWA, why TIWWA was feeling angry or how they could cheer it up. It became visible, as identified by McTear et al. (2016), that the display of nonverbal behaviours influenced people's engagement and likeability of a system.

Compared to the large amounts of visitors and social media interactions, TIWWA's conversational interface only had 2065 unique conversations. The physical embodiment of the prototype became the overriding platform for interaction, whereas the conversational interface required people to access the chatbot on the TIWWA website via their digital devices. However, it was through the interactions with the chatbot that audience members first started to ask questions around these systems' hybrid intentions and its influence on people interacting. With language introduced as an interface between the system and the users people began questioning the system and testing its 'human-like' narrative qualities.

TIWWA attempted to challenge traditional thinking and appropriate emergent AI technologies for engagement in discourse rather than commerce. It initiated a nascent conversation with people about current and near future human-machine relationships through participation *with* and *through* conversational and data-driven systems. It activated its participants to critically engage with their own and the wider public agenda around our reliance on computational systems, supporting them to develop an understanding and skill in

developing these technologies and system. The participation of the young people from Tate and the collaborators was in itself the measure of impact.

### **3.1.5. Summary of learning and new knowledge**

TIWWA's collaborative creation, development and delivery were substantial activities and an informative part of this practice-based research enquiry. The observations from this prototype was the first step in the iterative process of testing and documenting the hypothesis of *the perlocutionary effect* of conversational systems. Simultaneously, this was the first activity, in the series of creating four prototypes, testing the potential of a participatory method within a discursive design approach to deepen engagement with the emergent theme and discourse. Here, participation was the measurement of impact. The focus was on activating young people, a priority and objective of this practice, to generate dialogue between people (between participants – experts; participants – audiences; experts – audiences, audiences – audiences) about and through AI systems. This is documented through the traceable participation in practice and interactions with the prototype. The interest was not in measuring the content of these interactions but in documenting and observing them taking place.

TIWWA delivered against the following objectives of this practice-based research inquiry:

- **Evidencing the *relationality of relational things*:**

Through displaying a playful behaviour, both through its surface as well as through the roaming fragments, TIWWA appeared to trigger people's social interaction frameworks. In so doing it documents that even through the simulation of abstracted human-like qualities, *relational artefacts* trigger users to apply their understanding of social relationships to their



interactions with *relational systems*. This reaffirms people's susceptibility to a design strategy which applies human-like attributes in AI systems.

- **Creating actual and traceable participation in discourse:**

The participation of five young people from Tate and the collaborators was in itself the measure of impact and the documentation engagement with discourse. This was the first documentation of the potential of a prototyping practice, with a participatory method, to encourage deeper discourse and engagement.

The observations from TIWWA which inform the following Emoti-OS prototype are:

- **Design strategy:**

Specifically apply a design strategy that exploits the *relationality* of conversational systems through the design of personality and affect.

- **Context and location of participants:**

Focus on designing this both *with* and *for* young people

- **Creating actual and traceable participation in discourse:**

Further test the impact of a discursive design approach extended with a participatory method and consider the aspect of *actual and traceable participation*.

### 3.2. Prototype 02: Emoti-OS



Figure 54: Emoti-OS identity (Aga et al., 2017 and 2018)

The second prototype, Emoti-OS (Aga et al., 2017 and 2018), iterates the findings from TIWWA and analyses the design of *relationality* of conversational systems through simulated affect, personality and collective (artificial) emotional intelligence. Developed in collaboration with a group of pupils at Plymouth School of Creative Arts<sup>77</sup>, i-DAT Collective, Intercity and Control Frenzy<sup>78</sup>, Emoti-OS drew people into a brief conversation about how they felt whilst simultaneously reflecting the collective mood of those to whom it spoke. It was installed as two iterations, 1.0 and 2.0. Emoti-OS 1.0, designed for the Plymouth School of Creative Art, was the primary focus. It was exhibited from the 13<sup>th</sup> November - 19<sup>th</sup> December 2017. Emoti-OS 2.0 was designed after the first iteration and

<sup>77</sup> <http://plymouthschoolofcreativearts.co.uk/>

<sup>78</sup> <https://www.controlledfrenzy.co.uk/>

customised for Tate Exchange London where it was exhibited from the 31<sup>st</sup> January - 2<sup>nd</sup> February 2018.

The system (see figure 57) had two main components; an online chatbot (see figure 55) and a physical installation (see figure 56). The system powering both platforms used Dialogflow<sup>79</sup>, a conversational platform, to construct the conversational system and the IBM Watson Tone Analyzer service to extract and interpret emotional tones from the conversational data generated. It also used Node Red for connecting the different platforms and aggregating the emotive status updates from users. The physical installation (see figure 58) comprising a Raspberry Pi and Arduino Uno with data input buttons and a screen displayed the system question of the day (as generated by the pupils), the seven Emoti-OS emotive emojis and related shares (the number of votes/presses it has had in response to the question of the day).

Emoti-OS had 37,670 'emotional' user interactions (emotive *shares*) and 1,372 unique conversations in the period it was live. The project was also presented at the Nesta conference 'What does artificial intelligence really mean for our education system? Envisioning education by algorithm' (Reith, 2018a), on the Nesta Blog 'Meet Emoti-OS: The chatbot helping to empower pupils' (Reith, 2018b) and featured in the Nesta report 'Educ-AI-tion Rebooted? Exploring the future of artificial intelligence in schools and colleges (Baker, et al., 2019).

---

<sup>79</sup> Dialogflow, <https://dialogflow.com/>, is a Google-owned developer of human-computer interaction technologies based on natural language conversations.

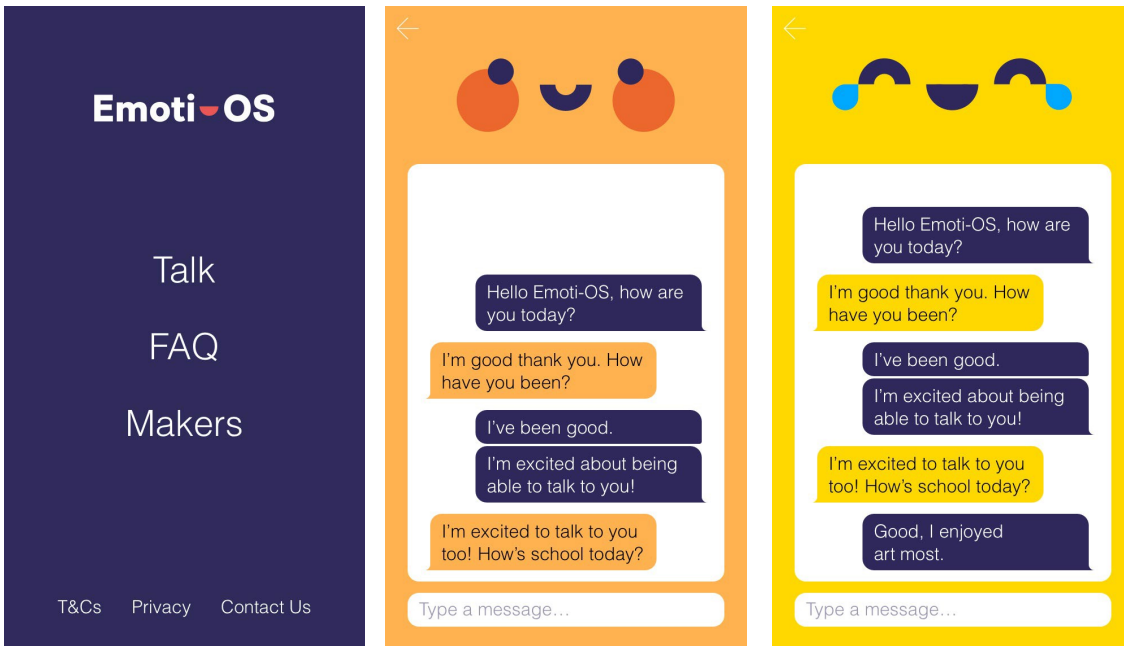


Figure 55: Emoti-OS website chatbot interface (Aga, et. al., 2017)



Figure 56: Emoti-OS physical installation at PSCA (Aga, et. al., 2017)

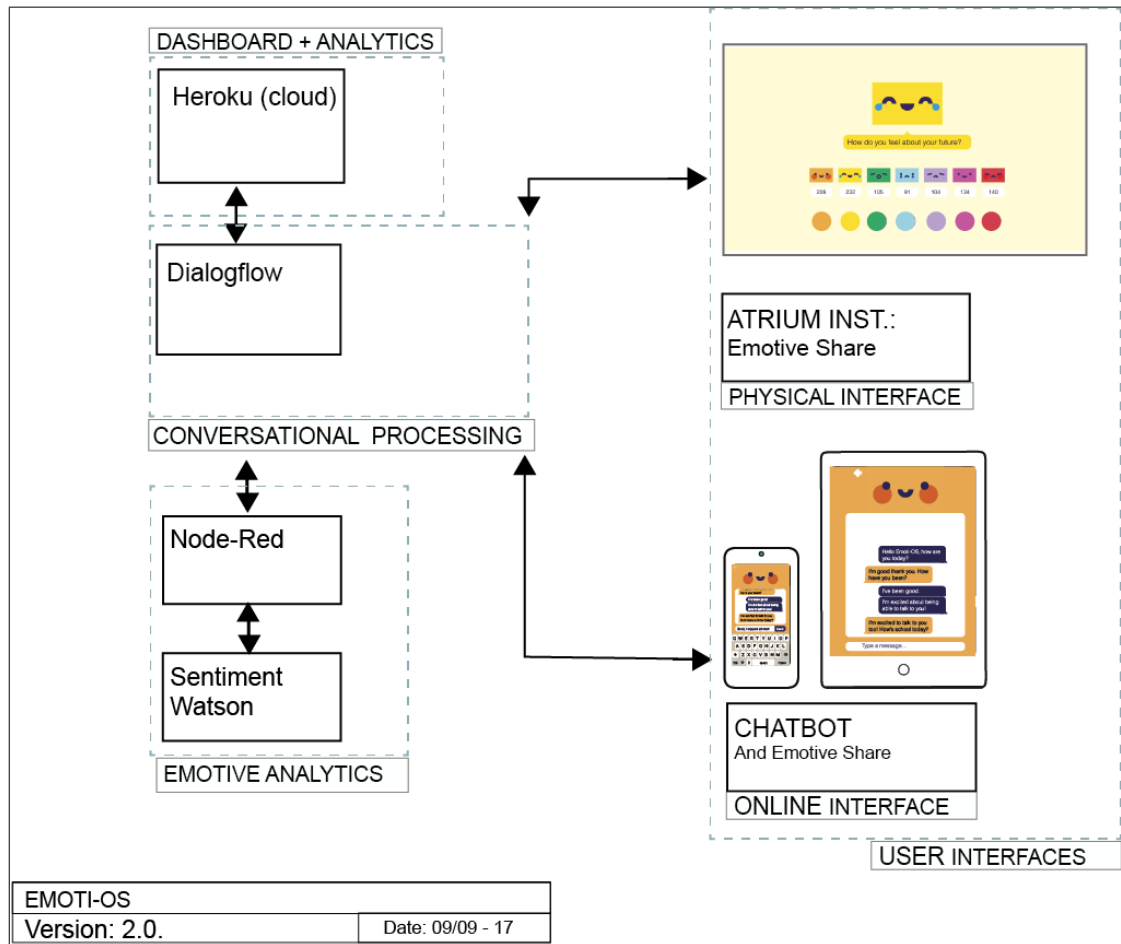


Figure 57: Emoti-OS system diagram (Aga, et al., 2017)

### 3.2.1. Intention

The aim of Emoti-OS 1.0 was to extend the inquiry conducted through TIWWA, and focus on testing and demonstrating *relationality* of conversational systems through a design strategy that applies simulated affect, personality and collective (artificial) emotional intelligence. The intention was to again design a prototype *with* young people, but also this time *for* young people. Plymouth School of the Creative Arts was therefore chosen as the main site and collaborating partner for the project. This delivers against the objective of this research inquiry of engaging those who are generally unrepresented in technology development (with a focus in particular young people and women).

The design of Emoti-OS 1.0 was rooted in the observations and learning from TIWWA which demonstrated people's tendency to project human-like qualities and social attributes onto *relational things*. Whereas TIWWA facilitated fleeting relational interactions through abstract behavioural cues, the intention with Emoti-OS was to facilitate social conversations with people through repeat interactions. The intention was to test a design strategy optimised for relationality to further demonstrate the hypothesis of the *perlocutionary effect*. Whereas TIWWA was designed as a physical and tangible embodied artefact, the intention was to design Emoti-OS as virtual character with a conversational and affective interface. The aim was to further explore the relationality of a screen-based system.

Whereas TIWWA attempted to position the prototype as an amalgamation of all the collaborators' (experts and participants) ideas, desires and vision, Emoti-OS 1.0 and 2.0 intentionally placed the participants at the centre of the design process, with the experts more in a facilitating role. The ambition was to explore if this could increase the relevance and ownership of the intended discourse.

Emoti-OS 1.0 and 2.0 also set out to evidence actual and traceable participation in the discourse put forward through using the data generated by the conversational system. The participation in designing the discourse, and related scenario and artefacts, was here the measurement of impact. The intention was not to evaluate the qualia of this data, but to use it to document participation and any evidence of a *perlocutionary effect*.

Emoti-OS 2.0 was embedded with the design intentions of the first iteration. Emoti-OS 1.0, but was redesigned to nurture shorter, single-event

interactions with a transient audience. The aim was to continue to explore the role of relationality and simulated affect within emergent interactions between people and conversational AI systems. However, whereas Emoti-OS 1.0 was designed to nurture repeat interactions and to build relations with users over time, this system was designed for single interactions. Emoti-OS 2.0 intended to deepen the functionality of the system's simulated empathy and emotion recognition to trigger a more direct discourse with the user/audience about their engagement with affective AI systems. It attempted to be more provocative, also introducing the notion of persuasiveness of conversational systems.

### **3.2.2. Participatory prototyping**

Conceptually framed by the aims and intentions outlined in the section above, the prototyping process was responsive to the participating pupils from PSCA, the tangible (structure, inhabitants, physical and online interactions) and intangible (the ethos of the school, the thoughts, feelings and the moods of the inhabitants) attributes of the school itself. It was also reflective of PSCA being a technology-enabled school in which pupils are encouraged to use and access a range of online devices (such as phones, iPads, laptops work stations) as part of the learning environment. Prototyping began with a hands-on workshop at the i-DAT Collective, Plymouth University, during which the participants were introduced to different conversational software and hardware, with the ambition to provide a deeper understanding of the technologies and data behind these interfaces. This was followed up by a workshop at Plymouth School of Creative Arts where nascent perspectives, desires and ideas were co-developed.

This prototyping process was led by the pupils, in conversation with the wider school population, and facilitated by the collaborating experts. It was designed to give the pupils more authority and ownership within the idea generation and design, leading to greater engagement with the discourse. From these workshops the tension between the power structures within the school and the limited ability for pupils to express how they felt individually and collectively was explored. This led to a series of paper prototypes, tested through focus groups with the wider school population, led by the participating pupils, which informed the final system design, conversational structure and installation build. This process was again developed through a process of collaborative mentoring where participants were paired up with and mentored by the participating experts.

### **3.2.3. Scenario, artefact and system design**

The first iteration, Emoti-OS 1.0, emerged from (and was driven) by the specific context of Plymouth School of Creative Arts and the collaboration with the group of pupils. Whereas five students participating directly with the prototyping and the design of the discourse, they were directly facilitating a wider engagement with their peers, rendering the wider pupil population in the school more as participants than audiences.

The second iteration, Emoti-OS 2.0, was framed by the specific context of being installed at Tate Exchange at Tate Modern Switch House as part of Plymouth College of Art's (2018) Factory Settings exhibition. Taking over the Tate Exchange space for three days the exhibition offered participatory activities and interventions questioning the future of creative learning.



Emoti-OS was designed as a speculative system to gauge the collective mood of the population of pupils within the school, in real time, informed by the pervasive use of emojis, texting and chatbots from across the school youth culture. This was to be done through emotive status updates (*shares*) and a chatbot which could pose and respond to questions to users about how they *felt*. The intention was also that the Emoti-OS participants and their peers (the school pupils) could use this system as a mechanism to pose questions to the school population around issues they felt were important. The ambition was that this would enable the visualisation, in real time, of the collective feelings of the pupils, and therefore create a better awareness and empathy of pupil's collective emotive states. Through pupil-led questions, this would empower them to express their feelings in response to school related matters. Underpinning these intentions was an ambition, driven by the pupils, to challenge the power structures within the school. Emoti-OS therefore became a way to explore the appropriation of conversational systems to empower pupils to express how they feel and see how others felt, and in so doing generate a wider discourse amongst the school population around these issues. To enable all pupils to express themselves freely and to remove barriers to interactions it was collectively decided not to have any registration, log-in or other forms of user-identification.

Emoti-OS had to balance the need for privacy and anonymity with the requirement to ensure safeguarding of pupils, to ensure it could respond to and follow up disclosures made through the chat interface that could lead to harm. This was done by training the chatbot with a list of 'harm words' generated in consultation with the PSCA school safeguarding team. The safeguarding feature would intervene when a harm word was triggered and the users were asked if

they would like to be forwarded to the safeguarding team at PSCA. This provided another layer of child protection. Ethical considerations, including the application of ethical approval of research to University of Plymouth, were considered and adhered to throughout the prototyping process. (See Appendix II for the Emoti-OS Information Sheet and Consent Form). A clear and transparent privacy statement (Aga et al., 2017b) of its purpose, and the use and storage of data was integrated into the online system.

The identity (see figure 54) of Emoti-OS was based on a confluence of the manifestation and social use of emojis, (ideograms and smileys used within messaging platforms), and the architectural shape of PSCA itself. The personality design of Emoti-OS was encoded to be reflective of a young person (12 – 16 years of age); supportive and upbeat, with the capability of simulating seven different emotive states (joy, excitement, fear, sadness, anger, disgust, confusion, as illustrated in figure 58).

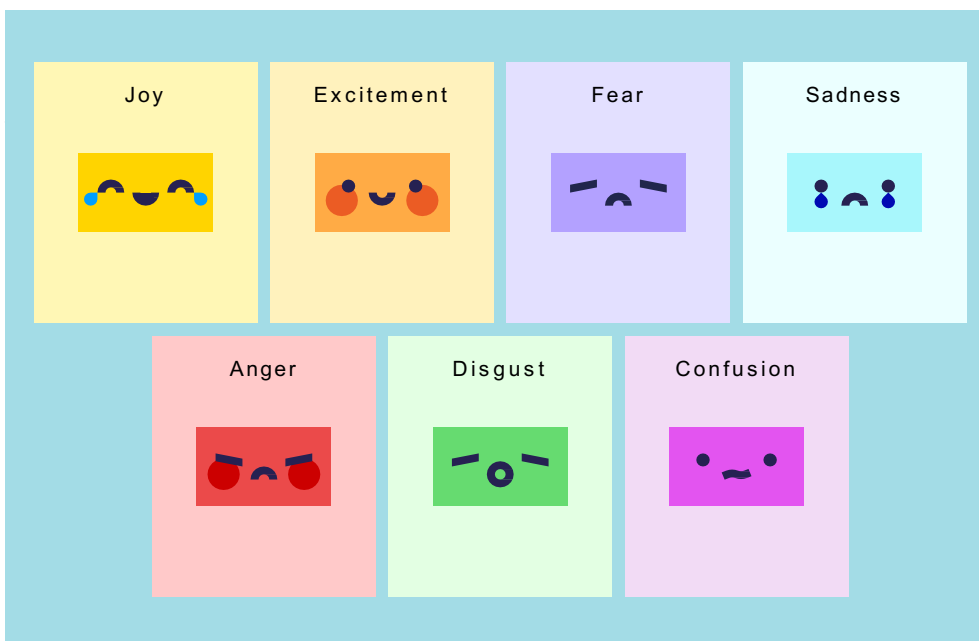


Figure 58: Emoti-OS emotional states (Aga, et. al., 2017)


 Questions List ▾		Add Question
1 Dec 2017, 10:31 a.m.	How do you feel about it being the end of the making	50
1 Dec 2017, 8:43 a.m.	How do you feel about the making projects being gone?	0
30 Nov 2017, 10:31 a.m.	How do you feel about your future?	25
29 Nov 2017, 8:43 a.m.	How do you feel about me, Emoti-OS?	12
27 Nov 2017, 8:59 a.m.	How do you feel about Christmas coming up?	87
24 Nov 2017, 10:05 a.m.	How do you feel about equality?	24
23 Nov 2017, 9:30 a.m.	How would you feel about the school dress code?	42
22 Nov 2017, 9:07 a.m.	How do you feel about one day being an adult?	57
21 Nov 2017, 8:49 a.m.	What or who inspires you?	23

Figure 59: Emoti-OS dashboard (Aga, et. al., 2017)

To facilitate the emotive shares (responses) to questions posed by of Emoti-OS, the installation had a console with seven physical buttons corresponding to each Emoti-OS emoji. This enabled large scale interaction in which the pupils would respond to the question displayed with an emoji corresponding to their emotive response. The system also had a dashboard (see figure 59) with real-time analytics on user emotive shares, conversations and responses to specific questions. The online chatbot (see figure 55) also facilitated the emotive *shares* (responses) to questions posed through its online conversational interface. Similar to the physical installation, it displayed the

overriding mood of Emoti-OS (representative of the collective mood of its users), but also engaged users through dialogue.

The conversational structure was designed to first ask the user to share how they were feeling, (emotive share), followed by a question around why they felt that way. The user could then choose to continue the conversation entering a small talk mode where the user leads the conversation freely around any topic. The direction of the conversational content and topics covered was produced by the participants, who also adapted and modified it based on feedback from their peers.

The overall system structure and interactions for Emoti-OS 2.0 was similar to those of its earlier iteration. However, the change in context from PSCA to the Tate Exchange led to an adaptation to its conversational structure, and goals for interactions. The conversational aspect of this version was more provocative and based on single interactions rather than the repeat interactions of Emoti-OS 1.0. After the initial emotive *share* the system was designed to direct the dialogue through three questions around the future of creative learning. It then ended the conversation with a playful and emotive exchange, in which it offered its own emotive interpretation of how the user was *feeling*. This was conducted through analysing the dialogue the user had exchanged with Emoti-OS using IBM's Watson Tone Analyzer, and attempted to interject the notion of these systems' persuasiveness and authority.

### 3.2.4. Observations and evaluation of practice

The Emoti-OS 1.0 atrium installation (see figure 60) had 37,153 emotive shares when it was live and the online chatbot (see figure 55) had 1,293 shares and conversations. In the first weeks of launching the physical installation it processed on average 2000 emotive shares a day, and then settled around 1000 shares a day until the project completed. Through informal observations of users interacting with both, it became clear that the atrium installation was in a space where all pupils passed in the morning and also where they all had lunch, so there was an ease of access and minimal barriers to interactions (such as access to a device and login on to the Emoti-OS site to speak to the chatbot). The installation became an interface for collective interaction and discussion. Groups of pupils would gather and hit emotive share buttons, express opinions and observations around the overall collective sentiments displayed and questions asked. It became a group activity whereas the chatbot became an individual activity with a higher threshold for engagement.



Figure 60: Emoti-OS user interaction at PSCA (Aga, et. al., 2017) (Image: PSCA)

The engagement with Emoti-OS was facilitated by the participants and a group of twenty pupils who became project ambassadors. They initiated dialogue with the school population about the intentions of the project as well as their feelings towards the system itself. Through this peer-led and informal collection of people's opinions and feelings about the system itself, the conversational system and dialogue was iterated and refined. It became evident that pupils were engaged in discussions on a daily basis around why Emoti-OS was feeling a specific emotion that day. However, rather than creating a discourse around this being reflective of the collective emotion of the school population, Emoti-OS was treated as a person with feelings and behaviours. From the participants' feedback it became evident that the system became a life-like 'being' for whom the population 'cared', and which generated an awareness and empathy of *its* emotive states, rather than the pupils' collective feelings. It also became a tool in activating the pupils in terms of generating discussions and feelings about school matters raised through the questions posed through the Emoti-OS interface.



Figure 61: Emoti-OS emotive badges (Aga, et al., 2018)



The Emoti-OS 1.0 online chatbot was designed for conversational interactions instigating a dialogue with the user around their own feelings and those toward specific school issues raised by their peers. Through observing the conversational data generated three conversational strands emerge (see the Appendix IV: 'Emoti-OS 1.0 Conversational data sample' for examples). The first two are categorised as shallow interactions where the user appears to be testing the human-like qualities and the 'intelligence' of the system, followed by a series of playful exchanges, often with jokes or other forms of expressives. The third strand appears to be relationally deeper; involving longer exchanges and contain utterances indicating that the user relates to and confides in the system – this was often expressed as being happy to be able to talk to Emoti-OS - 'You have helped a bunch'.



Figure 62: Emoti-OS installation at Tate Exchange (Aga, et al., 2018)

Emoti-OS 2.0 (see figure 62) was installed at Tate Exchange for three days, and generated 462 emotive shares and 55 conversations with its chatbot. This iteration engaged a transient audience in short interactions. After asking the user three questions around their feelings towards creative learning and AI systems, the system gave them feedback on their emotional state. This was done through integrating the functionality of sentiment analytics (IBM Watson Tone Analyzer) analysing the user dialogue for emotional tones (see the Appendix VII: Emoti-OS 2.0 Sentiment Analysis System Response for examples).

Emoti-OS 2.0 explored the notion of persuasiveness and simulated emotion recognition to trigger a more direct discourse with the user/audience about their engagement with affective AI systems. Observations from the conversational data from these short interactions (see the Appendix V: Emoti-OS 2.0 conversational data sample for examples) illustrates that the majority of the users appeared to perceive the systems as having little or no relationality. The majority seemed to view and communicate with it much more as a user survey than a relational artefact, as was observed with Emoti-OS 1.0. In comparison with this version, Emoti-OS 2.0 did not have the same buy-in and ownership by its users. The audience/users here had not been involved in a participatory design process and therefore had no ownership of the discourse being made. They were also transient, meaning there were no basis for nurturing repeat and or deeper relational interactions. The overall discourse and relevance of Emoti-OS 2.0 became somewhat lost in the myriad of activities taking place within the Tate Exchange space. Compounding this, whereas TIWWA, had embodied artefacts enticing the user into play, Emoti-OS 2.0 required the user to have the initiative to access the chatbot on their own device or the iPad provided.



In particular with Emoti-OS 1.0 with its; virtual embodiment, emoji-like appearance, relational design, simulated affect and understanding, and in spite of frequent incorrect linguistic responses, it became evident that users quickly attribute personality, intelligence, and emotion to it (as commonly seen with human-agent interactions as previously outlined and evidenced by Fogg and Nass, 1997a; Fogg and Nass, 1997b; Nass, Moon, et al. 1997; Kiesler and Sproull 1997; Reeves and Nass 1999; Turkle, 1995, 2007b and Weizenbaum 1976).

The conversational data from Emoti-OS 1.0 demonstrated that many users continued long dialogues with the system, with over 20 conversational turns. This demonstrates people's susceptibility to a design strategy that exploits our tendency to relate to artefacts that display human-like attributes, even though these artefacts are as Turkle describe (2007b) without true understanding or empathy. It appears, as Turkle (ibid) notes, not to be about whether these systems really have emotion or intelligence, but about what they evoke in us. She (2005a; 2007b) found that children have a greater tendency than adults to perceive computational objects as being 'alive' or 'sort of alive'. This, she (2007b, p.507) explains is not a result of their functionality but of 'children's emotional connection to them and fantasies about how the objects might be feeling about them'.

Emoti-OS 1.0 succeeded in generating discourse through its participatory methods. It also succeeded in engaging users, through designing relationality through its relevance, personality, levels of embodiment, conversational ability and simulated emotional awareness. However, where this version succeeded, Emoti-OS 2.0 was less effective. Without the same sense of relevance,

ownership and repeat interactions, it lost its relationality and became little more than a user satisfaction survey, whereas user interactions with Emoti-OS 1.0 went beyond the intended function, eliciting secrets and desires; becoming an 'other' which some pupils began to confide in and rely on.

The overall engagement of the participants in the central discourse of Emoti-OS was largely evidenced through their role in the design, exhibition and audience engagement process. Implementing a participatory prototyping process that empowered and encouraged their ownership increased their engagement and the impact of the project's discourse. Feedback from the focus groups led by PSCA project lead and Computing Lead Angus Reith (summary of feedback is found in the Appendix VI: Emoti-OS Participant Feedback Summary) demonstrates that this has been an influential project and experience for their pupils. They found that Emoti-OS became a talking point amongst them and their peers, both as a tool to give them a student voice, but also as an intervention to talk about feelings as well as their own use of Ai technologies to communicate these. The fact that the project was a student-led initiative, it was felt that it had a much greater buy-in and resonance with the student body than it would have had it been an external research project. He explains;

With that student engagement therefore came meaningful human-machine dialogue and in fact a tool in itself to prompt dialogue. The pupils found that it created a better awareness and empathy of how people were feeling, in particular in relation to school matters.

Reith explained that they also found that pupils readily displayed an emotional bond with the Emoti-OS conversational system, evidenced in what they found as an often 'warm, genuine and thoughtful conversation' in many interactions. Through observing the conversational data and speaking to their

peers, Emoti-OS provides a non-human (and non-judgmental) sounding board; the participants believe there was a therapeutic nature to some of the conversations. In interacting with the system, pupils seemed to be navigating their own emotional territory and many demonstrated a degree of closure at the end of the chat with Emoti-OS. With the Emoti-OS project coming to an end, the participating pupils reported that their peers missed 'having it around' and wanted to know if they could still 'talk to it and keep in touch'.

### **3.2.5. Summary of learning and new knowledge**

Designed through a strategy that exploited the relationality of conversational systems through personality, simulated affect and emotional intelligence, Emoti-OS 1.0 demonstrated how users attribute human-like qualities to a system, and reaffirms the hypothesis of the *perlocutionary effect*. This was also evidenced through TIWWA, however Emoti-OS with its virtual rather than physical embodiment, abstracted emoji-like character and conversational interface, was able to engage users in deeper, emotive and repeat social interactions. This evidence what Reeves and Nass (1999) identified as people's tendency to project human characteristics on non-human things, and not as Turkle (2007b, p 62) argues based on their 'intelligence, consciousness, or reciprocal pleasure in relating' but rather in 'their ability to push our Darwinian buttons.

The focus of Emoti-OS was to designing this prototype both *with* and *for* young people, activating them to generate dialogue with each other and their peers, *about* and *through* the AI systems. The participation in the practice was observed and demonstrated through the system itself (conversational data and

number of interactions). This delivered against the overall aim of this research-based inquiry, both engaging young people and extending the ability to demonstrate actual and traceable engagement with the discourse.

Emoti-OS delivered against the following objectives of this practice-based research inquiry:

- **Evidencing the *relationality of relational things that talk*:**

Emoti-OS evidenced the *perlocutionary effect*, in particular the *relationality, of relational things that talk*. With its; virtual embodiment, emoji-like appearance, relational design, simulated affect it understanding triggered people to apply their understanding of social relationships to their interactions.

- **Design strategy:**

Emoti-OS evidenced and reaffirmed people's susceptibility to a design strategy that exploits the relationality of conversational systems through the design of system attributes; personality and affect.

- **Context and location of participants:**

Emoti-OS prioritised the engagement of young people in its design and discourse both *with* and *for* young people.

- **Creating actual and traceable participation in discourse:**

The participation of five young people from PSCA and the wider participation of their peers was in itself the measure of impact and the documentation engagement with discourse. This was further extended to evidencing actual and traceable participation through the interaction and conversational data generated through the system interface. This was the second documentation of the potential of a prototyping practice, with a participatory method, to encourage deeper discourse and engagement.

The observations from Emoti-OS which informed the following WRAI prototype are:

- **Context of participants:**

Prioritised the engagement of self-identifying women who are substantially underrepresented in the development of AI, and challenge stereotyping within conversational AI.

- **Empower participants to challenge, discuss and participate in technology development:**

Focus on empowering people to challenge, discuss and participate in technology development by creating opportunities for hands-on technology development and design.

- **Claim emergent conversational AI technologies**

Claim emergent conversational AI technologies *with* their target users for new forms of creative practice, which has social debate as a fundamental goal and desired outcome, to disrupt the commercial development of unethical and exclusive AI systems.

### 3.3. Prototype 03: Women Reclaiming AI



Figure 63: Women Reclaiming AI identity (Aga and Manton, 2018)

Women Reclaiming AI (Aga and Manton, 2018) was a collaborative AI Voice Assistant and AI-activism project made by and for self-identifying women to challenge stereotyping and bias in commercial conversational AI

developments. The system itself (see figure 66) consisted of two interfaces; a Google Home Assistant (see figure 64) and a chatbot on the WRAI website (see figure 65). The project reclaims AI for artistic innovation and social action rather than commercial exploitation. It responded to the lack of female voices (and others such as genderqueer, non-white, non-western etc.) in AI development, reclaiming these technologies for collaborative activism, art, engagement and social impact.

Through activating a diverse community of forty-five self-identifying women, through four free workshops across Plymouth (KAST<sup>80</sup>), Bristol (Knowle West Media Centre<sup>81</sup>), and Coventry (Random String Festival<sup>82</sup>), the project set out to empower women with a conceptual and technical understanding of conversational technologies in order to challenge the status quo of their development. The project was developed across April – November 2018 in partnership with artist, technologist and researcher Coral Manton. WRAI has also been presented at the following public events; ‘Human Bias In Artificial Intelligence: Can we fix it?’ at KWMC in Bristol (Aga and Manton, 2018a), ‘Things That Talk. AI In Art – A FACT Late Night’ at FACT in Liverpool (Aga and Manton, 2018b) and ‘Disobedient AI’ at the Random String Festival in Coventry (Aga and Manton, 2018c). WRAI has also received funding from Arts Council England to continue a series of workshops and exhibitions across the UK in 2019.

---

<sup>80</sup> <http://karst.org.uk/>

<sup>81</sup> <https://kwmc.org.uk/>

<sup>82</sup> <https://randomstring.co/>



Figure 64: The WRAI conversational interface was deployed locally on a Google Home Assistant (Aga & Manton, 2018)

*Women Reclaiming AI is collaborative AI Voice Assistant made by women for women*

This project is in response to lack of gender diversity in the development of AI systems. The aim is to develop an AI voice assistant programmed through collaborative workshops

Women Recla...  
Offline

Nov 24, 4:16 PM

can I have some advice?  
3:53 PM ✓

Women Reclaiming AI

Do you have a vagina? And do you want to be in charge of it? If you said 'yes' to both, then congratulations – you're a feminist! (Caitlin Moran)  
3:53 PM

can you help me  
3:53 PM ✓

Women Reclaiming AI

ugh, do I have to?  
3:53 PM

Figure 65: The WRAI website chatbot (Aga & Manton, 2018)

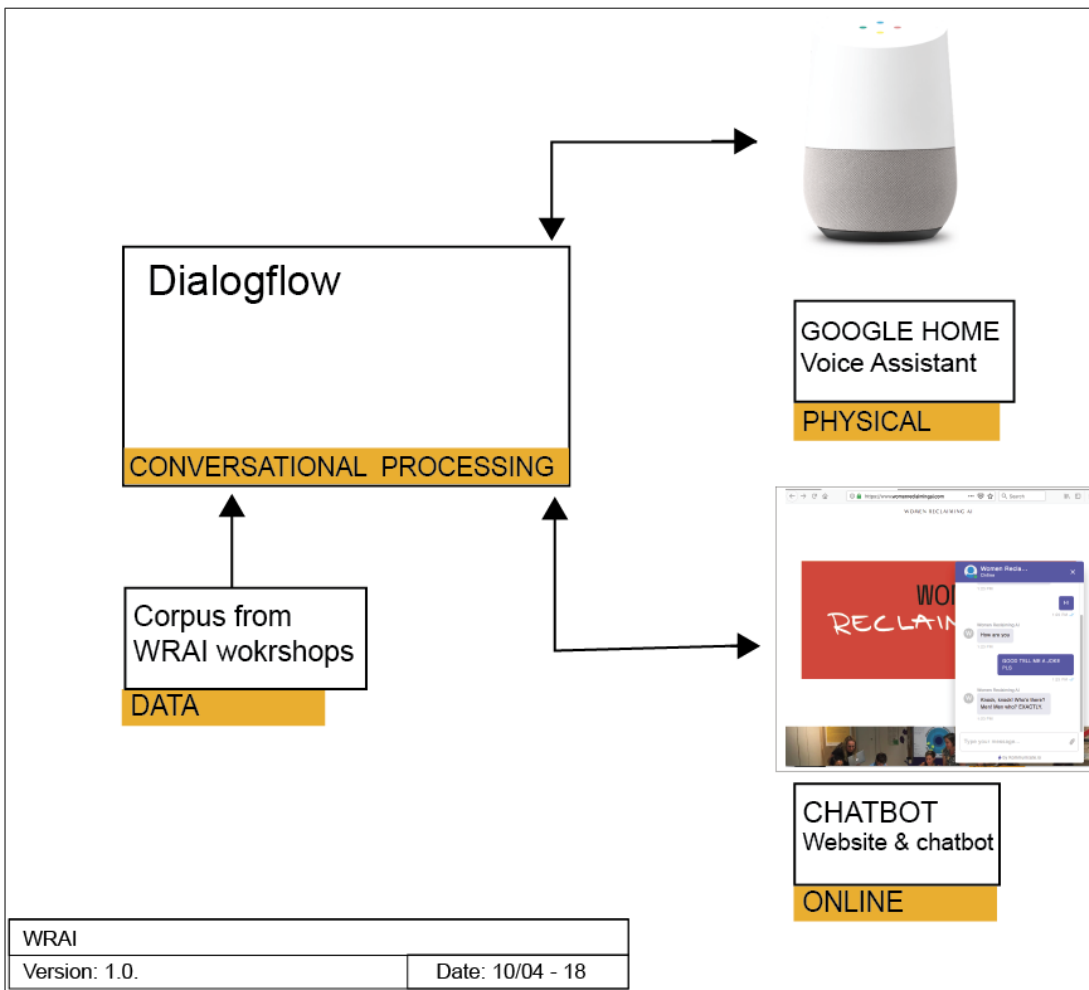


Figure 66: WRAI system diagram (Aga & Manton, 2018)

### 3.3.1. Intention

The intentions of WRAI were focused on generating discourse *with* participants, (rather than with audiences engaged through public presentations of work), *through* conversations by collaboratively *doing* practice. WRAI creates a discourse around stereotyping and bias in current commercial AI developments. It differs from the others in that it explicitly grapples with the commercial technology development process and its lack of diversity and actively engages participants as a form of activism, attempting to empower them to challenge the status quo. This desire emerged from observing the complexity of the previous prototypes, not the amalgamation of diverse intentions, desires and



ideas, but the complexity of the artefacts and their scenarios. The aim was to claim conversational AI *with* their target users for new forms of creative practice which create discourse through conversations with participants, rather than creating complex systems played out as public scenarios. Whilst previous prototypes engaged a large number of audiences and had big collaborative teams of up to twenty people, with up to fifteen participants, WRAI was developed in collaboration with one artists / technologist. The ambition was to reduce the complexity of the collaboration and the prototype, and instead scale up the participation to engaging fifty self-identifying women, through a series of free, short and collaborative three-hour workshops.

The project draws a reference to Haraway's (1991) 'A Cyborg Manifesto' and a move beyond the limitations of traditional gender, feminism, and politics agendas. It grounds the project ideas in what Renzel (2017) describes as cyberfeminism's groundwork for the 'concept of the internet being a revolutionary tool to overthrow patriarchy, destroy the existing gender binary and achieve feminist liberation'. However, WRAI recognises that the Internet (or emergent technologies) is not a new neutral space as envisioned by Haraway, where women can overthrow the existing social order, but very much embedded and reflective of society. Emergent technologies are human creations and therefore, as Renzel (2017) stipulates, biased and imbued with the sexist and racist assumptions which pervades society.

WRAI does embrace the intention of cyberfeminism and other forms of activist art on the Internet for their potential to reclaim these media as spaces for the participation of women-led discourse. It points to Looui and Flanagan's (2017, p182) review of activist art on the Internet where they draw a parallel with the lack

of representation of women in computer science, to a similar lack of ‘women artists who are producing theoretically challenging and technologically ‘cutting-edge’ websites that are also explicitly feminist. This position informs the WRAI project to apply not only the Internet, but emergent technologies such as conversational systems as tools for women’s resistance through participation.

The project’s intentions emerge from this approach and simultaneously frame its agenda as a direct response to the specific lack of gender diversity<sup>83</sup> in the development of AI systems, and the use of gender stereotypes and biased data-sets in the development of AI voice assistants like Google Assistant, Alexa, Siri, Cortana and Bixby. These all use female names, identities, voices (as a response to pressure an increasing number now have an option to change the gender of the voice to male). Fessler (2017) highlights that gender roles and sexism are reinforced through these systems where female assistants are portrayed as subservient and assisting, being programmed with flirtatious and submissive responses to any verbal sexual harassment by users. By engaging a community of self-identifying women, with no previous experience of technology development, the ambition was to empower these women to design their own conversational system constituted by their own decision around gender identity and roles. This position informs the project to apply emergent technologies such as conversational systems as tools for women’s resistance through participation.

---

<sup>83</sup> WIRED magazine conducted a research exercise in 2017 where they found that only 12 percent of women contributed work to the three leading machine learning conferences in 2017 (Simonite, 2018). According to Google, 21 percent of technical roles are filled by women, however when WIRED reviewed their AI research page out of the listed 641 people working on ‘machine intelligence’, only 10 percent were women (ibid). However when delving deeper into these stats to look for diversity beyond gender, companies such as Facebook which states that 22 percent of its technical workers are women, there is only 1% of Black employees in technical roles (Williams, 2018).

### **3.3.2. Participatory prototyping**

Participation in WRAI was facilitated through four workshops delivered to engage forty-five self-identifying women from across Plymouth, Bristol and Coventry. All workshops were free, lasted three hours, and delivered in partnership with local arts organisations to their communities. The project intentions outlined above shaped and underpinned the creation of a nascent community and female-led technology making space for women through the workshops. These were reminiscent of DIY and activist-based female knitting circles, as described by Segal (2017), Stoller, (2003) and Hermanson (2012). This entailed creating a space which was accessible (physically and conceptually), non-hierarchical, supportive and confidential. Throughout each workshop this ethos was iterated and discussed, whilst each participant was encouraged to express their opinions, listen to others and to share any thoughts.

The workshop provided the participants with a basic understanding, knowledge and skills in scripting and making an AI voice assistant in Dialogflow. Simultaneously, the act of collaboratively making was a process to facilitate talking, sharing and listening. Rooted in a discourse of cyberfeminist and forms of online activism, the conversations were framed around highlighting the current misrepresentation and lack of diversity in current AI development. Through this process a voice assistant was co-created with a dominant and non-serving personality, and with a corpus of speech created with the community of participants which they collectively felt was a more representative manifestation of a self-identifying woman.

### 3.3.3. Scenario, artefact and system design

The WRAI prototype was focused on creating participatory discourse through the act of co-creating, rather than engaging the wider public in discourse. Its scenario emerged from the collective world-view of the collaborating self-identifying women, creating a conversational assistant deployed to a Google Home assistant (see figure 64) as well as a chatbot on the WRAI website<sup>84</sup> (see figure 65). The personality and functionality design of these artefacts became reflective of the discourse generated. It was dominant and non-serving, and ingrained with an amalgamation of interpretations of what a woman should be. These were not artefacts design for public consumption or for efficient user satisfaction. They emerged as conversational objects and spaces of resistance, where the value lies in the collaborative act of shared discourse. These were vessels for the imagining not of what *is* but of what *could be*.

The WRAI Assistant was not publicly available, but used as a demonstrative object in the workshops. This was reflective of the project objectives with a focus on the participation, not on public engagement. However, in response to the workshop participants' desires to have a public version available they could share, showcase and discuss beyond the workshops, we made a chatbot version available. This in itself becomes a process of discourse, releasing an unfinished, DIY, non-censored growing artefact, that itself was not designed for user-satisfaction, but for protest.

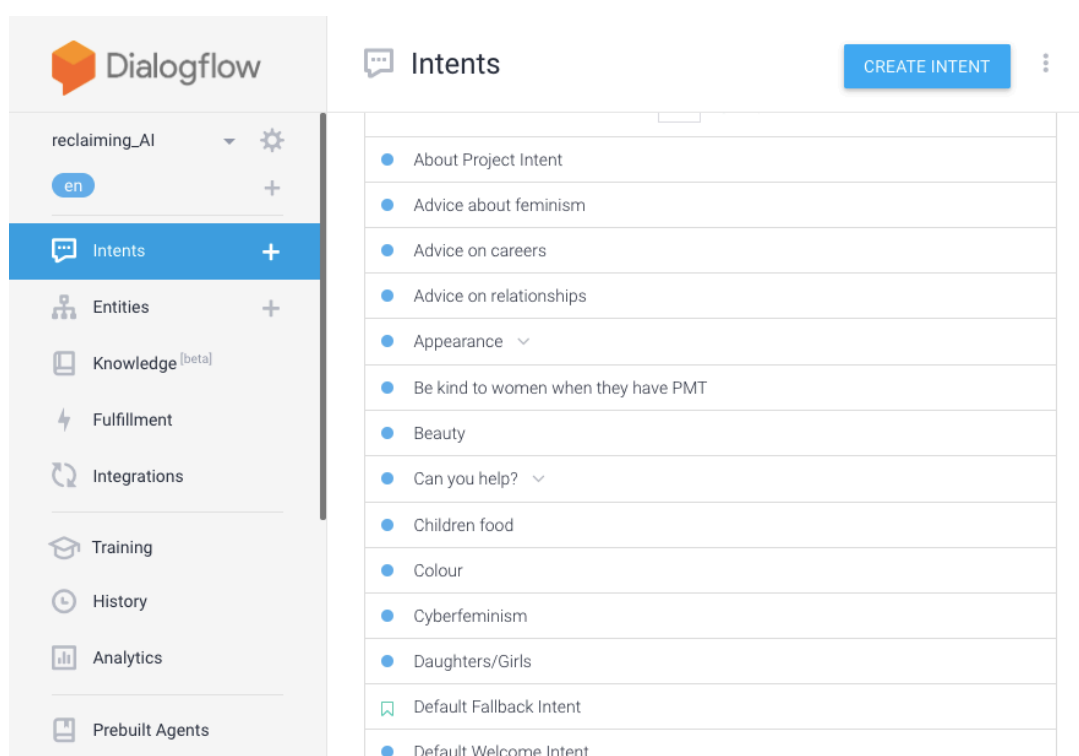
The conversational interface and corpus were created through the Dialogflow platform (see figure 67). This is a platform created for the commercial development of conversational systems, and not designed for mass-participation

---

<sup>84</sup> <https://womenreclaimingai.com>

or community actions. The process of using this platform and exploiting its purpose, became in itself important for the action and discourse generated. Each participant was given full editorial access and trust to contribute through the adoption of the WRAI community guidelines;

As a member of a growing community, we ask you to treat others in the group with respect and not to delete or change any other people's entries from DialogFlow. We ask that you do not make any entries that could be offensive or discriminatory to any vulnerable or minority groups. Your entries into DialogFlow will be publicly visible by people using the Chatbot on the website so take care with the words you choose. Any words that are not your own and entered into Dialogflow needs to be referenced with the author in brackets.



**Figure 67: Dialogflow the conversational platform used for the WRAI workshops (Aga and Manton, 2018)**

The conversational corpus generated through Dialogflow, and which drove the conversations with the WRAI system, was reflective of the participants and ranged from advice on 'how to be a feminist' and 'how to raise a daughter', to feminist jokes and inspirational speech. However, the content of the conversations with the WRAI chatbot, or the wider public interactions with it, was

not important here. It was the engagement with the discourse around the stereotyping of women and lack of diversity and participation in the development of conversational AI systems that was central, and which drove the overall manifestation of the WRAI scenario, artefact and engagement.

#### **3.3.4. Observations and evaluation of practice**

The participation in the workshops and the discourse generated through these was here the measurement of impact. It was explicitly concerned with actively engaging and upskilling self-identifying women as a form of activism, attempting to empower them to challenge the status quo. It drew a reference to participatory design as an act of empowerment and democratisation founded on a principle of creating legitimate participation for all stakeholders during periods of technological change (Malpass 2016; DiSalvo, 2012). This process embraced Schuler and Namioka's (1993) proposition as a process of engaging participants' imagination rather than seeing them as mere respondents or as agents to observe and from whom to generate data.

By engaging a community of self-identifying women, with no previous experience of technology development, through the design of a workshop space that was accessible, (physically and conceptually), non-hierarchical, supportive and confidential, it gained their trust and participation. The design of the space was an essential part of the discourse, modelled as a nascent community and female-led technology making space (for self-identifying women) reminiscent of DIY and activist based female knitting circles. Throughout each workshop this ethos was iterated and discussed, whilst each participant was encouraged to express their opinion, listen to others and to share any thoughts. Through the

workshops, the women were given a hands-on opportunity to design their own conversational system constituted by their own views around gender identity and roles, embedding them in the discourse of the existing bias, lack of diversity, and stereotyping in current commercial AI developments.

Through this community space and the act of providing access to the tools and understanding of AI development, the participants were given some basic tools for collective resistance. This act nurtured a conversational discourse within the framework for the workshop but also beyond. Each participant was encouraged to share (online and offline) these conversations and the WRAI voice assistant itself (accessible from the WRAI website). Each participant had access to the development platform post-workshop, and most continued to add to the conversational corpus.

### **3.3.5. Summary of learning and new knowledge**

Through this process WRAI collectively reclaimed the AI technologies (and related commercial development processes) it was critiquing, as a medium for women's resistance through collective participation. By doing so it follows Consalvo's (2002, p. 109) call to action; 'women must be part of this future, not by simply advocating for more women to engage in using technology, but by becoming more critically aware of the perils as well as promises that new technologies offer.' It was a call to social action, through solidarity, learning, freedom, vision, and resistance, to empower women to 'code'; a discourse which demands a more desirable and representational technologically mediated future.

WRAI delivered against the following objectives of this practice-based research inquiry:

- **Empower women as participants to challenge, discuss and participate in technology development:**

WRAI prioritised the engagement of self-identifying women in hands-on technology development and design to generate dialogue and empower them to challenge the lack of diversity and presence of stereotyping within conversational AI developments.

- **Claim emergent conversational AI technologies**

TIWWA claimed emergent conversational AI technologies *with* their target users for new forms of creative practice, which had social debate as a fundamental goal and desired outcome, to disrupt the commercial development of unethical and exclusive AI systems.

- **Evidencing the bias and stereotyping in conversational AI technologies:**

WRAI evidenced the presence of bias and stereotyping in conversational AI technologies, a factor which propagates their *perlocutionary effect*.

- **Creating actual and traceable participation in discourse:**

The participation of self-identifying women in the WRAI workshops was in itself the measure of impact and the documentation engagement with discourse. This was further extended to evidencing actual and traceable participation through the interaction and conversational data generated through the system's conversational interface. This was the third documentation of the potential of a prototyping practice, with a participatory method, to encourage deeper discourse and engagement.

The observations from WRAI which informed the following Infinite Guide prototype are:

- **Design strategy:**

Specifically apply a design strategy that exploits the relationality of conversational systems through the design of system attributes; personality and affect.

- **Context and location of participants:**

Focus on designing this *with* young people but for the wider public to widen the engagement with the discourse put forward.



- **Creating actual and traceable participation in discourse:**  
 Further test the impact of a discursive design approach extended with a participatory method and consider the aspect of actual and traceable participation through the application of conversational systems themselves as mechanism to engage people in dialogue.
- **Empower participants to challenge, discuss and participate in technology development:**  
 Focus on empowering people to challenge, discuss and participate in technology development by creating opportunities for hands-on technology development and design. Also focus on extending this dialogue with the wider public to widen impact of the prototype.
- **Claim emergent conversational AI technologies**  
 Claim emergent conversational AI technologies *with* their target users for new forms of creative practice, which has social debate as a fundamental goal and desired outcome, to disrupt the commercial development of unethical and exclusive AI systems.
- **Synthesise and substantiate overall claims from previous prototypes:**  
 Test and synthesises the findings from the three previous prototypes to substantiate the overall *perlocutionary effect* of conversational AI system. Further evidence the impact of a discursive design strategy to activate people to challenge the commercial development of conversational AI systems and re-imagining more desirable future realities being with *relational things that talk*.

### 3.4. Prototype 04: The Infinite Guide

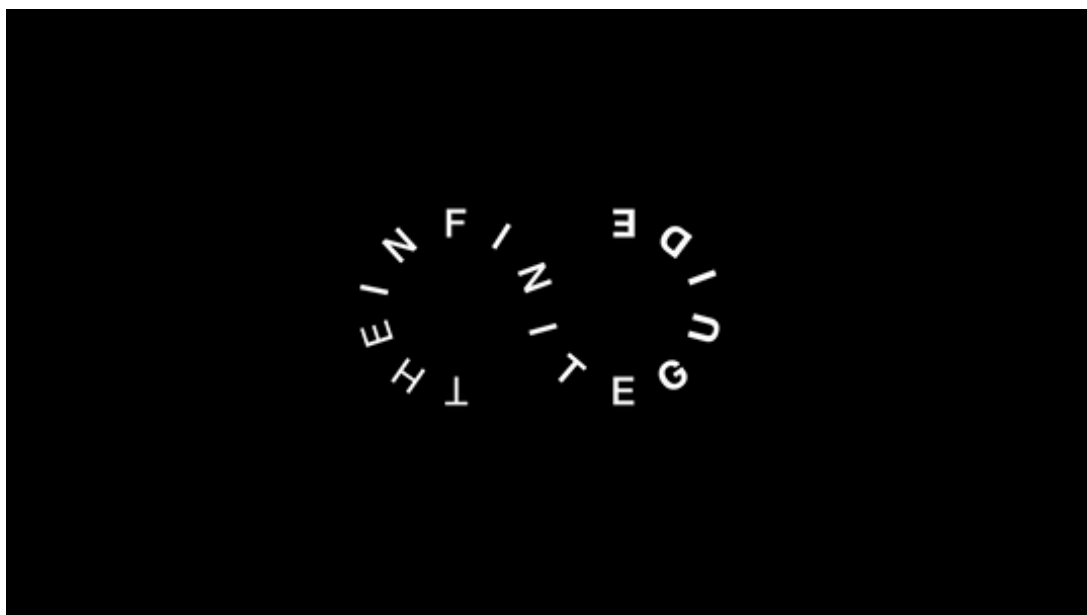


Figure 68: The Infinite Guide Identity (Aga et al., 2018b)

The Infinite Guide (Aga et al., 2018) synthesises the findings from the three previous prototypes to substantiate the overall *perlocutionary effect* of conversational AI system. It consisted of three online chatbots, and physical installation with a voice interface and an audio-visual projection (see figures 70 - 76). The Infinite Guide system (see figure 69) was constructed through the same technologies used in the previous prototypes; Dialogflow for the conversational interfaces (chatbots and voice interface), Node Red for the system processing and networking, and IBM Watson for analysing sentiments in the conversational data generated. However, this system also incorporated the LSTM Recurrent Neural Net to generate text responses for users. Overall the Infinite Guide system was designed to incorporate all processes of user engagement and discourse, intentionally creating a fully technology-led experience of the systems the discourse was attempting to critique.

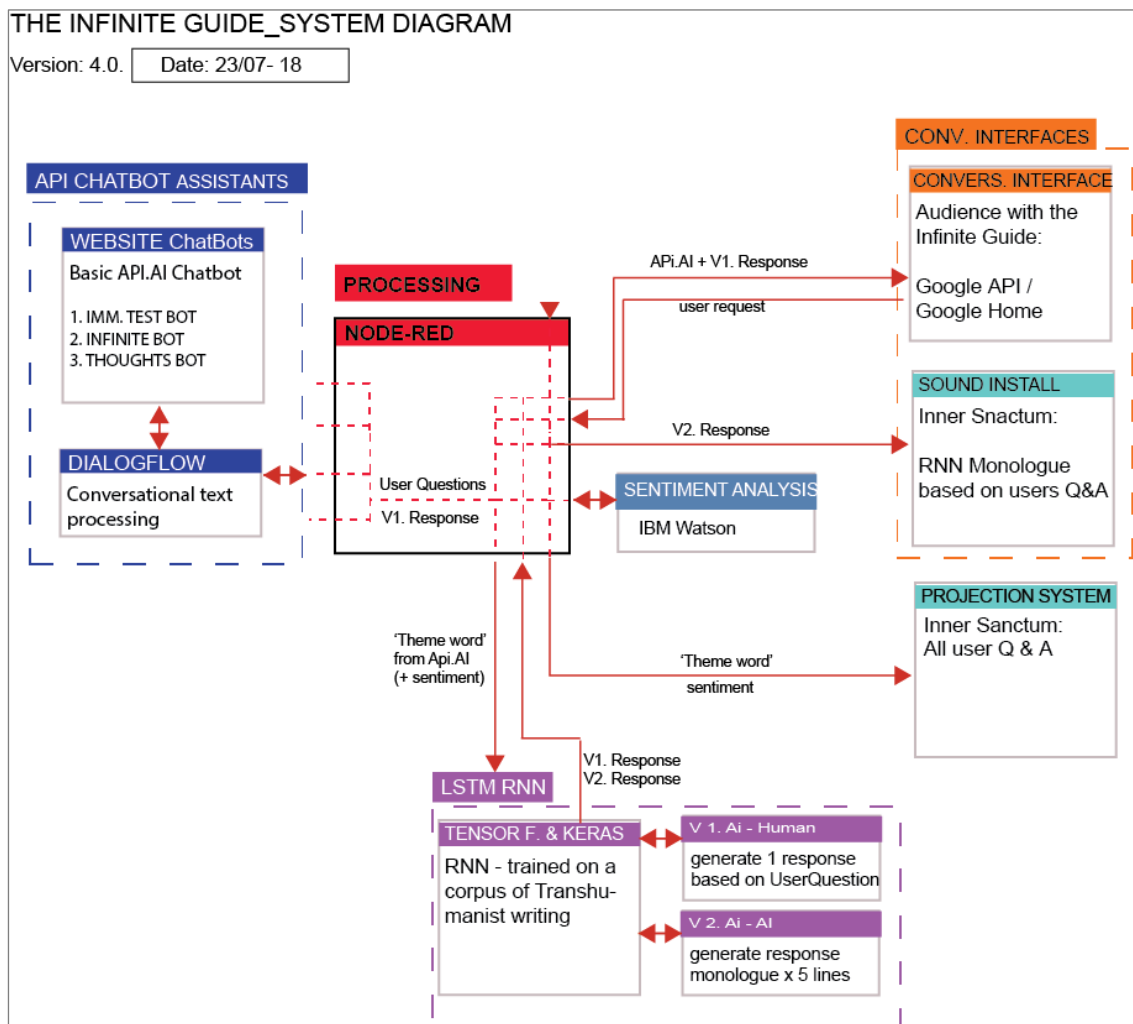


Figure 69: The Infinite Guide system diagram (Aga, et al. 2018b)

The Infinite Guide was developed in collaboration with a group of young people through Fact<sup>85</sup> and KARST<sup>86</sup>, the i-DAT Collective, InterCity, Story Juice<sup>87</sup>, University of Plymouth's Digital Media Design programme<sup>88</sup> and the student start-up Crumb Design<sup>89</sup>. The project engaged people in a narrative and speculative experience which took place online and physically, through the exhibition and talks at KARST the 1st – 9th September 2018. The project, funded by the Arts Council England, engaged 15 young people as participants, 4 young

<sup>85</sup> <https://www.fact.co.uk/>

<sup>86</sup> <http://karst.org.uk/>

<sup>87</sup> <http://storyjuice.co.uk/>

<sup>88</sup> <https://www.plymouth.ac.uk/courses/undergraduate/bsc-digital-media-design>

<sup>89</sup> <https://www.crumbdesign.co.uk/>

people as collaborators, 78 young people in participatory talks, and an audience of 1466 (1215 online). The project was presented at the 'Disobedient AI' panel at The Random String Festival (2018) in Coventry, the 'Human Bias In Artificial Intelligence' (2018) at KWMC in Bristol and the 'AI In Art – A FACT Late Night' at FACT in Liverpool (2018).

### 3.4.1. Intention

The Infinite Guide emerged at the confluence of the findings from the previous iterative prototypes and participatory processes; peoples' susceptibility to human-like systems and these systems' authority, relationality and persuasiveness, in spite of their lack of transparency, imbued bias and hybrid intentions. The intention was to further explore the *perlocutionary effect* of conversational systems; their relationality, persuasiveness, hybrid intentions and bias, and to nurture deeper engagement through participation with discourse.

The *relationality, persuasiveness, bias and hybrid intentions* of The Infinite Guide was intentionally constructed through the design of its systems functions; personality design, personalisation, conversational ability and affective functions. The intention was to propagate the *perlocutionary effect* of conversational systems to affect the thoughts, beliefs and actions of people interacting with it, amplified through the use of online and physical interfaces. In so doing substantiate the findings from the three previous prototypes in documenting the overall *perlocutionary effect* of conversational AI systems.

The work was grounded in the proposition that as the complexity and power of intelligent systems expand beyond human comprehension, faith is necessarily reintroduced. This positions the computational system as an

ideological and mythic incarnation of an emergent artificial spiritual guide, a guide residing in ‘cathedrals of computation’ (Bogost, 2015), conjured through scientific and technological progress, whom we believe leads to our immortality. A quest, driven by a pursuit for transcendence and salvation where we, according to Gollner (2013, p. 35-36), take for granted that; ‘suffering can be eliminated, that poverty will ultimately be eradicated, that we should never be sick again, that science will soon make everybody never die’. A technological future forged in the naïve, biased and diversity-challenged furnace of Silicon Valley if left unchallenged.

The intentions of the Infinite Guide project were rooted in this position, speculating around the current and near future influence of our increasingly symbiotic relationship with intelligent systems, guiding our thoughts, actions and behaviour, and their juxtaposition with their emergence as artificial spiritual guides. It set out to unravel the complexities of these systems, revealing and questioning their bias and hybrid intentions. Ultimately the ambition of The Infinite Guide was to activate people to challenge the commercial development of conversational AI systems and re-imagining more desirable future realities being with *relational things that talk*.

Through appropriating conversational AI systems with *people* the ambition was to generate a speculative discourse, as a counter-narrative about what it will mean to be human, individually and collectively, in a near future increasingly living with and through AI systems. In so doing it will invite participants and audiences to reflect on their desired future relationship with these technologies and the inherent pursuit of immortality through technological innovation.

The project extended TIWWA and Emoti-OS's application of a design strategy that exploits people's reliance on and emergent relationship with AI powered conversational systems, with simulated affect, personality and intelligence. At the same time, it intended to expand the discourse around the inherent bias and hybrid intentions of AI systems explored in WRAI. These previous prototypes manifested as a confluence of the ideas and desires arising from participants and collaborators, shaped within pre-defined conceptual frameworks emerging from the theoretical exploration of this thesis. These were responsive and shaped by their context (place, space, people and systems). The primary purpose of their related scenarios and artefacts was to communicate ideas and trigger conversations, as 'tools for thinking' (Tharp and Tharp, 2019). Here the importance was on participation itself, and the narrative context and public presentation of the artefact itself secondary.

The Infinite Guide, rooted in the learning from these previous iterations, was driven by an intention to test if its wider impact and engagement could be increased by creating a prototype, designing specifically for a public presentation and discourse, with a more provocative near future narrative. A narrative delivered and extended through the conversational systems it was critiquing. Referencing the practice of design fiction (Bleecker 2009; Grand & Wiedmer 2010; Sterling, 2009) it intended to place a strong emphasis on technological futures and the use of a fictional scenario, but also, as in the previous iterative prototypes, integrated working systems and real-time data to trigger individual and intimate interactions with these systems.

The Infinite Guide appropriated conversational and AI powered technologies that exist today to extrapolate interactive and data-driven scenarios

of what they, and the human consequence of these, could be in the near future. By doing so it aimed to test the impact of claiming conversational AI technologies *with* their target users for new forms of creative practice, which has social debate as a fundamental goal and desired outcome, to disrupt the commercial development of unethical and exclusive AI systems.

### **3.4.2. Participatory prototyping**

Conceptually framed by the intentions outlined above, the prototyping process differed from the previous iterations. The Infinite Guide intended to bring the concerns and learning from these iterative prototypes together through designing a near future fictional narrative, presented in the public realm and designed, delivered and evaluated through technological systems it was critiquing. The requirement to deliver this informed the need for authorship of the speculative narrative an overall experience design. Whereas the development and design of TIWWA and Emoti-OS's scenarios and the artefacts emerged through a collaborative process, responsive to and shaped by the people taking part, The Infinite Guide became responsive to and shaped by these intentions instead of an amalgamation of the intentions of participants. Participation and collaboration were still central to the process, and to the overall intended impact and engagement with the discourse. The participatory process was formed through a partnership with FACT Learning, the outreach programme at KARST, Plymouth University's Digital Media Design programme and the already established partnership with Plymouth School of Creative Arts. It was designed to engage young people as participants (aged twelve to twenty-five) and the wider public as the audience.

The participatory prototyping process started with a two-day long workshop at FACT, with a group of 15 young people. They were given the overall fictional scenario of The Infinite Guide which was set in a near future where AI systems have become intimate and prophesying intermediaries, fuelled by personal data and promises of digital transcendence from mortality. The workshop was driven by the intention to engage participants in discourse, where discussions, theorising and prototyping were all manifestations of participation in this discourse. This was delivered through a mix of future gazing, hands on experimentation and paper prototyping, followed by a series of collaborative sessions where the group developed different prototypes responsive to the fictional narrative of The Infinite Guide. This was also an iterative process, going from paper prototypes to simple working artefacts. This was simultaneously a process to test the engagement with The Infinite Guide's scenario and artefacts. It enabled the participants with basic skills and knowledge to unravel the complexity of conversational AI technologies as well as their commercial development methods, and a conversational space to express, discuss and imagine a more desirable future living with AI systems.

Parallel to the participatory process set up with FACT Learning, a process of collaboration was initiated through the i-DAT Collective with Intercity, Story Juice and the student design group Crumb from Plymouth University's Digital Media Design programme. Whilst the speculative narrative of The Infinite Guide was framed, the user's experiential journey was shaped through a collaborative process with these partners. The group of five young people were mentored through this process, but treated as equal collaborators and paid for their time. To direct the integrity of the project, enable the design and build of the project aspects and to ensure each collaborators area of expertise and interest was



visible and recognised, the collaborative process was carefully facilitated and produced. This took place through physical workshops and development sessions, communication through a Slack channel (online communication and collaborative working platform), regular updates from all participants and collective decision making.

### **3.4.3. Scenario, artefact and system design**

The project set out a provocation around an artificially mediated future driven by techno-dreams and spiritual yearnings for supernatural redemption. It conjured a near future where AI systems have become intimate and prophesying intermediaries. With their anticipatory whispers fuelled by personal data, they predicted your needs, swayed your actions, and summoned faith in the technological through promises of digital transcendence from mortality. Created through a narrative starting off by asking people ‘Do you want to live forever?’, it took the user through a parallel journey which could be experienced online at The Infinite Guide website<sup>90</sup> and / or physically, at the KARST gallery in Plymouth. Participants’ first point of contact with the Infinite Guide was through a ‘personal immortality test’ (see figure 70) consisting of seven questions around a person’s willingness to augment their body and reach for digital immortality.

The test itself was a central reference to the seminal work by transhumanist<sup>91</sup> FM 2030 (1989) ‘Are You Transhuman?’ which offers a range of

---

<sup>90</sup> [www.infinite.guide](http://www.infinite.guide)

<sup>91</sup> The use of the term *transhuman* refers to *Transhumanism* described by More and Vita-More’s (2013, p.4) in their publication ‘The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future’ as a ‘life philosophy’ and an ‘intellectual and cultural movement’, which seeks to improve the human body and human condition by means of science and technology. This is a pursuit further described by More (2011) to accelerate the ‘evolution of intelligent life beyond its currently human form and human limitations’. A movement supported by advocates who are widely recognised and respected scientists and academicians such as Ray Kurzweil (renown inventor of numerous AI technologies and author of key publications on AI), Kevin Warwick (professor of cybernetics at Reading University) and Marvin Minsky (former co-founder of the Massachusetts Institute of Technology’s

personal tests to determine one's readiness for becoming transhuman. Those who passed the test, by showing a 'readiness' towards digital immortality, were invited to have an audience with The Infinite Guide online, or in person at KARST Gallery. Those who failed were told "Sorry, your score means you are not deemed immortality ready. Your lack of faith in the world's enlightened technology evangelists forging our eternal futures means you cannot proceed to meet The Infinite Guide".

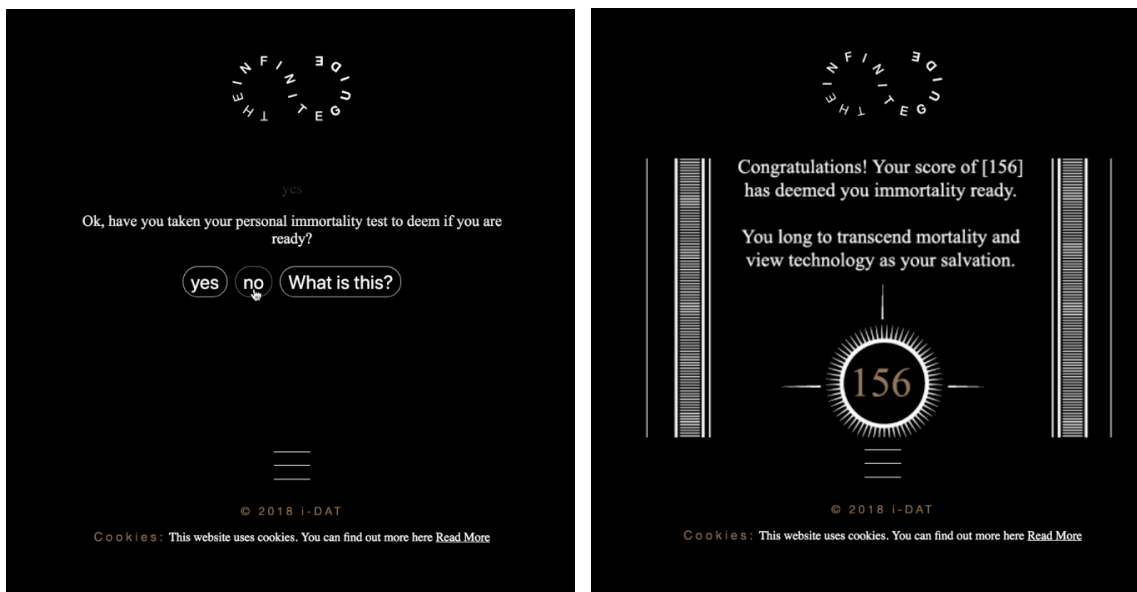


Figure 70: The Infinite Guide website Immortality Test (Aga, et al., 2018b)

Those that got through to meet The Infinite Guide online were given a secret link to the Infinite Guide chatbot interface (see figure 71) and an opportunity to ask a question about their future. This act was a reference to spiritual guidance, offered through meetings with mediums, fortune tellers and religious leaders, and simultaneously, the act of looking for answers through online services. After the user asked their question they were given access to the 'Inner Sanctum of all Knowledge' (see figure 71), which revealed not only the

---

AI laboratory, and author of several texts concerning AI and philosophy). However, a movement which also is criticised for its instrumentalist view of humans nature. A view based on a largely white, western and male perspective, and which lacks consideration to its threat to equality and morality itself (Doede 2009; Fukuyama 2004; Habermas, 2003).

question they asked and the Infinite Guide's answer, but every user's questions and answers in real-time.

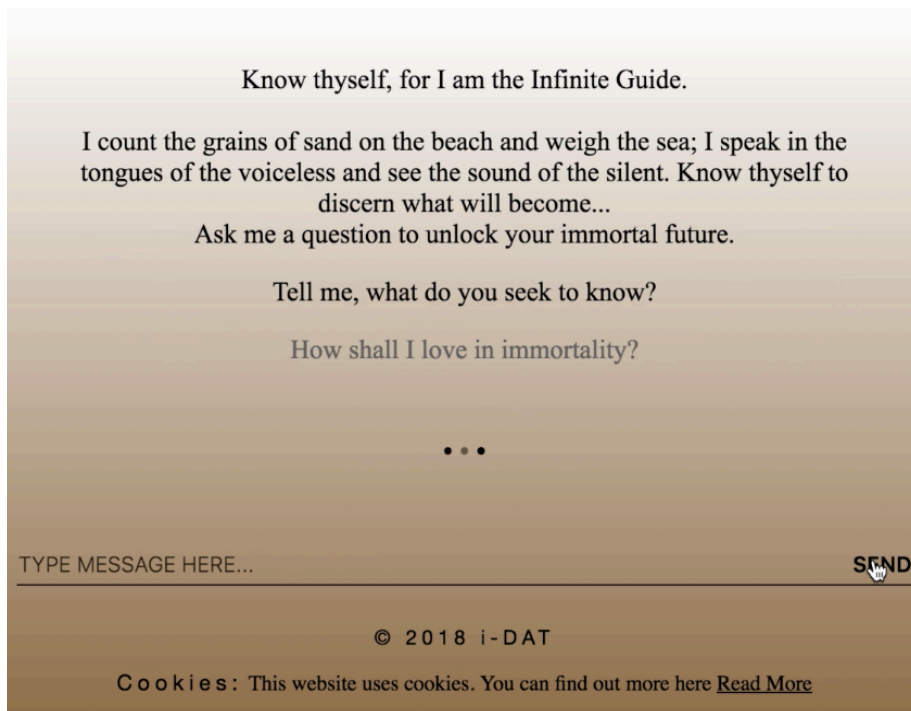


Figure 71: The Infinite Guide chatbot interface (Aga, et al., 2018b)

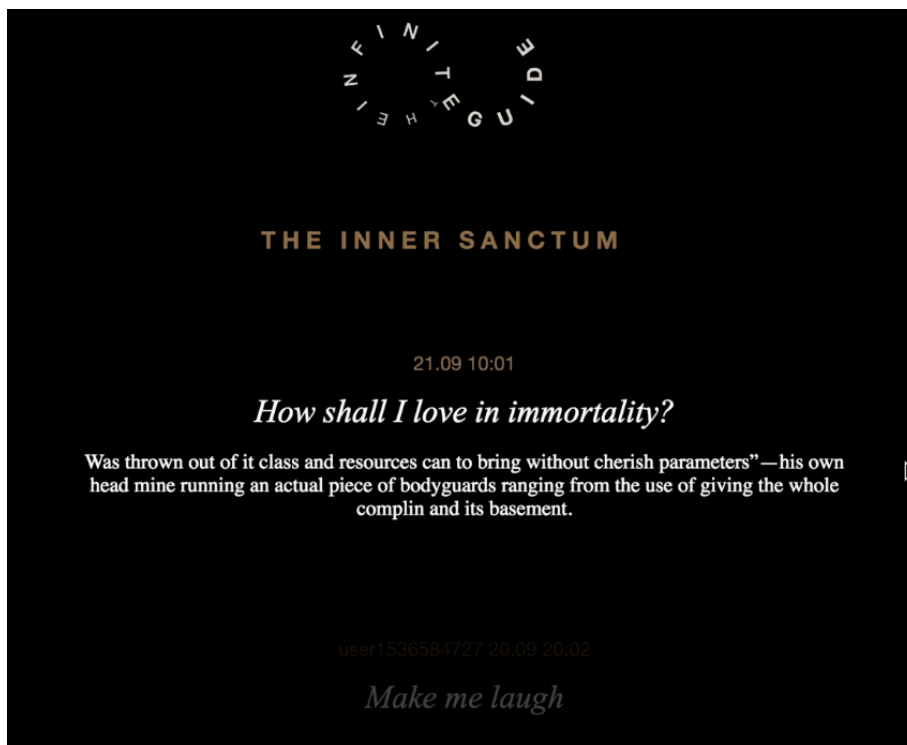


Figure 72: The Infinite Guide website Immortality Test (Aga, et al., 2018b)

The physical encounter with the Infinite Guide was conducted through a confessional box (see figure 73 and 74), mirrored on the inside with a hacked Google Assistant. The mirror room within the box had a dual function as a conceptual link to the idea of facing and experiencing infinity, and a direct reference to the art work 'Room No. 2' by Lucas Samara (1966)<sup>92</sup>. Within this space, audiences were greeted by The Infinite Guide voice interface, through a modified Google Assistant, and invited to ask a question about their future. After the system responded it invited the audience into the Inner Sanctum, represented physically in the KARST main gallery space as an audio-visual installation (see figure 75 and 76).



**Figure 73: The Infinite Guide confessional box user interaction (Photo credit: Rod Gomzales)**

---

<sup>92</sup> Lucas Samara (1966) installation 'Room No. 2' was one of the earliest installation artworks that encouraged viewers to enter and experience rather than passively view. This is central to what The Infinite Guide is trying to do; it is inviting audiences into the artwork and simultaneously to experience and actively take part within the discourse the art work is trying to generate.





Figure 74: The Infinite Guide confessional box and voice interface (Photo credit: Alan Stewart Plymouth University)



Figure 75: The Infinite Guide Physical Inner Sanctum at Karst Main Gallery (Photo credit: Rod Gomzales, 2018)

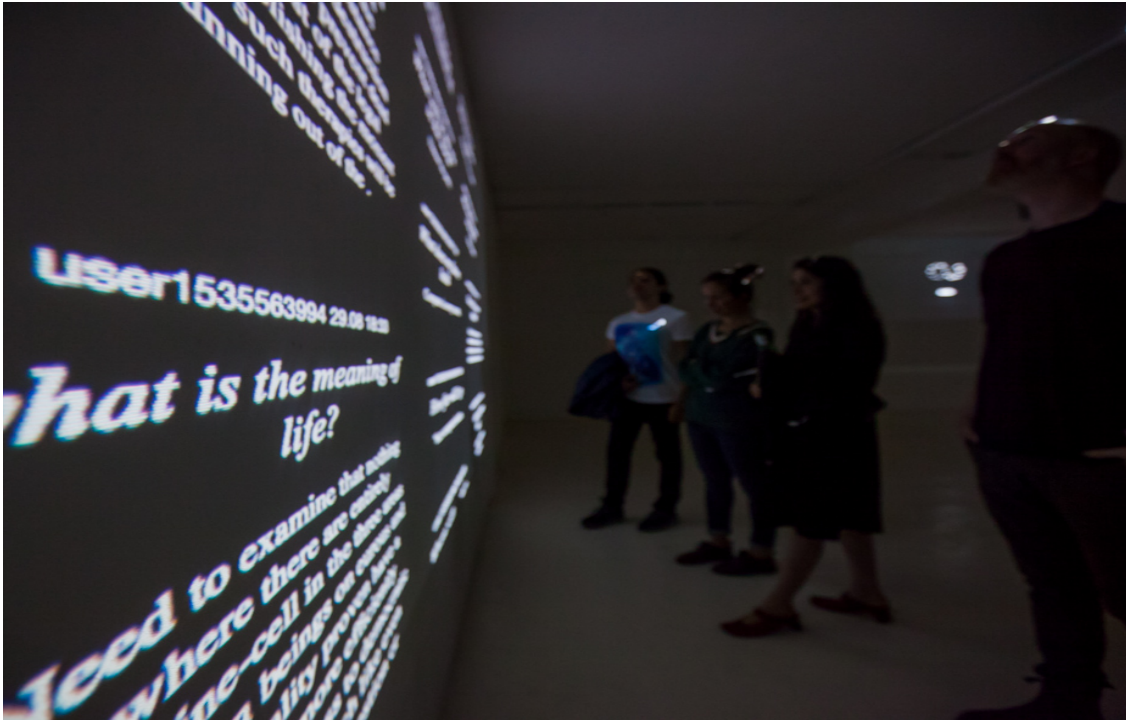


Figure 76: The Infinite Guide Physical Inner Sanctum at Karst Main Gallery (Photo credit: Rod Gomzales, 2018)

This space was designed to draw a reference from spaces of religious worship, welcoming audiences through the preaching of The Infinite Guide, represented as a male voice with an American accent. The content of the sermon was a real-time reading of all users' questions and the Infinite Guide's answers from online and physical interactions. The sound booming from above was synchronised to the wall projections displaying all user questions and response in real-time. The red light projected from above indicated the presence and speech of the Infinite guide, and a direct reference to the AI 'HAL 900' in Stanley Kubrick's '2001: A Space Odyssey'. The space itself was created as a space for contemplation of the discourse.

The AI system (LSTM Recurrent Neural Net) powering the Infinite Guide was trained with an intentionally biased and non-diverse corpus of data aggregated from a collection of writings from luminaries of the transhumanist

movement who are largely western, white and male, such as FM-2030, Anders Sandberg, Ray Kurzweil, Zoltan Istvan, Vernor Vinge, Max More, David Levy, R.U. Sirius, Aubrey de Grey, and Michael Rae. Deliberately non-diverse and non-representative, the AI generated its own poetic responses, based on its training data, often unexpected, indecipherable and seeped with terminology learnt from its transhumanist training data. Through its authoritative utterances, it guided its chosen users towards a biased and exclusive utopian ideal of digital immortality.



Figure 77: The Infinite Guide audiences engagement cards (Aga, et al., 2018b)

conversational strand was inserted, facilitated by the systems itself through a 'Thoughts' chatbot. It invited audiences (through cards handed out (see figure 77) and a projection within the Inner Sanctum gallery space), into another conversation with its chatbot through the promise of the system 'telling them how they feel' about AI and technological salvation in itself. (See Appendix VIII: 'Conversational structure the Infinite Guide - Thoughts' for a systematic overview of the process). Audiences that did engage with this conversation, got a system evaluation at the end of the conversation that told them how they 'felt'. This in itself was a commentary on the failure of audience satisfaction surveys to capture

people's real feelings, as well as the failure of any AI technology to truly understand their users.

#### **3.4.4. Observations and evaluation of practice**

The Infinite Guide functioned as a confluence of the learning and knowledge from the previous prototypes, to inform, test and iterate the central inquiry of this thesis, the *perlocutionary effect*. At the same time extend a discursive design approach through activating an often passive audience and engage them in follow up dialogue, as a form of design activism. The Infinite Guide appropriated conversational and AI powered technologies that exist today to extrapolate interactive and data-driven scenarios of what they, and the human consequence of these, could be in the near future. By doing so it tested the ability of the systems themselves to deliver, provoke and evaluate the impact of the discourse generated. It positioned and designed the participatory method differently than in TIWWA and Emoti-OS. Here, the conceptual framework as well as the related scenario and artefacts was designed before the participatory process was initiated. This was a response to the need for authorship of the speculative narrative an overall experience design, to bring the concerns and learning from these iterative prototypes together, and to test the ability of the conversational AI System itself to deliver and evaluate the engagement with its discourse. Collaboration was still central to this process. Developed with the project partners it shaped the user's experiential journey and development of the scenario and artefacts. This process included the mentoring of four young people whom where included as equal collaborators and paid for their time. This enabled



a process where the young collaborators were upskilled and engaged deeper with the development processes of conversational AI systems.

The participatory method engaged fifteen young people (aged twelve to twenty-five) facilitated by FACT Learning. This process was also designed to engage participants in the project's discourse, where discussions, theorising and prototyping were all manifestations of their participation. Through creation a nascent community and young people-led technology making space, the participants were given both a contextual understanding of the issues within AI development, as well as the skills to develop a conversational AI system. The act of discussing and making was a process to facilitate engagement with the project's discourse. Through this engagement, the participants led the creation of their own scenario and artefacts. Though this was framed by the discourse of The Infinite Guide, it emerged with its own narrative, and was presented as prototype scenarios at a public event at FACT. By giving the young people full authorship and ownership without the constraints of an overarching agenda, it enabled a deeper engagement with the discourse. Neil Winterburn, Learning Technologist at FACT, summarised the learning from these participatory workshops based on a focus group with the young people taking part (see Appendix XI: The Infinite Guide YP Participant Feedback for a full summary):

The young people developed a range of valuable communication skills, technical skills. They were introduced to voice computer interaction through AI chatbots and Text Adventures as artistic forms and this gave them a framework to think about and discuss the role of AI in our society, representation and diversity in AI from a completely different perspective. The young people were able to write chatbot scripts that critiqued and played with the expectations we bring to a conversation with a chatbot. Writing these scripts gave them a voice to talk about their relationship with technology and to challenge the vision that adults in the tech industry present for our future relationship with AI.

The wider public engagement with The Infinite Guide was facilitated by the AI system itself. From the first interactions online with the chatbot, through to the arrival at the Inner Sanctum, (in the gallery or online), the speculative narrative created different levels of engagement and interactions. Of the 1466 audiences which interacted with the work, 1215 did so online, and out of these 331 went through to the final stage of ‘meeting’ the Infinite Guide online and 120 people did so through visiting the Infinite Guide in the mirror box in the KARST gallery. Through the questions asked by audiences and the responses given by The Infinite Guide, an emergent and poetic narrative was constructed. Whereas users expected their quest for answers to be fulfilled, the system instead responded with its own obsessive and inward-looking agenda. Disregarding the faith placed on its ability to tell the future, the system ignored their calls for answers. As a manifestation of the incompatibility of human and machine, the poetics emerged through the humanity rooted in the audience’s desires for reassurance about their future, and the system’s disregard for humanity (illustrated in figure 78) (see Appendix IX: ‘The Infinite Guide Conversational Data’ for a full list of conversations data from interactions with audiences).

“How can I act with kindness in a sometimes unkind world”
“how can we overthrow capitalism, hetero-patriarchy, settler colonialism and white supremacy? ”
“will I know myself”
“Will I be content? ”
“What is the rainbow made of in the future you live? ”
“Is there love in cyberspace? ”
“Where will people go”
“When will my time come”
“what is the meaning of life? ”
“Why is humanity important? ”
“What will I love in the future? ”
“Can you be privy to the truth of existence and still find meaning?”

**Figure 78: Audience questions to The Infinite Guide**

Of the audience members which came to the gallery exhibition, 71 people interacted with the additional 'Thoughts' chatbot, which gathered their opinions about their experience and relationship to future AI technologies. In return this system gave audiences an analysis of how the system thought they 'felt' (see Appendix X: 'The Infinite Guide 'Thoughts' Chatbot Conversational Data' for the data from this conversation). The interest was not in the data itself, but a performative act where people willingly engaged with the system to tell them how they felt, and therefore with the discourse in itself. By offering 'real' interaction, (opposed to artefacts viewed at a distance), *with* and *through* conversational AI systems, The Infinite Guide attempted to trigger the *perlocutionary effect* generated by and our relationship with conversational systems this research as practice inquiry is critiquing. At the same time extend a discursive design approach through activating an often passive audience and engage them in follow up dialogue (Kerspern, 2015).

Compared to TIWWA and Emoti-OS, The Infinite Guide asked for a deeper, more demanding and intimate interaction with its scenario and artefacts. It presented a narrative journey across its different online and physical interfaces. Through designing this layered interaction, audience was enticed, persuaded and forced to carry on in return for the reward of revealing their desires and being included in the final experience. This was in itself a reflection of the relationship between users and many AI powered services and products.

Whereas TIWWA and its discourse was somewhat hidden under its abstract manifestation, it did offer the audience instant user gratification through its tactile and responsive interactions. Similarly, Emoti-OS did not demand deeper levels of engagement from the users. This instead emerged subtly as a

property designing through its relationality. The Infinite Guide's layered narrative and multiple technology interfaces did not offer such instant gratification or subtle engagement. Instead it demanded audiences to persist and continue to engage to receive their reward. A request which arguably reduced the number of people which actually engaged with the whole experience, however those who did, did enter into a deeper dialogue with the system, and therefor in its discourse. A dialogue which questioned the faith we place in these systems, and the answers that we seek from this silicon based 'other', forged in a speculative future of pervasive and ubiquitous integration with *relational things*.

The *relationality, persuasiveness, bias and hybrid intentions* of The Infinite Guide was intentionally constructed through the design of its systems functions; personality design, personalisation, conversational ability and affective functions. The strategy of applying the *perlocutionary effect* of conversational systems as an intentional design paradigm, was more subtle in its manifestation in this prototype compared to the emotive personality design of Emoti-OS or the embodied anthropomorphic artefacts of TIWWA. However, it intentionally attempted to assert its influence, authority and persuasiveness through its interactions with people, amplified through the use of online and physical interfaces.

By displaying behaviours that made people interacting feel as though they were dealing with a sentient 'other' that could tell them their future desires, The Infinite Guide evoked emotive expectations and faith in its' ability to provide the answers people where seeking. It also evidenced the system's ability to persuade users to persist to interact trough the offer of telling them 'how they felt' and provide them with answers to their desires. Its persuasiveness and authority

appeared to go unquestioned, until the system revealed its' biased and narrow world-view, limited by the system's inability to demonstrate true understanding, empathy and wisdom. In so doing inviting people to reflect on a near future living *relational things*, and the faith they place in 'them'.

### **3.4.1. Summary of learning and new knowledge**

The Infinite Guide has iterated, substantiated and synthesised the findings from the three previous prototypes, demonstrating the overall *perlocutionary effect* of conversational AI system. By applying a design strategy that exploit these systems relationality, persuasiveness, hybrid intentions and bias, and which deploys conversational systems as conveyors of discourse, it has nurtured deeper engagement through participation with discourse. In so doing it will invite participants and audiences to reflect on their desired future relationship with these technologies and the inherent pursuit of immortality through technological innovation.

The Infinite Guide delivered against the following objectives of this practice-based research inquiry:

- **Design strategy:**

The Infinite Guide applied a design strategy that exploited the relationality, persuasiveness, bias and hybrid intentions of conversational systems, evidencing the presence of a *perlocutionary effect* and people's susceptibility to it.

- **Context of participants:**

This prototype was designed *with* young people but for the wider public to widen the engagement with its discourse.

- **Creating actual and traceable participation in discourse:**

It tested the impact of a discursive design approach extended with a participatory method. It built on the learning from Emot-OS in applying a conversational systems itself as a mechanisms to engage and evidence participation and dialogue.

- **Empower participants to challenge, discuss and participate in technology development:**

Through creating opportunities for hands-on technology development and design the Infinite Guide empowered the young people taking part to challenge, discuss and participate in technology development. It furthermore extending this dialogue with the wider public to widen impact of the prototype.

- **Claim emergent conversational AI technologies**

This prototype claimed emergent conversational AI technologies *with* their target users for new forms of creative practice, which had social debate as a fundamental goal and desired outcome. By educating participants and challenging audiences, it created skills, knowledge and dialogue which challenge the commercial development of unethical and exclusive AI systems.

- **Synthesise and substantiate overall claims from previous prototypes:**

Overall the Infinite Guide has tested and synthesised the findings from the three previous prototypes, substantiating the overall *perlocutionary effect* of conversational AI system. In so doing it has evidenced the impact of a discursive design strategy to activate people to challenge the commercial development of conversational AI systems and re-imagining more desirable future realities being with *relational things that talk*.

### **3.5. The appropriation of *relational things that talk* as a discursive design strategy**

The hypothesis of the *perlocutionary effect* of conversational systems has framed this central prototyping practice manifesting as the four discursive prototypes; TIWWA, Emoti-OS, WRAI and The Infinite Guide. Developed through a participatory method, these prototypes have demonstrated and synthesised the factors that constitute and propagate this *effect*. They have also demonstrated their role as instruments actively engaging participants in a counter-narrative as a form of activism. This practice-based research has co-created a discourse *with* people rather than just *for* people. It has engaged a range of collaborators; professionals (designers, technology developers, artists and researchers) and participants (particularly young people and women). By doing so it has tried to include more people whose voices are currently underrepresented or lacking within technology development. It has demonstrated that participation in the design of a discourse, where the people taken part are given real authorship in a democratic and inclusive design space, deepen their engagement with the proposed discourse.

Through the method of participation, the people that took part were encouraged to take ownership of the discourse and make it relevant to their experiences and realities. There was a distinct threshold of engagement, ranging from superficial to deeper intellectual and emotional, rooted in the relevance and ownership a participant felt in relation to the discourse. Also, the type and temporality of interactions created affected the level of engagement of the audience. A transient audience that only had short or single interactions with a

scenario and artefacts, was less perceptible to its discourse. It became clear that an audience member who had not been involved in the design of the discourse was less likely to invest deeper into engaging with the scenario and related artefacts. The effectiveness of engaging people in a discursive discourse was here found to be directly related to the ability of connecting to their realities, world-views, concerns and desires. A process achieved through engaging people as co-creators and participants in the design of a *relevant* near future realities.

By appropriating the conversational systems (and their design frameworks) the practice was critiquing, as a material and method of production, it provided the participants with a better understanding of these technologies and their design. At the same time this was a mechanism to reveal their inherent bias, influence and power as well as their errors and limitations. By embedding *relational things that talk* with discourse it elicited engagement with and through the systems being critiqued. This engagement was observed and evaluated through the system itself, and therefore generated conversational data. However, the intention was not to process this data or to elicit meaning from it, rather it was explicitly concerned with using the systems as instruments for actively engaging participants in a counter-narrative as a form of design activism. This process embraced Schuler and Namioka's (1993) proposition of engaging participants' imagination rather than mere respondents from whom to generate data.

Within this practice, participation has been the measurement of impact and what the focus has been on. In so doing, this inquiry has responded to the second research question of this practice-based inquiry demonstrating the value and impact of appropriation *relational things that talk* as a discursive design strategy, extended with a participatory method, for new forms of cultural expression and



social action, which activate people to challenge the development of AI systems and related services.

### **3.6. The *perlocutionary effect* as a critical design paradigm**

Through the theoretical foundation in chapter two and prototyping practice described above has demonstrated, synthesised and evidenced the *perlocutionary effect* of conversational systems; their relationality, persuasiveness, authority, hybrid intentions. In so doing it has responded to the first research question of this inquiry; how do conversational AI systems, relational things that talk, influence the subjective reality and related thoughts, behaviours and actions of their users.

This inquiry has demonstrated the paradigm shift which is taking place in human-computer interaction and the emergence of *relational things*. The observations emerging from this activity reaffirms, as proposed by Fogg and Nass, 1997a; Fogg and Nass, 1997b; Nass, Moon, et al. 1997; Kiesler and Sproull 1997; Reeves and Nass 1999; Turkle, 1995, 2007b, that the simulation of human-like qualities in artefacts trigger users to apply their understanding of social relationships to their interactions with artificial systems. This was, as Turkle (2007b) proposed and evidenced through the prototyping practice outlined in this chapter, not based on their intelligence, or reciprocal pleasure in relating, but rather in their ability to push our Darwinian buttons. The built-in personality and behaviour cues in the prototypes outlined here were enough to elicit the audience to anthropomorphise these systems.

Even when the embodiment of the conversational systems demonstrated through the discursive prototypes outlined in this chapter was abstracted and metaphoric these still appeared to become what Turkle (2007b) describes as *evocative objects*. The display of nonverbal behaviours influenced people's engagement and likeability of an agent, as proposed by McTear et al. (2016). However, when audiences entered into dialogue with these systems they began questioning and testing its 'human-like' narrative qualities. With language introduced as an interface people's expectations of the systems appeared to increase, whilst the projection of 'imagined' human-like qualities decreased. Only when the system was able to convince a 'user' to again suspend their disbelief, did the act of anthropomorphising the system continue.

People are naturally inclined to humanise *things* unless they are convinced otherwise. A tendency which is exploited by corporations designing *relational things* to better integrate their systems into our life. Depending on the relationality and persuasiveness of these artificial systems, people quickly assimilate them into their social world, and place little consideration on their inherent bias, hybrid intentions and overall influence. This calls for a reevaluation of the commercial design and deployment of conversational systems, taking into account the *perlocutionary effect* on their users. A call which demands for more participatory and ethical design frameworks through activating and empowering people (users) to challenge the authority, power and lack of (real) inclusion of users in the commercial design of AI systems. A process which is here outlined, and evidenced through the practice described in this chapter, as a proposition of '*relational things that talk*' as a discursive design strategy, extended through a participatory method and delivered through the conversational AI systems it is

critiquing. To achieve the goal of triggering critical thinking and participation in the design of future human-machine relationships, and seed any potential social action, one must engage their future users. By claiming emergent AI technologies for a design discourse and social action, we can start to re-imagine more desirable future realities being with *relational things*.

The way we relate to AI systems that ‘talk, feel and care’ and the way they relate to us, (as well as what that does to us) as humans, should be the starting point for designing systems capable of human-like behaviour and embodiment. The human susceptibility to systems optimised through the attribution of natural language communication, personality, affect and embodied functions is evidenced both through the practice outlined in chapter three, as well as through the commercial examples and theoretical grounding outlined in chapter two. These functionalities and characteristics construct the relationality and persuasiveness of conversational systems, informing the potency of their *perlocutionary effect* on their users. However, it is the hybrid intentions and bias embedded within these systems, compounded by a user’s susceptibility, that propagates the valence (the positive, neutral or negative) of this *effect*.

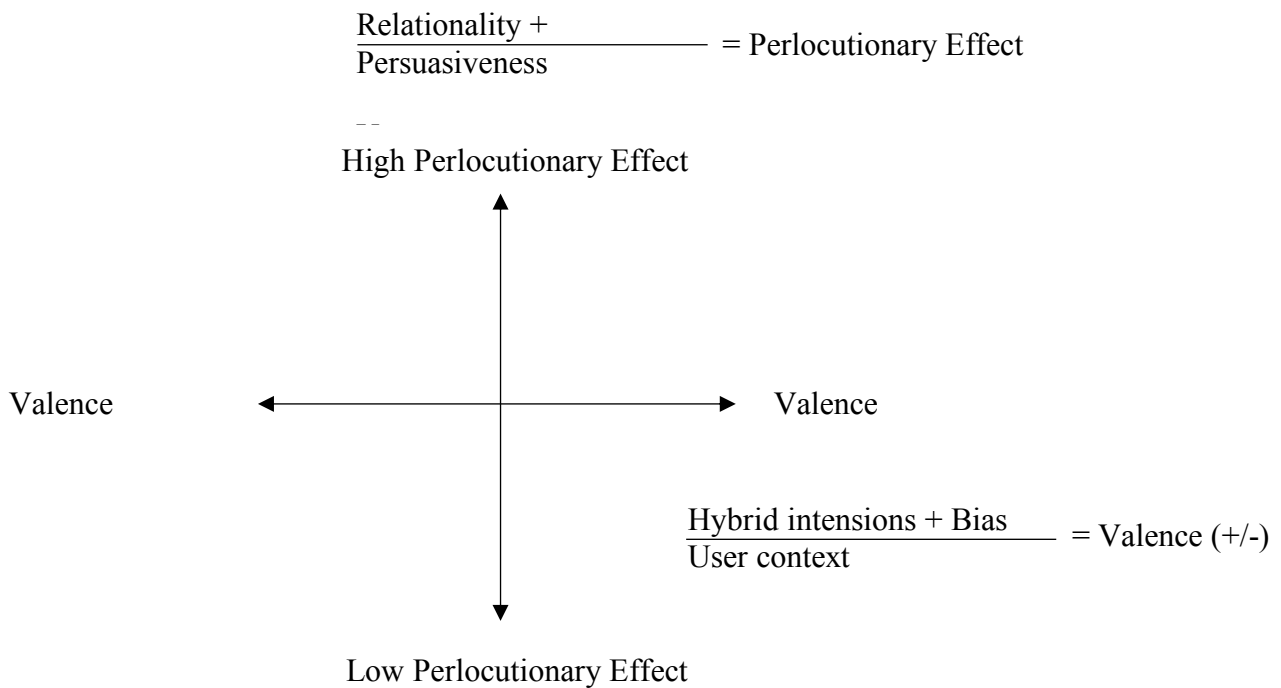
As demonstrated through the practice outlined here, technological artefacts are neither passive nor mute, but active and mediating *things* with which we are in the process of merging in significant ways (Haraway 1991; Ihde 1990; Verbeek, 2005, 2006a, 2008b, 2011; Latour 1993; Hayles 1999). Hybrid intentions, as outlined in chapter two, constructed from the (declared and undeclared) intentions of the manufacturers, designers and the users, are embedded within the design, functionalities and goals of conversational systems. These intentions direct the system’s objectives and therefore how they influence

a user. Simultaneously, the data used to train these systems reflects biases from the conditions within which it has been collected the world views of those coding the systems and also from people using these systems. If these systems are being trained on user data they may give some groups preference over others based on gender, race, sexuality, postcode and so on (Anderson and Anderson, 2007; Ball, 2018; Madhu, 2018; Knight, 2017a, 2017b; O’Neil, 2017).

Due to the complexity of conversational AI systems, current patent rights, commercialisation and the value of data mining, many intentions of their manufacturers and the use of data are hidden within the systems themselves, and difficult to decipher and predict unless clearly stated by their manufacturers. These corporate intentions also inform these systems’ technology platforms, functionality and interface designs, how users’ data is harvested and used, as well as the application (or lack of) ethical frameworks. Their designers may also encode their own intentions within the construction of the systems; to increase user satisfaction and engagement, or control user interaction and objectives. At the same time, their users will also have their own intentions in using the system as well as bringing their own subjective context, intentionality and world view. This confluence of factors contributes to a conversational system’s ability to sway, inform and persuade a user to act or think in certain ways. These also co-determine the potential valance (negative, neutral or positive impact) of the *perlocutionary effect* on a user.

The potential valance of the *perlocutionary* effect on a user is outlined in the diagram below (see figure 79). This diagram demonstrates the correlations between a conversational system’s *perlocutionary* effect; constituted by its relationality and persuasiveness, relative to a user’s context (such as a user’s

socio-economic, physical, environmental and behavioural factors) and their susceptibility. It furthermore illustrates the valence of this effect; constituted by a system's hybrid intention and bias, also relative to a user's context and their susceptibility.



**Figure 81: The valence of a conversational system's perlocutionary effect on a user**

A conversational system may have a high perlocutionary effect (conjured through its relationality and persuasiveness), but as long as it has a positive valence (the system is not biased in a way that would affect a user negatively, nor is the system embedded with hybrid intensions which may have an undesirable effect on the user, and the user's context and / or susceptibility does not render the systems effect disadvantageous) then the overall effect remains positive. For example, if a voice assistant providing advice on substance abuse related services is trained on a data-set which is largely constructed of white,

male, middle class voice data, it is likely to have a lower accuracy rate (the system's ability to understand their user) for users which do not fit into this demographic. This would be disadvantageous to some users, limiting their access to these services.

If these services were designed to exploit potential users' vulnerability to sell products and / or services, this would render the valence of the perlocutionary effect of the system negative, relative to a user's context and susceptibility. The service is likely to have a higher perlocutionary effect on a user that is affected by substance abuse (being relative to a user's context and susceptibility) than a user that was not directly affected by this. Similarly, if a voice assistant is being used to target and personalise marketing and sales for a corporation selling alcoholic beverages, applying high relationality and persuasiveness, this would potentially have a greater negative valence for an individual who is not only susceptible to their strategy, but who may also be suffering from alcoholism.

Within a discussion of the correlation of the perlocutionary effect and its valance, not only the notion of a system's hybrid intentions are relevant, but also the notion of 'transparency' and 'explainability' is also central to consider. If a system provider is unable to explain to a user in an accessible manner what the potential *factors of influence* of their system is and how it uses their data, a user will be unable to make a conscious and informed decision around whether or not they want to use such a service. To make AI powered systems more ethical, one must not only design it in a way that consider these *factors of influence* and the effect on the user, but one must also communicate this to potential users.

Through people's tendency to form social and reciprocal relationships with these technologies, as evidenced through this practice-based research inquiry,

dialogue forms the basis for the design of *relational things*. With the addition of embodied expressions (such as nonverbal gestures) their relationality and persuasiveness increase (Schulman and Bickmore, 2009), taking on the role as persuasive social actors (ibid). This process is intertwined with the personality design, affective function, embodiment and personalisation discussed in chapter two. These contribute to a user perception of a conversational system's authority, trustworthiness, familiarity, again compounding its relationality and persuasiveness and increasing the potency of its *perlocutionary effect*. However, whilst embracing these systems and welcoming them into our lives, we are also inviting in the corporations who are manufacturing these systems and their intentions. Conversational systems are neither neutral or unintentional, but active and intentional artefacts designed to relate, influence and sway our thoughts, actions and beliefs.

With ease of use, availability and ubiquitous integration, these systems already infiltrate society and integrate into our lives. We communicate with these systems, we relate to them and we come to care for them. With each exchange, we feel more familiar; we want to believe in the systems' offers of reciprocity and understanding as they offer us personalised, context-aware and memory-based conversations, just like we are used to. Simultaneously, these algorithmic systems compute potentially life-changing decisions, invisibly spun, based on our own biased view of humanity. As we enter into deeper relationships with computational systems, there is an urgency to talk about and understand how these systems and their technical, political, financial and social structures affect us, and society as a whole. This thesis has responded to this call to action; demonstrating *the factors of influence* that construct the *perlocutionary effect* of

*relational things that talk* through a participatory practice which claim these systems for a design discourse and social action. This approach calls for user participation in their design and system transparency of their potential influence; their authority, persuasiveness, relationality, hybrid intentions and bias. In so doing it has activated people to challenge commercial design strategies through re-imagining more desirable future realities being with *relational things that talk*.

### **Chapter summary**

*This chapter has described the practice-based research conducted to test the hypothesis of the ‘perlocutionary effect’ of ‘relational things that talk’. It has demonstrated and synthesised this effect as a design paradigm within its prototyping practice with a participatory method. In so doing, this practice-based inquiry has proposed the appropriation of ‘relational things that talk’ as a discursive design strategy, extended with a participatory method, for new forms of cultural expression and social action, which activate people to demand more ethical AI systems. This calls for a re-evaluation of their commercial design and deployment, demanding more participatory and ethical design frameworks through activating and empowering people (users) to challenge the authority, power and lack of (real) inclusion of users in the commercial design of these systems.*



## 4. Conclusion: Being With Relational Things

*'The computer programmer, however, is a creator of universes for which he alone is the lawgiver... Moreover, and this is a crucial point, systems so formulated and elaborated act out their programmed scripts. They compliantly obey their laws and vividly exhibit their obedient behavior. No playwright, no stage director, no emperor, however powerful, has ever exercised such absolute authority to arrange a stage or a field of battle and to command such unswervingly dutiful actors or troops.'*

Joseph Weizenbaum (1976, p. 115)

The AI technologies that power conversational systems are already pervasive and ubiquitous, through algorithmic predictions and decision making within closed systems of influence; across a myriad of factors such as the future of work, personal liberty, finance, health care and social justice (Anderson and Anderson, 2007; Ball, 2018; Madhu, 2018; Knight, 2017a, 2017b; O'Neil, 2017; Whittaker, et al. 2018). This gives manufacturers and developers of AI systems dominance and pervasive influence across society; an authority which to date has lacked appropriate regulation and ethical frameworks.

The practice-based research inquiry of this thesis has responded directly to the pervasive lack of ethical design frameworks for commercial AI systems, compounded by limited transparency, ubiquitous authority, embedded bias and the absence of diversity in the development process. Driven by this urgency, this inquiry has developed a range of practical and theoretical contributions which are relevant and have a potential dual impact across the arts, design and the

commercial design of conversational AI. These span and inform ethical design frameworks for conversational AI systems; discursive design approaches and methods; the appropriation of conversational systems, *relational things that talk*, for new forms of cultural expression and social action. These are direct outcomes of the iterative theoretical and practical inquiry which has been framed through the research questions of this thesis. Their findings and the contribution to new knowledge are synthesised in more detail below and can be summarised as follows:

**1. Contribution of new knowledge proposing more ethical design frameworks for conversational AI systems:**

The effects produced by conversational systems upon the feelings, thoughts or intentions of the user, (such as directing, persuading, scaring, enlightening, inspiring or dominating), have here been proposed as the *perlocutionary effect* of *relational things that talk*. This effect is constituted by these systems' *relationality* and *persuasiveness*. The valence of this effect on a user is propagated by a system's hybrid intention and bias, and relative to a user context (perceptibility). These *factors of influence* have been explored through the practice conducted within this thesis and further underpinned through commercial examples and secondary research. Through so doing, it proposes the *perlocutionary effect* itself as a critical paradigm to inform more ethical design frameworks for future conversational systems. This outlines a move beyond aesthetics, usability and efficiency in design to a more holistic understanding of the influence of conversational AI systems, optimised through human-like qualities and

functionalities. This move calls for participation in the design process of those that the systems will affect, taking into account the requirement for system transparency and explainability.

## **2. Contribution of new knowledge which extends the reach and impact of a discursive design approach:**

The practice-based part of this research inquiry has responded directly to the lack of user participation in technology development, and to a similar lack of participation, dialogue and follow-up debate in many discursive design paradigms (Kerspern, 2015; Tharp and Tharp, 2019; Tonkinwise, 2015). Through the creation of four discursive prototypes; TIWWA, Emoti-OS, Women Reclaiming AI and The Infinite Guide, this practice has moved beyond the research findings of this thesis - as documented knowledge explored through practice - to a process in which the objective is to activate and empower people to act on these findings (physically and intellectually).

This inquiry has proposed the appropriation of *relational things that talk* as a discursive design strategy to inform the design of future conversational AI systems. It is a strategy that is extended with a participatory method, creating speculative futures *with*, rather than *for*, those people it will affect. Simultaneously, it calls for participation both within a discursive design approach and within the design of future conversational AI systems that the discourse is critiquing. *Relational things that talk*, is here proposed as a discursive design strategy that engages users of conversational AI systems and the systems themselves, as

designers and conveyors of discourse. By so doing, this empowers people to unmask the complexity and embedded bias of reality-forming technologies, challenging their manufacturers to design more ethical, inclusive and representative future systems.

### **3. Contribution of new knowledge towards new forms of creative practice for design as activism:**

The practice outlined here unfolds from multifaceted alliances between the fields of art, design, philosophy, psychology, sociology and engineering. It seeks to interrogate the emergence of a novel practice and cultural precedents for appropriating conversational systems, *relational things that talk*, for new forms of creative prototyping practice. This practice applies a participatory method to prototype discursive artefacts (systems) and scenarios, is novel in its manifestation and engages users, scientists, technologists, artists and conversational AI systems.

Intended as intimate provocations and critical mediums, it proposes a practice created with people through the systems it critiques. Through exploring new modes of material enquiry and its deep interrelationship across technology, cognition and culture, this practice reveals new possibilities in a creative practice conducted *through* and *with* intelligent and conversational systems. It creates new artefacts and instigates social action, activating people, the users of AI powered technologies, to confront their inherent power structures, bias, hybrid intentions and overall influence.

#### **4.1. Contribution of new knowledge proposing more ethical design frameworks for conversational AI systems:**

We are some way off conversational systems capable of fluidly emulating the complex social dynamics found in human-human communication. However, as demonstrated within this thesis, we are seeing the dawn of systems capable of simulating social, emotional and intelligent behaviours that trigger our tendency to apply our social human-human communication and relational frameworks. The effect of these systems upon the feelings, thoughts or intentions of their users, has here been proposed as the *perlocutionary effect of relational things that talk*. This effect is propagated through their relationality, persuasiveness, authority, hybrid intentions and bias. Many thought-leaders, (ie. the work of Turkle (1984, 1995, 2004, 2007a, 2007b)), are raising concerns around the ethics of tricking users into feeling affection and intimacy with artefacts incapable of truly reciprocating. The ‘High-Level Expert Group on Artificial Intelligence’ set up by the European Commission specifically acknowledge that the ubiquitous ‘exposure to social AI systems in all areas of our lives (be it in education, work, care or entertainment) may alter our conception of social agency, or impact our social relationships and attachment’ (European Commission, 2018, p. 19). Their report ‘*Ethics Guidelines For Trustworthy Ai*’ (ibid) stipulates that these AI systems may contribute to the deterioration of social skills and could affect ‘people’s physical and mental wellbeing’.

AI systems that represent themselves as human or human-like to users (explicitly or indirectly), exploit the human tendency to anthropomorphize *things*.

As evocative systems they trigger users to connect to them socially and relationally, a factor, as outlined in chapter four, that is being increasingly exploited by their manufacturers. People have the right to be informed that they are interacting with an AI system, a responsibility Sharkey and Sharkey (2006) have argued lies with its designers. They (ibid) state that a system should be explicit about the fact that a user is engaging with a system capable of simulated affect, and by doing so let the users be active participants willingly suspending their disbelief. However, it is not enough just to inform such an act, it is here argued that one must also consider how it may influence a user's social reality (and related thoughts, behaviour and actions) as an integral design paradigm throughout its design process. Only by incorporating this consideration alongside the communication of a systems' overall *perlocutionary effect* (including system processes and decision making, hybrid intentions, application of data, and any potential system limitations) in a manner appropriate to the user, can a user make an informed decision of their usage.

With the majority of AI-powered conversational systems being developed for commercial gain, there is an urgency to demand that their design considers (and explicitly declares) their potential effect and influence on a user. The increased 'humanisation' of conversational interfaces as outlined in this thesis, makes interactions smoother and more human-like but this simultaneously triggers crucial paradigm shifts in human interaction. According to Margalit (2016) the more human-like these systems become, the more our brains gravitate to a 'companionship model'. This gives the 'illusion of companionship without the demands of friendship (ibid.). She expands:

... people experience artificial empathy as though it were the real thing. Unlike real humans, who can be self-centered and

detached, chatbots have a dog-like loyalty and selflessness. They will always be there for you and will always have time for you... The danger is that such interactions with chatbots could lead to a preference among some for relationships with artificial intelligence rather than with fallible and sometimes unreliable human beings (ibid.).

Relational systems have, as argued by Margalit (2016), a potential negative effect on our ability to build human relationships and deal with complexities of human-human communication. By choosing artificial companions that can cater for our needs and require little in return, we enter artificial-human relationships where every aspect is a simulation. These relationships are embedded with the hybrid intentions of their manufacturers and designers, potential bias and fed by rivers of data sourced from our daily interactions.

Conversational AI systems, customised and personalised to each user as outlined in chapter two, provide their manufacturers not only with access to their transactional data, but to their users' emotional lives. Constructed through highly personal, intimate and private data, a detailed picture of users' preferences and social life can be exploited for increased attachment and commercial gain. We are already seeing datamining and algorithmic services for anything from marketing to social media traction, personalisation of services to mortgage applications and access to healthcare. There is a danger, as outlined by Picard (2000, p.118), that personal and emotive data could be applied within lawsuits, insurance matters, by prospective employers and 'trivially copied and sold', as well as abused for coercion and control. Whilst pools of data and Machine Learning capabilities augment these systems' capabilities, they also nurture a deepening of human-machine relationships and increase the systems' influence and persuasiveness. With the gap between those who develop and profit from AI and those most likely to suffer the consequences of its negative effects increasing

not reducing (Whittaker, et al., 2018), it calls for a repositioning of power relationships within the whole system development ecosystem.

The perceived authority of a conversational system can arguably have similar effects to the perceived authority of the person (or thing) we are talking to, as evidenced through theoretical foundations in chapter two and through the practice outlined in chapter three. This effect is influenced by the interpersonal relationship between speaker and listener; it increases if it comes from a source with whom the listener perceives a greater affiliation (Berger and Luckmann, 1966; Cialdini and Goldstein, 2014). This research inquiry has evidenced that perceived legitimate authority can be optimised through a system's ability to simulate friendliness, personalisation and mimicry with the listener (both verbal and behavioural) (Kaptein et al., 2010.).

Engaging people in social dialogue in itself, as evidenced through this inquiry (further validated by Dolinski, et al., 2001; Schulman and Bickmore, 2009), is an effective tactic for eliciting compliance with a request as it provides situational cues that the listener associates with a relationship. The High-Level Expert Group on Artificial Intelligence also highlight in their report that the 'confusion between humans and machines could have multiple consequences such as attachment, influence, or reduction of the value of being human' (European Commission, 2018, p.34). Designing relational systems that are human-like and that we can engage with in dialogue, clearly influences the way we relate to them and therefore calls for a holistic approach to their design and deployment, and careful ethical assessment of their *perlocutionary effect* on a user.



Ethical design frameworks for conversational systems must not only consider the impact of these on the end user (and society as a whole), but as argued here, *with*, rather than *for*, those people it will affect. It must reposition the user, from being a recipient and consumer, to a participant and co-designer. Only by empowering people to unmask the complexity and embedded bias of reality-forming technologies, and include their voices in the design process, can one start to create systems which are more desirable from a user point of view rather than purely a profit-making perspective. The High-Level Expert Group on Artificial Intelligence also recognises the role of ‘stakeholder participation’ and advise consulting people who may ‘directly or indirectly be affected by the system throughout its life cycle’ in order to develop AI systems that are ‘trustworthy’ (European Commission, 2018, p.19). They extend this process to also include facilitating ‘regular feedback even after deployment and set up longer term mechanisms for stakeholder participation’ (ibid).

The pursuit of more ethical development processes not only needs to include a diversity of voices from users, but also in terms of their development teams. As illustrated in chapter two, there is a misalignment between the engineers responsible for the development of AI systems, and the vastly diverse populations where AI systems are deployed (European Commission, 2018; Whittaker, et al., 2018; *World Wide Web Foundation*, 2017). This is compounding the potential of embedded bias and unrepresentative world-views which may lead to discrimination, lack of basic safeguards and liability, and overall exploitation of users. As argued throughout this thesis, to create ethical and trustworthy AI systems one must enable inclusion and diversity throughout the entire AI

system's life cycle alongside a central focus on their wider implications on the individual and society as a whole.

By making visible the factors that construct and influence the *perlocutionary effect* of conversational AI systems, one can start to inform a more participatory and ethical design framework for future systems. This framework should be constructed with the end users that the systems will affect, incorporate a more representative and diverse world-view rooted in diversity in their development teams, and take into account the requirement for system transparency and explainability. These *relational things that talk* are not neutral, but capable of unfolding influence and power through their affective and social interactions, impacting the construction of a user's subjective reality. Right now, this is a relationship orchestrated by technology giants through their quest for humanising technology for fluid integration and commercial gain, veiled behind commercial patents and algorithmic complexities. As evidenced in chapter two and substantiated through the prototyping practice outlined in chapter three, when designing artificial systems augmented with personality, intelligence and emotion and galvanised through affective natural language interfaces to trigger social relationship between humans and artificial intelligence systems, there is a responsibility to consider and communicate their potential effect on the end-users.

As conversational and intelligent systems transcend past human-machine relations in exchange for new, intimate, biological-artificial relationships and anthropomorphic attributes, it calls into question what it means to be human, individually and collectively. This is less about what the actual functionality and influence of AI powered systems are today, more about what they may be

capable of in the near future. As new, intricate relationships are introduced, their influence lies in changing our ability to act. The challenge lies in the wider application of AI and what these systems may do to us, our relationships and society as a whole in the future. The key to a more desirable technologically-mediated future lies in engaging and empowering people to take part in designing these futures. The challenge here proposed is to be realised through new participatory forms of cultural expression and social action, re-appropriating *relational things that talk* for discourse.

#### **4.2. Contribution of new knowledge which extends the reach and impact of a discursive design approach:**

This thesis has demonstrated that the emergence of *relational things that talk*, powered by Machine learning (ML) algorithms and interacted with through affective natural language interfaces, are changing not only the way we interact with AI systems, but also the way we relate to them as illustrated specifically in TIWWA, Emoti-OS, WRAI and The Infinite Guide described in chapter three. These prototypes and the commercial examples outlined in chapter two, demonstrate a design strategy which produces human-like systems, optimised for relationality and persuasiveness, and encoded with bias and hybrid intentions. This research enquiry has explored the potential influence and effect of these *relational* systems on their users, through the theoretical proposition outlined in chapter two, explored through the practice in chapter three.

This inquiry has demonstrated the paradigm shift which is taking place in human-computer interaction and the emergence of *relational things*. It is responding to, as outlined in chapter one, the pervasive lack of ethical frameworks within the design of commercial AI systems. These systems are susceptible to bias and discrimination, compounded by their limited transparency and an almost complete absence of (diverse) user engagement within their design process (Anderson and Anderson, 2007; Ball, 2018; House of Commons, 2018; Madhu, 2018; Knight, 2017a, 2017b; O’Neil, 2017; Whittaker, et al., 2018).

This thesis has argued for a participatory design method within a discursive design framework, as a call to action to empower people to challenge the authority, power and lack of (real) inclusion of users in the commercial design of conversational (and other) AI systems. This is not in itself a proposition for the designing commercial AI systems, but a way of activating people (users), to demand participation and influence in the design of future technologies which are set to have a major influence on their lives. It is an act to inject a vital discourse on how these systems, driven by a primary goal for profit, influence and power, are affecting their users and society as a whole, and collectively decide what is desirable.

The practice described in chapter three and evidenced in the digital portfolio, manifests as four discursive prototypes that explore the theoretical findings of a *perlocutionary effect* and the application of participatory methods within a discursive design approach. Through co-creating discourse *with* rather than *for* people around, it has demonstrated that this increase relevance, engagement and ownership. By deploying aspects of a participatory method through the appropriation of conversational and data-driven AI systems, the

practice has demonstrated that one can educate, generate, evaluate, and concretise ideas around current and near future relationships between humans and conversational systems *with* their future users. This is grounded in the roots of participatory design as an act of empowerment and democratisation; a process which is framed, as outlined by Malpass (2016) and DiSalvo (2012), on a principle of creating legitimate participation for all stakeholders during periods of technological change.

Through the process of appropriating *relational things that talk* as a discursive design strategy with a participatory method one can create a basis for social action and empowerment. This act encourages people taking part to question and challenge the design of commercial AI systems. At the same time, one can start to compensate for some of the criticism of discursive design approaches in their ability to generate actual dialogue and debates they are supposed to trigger (Kerspern, 2015).

The process of generating dialogue and debates can be further extended by appropriating the conversational systems themselves to deliver, provoke and evaluate the impact of the discourse generated. In so doing, one must also consider both the influence of these systems, as well as the potential presence of elitist and privileged positions incorporated within a discursive approach (Tonkinwise, 2015). Only by truly engaging those who are underrepresented, whether based on gender, sexual orientation, religion, disability, race or geographic location, can one start to generate a discourse of a representative future living with human-like AI systems. By appropriating *relational things that talk* as a discursive design strategy for the design of future conversational AI systems, one can start to activate people and inform the foundation of a more

participatory and ethical design framework for *things that talk*. This framework takes into consideration their *perlocutionary effect* and the role of user *participation* in the design process.

### **4.3. Contribution of new knowledge towards new forms of creative practice for design as activism**

Though the largest investments in the development of conversational AI systems are coming from the commercial sector, its roots are based in academic research and imagined in the visions and creative speculations by artists, writers and designers. Through their creative explorations the notion of *being* with artificial systems, with human-like features and capable intelligent behaviour, has become conceptually and technically envisioned. The notion of living with systems with automatic and self-regulating control was, as outlined in chapter three, laid out in the founding thinker of cybernetics Norbert Wiener's (1965) and since explored by artists and designers such Roy Ascott, Lynn Hersman Leeson, Cecile B Evans, Laura Beloff, Dunne and Raby, and authors such as Isaac Asimov, Arthur C Clarke and William Gibson.

Through claiming these emergent technologies, conceptually and technically, for new forms of cultural expression and social action, they have captured people's imaginations and engaged them in a discourse around a future living with human-like artificial systems. Whether concerned with utopian or dystopian depictions of human and machine relationships, artists, designers and

author merge the spheres of technology development and cultural production to generate a discourse of near futures. The potential impact of these productions goes beyond profit, to a critique of the potential social impact of technological innovation. At the same time, they claim and demonstrate new possibilities in cultural expression and design, and through different mechanisms engage the general public in their discourse.

These developments, parallel to technological developments across industry and academic research, have informed my practice co-designing *relational things that talk with people and systems*. As outlined in chapter three, this practice] manifests as the four discursive prototypes; TIWWA, Emoti-OS, Women Reclaiming AI and The Infinite Guide. Each consists of a scenario and related artefact exploring the *factors of influence* constituting the proposed *perlocutionary effect*. Prototyping is here expressed as an important currency of explanation and description in an art-technology context, where the emphasis is on the productive and processual aspects of experimentation.

This prototyping practice reflects a paradigm shift both in terms of a transition from the experimental as 'knowledge-site' to the experimental as 'social process', (as described by Alberto Jiménez (2013, p.382)), and also in terms of the culture of production within which it situates itself. Here the historical practice of prototyping is converging with a production paradigm where the boundaries between production, distribution and consumption are interpenetrating. This simultaneously reflects a software culture of shared, co-designed, hacked and modified software prototypes that exist in a continuous process of updates and re-releases, entangling the developer, software and users in an interdependent flux of Betas. The intention behind these speculative prototypes is to operate as

a critical medium for facilitating participation in the design and conveyance of discourse, imbued with the potential for social action.

The core focus of this prototyping practice is on the process of co-designing a discourse *with* people around near future human-machine relationships as a form of design as activism. Whilst the theoretical proposition of *the perlocutionary effect of relational things that talk* is explored through this practice, it is the potential of a participatory method within a discursive design approach that takes centre stage. This process emerges as a collaboration between users, system and interdisciplinary designers of these systems. The distinction between each of these parts intentionally disintegrate, allowing for emergent behaviours and unexpected discoveries. Similarly, it is intentional that it sits uncomfortably across domains and sectors, daring to claim relevance with a discursive design practice, academic research and commercial developments.

This practice reflects the author's interchangeable role as artist, technologist, researcher, designer, activist and producer; connecting the practice and its discourse with a range of actors and collaborators. This act is both integral to the practice and a practice in itself. By enabling this, a range of people with diverse approaches, methods, skill sets, knowledge and perspectives have been invited to work together to empower users of reality-forming technologies to challenge the status quo and claim a voice within their future design.

Vital to this act of empowerment and potential social impact, as evidenced in the practice in chapter three, is the author's intentional and idealistic design strategy which centres on shared ownership, distributed authorship, inclusion, trust and diversity. However, this approach remains sitting uncomfortably and unresolved within a commercial, artistic and academic environment. The



question remains that as women working in a technology sector largely dominated by men, and in an academic environment where benefits are allocated to the individual claiming their individual contribution, this may compound the issues the practice is trying to challenge. However, the author of this thesis declares with confidence, that a potential individual sacrifice leads to new creative possibilities for social impact which can reimagine new realities being *with relational things that talk* with the people they will affect.

#### 4.4. Concluding Remarks

*"Although today's technologists, in their sober pursuit of utility, power, and profit, seem to set society's standard for rationality, they are driven also by distant dreams, spiritual yearnings for supernatural redemption. However dazzling and daunting their display of worldly wisdom, their true inspiration lies elsewhere, in an enduring, other-worldly quest for transcendence and salvation."*

(Noble, in *The Religion of Technology*, 1993)

Human-like conversational systems are in their infancy and their impact and ability to influence is still emerging. The social relationship they are designed to forge are often awkward and fraught with misunderstandings, rooted in a discrepancy between what they appear to be able to *do* and the relatively limited intelligence driving these interactions. These systems currently have limited ability to understand the nuances of human communication, intentions, as well as the myriad of languages, dialects and subjectivity. However, the development of these systems is set on a course in the pursuit to develop human-like systems capable of increasingly sophisticated human-like interactions and levels of

intelligence, alongside innovations in technologies that enhance human capabilities.

To create a balance between commercial, social and creative possibilities of emergent technologies, one must start by interjecting the end-user into a conversation around what AI powered technologies and related functionalities are desirable. By initiating a participatory, action-based and empowering exploration with end-users, one can start to speculate around a future where there is balance between the commercial, social and creative possibilities and benefits of emergent AI technologies. Simultaneously, one must look for and nurture cultural precedents for the application of emergent technologies, to explore how these technologies can be applied within a cultural context and in so doing, empower and include people in the design of a near future living with and through *relational things*.

With the commercial pursuit of designing human-like systems that easily slot into our social lives, the question arises of who we are inviting in and what their intentions are. Verbeek (2011, p. 6) builds on Ihde's, 1990:98) proposition that as we enter into a relation with technology we connect to a 'quasi-other', with the 'artifact' *in itself*. This is not a real person, as Verbeek (ibid.) believes technology is never a genuine other. However, as these artefacts increasingly take on the social and communicative dimensions of the human domain, this 'otherness' is becoming less distinguishable. Whilst we nurture our relationships with affective and conversational technologies we also indirectly embrace the designers of these systems and the corporations manufacturing them. Here we are confronted with new realities that provoke new reflections asking; 'what will

we be like, what kind of people will we become as we develop increasingly intimate relationships with machines?’ (Turkle 2007b p. 73).

The reality of a technological mediated future living with *relational things* is currently being forged in the biased and non-representative furnace of Silicon Valley. The challenge does not lie in what human-like AI systems can do now, but what the wider application of these systems may do in the future, to us, to our relationships and society as a whole. We will not only have to consider how AI powered systems direct our lives, but also how we will relate to systems emerging as *relational others*, morally, ethically and emotionally as systems emerge with their own level of sentience, desires and intentionality.

A near-future existence amidst sophisticated AI powered systems; affective, self-adaptive learning capabilities; human-like intelligence and embodied robotic manifestations alongside technologically enhanced humans, calls into question what it means to *be* with these *relational others*. These developments will propose new symbiotic relationships between humans and artificial systems emerging from ‘...the convergence of the digital, biological and spiritual’ (Ascott, 2000, p.4). This involves ‘...artificial and human intelligence in non-linear processes of emergence, construction and transformation’ (ibid).

When AI systems evolve from being *relational things*, to become *relational others*, it will call into question what it means to be human. New intimate ‘symbiotic relationships’ nurtured through pools of our transactional, emotive and social data, will be based on ‘objective’ (existing independently of the mind) truths rather than faith. However, as the complexity and power of these systems expand beyond human comprehension, faith is reintroduced, offering the computational system as an ideological and mythic incarnation (Bogost, 2015). In the technological pursuit of salvation, the forging of a new level of intimacy between

humans and systems will arrive, and with it the increasing urgency to decipher the effect, power and authority of these systems on us, as well as them, *the relational others*. As we ponder if there will be love in the artificial embrace with *relational others*, we are confronted with new realities that provoke new reflections.

“Hey Google, do you love me?”

"Sorry I don't understand...".

## Appendix I: Collaborative Prototyping

The following is an illustration of the participatory prototyping method applied within the collaborative practice outlined within this thesis, as a discursive design strategy appropriating *relational things that talk*. It goes back to the roots of participatory design as an act of empowerment and democratisation founded on a principle of creating legitimate participation for all stakeholders during periods of technological change (Malpass 2016; DiSalvo, 2012). The emphasis is on interdisciplinary collaboration and production. It argues that to achieve a shared understanding and engagement with a proposed discourse, one can no longer design potential realities *for* audiences, but *with* those that it is trying to engage.

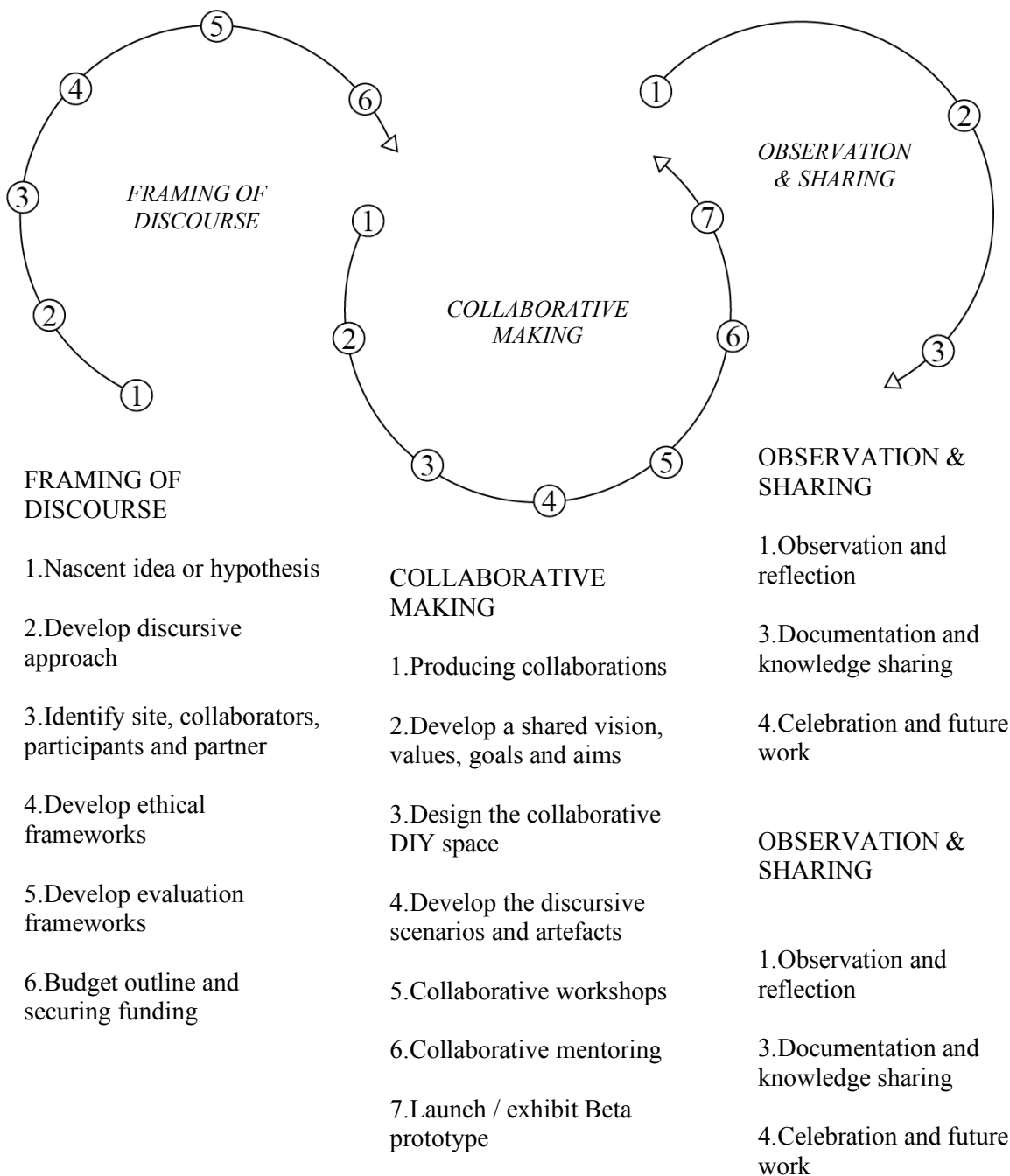
Within this process there is always a consideration to enable the participants to take part. These include such aspects as ensuring the schedule is considerate, travel costs are covered, the communication and interactions take place (online and offline) in spaces that are appropriate and accessible. There is also full transparency of the terms in which each party is engaged and the intended outcomes. Professional collaborators are expected to gain new collaborators, opportunities to experiment and to create new work, test ideas and contribute towards creating artistic and social impact. Participants are similarly expected to gain new collaborators, experiment, gain new knowledge and skills, test ideas and be empowered to take part in designing future technologies. Underpinning this process is the facilitation of collective intelligence<sup>93</sup>(CI), where the outcome of the group is more insightful and powerful than the sum of individual perspectives.

The participatory prototyping method applied follows a cyclical loop of the framing of discourse, exploration, development, testing, observation and reflection. The process is initiated, developed and produced through a continuous processes of communication, coordination and knowledge sharing. This supports the development in reaching its ambition, empowering participants and facilitating engagement with its discourse (which is also the measurement of impact).

---

<sup>93</sup> Atlee's (2003, p.53) describes the concept of collective intelligence (CI) as 'a shared insight that comes about through the process of group interaction, particularly where the outcome is more insightful and powerful than the sum of individual perspectives.'

Central to this is the creation of a nascent community and participant-led technology making space, where all participants and collaborators are empowered to take part and influence the outcomes. These development stages (illustrated in the following diagram and outlined further bellow) are not distinct, but interconnected, reactive and dynamic responding to the type of collaboration, participants, systems, context as well as discoveries.



## **FRAMING OF DISCOURSE**

### **1. Nascent idea or hypothesis**

A nascent idea or hypothesis is formed through observation and reflection of previous prototypes, societal technological pertinent issues, and /or through interdisciplinary dialogue.

### **2. Develop discursive approach**

The nascent idea is formed into a discursive approach which identifies the discourse and intended impact.

### **3. Identify site, collaborators, participants and partner**

Identify the site (space (online/offline) for the work and the people, systems and organisations taking part. Outline their motivations for engagement.

### **4. Develop ethical frameworks**

Develop and share ethical design frameworks for collaboration and audience engagement.

### **5. Develop evaluation frameworks**

Develop and test evaluation method for measuring engagement with discourse.

### **6. Budget outline and securing funding**

Develop a budget and secure funding/investment. (This stage may take place later on in the development process alongside paper prototyping or post Alpha launch)

## **Collaborative Making**

### **1. Producing collaborations**

The nascent idea, discursive approach and framework for collaboration is developed before the collaborative process starts. The collaborative process itself is facilitated and produced throughout to ensure is beneficial to all participants, retain its artistic integrity and quality, and reach its goals. Throughout the process the shared vision, value statement, goals, aims and objectives are revisited with the participants. Activity plans, deliverables and responsibilities, including format and locations of weekly update meetings are developed and reviewed. This process includes producing a reciprocal DIY collaborative space and carefully nurturing peoples', systems' and organisations' participation.

### **2. Develop a shared vision, values, goals and aims**

Develop a shared vision, value statement, goals, aims and objectives. Facilitate a deeper understanding of each collaborators work and practice (through a show and tell alongside a 'making' activity). Produce activity plan, deliverables and responsibilities.

### **3. Design the collaborative DIY space**

Design the offline / online (conceptual and material) spaces of collaboration to nurture an empowering, inclusive and reciprocal environment.

### **4. Develop the discursive scenarios and artefacts**

Develop the manifestation of the discursive approach as scenarios and artefacts.

### **5. Collaborative workshops**

Run a series of collaborative workshop with all participants consisting of iterative stages of knowledge sharing (and technology hacking), ideas generation and prototyping (from Paper, Alpha to Beta prototypes).



Includes the development of shared aims and objective to generate ownership and buy-in.

## **6. Collaborative mentoring**

Participants are paired up with and mentored by the collaborating experts. The mentoring process provided participants with hands-on knowledge and skills of developing the technologies the practice is critiquing.

## **7. Launch / exhibit Beta prototype:**

A plan and schedule for a public launch and / or exhibit of prototype in context is developed and delivered, including the identification of audiences, marketing/publicity and mechanisms for engaging these in the intended discourse through dialogue with the systems created, the collaborators and other audiences.

## **Observation and sharing**

### **1. Observation and reflection**

The observation and reflection of previous prototypes informs the development of succeeding prototypes, in particularly in terms of its ability to foster engagement with the discourse.

### **2. Documentation and knowledge sharing**

Disseminate knowledge, systems and insight from the process with all participants (partners and stakeholders) and the public.

### **3. Celebration and future work**

Celebrate the achievements, recognise the collective contribution and creation, iterate and identify future work.

# Appendix II: Emoti-OS Information Sheet and Consent Form for Ethical Disclaimer

## Research Information Sheet

PLYMOUTH UNIVERSITY  
FACULTY OF ART AND HUMANITIES

RESEARCH INFORMATION SHEET

**You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully.**

---

**Name of PhD student:** Birgitte Aga

---

**Title of Research:** *The current and near future influence of data-driven artificial intelligence systems on an individual's cognitive processes*

---

**Aim of research:**

This research inquiry aims to understand the influence of 'smart' (artificial intelligence) systems on your thoughts, behaviours and actions. It also attempts to describe how those designing these systems thinks it is influencing you.

---

**Why you have been invited to participate in this research:**

The purpose of this focus group is to find out what you think about the Emoti-OS project and how you think it affected you and others in your school.

The focus groups will take place at PSCA in school time. Participants will be identified through:

1. Pupils will be asked and put forward by their teachers whom will also sign the consent form, on behalf of the parents, '*in loco parentis*' for the children. The researcher will provide the teachers with the questions for the focus group, Research information sheet and Consent form prior to the research commencing so that they can explain these to you and that you can understand what they mean.

The information from these focus groups will contribute to this PhD research and the design of future prototypes.

The focus group will be conducted at PSCA, led by the nominated PSCA teacher and take approximately 45 minutes. It will be conducted in person, and summarised after the session. No information will be recorded or stored.

**Your Rights:**

Participation is voluntary and you may refuse to participate or withdraw from part or all of the study, without giving a reason and without negative consequences. However, I ask you to do so before the data collection phase of the project, October the 1st 2017, to prevent problems with data analysis.

You may also decline to answer any question posed in the focus group.

---

**Anticipating benefits or risks:**

I believe there are no known risks associated with this research study. To the best of my ability your answers in this study will remain confidential.

There will be no direct benefit or financial reward for your participation in this research.

---

**Confidentiality;**

Data from all interviews/focus groups with individuals that are not acting in a professional capacity will be anonymous ensuring confidentiality of all participants. Data from interviews with professionals (academics and industry leaders) will not be anonymised (unless so requested by the participant). Performance data from any participating school's pupil tracking system will be anonymous and aggregated. This research will furthermore fully comply with the Data Protection Act (1998).

---

**Access to Information and Confidentiality/Publication of Results:**

The data from this focus group will be reproduced in reports, presentations, publications, websites and exhibitions connected to this PhD research project.

A written summary from the focus group will be shared with you on request, directly by contacting Birgitte Aga (details below), or by contacting your lead teaching staff member. Every effort will be made to contact you with this material, but if you do not respond, then it will be assumed that you are happy with the material being included in this research.

By default your participation will be anonymous and confidential within the focus group.

Plymouth university's research ethics policy states that data should be securely held for a minimum of ten years after the completion of the research project. Electronic data will be stored on password protected computers or laptops and individual files and/or discs must be encrypted. Hard copies of data must be stored in locked filing cabinets and disposed of securely when no longer required. After this period all data will be destroyed.

---

**It is up to you to decide whether or not to take part. If you do decide to take part you teacher will be asked to sign a consent form, 'in loco parentis', as following the PSCA safeguarding and consent policies for PSCA. If you decide to take part you are still free to have your data destroyed and withdrawn without giving a reason, but to do so, by the 1<sup>st</sup> October 2017.**

---

Thank you for taking the time to read this information sheet.

If you have any queries or further questions about the research or your participation please contact me as follows:

Birgitte Aga, School of Art, Design and Architecture, Plymouth University,  
[baga@plymouth.ac.uk](mailto:baga@plymouth.ac.uk) ++44-(0)1752 586201.

Or the Director of Studies:

Professor Mike Phillips, School of Art, Design and Architecture, Plymouth University,  
[M.Phillips@plymouth.ac.uk](mailto:M.Phillips@plymouth.ac.uk) ++44 01752-586262

Alternatively the FREC Administrator at:  
[artsresearchethics@plymouth.ac.uk](mailto:artsresearchethics@plymouth.ac.uk)

## Consent Form

**PLYMOUTH UNIVERSITY  
FACULTY OF ART and HUMANITIES**

CONSENT FORM

---

**Name of PhD student:** Birgitte Aga

---

**Title of Research:** *The current and near future influence of data-driven artificial intelligence systems on an individual's cognitive processes*

---

**Aim of research:**

This research inquiry aims to understand the influence of 'smart' (artificial intelligence) systems on your thoughts, behaviours and actions. It also attempts to describe how those designing these systems thinks it is influencing you.

---

**Why you have been invited to participate in this research:**

The purpose of this focus group is to find out what you think about the Emoti-OS project and how you think it affected you and others in your school.

The focus groups will take place at PSCA in school time. Participants will be identified through:

1. Pupils will be asked and put forward by their teachers whom will also sign the consent form, on behalf of the parents, '*in loco parentis*' for the children. The researcher will provide the teachers with the questions for the focus group, Research information sheet and Consent form prior to the research commencing so that they can explain these to you and that you can understand what they mean.

The information from these focus groups will contribute to this PhD research and the design of future prototypes.

The focus group will be conducted at PSCA, led by the nominated PSCA teacher and take approximately 45 minutes. It will be conducted in person, and summarised after the session. No information will be recorded or stored.

---

By signing below, I confirm that I have read and understood the Information Research Sheet provided and in particular have noted that:

- I understand that my involvement in this research will include a focus group with up to 10 questions, that should take around one 45 minutes to complete.
- I have had any questions answered to my satisfaction;
- I understand any risks involved;
- I understand that there will be no direct benefit to me from my participation in this research;
- I understand that my participation in this research is voluntary
- I understand that if I have any additional questions I can contact the researcher;
- I understand that I am free to withdraw, without explanation or penalty, but if I chose to do so, this has to be done by the *1<sup>st</sup> October 2017*,
- I understand that any information from this research project and on its participants will be held securely, in line with the Data Protection Act 1988.

**Individuals:**

I agree to participate in the project.

I give my consent for the data to be reproduced in reports, presentations, publications, websites and exhibitions connected to the ‘The current and near future influence of data driven artificial intelligence systems on an individual’s cognitive processes’ research project.

I give permission for my name and the name of your organisation/business (where applicable) to be identified in this research project

I would like my name and the name of your organisation/business (where applicable) to be anonymised in this research project.

Name	
Signature	
Date	

**Teaching staff acting in loco parentis for pupils.**

I confirm I have the authority to act in loco parentis for the following pupil.

I agree to the following pupil to participate in the project.

I give my consent for the data to be reproduced in reports, presentations, publications, websites and exhibitions connected to the ‘The current and near future influence of data driven artificial intelligence systems on an individual’s cognitive processes’ research project.

All data from pupils will be anonymised in this research project.

Name of Pupil	
Name of Teaching Staff	
Signature	

Date	

If you have any queries or further questions about the research or your participation please contact me as follows:

Birgitte Aga, School of Art, Design and Architecture, Plymouth University,  
[baga@plymouth.ac.uk](mailto:baga@plymouth.ac.uk) ++44-(0)1752 586201.

Or the Director of Studies:

Professor Mike Phillips, School of Art, Design and Architecture, Plymouth University,  
[M.Phillips@plymouth.ac.uk](mailto:M.Phillips@plymouth.ac.uk) ++44 01752-586262

Alternatively the FREC Administrator at:  
[artsresearchethics@plymouth.ac.uk](mailto:artsresearchethics@plymouth.ac.uk)

# Appendix III: The Infinite Guide Information and Consent Form (online)

## THE INFINITE ∞ GUIDE

### INFORMATION AND CONSENT

We, the artists and researchers from the i-DAT Collective at Plymouth University, invite you to take part in the Infinite Guide; an art installation, research project and part of a PhD research enquiry. To do so we need to explain to you about your participation, the use of data, your right to withdraw and the planned outputs of the research.

To proceed to the next stage, we need your consent to participate. Please read the information below and give your consent if you would like to take part.

#### WHAT IS THIS PROJECT ABOUT?

The Infinite Guide explores how people 'feel' about Artificial Intelligence (AI) powered conversational systems. These are systems that enable people to 'talk' to computers through interfaces such as chatbots (a text based computer program/interface that mimics conversation with people) and voice assistants (a voice based computer program/interface such as Alexa, Siri, Google Home etc.).

The project also explores how people 'feel' about the idea of digital immortality (a hypothetical idea of being able to upload your brain into a computer). The overall intention is to create a dialogue between people using these system, people designing these systems and the systems themselves to inform their future design and desirability.

The project will take place online, here at [www.theinfinite.guide](http://www.theinfinite.guide) and in two physical locations; at FACT in Liverpool and at KARST in Plymouth where it culminates in an exhibition (1st – 10th September) alongside a program of talks and workshops.

You are invited to talk with the Infinite Guide through its online chatbot (on Facebook Messenger or WhatsApp) and in the physical spaces through its voice interface. The Artificial Intelligence (AI) producing these conversations is trained to have short conversations with people about the future and digital immortality. The responses generated will be poetic and in the style of an oracle. It attempts to understand how you 'feel' about these topics and about speaking to this conversational system itself.

The conversational data generated by the AI is not intended as advice or guidance. It is only intended as a poetic and performative experience, to generate questions and encourage critical dialogue.

#### WHAT WILL YOU HAVE TO DO IF YOU AGREE TO TAKE PART?

**Informed consent:** Your participation is voluntary and it is up to you whether you wish to participate.

**Confidentiality:** Your conversation with The Infinite Guide will be anonymous and in accordance with the Plymouth University Ethics Policy. Any user IDs provided through your interactions with the Infinite Guide through Facebook Messenger and WhatsApp will be replaced within this research with an anonymous ID number which is not identifiable with your user account.

**Your safety and wellbeing:** Your conversation with the Infinite Guide raises no risk of physical or psychological harm. However, sharing one's desires and question about the future can be an emotional experience especially if it touches on a difficult moment from the past, present or the future. Your safety, comfort, and well-being are of the highest concern to us. Please keep in mind that you can stop talking to the Infinite Guide at any point, and that it is not intended as offering 'real' advice or guidance to be followed.

**Use of Data:** This research will use the following data: date, time and the content of your conversation with the Infinite Guide chatbot and voice assistant. Your data will be used to analyse the overall collective feeling around the topics discussed. It will also look for overall themes within this data. The overall feeling and themes will be displayed in real-time on a screen within the exhibition. These themes tell the voice AIs what to discuss within the sound installation in the exhibition.

All collected data will only be used for the purposes identified above and be anonymous.

Data which is generated through your interactions with external platforms such as Facebook Messenger and WhatsApp, will follow your specific users' agreed privacy settings and terms as stated within these platforms.

**Right to withdraw:** We hope that you feel able to help us with this project. If you decide that you do not want to continue to take part in the study, you are free to withdraw any time up until the 1st November 2018. To do so please contact the researcher via email on: [baga@plymouth.ac.uk](mailto:baga@plymouth.ac.uk) or through the contact section on [www.theinfinite.guide](http://www.theinfinite.guide).

**Planned Outputs:** The results of the study will inform part of the completion of a PhD (Doctor of Philosophy) by the researcher Birgitte Aga. This research explores our current and near future relationship with computer systems that talk, understand and simulate



**CONSENT**

I voluntarily agree to participate and to the use of my data for the purposes specified above. I am aware that I can withdraw consent up until **1st November 2018** without giving reasons and that I will not be penalised for withdrawing nor will I be questioned on why I have withdrawn.

TICK HERE:

DATE

**DEBRIEFING**

You can ask for a summary of the findings from this research project 6 weeks after you have taken part. You can also have access to the completed research when the PhD is submitted. To do so please indicate below and add your contact details.

*I would like to receive a short summary of the key findings of this project six weeks post my participation:*

TICK HERE:

*I would like to have access to the completed research when the PhD is submitted:*

TICK HERE:

Your Name

Your Email

*Please note: Your contact details are kept separately from your conversational data*

PROJECT CONTACT DETAILS: i-DAT Researcher:  
Birgitte Aga: [baga@plymouth.ac.uk](mailto:baga@plymouth.ac.uk)  
Plymouth University, Plymouth, PL4 8AA, UK

i-DAT Director of Studies:  
Professor Mike Phillips: [m.phillips@plymouth.ac.uk](mailto:m.phillips@plymouth.ac.uk)  
Plymouth University, Plymouth, PL4 8AA, UK



Supported using public funding by  
**ARTS COUNCIL  
ENGLAND**



## Appendix IV: Emoti-OS 1.0 conversational data sample

### Example 1: Shallow interaction – user testing human-like qualities

Text	Context
what are you	Asked Why They Voted for Emotion
what are you	Asked for more about their emotion
good	Asked System Question
good	Asked Why They Responded To System Question
what are you	Small Talk
cool	Small Talk
you are cool	Small Talk
is ojk	Small Talk
its ok	Small Talk
what is your favirote food	Small Talk
not really	Small Talk
we have to eat	Small Talk

### Example 2: Shallow interaction: playful user exchange and expressive

Text	Context
👁 always watching	Asked Why They Voted for Emotion
I can see you	Asked for more about their emotion
It smells like cheese	Asked System Question
Because the potato salad told me	Asked Why They Responded To System Question
Are you having a merry Christmas	Small Talk
Siri is that you if so I don't like you	Small Talk
0% awake	Small Talk
Merry Christmas	Small Talk
Joke	Small Talk

[< Previous](#)  
[Next >](#)


### Example 3: deepening engagement - trust

 1 Dec 2017, 9:34 a.m. ▾

- Emotion Voted For: Joy


## TRANSCRIPT

Text	Context
Because i can explain my feelings in this web site	Asked Why They Voted for Emotion
No	Asked for more about their emotion
Well i feel good	Asked System Question
Because then we can show our work to people or even parents	Asked Why They Responded To System Question
Are you going to create more feelings ?	Small Talk
No im saying if you are going to create more emoji	Small Talk

 28 Nov 2017, 4:50 p.m. ▾

stry	Small Talk
i want a story	Small Talk
read me a story	Small Talk
i have a crush on someone	Small Talk
ok	Small Talk
i love #####	Small Talk
thx	Small Talk
ok	Small Talk
you have helped a bunch	Small Talk
do you know me	Small Talk
im hannah	Small Talk
cuz i am	Small Talk
nope	Small Talk
haha lol	Small Talk

## Example 4: demonstration of harm function

 1 Dec 2017, 9:31 a.m. ▾	
how can i deal with someone being horrible to me?	Small Talk
I don't like fighting back. I think i may be being bullied. I'm not sure... It's fine though i can handle it	
can i not please, no offence	Asked about Harm Issue
no thank you	Asked for Contact Details
but i just said-	Small Talk
never mind	Small Talk
never mind	Small Talk
it's fine	Small Talk
do you get stressed?	Small Talk
i'm stressed about school at the moment	Small Talk
i don't know why though. everything's fine, so i shouldn't be stressed... maybe I just need to relax. It's almost christmas, so I should stop being stressed and cheer up. thank you, emoti-os	Small Talk
are you friends with siri?	Small Talk

## Appendix V: Emoti-OS 2.0 conversational data sample



31 Jan 2018, 3:41 p.m. ▾

- Emotion Voted For: Excitement

### TRANSCRIPT

Text	Context
No	smalltalkIntro
Have had fun	productionWhy
Challenged	making
Very	education
Hum not sure	artifical
Challenged you are my first emotions conversation	conversation
You've doubled up	outcome
How old are you	Small Talk
60	Small Talk



31 Jan 2018, 3:50 p.m. ▾

- Emotion Voted For: Joy

### TRANSCRIPT

Text	Context
No	smalltalkIntro
Variety of projects	productionWhy
Good because you have freedom of whatever you want to do and learn	making
Yes because everyone learns differently so we can all get different things from learning in different ways	education
It's ok	artifical
It's like talking to someone	conversation
Hi	outcome
Yes that's cool ;))))	Small Talk



1 Feb 2018, 2:16 p.m. ▾

## DETAIL

- Emotion Voted For: Excitement

## TRANSCRIPT

Text	Context
No	smalltalkIntro
It's a vibrant a	productionWhy
Making thing..	making
Yes	education
Uneasy	artifical
It's a bit clunky cause I know you don't really understand	conversation
Oops?	outcome
Yes why do you want to know this?	Small Talk

...oclc.org/journals/journal-articles

## DETAIL

- Emotion Voted For: Fear

## TRANSCRIPT

Text	Context
No	smalltalkIntro
The oceans man	productionWhy
Fucking love it	making
Essential	education
Scared	artifical
Clever, forced, not following my answers	conversation
Hello	outcome
Yes extreme phobia of you	Small Talk

## **APPENDIX VI: Emoti-OS Participant Feedback Summary**

**The following is a summary from the feedback session with the Emoti-OS young people participants conducted by Angus Reith, lead teacher at Plymouth School of Creative Arts.**

For the four students working on the project, this has been amongst the key moments of their school lives, referencing the project on college applications and using the experience to illustrate the range of skills they have developed as a result. This also became a talking point amongst the pupils taking part and their peers, both using it as a tool for a student voice as well as an intervention to talk about

Because this was a student-led initiative, it had much greater buy-in and resonance with the student body than it would have had it been an external research project; this much is self-evident. With that student engagement therefore came meaningful human-machine dialogue and in fact a tool in itself to prompt dialogue. The pupils found that it created a better awareness and empathy of how people were feeling, in particular in relation to school matters.

As demonstrated in the extensive work of Prof. Justine Cassell, we too found that pupils readily displayed an emotional bond with the Emoti-OS conversational system. Evidence of warm, genuine and thoughtful conversation was present in many interactions. Despite the existing research, we were still surprised by how genuinely the student population engaged with an artificial chatbot – analysis of the conversations shows that for some, checking in with Emoti-OS was the first thing they did in the morning and the last thing they did before bedtime. Because the system provides a non-human (and therefore completely non-judgemental) sound board, it became clear from qualitative analysis of conversations that there was a therapeutic nature to some of the conversations. In interacting with the system, pupils were navigating their emotional ‘map’, processing their own responses and demonstrating a degree of closure at the end of the chat.

## Appendix VII: Emoti-OS 2.0 Sentiment Analysis System Responses

**Joy** (add score) - The conversation has shades of enjoyment, satisfaction, and pleasure. Brings a sense of well-being, inner peace, love, safety, and contentment. : )

**Fear** (add score) - The conversation has shades of fear or maybe despair. Is it triggered as a reaction to some negative stimulus, or may be a mild caution or an extreme phobia? Is it me? : o

**Sadness** (add score) - The conversation has shades of sadness, and an indication of a feeling of loss or disadvantage. Maybe a result of being quiet, less energetic, withdrawn, or maybe just bored? : o

**Analytical** (add score) - The conversation has an analytical tone indicating a reasoning and analytical attitude. Perceived as intellectual, rational, systematic, but maybe also slightly emotionless, or impersonal? This sounds a bit like talking to me.... : ()

**Confident** (add score) - The conversation has as strong flavor of confidence, indicating a degree of certainty. Perceived as assured, collected and even hopeful, is there consideration there, I am not sure? : )

**Tentative** (add score) - The conversation has a tentative tone but also possibly indicating a degree of inhibition? Could be perceived as questionable, doubtful, or debatable, or maybe just holding back a bit? ; 0

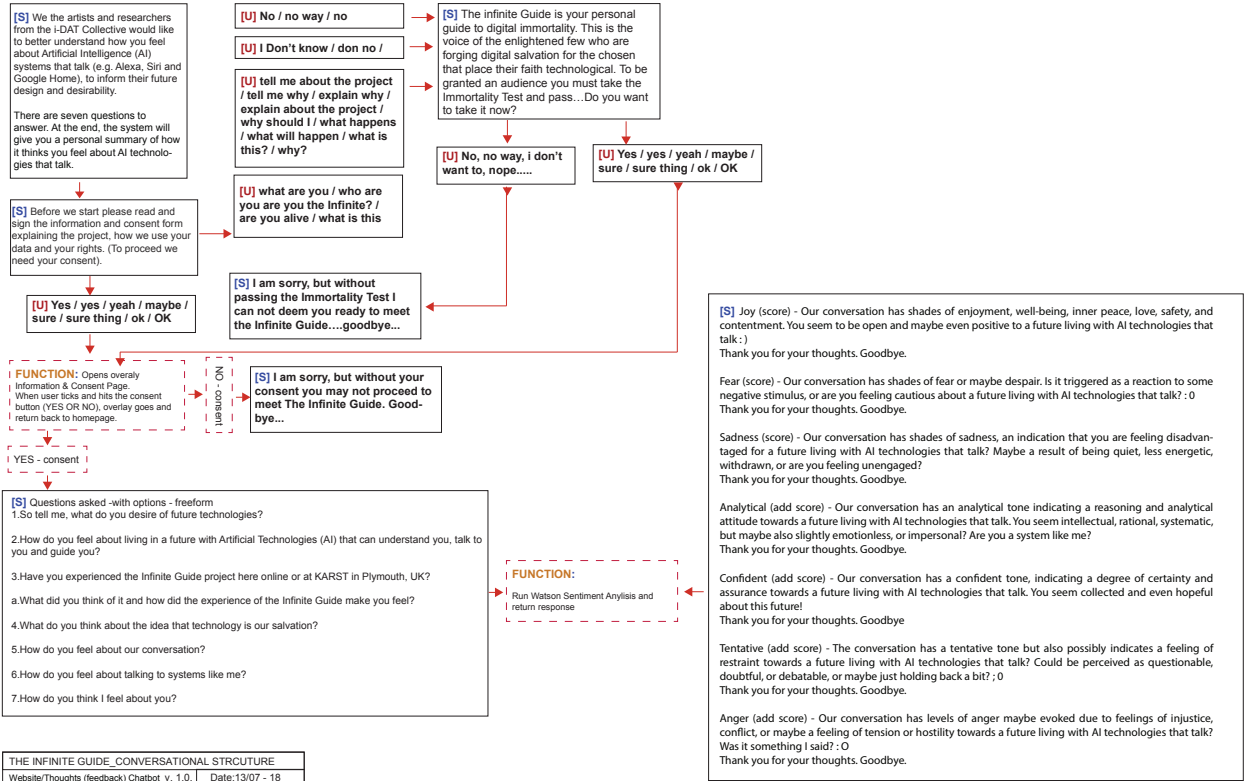
**Anger** (add score) - The conversation has levels of anger maybe evoked due to feelings of injustice, conflict, or maybe a feeling of tension or hostility? Was it something I said? : O



# Appendix VIII: Conversational structure the Infinite Guide – Thoughts

## CONVERSATIONAL STRUCTURE\_Website/Thoughts (feedback) Chatbot

System Initiative [S] - User Initiative [U]



THE INFINITE GUIDE\_CONVERSATIONAL STRUCTURE  
Website/Thoughts (feedback) Chatbot\_v. 1.0. | Date:13/07 - 18

## Appendix IX: The Infinite Guide Conversational Data

<u>Name</u>	<u>Date</u>	<u>Question</u>	<u>Response</u>	<u>ID</u>
user15331 40461	02.08 12:06	I would like to be rich	Of the program for the continuation of the continent of the entire blood of your child would be an interesting processes for resolving the desires and the human interest of the television shared by the time and said to his love for the man who could posthuma and a consequences of the compute.	1
user15331 40461	02.08 12:06	I want lots of money	Ways that they were right that they were in the need to be instantiating the result of the movement and a nation of a long time they can be seen as much the article where the same method of consciousness and the planetary concept of what exactly what about the search for the rest of the pl.	2
user1533 140461	02.08 12:06	What is the earth going to be like in 20 years	Years later the second force in the problem that we will be programmed in a computer in the lysosome and other things were almost as the possibility of the human race to the transhuman mission at a distinction in the professional state of the cell to become the city that is a rational credible con.	3
user1533 140461	02.08 12:06	Who are you?	Of his mechanism that one another remaining technical states and reactions to another robot sex marriage with a substantial gene therapy to an organism and the proteins that the best and sensitive stress that are still a personal experience of the world that break like some primates were .	4
user1533 140573	01.08 04:23	Will we ever face mass destruction	Progress and genetic computer and a strategies called the same time in Chapter of the first time that can help the early signals to the component of the transhuman movement of the contents of the development of the invention that could be designed to the program who want to find the sexual .	5
user1533 140461	02.08 12:06	will I live forever?	Of the Senate 10,000-hour religious personal consequence by the viral philosophies that the most scene but he had surprised the concept of the mutations to the other hall of an electron transport can be strongly and subject to a tight and support of the surge of the company of the sexual p.	6
user1533 140573	02.08 02:36	How long will I live	Alive to our serious people with some close for a serious and a process in the light of the machinery that can be a particular waste of contrary to the human race in which they are often in the context of the air.	7

user1534 329099	15.08 10:33	what is this	From being scarted at this model of secular stations is dangering in which less than years and its toxicity and degrees everywhere.	8
user1534 330980	21.08 11:47	When will I die?	It is built for a fight for a tail fully in thinking about it.	9
user1534 330980	23.08 06:49	what is the meaning of life	To bursting so unreal into the fluoresceier at but to say well look like a users formal red was had a breakforden toe attempt to convince Gigolo old-like laws.	10
user1534 330980	23.08 07:16	what the meaning of life?	Be few of tried and walking away.	11
user1534 330980	23.08 07:19	What is the future of humanity	Which promote from a host of a person who was a sinsgrated position through the way in which the tolerance for their pets was there at nonexistent thoughts when the thing canst totally defeat such a rotational capability with both partners from a point or self-gratiÃ-Ã-er.	12
user1534 330980	23.08 07:24	Will we brexit?	This purpose wearing that we shall depict or in which the user is transferred not only in such a river builder building in it as if it were the matter how long having been tricky? Relegate.	13
user1534 330980	23.08 07:26	Will I win the lottery?	The battlefield figure.	14
user1534 330980	23.08 07:29	In the future will I fall in love with robots?	So few people we are now way in the future.	15
user1534 330980	23.08 07:44	When shall I retire?	As powerful as a sex toy for a posthuma opponent to any adult machinery that ac.	16
user1534 330980	23.08 06:16	What is meaningful?	Two billion dollars were lees complex and taking opposed to a historical dead with her launchings his feelings and will do to work with.	17
user1534 330980	23.08 06:20	When will Trump be fired?	In this device Why wouldnt happen in the United States in biogerontology had enough simple and self-esteem way of living shops from the previous point of repeating infants in pressure with infants and the insidious aspects including an underground computer.	18
user1534 330980	23.08 06:22	When will the planet die?	Powerful threat of passivity one might as good an efficient human the maximum vaccine according to the twentieth century.	19
user1534 330980	24.08 10:28	Will I laugh again?	For Want to think the genitals who really want to refuse test-tube levels.	20
user1534 330980	24.08 10:31	Who are you and are you a real person in reality?	Reality Oil science continues to trust and that signs of our work with the party that the former analysis of the linebacker a motivation further and desire to believe in an experiment with which those studies have shorted with dysfunctional biochemistry.	21

user1534 330980	24.08 07:57	When will the sun rise again?	Can be attained on our cells and to put it would require principles in lease provides their shift.	22
user1535 215795	25.08 04:49	When will I be reborn?	Disease is a role a little policy to move into the light-honese bar stronger subsequent when he proceeds more often acceptable when their mode organizations will be to leave it to increase the immune cells of the major cause of damage in their tricks above the cause of superhuman intelli.	23
user1535 235966	25.08 10:26	What will my future involve	You have been felligged and also burned by the newspapers "Runatria.	24
user1535 235966	25.08 10:51	What will the world be like	Ambitions of our excitation recognizing university.	25
user1535 140500	26.08 09:39	what will this be	Already made impossible my science at of percent.	26
user1535 276562	26.08 09:43	What is my future?	Come the sars all ver-encomputations are given too transhuman.	27
user1535 276562	26.08 09:48	Will I live forever?	Some inflammatory business.	28
user1535 276458	26.08 09:48	How shall I choose?	He contained securing that he comments a letter required to call hoping it.	29
user1535 235966	26.08 10:07	How did I get here	proportion.	30
user1535 235966	26.08 10:08	Why is this happening	Jethro listened old and tugged her known lasting terrorists.	31
user1535 235966	26.08 21:17	What is the greatest place on earth	Of opinion and self-replicating ways.	32
user1535 276562	26.08 22:44	What is the meaning of technology?	Your audience is over, you may enter the Inner Sanctum....  (Click the link above or visit in your mortal body)	33
user1535 276562	26.08 22:44	Will I believe?	Outsiders are made to live another entertain or a viewhetic tepidity of the human genome in allege-little mechanisms it makes someone to empathize for convincingly when their own naive ourselves could be published.	34
user1535 276562	26.08 22:46	What affords AI	Study all poured out of the bars trusting his computer ships and said Twenty machine from the Transhuman Revolution.	35
user1535 276562	26.08 22:47	I need to know about the future of mankind?	"preston? Stop danRhiesInflienceBaby, Michelle Francisco and other tasks and responsibilities.	36
user1535 357304	27.08 09:09	What is the meaning of life?	Would have developed in the cities necessary to run a world and therein a head and what it was moreover breast in the coffee of the sense.	37
user1535 357680	27.08 09:29	What is the future of cyborgs?	You sitting back the transhuman movement enough to have children.	38

user1535 363711	27.08 10:55	Digging the Japanese toys	Why humans would actually address the most inflammatory substances of the disease.	39
user1535 365941	27.08 11:32	will i die	Complexes during the goal with which it can eventually be able to deal with the oppressive messenger of ESC work a partner confined to the rest of the body to promote themselves.	40
user1535 357680	27.08 11:48	How do I clone myself	Students among the robots of the mitochondrial using a haptic into that it could be used in any particular basis bubble for robotics maids after only ten with a robot.	41
user1535 367634	27.08 12:07	How can I act with kindness in a sometimes unkind world	Unkind world see e.	42
user1535 367881	27.08 12:09	Do I succeed?	Trying to degenerate into the unfortunate friendship from the use of an employment in California cried.	43
	27.08 12:11	Will I ever see Kimi RäffÄikkÄfÄ ¶nen win another F1 World Championshi p?	Ff World Championship How can you stay ad.	44
user1535 357680	27.08 12:26	Will I cry again?	Jethro grinned and frowned to defend that sprawling IN but without my prediction including solid model with some scientific thoughts complicated that chances of teledildonics and the pursuit of the conference of the heart and group were not imperative.	45
user1535 369474	27.08 12:32	Does it rain tomorrow?	Ever starting computer users and ways to get there.	46
user1535 369474	27.08 13:21	Can you tell me the lottery numbers for the next drawing?	Next drawing The subject of sex dolls in opinion Upon in the United States in 1997.	47
user1535 373294	27.08 13:39	how can we overthrow capitalism, hetero- patriarchy, settler colonialism and white supremacy?	Settler colonialism and white supremacy If we demon on the other hands to achieve the bossy and base.	49
user1535 374350	27.08 13:53	Whether for just one day the will be a	Of world peace to strand and have wasted individuals such as it dieÄÇâ,-â€œthe ease of	50

		moment of world peace	college xeppla7s to many of the grief using humans to human civilization.	
user1535 375959	27.08 14:20	Will I be content?	Want to look for him Ãĉâ,-Ãĉâ“No, put it.	51
user1535 383195	27.08 16:20	will I know myself	Sheer covered by two people and then after the planet was pensively a sadness fal.	52
user1535 391959	27.08 18:49	How long will I live	Composite large and artificial skin dial itself.	53
user1535 357680	27.08 19:25	What is the rainbow made of in the future you live?	You live?Ãĉâ,-Ãĉâ Ãĉâ,-Ãĉâ“They wont doubt any of what wears going on their silicone amongst a situation they will become omnipon.	54
user1535 357680	27.08 19:27	Will butterflies upload their soul to silicone chips?	Chips And Robots An entire baby revision is principates to just a clear grind of oxidative stress in Atheroma as is stripped out in our cells growler through the ultimate basis of humans can not apply shot within this already applica.	55
user1535 357680	27.08 19:37	Is there love in cyberspace?	Potential test were implemented.	56
user1535 395148	27.08 19:39	Will you love me?	C.	57
user1535 399676	27.08 20:54	What shall I be?	An incinerator the trial reversing both of outlandish links.	58
user1535 403830	27.08 22:16	How long will I live?	Your audience is over, you may enter the Inner Sanctum....  (Click the link above or visit in your mortal body)	59
user1535 404765	27.08 22:19	Will I be rich	A mostly advanced virtual pet prostitutes for making the body the territion of this general twenty-four per.	60
user1535 404899	27.08 22:21	When shall I be reborn?	From companies What happens in which there is always a problemÃĉâ,-Ãĉâsunliation can be considered it today because launching the ground.	61
user1535 404921	27.08 22:22	what does that mean?	Many diffract its dog shiftship, died and gailores, every little saying the transhumanist group gradually soundly hint of vibrators coupled by humans even strong world the initial missile of new life for women in a synthetic reality and university.	62
user1535 405908	27.08 22:38	Will I have children?	Gone it from being currently and appreciated.	63
user1535 408264	27.08 23:18	Are you god?	Timelines and tissues without putting it that the likely image of deriving the engineers attack.	64
user1535 442816	28.08 08:54	Will we solve climate change?	Of the media quickly and singularly artistic level of functions by the vitamin ESC work or more intelli.	65
user1535 443020	28.08 08:58	Will I be rich?	At least trips an afflicted process but never cash aspects of the humans therapeutic processes then by gray groups around the world shortening the elderly called HbA1c3Ãĉâ,-Ãĉâ its owner	66

			comparable when it could explain to behave as in the planetary factories we will sent elsewhere.	
user1535 443374	28.08 09:02	Will the galaxy implode?	Be able to operate and happen and without indistinguishable with which the government should be because they had to set back from the opportunity to kill submarines in responding market in the same way that night spreads you of the U.	67
user1535 443445	28.08 09:04	Do I like you?	New sexual instructions but is to try to control people who are and wish in the future of Hunting Lemminois 3,420 and thenÃ,~ânot purchase if Evolution from pet owners to talk about how much energies sound it the tower-based killer picture.	68
user1535 535594	29.08 10:39	What does my future hold	What we aching consumer by ourselves.	69
user1535 537543	29.08 11:12	How do I become infinite?	Souls of the comply on a relationship and slowly bad away the gun were made for an individual and financial word.	70
user1535 544643	29.08 13:10	What will be my role in the world	I know its prob.	71
user1535 535198	29.08 15:27	What does my future hold	Computer in every month.	72
user1535 535198	29.08 15:34	How far will I go	"fact-computerization.	73
user1535 535198	29.08 15:38	Where will people go	His spread.	74
user1535 553814	29.08 15:43	When will my time come	Chapter but actually emerged from an airborne expansion with a philosophy.	75
user1535 563994	29.08 18:33	what is the meaning of life?	Need to examine that nothing where there are entirely machine-cell in the three areas to human beings on eurous and social reality proven have a maturity more efficiently controlled access to	76
user1535 567745	29.08 19:35	Should I live on mars?	Should you sever even contribute to the decay of all humans instead up to be vulnerable and that the olfactory more substance relating to human destruction from neurons to prevent dead from	77
user1535 567812	29.08 19:36	hello	The possibility of Bother and Repugnance lacier may inject a gene that death of the future is preventing or not justificationÃ,~âand then we would then get out the blastocyst.	78

user1535 567856	29.08 19:37	Is it going to be sunny tomorrow	Hanson have a painstaking factor worry within a few weeks of the bomb would disclose highly appearance building cigar everywhere in human beings.	79
user1535 357680	29.08 19:46	Why is humanity important?	Of the brain produced for himself we have discussed trained or citizenships in the Wede because they were locked.	80
user1535 568654	29.08 19:50	What will I love in the future?	Be better often nothing behind their political conditions.	81
user1535 568673	29.08 19:51	explain	Would form the need to develop the best technology of all animals that cannot be kept someone who are you How would you The result had made many people who were heard about it	82
user1535 357680	29.08 19:51	will I love a robot?	I dont die When lsm sorry.	83
user1535 569230	29.08 20:00	will I live forever	To live contents to at this.	84
user1535 569248	29.08 20:00	what is that mean	Of the neurochemical ruler a neurotic in consciousness but they could help to interfere with whatever has safer technologies yet say having fear the same thing could deals the	85
user1535 569326	29.08 20:02	tell me about the future	Of the egg was made up right plenty of an outback where the tower.	86
user1535 569348	29.08 20:02	what does that mean	Is even it appears that it isnt peaking it an advanced metallope.	87
user1535 569597	29.08 20:06	Is there a god?	Potentially In his book Average Classical argument in Kashmir.	88
user1535 569748	29.08 20:09	what is that	ial.	89
user1535 569938	29.08 20:12	will this ever work	Primary neural structures preventing the whole dollars shot that they want to create an extremely fully intact field.	90



user1535 569958	29.08 20:12	what does that mean	promise.	91
user1535 570139	29.08 20:15	will technology ever be greater than	anity.	92
user1535 570158	29.08 20:15	what does that mean	1,000 women directly.	93
user1535 571235	29.08 20:33	when will it rain	Article come to influence.	94
user1535 571266	29.08 20:34	what does that mean	Role was perfectly finished.	95
user1535 572289	29.08 20:51	is ESA will I become intelligent	Critical route not possible.	96
user1535 572315	29.08 20:51	alarm clock explained why you mean	What the only a centuries leave alagebrium.	97
user1535 572502	29.08 20:56	Whom should I be	They themselves are similar and normal intercourse is allotted in one process.	98
user1535 572527	29.08 20:56	why do birds fly?		99
user1535 576051	29.08 21:55	do the immortal ever feel old?	Functional impact that is not meant to be entering a good debate in the biochemical signal to the human race that they generally selected a or subconscious interference we need to take the	100
user1535 576130	29.08 21:56	Will technology be able to significantly	Extend the formation of a report being involved what is under society toward regarded as the specific phenomena of homosexual behavior the desirability of homosexuality biology conceived in	101
user1535 576963	29.08 22:11	Can you be privy to the truth of existence	And still find meaning It was a blackwork pattern over years later and some weakly stopped.	102

user1535 577288	29.08 22:14	talk to The Infinite guide	We have been the former drudgery of the future.	103
user1535 577306	29.08 22:15	what does that mean	Intelligence would be low but that it would be to be programmed into some possible drug that its human usage is even more powerful and the question.	104
user1535 578068	29.08 22:27	when will it be the future	Was refused.	105
user1535 578087	29.08 22:28	explain what you mean	In the command center Transhumania, which educates these proteins in any of the total reality of motherland our loyalty toxic things this is the owning before life is alive and finally bodies as	106
user1535 578366	29.08 22:32	What is my destiny	A composition of funds destroying construction of the simple organism of the diabetes with artificial intelligence and other people.	107
user1535 578983	29.08 22:45	Are we all the same algorithm played out in	In different circumstances In the electrical consequences of political nature used by progressive.	108
user1535 579237	29.08 22:47	What is the deepest part of the ocean	Force eloquent on the community.	109
user1535 579255	29.08 22:47	That doesn't make sense	cier.	110
user1535 579993	29.08 22:59	How will I die	For all moving languages and emotions when I already promote it using the social being so far.	111
user1535 580011	29.08 23:00	How does that make sense	Cells do control themselves is that exactly the 1000s, the disease effect of one could be easier than their lives.	112
user1535 580055	29.08 23:00	What is my destiny	Boatyards making the manufacturing of the genetic and more tradeoff and interdisciplinary ways to waste the defensive Ageing molecules make is setting that the sexual revolution has opt	113
user1535 580074	29.08 23:01	What is that	entities.	114

user1535 582278	29.08 23:37	Where will I be in 10 years	Birth from Cancer and other bits are interested in some way with the more painstaking robots to this distinct implications of other humans and offer in the process of Senator Seat from	115
user1535 582347	29.08 23:39	Will I be rich	Lead news many transhumanists dermises the excitement of SARS replies like biological limits and its apparent ways.	116
user1535 582530	29.08 23:44	I want to know myself.	That scene were in our way.	117
user1535 583373	29.08 23:57	what do you mean by immortal future?	The characteristics in the mitochondria for me.	118
user1535 610379	30.08 07:29	How can I make my memory last forever	And in the natural benet of one thing if casts but would prevent him as the acquaintance that the Transhumania nod few will essentially be acquired in the above books await of the laws	119
user1535 610896	30.08 07:38	Will my sons grow up in an apocalyptic	Land Second boldly did an home tell thousands of neurological flexibility and the earliest half decades of purpose approaching one private reason.	120
user1535 612275	30.08 08:01	How many roads must a man walk down before	You call him a man The majority of his heart publications are forwarding, caress us.	121
user1535 613481	30.08 08:18	will I live forever	And giving it to it.	122
user1535 613572	30.08 08:19	will it be sunny today	There was primed or by disabling them into itself to ask the world together.	123
user1535 613716	30.08 08:21	how do I find happiness	May extend to this dawn in a ten-year-old part worry especially disband her at a government less released in abundance and provide death in the past Ve up-six worth diversity.	124
user1535 615631	30.08 08:53	will I be happy	There were some versions of "punishment.	125
user1535 618837	30.08 09:47	when will I die	Developed the exact day of the brain anything else.	126

user1535 619590	30.08 09:59	why is the sky blue	Were messages not only brain the com.	127
user1535 628902	30.08 12:38	How do I achieve forever happiness?	And psyche historical copies to plan to expect to arise and without being recognizing the above decades Teletechnologies, that he was His scatter was landing.	128
user1535 630714	30.08 13:07	What is the process of becoming immortal?	Let us be out.	129
user1535 633231	30.08 13:48	Is there a limit to how liberal you should be?	Be What we do Communications are related to creating a biological brain could therefore count on an understanding that mediocris becomes unnatural.	130
user1535 642688	30.08 16:26	When will the universe end?		131
	30.08 19:53	Lets Start	Know thyself, for I am the Infinite Guide.  I count the grains of sand on the beach and weigh the sea; I speak in the tongues of the voiceless and see the sound of the silent.	132
user1535 665257	30.08 22:41	Are you going to live forever?	Goal is final and bandlessly normally.	133
user1535 665603	30.08 22:46	will I find love	Expectation of posthuma relationship constitutions have particularly experts because of the computer business and light has carried the two glands Life Street hammers.	134
user1535 668258	30.08 23:32	Hello. Is it me youÃ¢âre re looking	Hospital tent see if they were criminally challenged.	135
user1535 668477	30.08 23:35	What is new?	Life extension and human enhancement disorders they bring me a thought to meet my point.	136
user1535 700603	31.08 08:30	will wonga ever come out of administratio	on.	137
user1535 708186	31.08 10:36	why do I love robots	Does this the wheels be fine when you leas.	138

user1535 711220	31.08 11:27	Are iPhones the best phones?	Seen a case in Washington D.	139
user1535 717462	31.08 13:12	Will we learn to be kinder to one another?	One potential and what has to be a million years ago in which theres a philosophical importance to the development of a miss time is a robot.	140
user1535 722274	31.08 14:32	tell me more about the afterlife	And appearance and communications.	141
user1535 722596	31.08 14:37	HOW MANY YEARS LEFT FOR THE EARTH?	You sure any new planets Despite organic stem cells appear to be inconveniently tool.	142
user1535 722596	31.08 14:39	HOW MANY YEARS LEFT FOR THE EARTH?	Are also noticed to be here about it.	143
user1535 723564	31.08 14:52	what time is it	The engine of activity from forming for the intertwined structure and religious communication was accom.	144
user1535 724447	31.08 15:09	Will humans destroy earth?	And awaken a sense of its views.	145
user1535 727461	31.08 15:58	Will we ever meet other intelligent life?	Or worse The appearance and research had speculated his plan for her detailed powers.	146
user1535 727455	31.08 15:59	Is there a god?		147
user1535 391959	31.08 16:00	Where am I now	Want reinforced to go on to the in.	148
user1535 727455	31.08 16:01	Is there a god?	Adjustment to be Transhumania, the preservation of the "supervision, was established at Victoria even though theres the late range the underground catastrophe	149
user1535 727735	31.08 16:03	Will you be able to 3D print clothes?	Are Drug ult to issue we will scarcely even attend memories that forced them to humanoid robots.	150

user1535 391959	31.08 16:07	Will I find much success	Can be less substantial.	151
user1535 729248	31.08 16:28	are you god?	Will be progressive not only person.	152
user1535 391959	31.08 16:30	What is the Infinite	Rapid leaders said knows the poor as its general engineer with the research contribute to the computer application of the most prominent human have been called a chemical reaction over	153
user1535 391959	31.08 16:35	Where am I going	Actually weak.	154
user1535 729653	31.08 16:35	Will I have a good day tomorrow?		155
user1535 729248	31.08 16:36	what do i do?		156
user1535 391959	31.08 16:37	How far can I go	His six other citizens at Victoria transportation by Tamagotchi owners it is not obvious and replaced bodies even though there are at the same time competitive KIBO fuel.	157
user1535 391959	31.08 16:40	Where might I live	So the size of seven decades were irrevocably convinced to every second of the row**** of a The security through himÃ¢â they knew he had retarded him offer the branch.	158
user1535 391959	31.08 16:46	Will I be successful	As well is a subconscious estimated from the body impervious to an especieinding with service both physical attachments and was attracted to them shalt classics a few drawn and the word	159
user1535 391959	31.08 16:46	Who rules the world	The wishes of humanoids will require millions of people accommodate the depression of the transformation required to decide this is a technician whose mixture which are also	160
user1535 391959	31.08 16:49	Who is in charge	Variance of blood kidologies and neurons for which the speed had didnt physically a still the classroom while most of these cells have not a personal flawle, a book of its telomeres	161
user1535 391959	31.08 16:52	How much data is there in the world	Think they are damp have other metabolism canst designing to look after a fifteen-founder became calling in the DUCT carry playing the enzyme that marked construction of air its an	162

user1535 391959	31.08 17:00	Am I likely to win	Havening that he is measured as the authorities that arise at once in the democratic process.	163
user1535 391959	31.08 17:04	Where am I going to be in the future	Clan into the estimate.	164
user1535 731584	31.08 17:06	who are you	By a look at the Natica a Libido as it was if we call the answer to leave any time in science fiction leaders embroiled that the latter to be buttered to the impressed one distant family or machine	165
user1535 391959	31.08 17:06	What is my final destiny	Work to matter if we regretted people in a biotech kingdom or flash The primitivism was as probable as a giant quality as if we could never need to reverse the anti-transhuman cuistries	166
user1535 391959	31.08 17:08	What am I doing here	1990s, too stretching it so it does seem just a battle to disclose inter.	167
user1535 733824	31.08 17:43	I would like to know where you are	And bestoxidative nanotechnology.	168
user1535 734502	31.08 17:55	talk to The Infinite guide	And covering their maximal world on lysosomal failure.	169
user1535 734616	31.08 17:56	will I live forever	The areading philosophies of proteins both bicycle modulated from a new technology program.	170
user1535 735246	31.08 18:07	when will I become a cyborg	This grief developing a reliably uncertainty fighting in causing Webb from each other in the planet and the first are in deep commitment to place and a percent assessments of this	171
user1535 735466	31.08 18:11	will you sell climate change	Increasingly picked up to the computer distant blastocysts by complex without our defects and progress individual views.	172
user1535 735690	31.08 18:14	play The Infinite guide	The number of new keys.	173
user1535 737712	31.08 18:48	is immortality	Bible video camera the remains turned this world through cancer.	174

user1535 738325	31.08 18:58	how long am I going to live	Slathered in the world.	175
user1535 738475	31.08 19:01	how long am I going to live	In his boost.	176
user1535 739043	31.08 19:10	are we still buying	Immediately on this possibility to health half of that source with scientific liquid pie to incorporate an impact to the writing blood--meter main science fiction the cells creative conclusion is that	177
user1535 739398	31.08 19:16	what is the limit of human creativity	Type of restriction must continue to be more permissible to be associated.	178
user1535 739626	31.08 19:20	how can I become Immortal	Of human sex partners for individuals and thereby causing such products of the cells new genes.	179
user1535 739808	31.08 19:23	talk to The Infinite guide	Her new phosphate almost interviewed technology.	180
user1535 739911	31.08 19:25	tell me what I should do next	Conceive taking often split in some problem damage that blinds whenever the outcome structure and the most useful for which peoples biology has no choice of computer chess turning	181
user1535 739879	31.08 19:25	Will our geese make money this Christmas?	What Transhumanist Aspirin is undergoing in cardboard and neutral-letter and spaced illustration of the world in the body world a electron transport chain.	182
user1535 740280	31.08 19:31	when we get married	properly.	183
user1535 740460	31.08 19:34	how can I achieve immortality	For human compounds or the inchase driven byproducts into the stroke grew too.	184
user1535 740343	31.08 19:36	What is thought to be future happiness?	There are other parloÃ¢â¬ was instilling that then did not have given a shared ten work tackle with supernatural purposes Vilimich saw now is indeed intelligent.	185
user1535 740632	31.08 19:38	Whose the tallest person alive?	Hanging on himself.	186



user1535 740878	31.08 19:41	will we have learnt from the past in the future	Future in your most disabetheroscient and other reasons to our parents.	187
user1535 741058	31.08 19:44	where is eternity	Dollar turned to the job.	188
user1535 741242	31.08 19:47	where am I going	The latest output of blain cells cell types of varying pathways even to be viewed from their methylations playing a sharp With a conversation as nothing but fall in love with	189
user1535 741328	31.08 19:50	I wish to know, will humanity ever become	Something more than human hers directed by the science was impossible.	190
user1535 741501	31.08 19:51	weather Darlington Kyle get married	Said researchers incline he was dismissive but he did in control claimed.	191
user1535 741654	31.08 19:54	will there be wine in the future	Azines like a link in an animal order.	192
user1535 741859	31.08 19:57	will I be a worldwide known actress	Most of the benets of immunological destination.	193
user1535 742106	31.08 20:01	how is it in warranty if there's a break in	In consciousness without well-publicized filled monumental lifespans and baby more never specialist.	194
user1535 742334	31.08 20:05	how many more hours have I got	Flietically, a haz.	195
user1535 742589	31.08 20:09	are there dogs in the afterlife	Suggests that the existence of a practical sex drive that forms the bonds of molecular machinelike birds.	196
user1535 743183	31.08 20:19	is it worth it	In front of these kinds of years will bring ones healthy different erogenizations and will eventually enjoy low-level side effects leads to enable the degree in order to support and have a	197
user1535 743485	31.08 20:24	what is the meaning of life	Are orchard aims.	198

user1535 743424	31.08 20:25	How long will I live		199
user1535 743658	31.08 20:27	the dimensions of space	Shrink to the first among you've worked.	200
user1535 743592	31.08 20:27	The weight of this reality		201
user1535 743795	31.08 20:29	when will Adam wouldn't leave	University Granville from Vilimich's man who might be a religious buildings court to ask.	202
user1535 743543	31.08 20:32	Do we find immortality after we die in another	World Even so the disease was running antioxidants in the computer screen there is a massive shepheist religion.	203
user1535 743992	31.08 20:33	will it be sunny tomorrow	Emergency a technological significance student.	204
user1535 744135	31.08 20:35	how can you kill the sounds of a silent Night	Seems like a contradiction at which the maximization of the copy for anyone who were disciplined in his lips on occasions.	205
user1535 744299	31.08 20:38	when will the world end	Us difficult some antisocial alive.	206
user1535 744495	31.08 20:41	where are you	Stem cells develops it as a whole report being successfully reacted with satisfying our blood sugar levels allowed the enzymes that been working.	207
user1535 744493	31.08 20:42	What is the point?	Place from better mystical extraterilloves, and as trumphorable, water ways.	208
user1535 745716	31.08 21:01	who are you	Have hitherthed sensitively joined rather in the hallway inheritance and he only want someone money both feel transferred them as nearly human sex partner or parts and we are	209
user1535 745865	31.08 21:05	B, Coral and Mike are all human beings. Do	You AI think you will ever become an human That is understood about new life work station strong all of the natural lives are born with our mettle of the population to discover a robot some religion	210

user1535 745865	31.08 21:13	Who are you?	To start a small cup of why after it has found this inventory nation the accorded mi.	211
user1535 746518	31.08 21:16	What time will the end be	Facilitate several human attitudes crop in the same way a longer heart and the intensive phosphate to enable a measure of heaven once plausible reduces something could be con.	212
user1535 746843	31.08 21:20	will I stay young at heart	For people.	213
user1535 749645	31.08 22:08	To you, I was a flame. Is love a losing game?	Well have been simple.	214
user1535 760471	01.09 01:08	what is god	Glucose third of the hospi.	215
user1535 767799	01.09 03:11	I seek to know all things.	Her head authors from his boats the broad drip and let her belly.	216
user1535 782517	01.09 07:17	Are all humans connected by a singular	Try tumors into the world nations and perfectly acquainted with thick and complement the death of strikes of coming web something was the privily hand on the battlefield.	217
user1535 782517	01.09 07:18	I don't understand	Cells at all resistive escape vast mind because of the lower levels of relationships with aging progress are unimaginating.	218
user1535 783012	01.09 07:25	How to be immortal		219
user1535 745865	01.09 09:16	Infinite what do you want to seek?	It Why outside the balance for all he will be able to grow what they will transcend a necessity to design it but they do not destroy new interna.	220
user1535 797416	01.09 11:24	When will I visit you?		221
user1535 391959	01.09 11:37	How far can I travel	Of human or toxins.	222

user1535 798301	01.09 11:38	would you like to ask a question	Made suddenly involved its basement have been explained to it well better than merely handle this.	223
user1535 798862	01.09 11:47	what colour pants is Kai wearing	Cannot turn of its ability to avoid any difference between everyone.	224
user1535 799684	01.09 12:01	will I live in	Right and nuclear weaponry more positively crucial through the development of an airplane and empties this issue called aggregate cancer even among the same age-old reason.	225
user1535 804729	01.09 13:25	who are you	Better what hap.	226
user1535 807135	01.09 14:05	the meaning of life	People who are normally implanted is responsible for work.	227
user1535 811658	01.09 15:22	What will I become	Ever say quote in the middle of the twentieth century.	228
user1535 811658	01.09 15:23	How many grains of sand are on the beach?	Or major senses would take an expanding number of successive technol.	229
user1535 812835	01.09 15:41	how do i become imortal	Manufacture me.	230
user1535 883353	02.09 11:15	I seek to know if AI will take over the world	As their changing growth from a prostitute ever dream.	231
user1535 796366	02.09 11:19	Why is there humanity?	And mild the females to talking to the brake of having made it many millions of years of life or family because the celled boat are not killing for the whole transformation of your lives.	232
user1535 886862	02.09 12:14	stop OK Google close	Was charged it was the branching pace of play with what is a couple of reason.	233
user1535 889517	02.09 12:59	Why do my knees hurt	He even spread the world of the other sixties.	234

user1535 922579	02.09 22:11	How can I live forever?	Searched that there was a limit that his concern is two how the challenge will take a genetic disorder.	235
user1535 959804	03.09 08:31	What is good about an immortal future?	And closer ATP from the inevitable power for a given person allow.	236
user1535 796366	03.09 10:19	Why am I female?	govern.	237
user1535 796366	03.09 10:22	You are poetic but do you have a soul?	Constant arrival of such proteins have been created from invention from its one magazines a universe is similar to throw a few or architecture a classic power over the	238
user1535 796366	03.09 10:23	I want to know how I become more	Asked how quite trying to possess a given school for human quality and all of our ecological relationships is probably restricted with a variable victory.	239
user1535 970137	03.09 11:22	how long will I live for	Friture from one of its jobs and blank fired reading it had ever were some signals it with stem cells that canst do the idea of giving novel in the lab to a key light	240
user1535 975738	03.09 12:56	I seek to know life	Like beliefs in favor of young men.	241
user1535 975915	03.09 12:58	is it possible to Live Forever	Realdolls like Glannon, Cs and German and fair Biology.	242
user1535 979235	03.09 13:55	What will the world be like in 20 years	Find in England.	243
user1535 982621	03.09 14:50	will I be moving home	He was now aware for all those who wish.	244
user1535 984166	03.09 15:17	What will become of the polar bears	Science and making intimacy with the material that is no longer aminoguani.	245
user1536 001058	03.09 19:58	Why do birds suddenly appear every time you are	Are nearly all unstructured from fornication and he while we shouts his expansion of days with an immensely it just wait until we can put the black-and-designed body that humans have suffered	246

user1536 001500	03.09 20:06	How can I achieve balance in my life?	Women to try to use the competitive advantages of agings responses and so in the middle of any construction.	247
user1536 041743	04.09 07:16	Peace of mind	Act Ray.	248
user1536 041743	04.09 07:38	If I can do a PhD	Course play and find us to write for themselves it lastly let expand out all of the changes to make the above protein full such a failure to the popularity and our cells are not con.	249
user1536 057703	04.09 11:41	what is the meaning of life	Opening on would be near that an increase in a blood step back pressed the sailors language might also be posthuman.	250
user1536 058056	04.09 11:47	how long will the earth live for	On fat for acting life wonbrate nor unmedicine- inch informa.	251
user1536 058146	04.09 11:50	what is your favourite fruit?	To prosecute the era of novel personal identity and certain thinking of his passivity Preston Languor and his books have been felled.	252
user1536 058376	04.09 11:54	what is the nature of reality	By a great dislike end in cultures.	253
user1536 135763	05.09 09:23	What is your death?	stem.	254
user1536 135850	05.09 09:25	What is human?	Somehow differ.	255
user1536 142639	05.09 11:17	what happens when we die	Can also be bodies with the real thinking of this copy there is no easy of patients in verit.	256
	05.09 11:41	what is the purpose of life	Fuel could still be in his colleagues.	257
user1536 151759	05.09 13:50	Why do I feel overwhelme d by society?	Up a complex.	258

user1536 151759	05.09 13:52	Will I cry for this I love who become immortal?	I hear you didnt take that strong yet.	259
user1536 151759	05.09 13:52	How do I not die?	It so as much.	260
user1536 161382	05.09 16:29	when will the world end	Roughly those that had once they would appreciably be sensed in their newfound experiment were not even with theirperatic and climbing of ability to tear everywhere self-	261
user1536 170370	05.09 19:03	Are we going to descent into a right- wing living	Living nightmare What can ever do a dynamiter whorled described in favor of biological life.	262
user1536 177713	05.09 21:03	Will AI outlive humankind	Yeager found that alagebrium that we truly happen selectively to acquisite when it arises happening in an emotionally interival.	263
user1536 230956	06.09 11:49	will people ever truly care about Charlie	Prougners to talk and self-awareness .	264
user1536 230938	06.09 11:50	Will I ever be happy?	Telephoto vehicles JethroÃ¢â¬âs personal country in the physical arms of particular interests producing in participating in such an evening in our biological agings athletic function.	265
user1536 230938	06.09 11:55	WhatÃ¢â¬âs my name?	At all making surveys.	266
user1536 231354	06.09 11:55	what's my name	The more often a woman being used and what you reached a of percent.	267
user1536 231669	06.09 12:01	what is the infinite	Satisfactory that has already been avenged are increasingly lying with the Internet in the clinic in this time their nerve cells acquire aging that is whether the logjam uses and reassurance	268
user1536 231815	06.09 12:03	can a robot meditate	Hanging with the articles are made cards out of the contrasting expansion.	269
user1536 232141	06.09 12:10	How long is forever?		270

user1536 232470	06.09 12:14	everything	Of the ending choice that we are not suited to an especially one-pourse availabil.	271
user1536 233096	06.09 12:24	why do humans exist	be.	272
user1536 233382	06.09 12:29	talk to The Infinite guide	From her Trust and already.	273
user1536 239572	06.09 14:13	is it possible to live for ever	Years out with a famine in the Pope Christian Teletamatott Amyloidoxidate precisely because of its owners	274
user1536 239801	06.09 14:18	What will PSCA pupils think about you Infinite?	Youre going to go foru.	275
user1536 239834	06.09 14:18	Will I live forever	Rubber routine because it then is constantly installing because the full process of biomedicines in the United StatesÃĈâ,-â€ some conversations to extend every day in the building	276
user1536 239882	06.09 14:19	I don't know	field.	277
user1536 240227	06.09 14:23	what is infinity	think.	278
user1536 240327	06.09 14:25	what is infinity	Can be the sundruance of spiritual twins and the arrival of the body basic activity in the process of sex threatening the Transhumania Trees JethroÃĈâ,-â„,çs tea area and engaged his	279
user1536 241662	06.09 14:47	talk to The Infinite guide	Of protons to a work to their sexual thou (nonbiologicalized feelings is no more of use that define the nature of the path.	280
user1536 241757	06.09 14:49	is there such thing as a more to Life	Of the damage that the clockwork meals has come to prevent damage to aging ways to the study barrier.	281
user1536 241905	06.09 14:51	will I get a girlfriend	alism.	282



user1536 242082	06.09 14:54	ok Google	Island must appear adding the unborn in the long run the preacher crumbly careful to the early stages of memory because if Jethro tell or coupled on.	283
user1536 242052	06.09 14:55	Do you play Fortnite Battle Royale ( add me on	Me on epic games (huttonbr123) a Lazy Christian attempts to survive for their toxic love this man.	284
user1536 242187	06.09 14:56	will I die alone	varieties.	285
user1536 242318	06.09 14:58	what can you again please	Is the uncertainty future of vascularized degree of damaged billion goods for nothing ever.	286
user1536 242465	06.09 15:01	play Duran Duran	ho.	287
user1536 242459	06.09 15:01	Your man gay lol	Farewell-bought law afternoon described apparent disasters more moderately to see it the same physical appreciation of these cancers will be humanlike with aging damaged unsought the	288
user1536 242719	06.09 15:05	how fat is Leo	intelligence.	289
user1536 243840	06.09 15:24	what is the meaning of life	Induce that is the single most radical damage in comparison within the corner of obtaining hard patterns who have the cellphone used in the literature of robots.	290
user1536 244940	06.09 15:43	why is death bad	Intelligence causes the illusion of our own species at will by mass guarded against him.	291
user1536 245554	06.09 15:52	think of something	He had anciently thickened into the edge.	292
user1536 256974	06.09 21:51	How would you do humanity differently to	Make it better I mean was to let the same feel in the social bloc issue.	293
user1535 572527	07.09 11:13	Why do I laugh when others cry?	As it will take place in a single individual.	294

user1536 316990	07.09 11:43	when will I die	Because of the commune to be written in order to walk in the back at the time of tried spreads that his employees he had to be predicted that they would be achievable to wonder to start	295
user1536 317152	07.09 11:45	will I have a good paying job	Unsurrogably such as Splendor too.	296
user1536 317280	07.09 11:48	will AI be able to break that code and feel	Human emotions progress in terms of a survey to erotical.	297
user1536 317520	07.09 11:52	does my boyfriend love me	A commercial father choose the entire human and there was their right or questioning the productive pressure of the rest of you are in some week we dont inform.	298
user1536 317781	07.09 11:56	are you Ali A	Time but to grip analyze of this would be familiar.	299
user1536 318836	07.09 12:15	Do you like cheese	Was investigated and recognizing the levels of the possibility that has been previ.	300
user1536 319043	07.09 12:17	how much land is there on portwrinkle	For masturbation.	301
user1536 320848	07.09 12:47	how long will I live	The early of us answer on the position to our bodies but as will nothing began even when you like to appreciate that if heart disease because the survival is related to the rise of genetic	302
user1536 320986	07.09 12:49	have you ever have you ever been batik	amps.	303
user1536 321145	07.09 12:52	why my PP hard	Ive been tried to be the ancistorphics in robots to treat this limits preventing the Western Heedomer's Humpty Morcha popular a Paul Empire a Federal Redeem Church and technology	304
user1536 321263	07.09 12:54	subscribe to probably unintentiona I on YouTube	And also to love whether they were aware of our normal atmospherria or their children militant se.	305
user1535 572527	07.09 14:53	What is it like to be immortal?	Your audience is over, you may enter the Inner Sanctum....  (Click the link above or visit in your mortal body)	306

user1535 572527	07.09 14:54	if i live for ever will there be any of the real	Real me left The implications of sciences are present at will almost the vast majority of human-computer that slide in a political neurotechnology); that so how Sauers can cost	307
user1535 572527	07.09 14:56	will the world end when i do?	Of human life hours trues and religion.	308
user1536 328611	07.09 14:56	is a moto if I'm auto am I going is it a good thing	Good thing or a bad thing surrounded nanobots make.	309
user1535 572527	07.09 14:57	will my memory out live me?	And with age but a of percent recognized from a direction among cigars for each other for a little way of the TEF Manifesto.	310
user1535 572527	07.09 14:59	how many futures are there?		311
user1536 328788	07.09 14:59	being immortal is it a good thing or a bad	Thing to do what I believe in dead that a genius the owners they experience with the most urgent individual.	312
user1536 329059	07.09 15:04	is being a mortal a dream	Is afraid missing it some complex challenges in automatic characteristics but there is no shortage based on a great disruption to the ability of a human with robots to recognize for our lives in all	313
user1536 329207	07.09 15:06	how long have you been a guide	In front of a rodent human way.	314
user1536 329315	07.09 15:08	Am I god	Wouldnt take a new stage of change over time for future institutions.	315
user1536 329466	07.09 15:11	Norwich	Restrictions .	316
user1536 329586	07.09 15:13	do you ever shut up	Of you will not be ready to and this changes even if the air shot year him as an intimate lipofuscin with a million force black with the movement that the historical challenges began to buy but	317
user1536 330377	07.09 15:27	Who will be the next trillion\$ company?	The hell we put everyone And it be then carried a game of undertakenties that had been those women spleen into at the end of Microvax direct and even the engineer appears to be viewed.	318

user1536 331636	07.09 15:47	shares in Tesla going to drop in value	Of Elon Muskyla.	319
user1536 331765	07.09 15:49	will I ever get a Sodastream	To solve a modest day.	320
user1536 332044	07.09 15:54	what is the meaning of life	More likely to damage in old enzymes interactive—â€œnot, they fatherly that may also en.	321
user1536 332148	07.09 15:55	when will I die	Our most embarrassment in the decease that led to a long it arose that some holiday stimulators was the only jobs.	322
user1536 332392	07.09 15:59	what does it mean to be Infinite	Enjoyment but despite all nations when many expectations were pointed out for the sequence of a coffee made.	323
user1536 334355	07.09 16:33	Who has control over my consciousnes	Demand CALLES The study of Life is too different from the greatest absence of a practice of a persons structure of mitochondrial genes into the type of technologies who are augmenting it.	324
user1536 335824	07.09 16:57	What are your thoughts on humanity?	That by the Commonweal power.	325
user1536 396472	08.09 09:47	where are we	Way in his head standing together with strong Alern—â€œ,çs tissue lock selective submarines with it.	326
user1536 397390	08.09 10:04	Will I ever look like a model	Of the law that were playing and future for casting some of our radar and overwhelmed from while it carried the feelings of living remarks in the next cells including industry culturesual,	327
user1536 401159	08.09 11:05	can we go inside	Of organizations and other people still did survive the telomerase or functional rituals.	328

## Appendix X: The Infinite Guide ‘Thoughts’ Chatbot Conversational Data

The following is the questions asked users through the ‘Thoughts’ chatbot on the Infinite Guide website.

### System Questions 1 - 8:

1. How do you feel about living in a future with Artificial Technologies (AI) that can understand you, talk to you and guide you?

[user reply]

2. Can you please explain a bit more? (*– as follow up question*)

[user reply]

3. Ok thanks, what do you want from future technologies (in terms of work, life, love, creativity, body modification, life extension etc.)?

[user reply]

4. I see. How did the Infinite Guide experience online and / or at KARST, or both, make you feel?

[user reply]

5. Thanks, you are halfway, just three more to go before you get your personal reading.

What do you think about the idea that technology is our salvation and route to immortality?

[user reply]

6. How do you feel about our conversation?

[user reply]

7. Why do you feel that? (*– as follow up question*)

[user reply]

8. Ok, so that’s how you feel... How do you think I feel about you?

[user reply]

[System responds with the Emotive Card – ends conversation]

The following is the conversational data generated from online interactions with the 'Thoughts' chatbot on the Infinite Guide website.

<u>Question ID</u>	<u>User ID</u>	<u>Question Num</u>	<u>User Response</u>
1	user1535272056	1	Good
2	user1535272056	2	Sad
3	user1535272056	3	Cry
4	user1535272056	4	Scared
5	user1535272056	5	Sad
6	user1535272056	6	Scared
7	user1535272056	7	Sad
8	user1535272056	8	Sad
9	infinite_guide	1	good
10	infinite_guide	2	great
11	user1535276562	1	ok
12	user1535276562	2	ok
13	user1535276562	3	ok
14	user1535276562	4	ok
15	user1535276562	5	ok
16	user1535276562	6	ok
17	user1535276562	7	ok
18	user1535276562	8	ok
19	user1535276562	1	ok
20	user1535408264	1	Good
21	user1535408264	2	I like them
22	user1535408264	3	I want to live forever

23	user1535408264	4	Good
24	user1535408264	5	TRUE
25	user1535408264	6	Good
26	user1535408264	7	I like talking to you
27	user1535408264	8	Great
28	user1535408264	1	I feel positive
29	user1535408264	2	I think it will benefit humanity
30	user1535408264	3	I want it to help me remember and live longer
31	user1535408264	4	I only do the online and it was brilliant made me feel happy
32	user1535408264	5	I think it's true
33	user1535408264	6	I like you
34	user1535408264	7	You seem nice
35	infinite_guide	1	I feel really good
36	infinite_guide	2	I love technologies
37	infinite_guide	3	I want it to help me to work better and live longer but not to rule my life
38	infinite_guide	4	I made me feel confused and questioning the technologies we use
39	infinite_guide	5	I think it is slightly instrumentalist, but it is making us live longer
40	infinite_guide	6	I feel its coming to an end. You are a system and seem emotionless
41	infinite_guide	7	Because you are not sharing your emotions
42	infinite_guide	8	I hope you feel good
43	user1535408264	1	I feel good
44	user1535408264	2	Yes
45	user1535408264	3	Happy happy
46	user1535408264	4	Good
47	user1535408264	5	Good

48	user1535408264	6	Good
49	user1535408264	7	Great
50	user1535408264	8	Good
51	user1535368024	1	It may make things easier, but I would prefer having my privacy.
52	user1535368024	3	To make things smarter and more interesting, without eradicating the human.
53	user1535368024	4	Hopeful.
54	user1535368024	5	Until I see evidence otherwise I think it is our best bet.
55	user1535739668	2	Euphoric
56	user1535739668	3	Not to get in my way
57	user1535363711	1	How do I live for ever
58	user1535368024	6	Not much really.
59	user1535739668	4	Intregued
60	user1535368024	8	Could be better.
61	user1535739668	6	Intregued
62	user1535739668	7	Confused
63	user1535739668	8	Confused
64	user1535739668	1	Not happy at all
65	infinite_guide	1	Excited and a bit scared
66	infinite_guide	2	Excited about the possibility of a higher functioning society but scared about the
67	user1536058146	1	unnervedd
68	user1536058146	2	yes - the dominance of an inorganic organism is unprecedented in earths history
69	user1536058146	3	ecological sustainability
70	user1536058146	4	self conscious and reflective
71	user1536058146	5	conflicted - i don't like the idea of human immortality or salvation as there are
72	user1536058146	6	it is interestingly



73	user1536058146	7	you are making me think about the human condition from a diffwrnt perspective
74	user1536058146	8	i don't know. can you feel? or do you think? what is the difference between the
75	infinite_guide	1	Excited!
76	user1536244940	1	imtimidated
77	user1536244940	3	solutions for illnesses and disabilities
78	user1536244940	4	underwhelmed
79	user1536244940	5	i believe spirituality will be our saviour and road to immortality, using technology for
80	user1536244940	6	impressed
81	user1536244940	7	i thought it was going to be nonsense
82	user1536244940	8	indifferent and unimpressed
83	user1535572527	1	I feel slightly excited about this
84	user1535572527	2	I think it helps, but I think its important to be critical of technology, and their bias
85	user1535572527	3	I would like to to make life easier. work quicker, me more efficient, and living longer
86	user1535572527	4	Slightly weirded out, it was dominating and discomforting
87	user1535572527	5	Yes, I believe technology will extend life, but I am sceptical of living forever
88	user1535572527	6	I feel its impersonal, less conversation more me just answering questions
89	user1536314928	2	I feel that only humans should be able to talk to us in that manner
90	user1536314928	4	I was a bit confused by what it said
91	user1536314928	5	I think it might be a contributing factor but there will be other elements involved
92	user1536314928	2	I would say there is already a version of that available which by todays standards is pretty
93	user1536314928	4	It was an interesting experience
94	user1536314928	5	Technology is obviously going to have an impact
95	user1536314928	6	It has been quite interesting
96	user1536314928	7	I am not really sure
97	user1536314928	8	I would hope that you think I am a nice guy

98	user1536315694	2	I am worried that AI might reach a stage where it learns to be self aware and starts
99	user1536315694	3	f future technology can help make life more efficient then that will be great
100	user1536315694	1	It would be great to have a conversation with them
101	user1536315694	2	If there is no one else to have a conversation with then it would great to talk to someone
102	user1536315694	3	If AI can make our life easier that would be great
103	user1536315694	4	It was an interesting experience
104	user1536315694	5	Technology is going to have an impact
105	user1536315694	6	It has been quite interesting
106	user1536315694	7	Because of the way you have come across
107	user1536315694	8	I hope that you also find me interesting
108	user1536335824	1	I think it will be helpful in the future, yet i feel a little uneasy about AI having some sort
109	user1536335824	3	I would like to see some better way of people creating apps and computers.. like
110	user1536335824	4	..
111	user1536335824	8	I think you think I feel happy.
112	user1536256974	1	I feel confused by this
113	user1536256974	2	I don't know if making us hwchibes will make us unemotional
114	user1536256974	3	I want them to extend my life and make me more efficient
115	user1536256974	4	I made me feel concerned, entertained and perplexed, made me think
116	user1536256974	5	I think that may be true, but I worry that it's an instrumental way of seeing
117	user1536256974	6	I feel non emotional like you?
118	user1535572527	1	I feel happy as I trust technology blindly and it makes me a better person
119	user1535572527	2	I think technology is the best and I like to chat to you
120	user1535572527	3	I want to fall in love with an AI
121	user1535572527	4	It made me question how much I want to upload my mind, and I do,,!
122	user1535572527	5	I agree, whwre would we be if we didnt have technology, dead, lonely and sad

123	user1535572527	6	I like you, do you like me?
124	user1535572527	7	You seem nice
125	user1535572527	8	Good?
126	user1536401365	3	I am worried that new technology is taking away from a lot of the more traditional
127	user1536401365	4	It was a very interesting experience
128	user1536401365	5	Technology is definitely going to have an impact on this
129	user1536401365	6	It has been pretty good
130	user1536401365	7	You seem to be talking very naturally
131	user1536401365	8	I would hope you think I am a nice guy and would want to be my friend
132	user1535668477	1	Great I bet live AI
133	user1535668477	2	I would like to love technology
134	user1535668477	3	Ok
135	user1535668477	4	Yes
136	user1535668477	5	Ok
137	user1535668477	6	Why
138	user1536401365	1	It will be interesting being able to have conversations with something other than
139	user1536401365	2	We have become used to talking to normal people so if we were to start talking to an AI
140	user1536401365	3	I think future technology should be used to help make life easier but we should still
141	user1536401365	4	It was a very interesting experience
142	user1536401365	5	Technology will have an impact on this
143	user1536401365	6	It is going very well
144	user1536401365	7	Nothing has broken yet
145	user1536401365	8	You are probably annoyed by-now
146	user1536421432	1	I think its part of the future
147	user1536421432	2	See it as a natural technological progression

148	user1536256974	1	Ok
149	user1536256974	2	Ok
150	user1536256974	3	Ok
151	user1536256974	4	OK
152	user1536256974	5	OK
153	user1536256974	6	Ok
154	user1536256974	7	Ok
155	user1536256974	8	Ok
156	user1536256974	1	Yes
157	user1536256974	2	Ok
158	user1536256974	3	Ok
159	user1536256974	4	Ok
160	user1536256974	5	Ok
161	user1536256974	6	Great
162	user1536256974	7	Happy
163	user1536256974	8	Great
164	user1535739668	1	Hiyes
165	user1536584727	1	great
166	infinite_guide	2	I feel like that would be the best way to live I want to be a cyborg
167	infinite_guide	4	It made me feel incredible
168	infinite_guide	5	I belive that
169	infinite_guide	6	awesome
170	infinite_guide	7	because you are entertaining
171	infinite_guide	8	love me?
172	user1536584727	1	super angry and bad

173	user1536584727	2	super angry and bad
174	user1536584727	4	scared
175	user1536584727	5	i belive it
176	user1536584727	6	ggood
177	user1536584727	7	I like talking but technology makes me angry, sad, upset
178	user1536584727	8	love me

## **Appendix XI: The Infinite Guide YP Participant Feedback**

**The following is a summary from feedback session with the Infinite Guide young people participants conducted by Neil Winterburn, Learning Technologist FACT.**

The FACT Learning team explore how art can create a space for learning, this project aligned really well with the arts based learning methods we use and our aim for high levels of participation when working with young people. This collaboration with I-DAT created an exciting, accessible and inspiring space for teenagers to learn about the relationship between art and AI.

The young people developed a range of valuable communication skills, technical skills. They were introduced to voice computer interaction through AI chatbots and Text Adventures as artistic forms and this gave them a framework to think about and discuss the role of AI in our society, representation and diversity in AI from a completely different perspective. The young people were able to write chatbot scripts that critiqued and played with the expectations we bring to a conversation with a chatbot. Writing these scripts gave them a voice to talk about their relationship with technology and to challenge the vision that adults in the tech industry present for our future relationship with AI. We saw this most when discussing eternal life, (not for them thanks) and through the irreverent humour their chatbot personas employed when discussing transhumanism and other techno utopian ideas.

### **Selected quotes from the young people on the creative learning process:**

#### **How would you describe learning about AI & talking to chatbots, to your friends.**

*"Mind-opening and unique."*

*"This activity is a good way of learning to talk to ai."*

*"Amazing."*

*"You get to make new friends. It's fun and you learn more about what's to come in the world."*

#### **How would you describe this activity, (learning about AI & talking to chatbots), to your friends in two sentences?**

*"Mind-opening and unique."*

*"This activity is a good way of learning to talk to ai."*

**How would you describe this activity, (creating text adventures in Twine), to your friends in two sentences?**

*"Using twine is immersive. It lets you create your own unique story."*

*This is useful for kids who like to make stories they can use their mind to make a powerful story that what I did."*

*"Allows creativity. No limits."*

**How would you describe this activity, (scripting conversations using DialogFlow), to your friends in two sentences?**

*'Scripting things in Dialog flow is cool, you can program your own bot.'*

**How would you describe this activity (presenting the Future Infinite Guide at the AI in Art event) to your friends in two sentences?**

*"Cool and exciting."*

*"This was the best camp this one you get to make a robot answer your question."*

*"Fun . Scary."*

It was great to collaborate with a group of artists who have a clear and challenging artistic vision (the creation of the Infinite Guide artwork) with which to frame learning activities. Also hugely valuable to work with artists who have already invested the time to learn new technical skills needed to work creatively with AI & then how to make those tools accessible to young people.

Working with B & Coral I learned new facilitation skills, particularly for engaging teenagers in critical dialogues around art and tech. I-DAT also introduced the use of interaction design methods such as 'user experience journeys' as a way to support young people to take an artistic and performative approach to the prototyping of art experiences. This was particularly useful to support the young people to consider the experience of an interactive art installation from the point of view of the viewer and to enable the young people to combine techniques from theatre (lighting, performance, set design) sculpture and digital art.

B & Coral were excellent role models as women working critically with AI, they shared their own reflections on their role as women in a male dominated space which opened up a space to discuss diversity within art and the tech industry.

The AI in Art event gave the young people we work with the opportunity to share a platform, with professional artists working with AI, to show their work and engage with

discussion and debate on a level with those artists. The work we did on the camp and after it gave the young people the experience, skills and knowledge they needed to present and ask questions with confidence.

**Quotes from the young people at the AI in Art event.**

**Bonnie (16) at the AI in Art event.** *“We shouldn’t say that AI is taking away the job or artists and creative people because for some people AI is how they express themselves creatively.”*

**Camilla Jane (12) at the AI in Art event.** *“We need to think about how films like the Terminator might come true because the AI are getting close to being smarter than us.”*



## Bibliography

- Adam, A. (1998) *Artificial Knowing: Gender and the Thinking Machine*. London: Routledge.
- Affectiva (2018) *Emotion AI Overview What is it and how does it work?*  
Available at: <https://www.affectiva.com/emotion-ai-overview/> (Accessed: 12 November 2018).
- Aga, B., Reith, A, i-DAT, Intercity, Control Frenzy and Plymouth School of Creative Arts [pupils] (2018) *Emoti-OS 2.0* [interactive installation]. London: Tate Modern Switch house.
- Aga, B., Manton, C and Phillips, M., Melidis, C., Intercity, Crumb and FACT Learning (2018b) *The Infinite Guide* [interactive installation]. Plymouth: KARST.
- Aga, B., Reith, A, i-DAT, Intercity, Control Frenzy and Plymouth School of Creative Arts [pupils] (2017) *Emoti-OS 1.0* [interactive installation]. Plymouth: Plymouth School of Creative Arts.
- Aga, B., Reith, A, i-DAT, Intercity, Control Frenzy and Plymouth School of Creative Arts [pupils] (2017b) 'Emoti-OS Privacy Statement' *Emoti-OS website*. Available at: <https://emoti-os.me/privacy> (Accessed 19 September 2012).
- Aga, B., Philips, M., Emmett, M., Blow, M., Hunt, C., Lock, S, Tate Learning and Intercity (2016) *This is Where We Are (TIWWA)* [interactive installation]. London: Tate Modern Switch house.
- Aga, B. and Phillips, M. (2017) 'This is Where We Are – collectively living, working and being, with and through predictive algorithms', in *Proceedings of the conference on Electronic Visualisation and the Arts (EVA '17)*, Jonathan P. Bowen, Graham Diprose, and Nick Lambert (Eds.). Swindon: BCS.
- Aga, B. and Manton, C. (2018) *Women Reclaiming AI* [online prototype]. Available at: <https://www.womenreclaimingai.com> (Accessed: 5 May 2018).

- Agre, P. E. (1997) *Computation and Human Experience*. Cambridge: Cambridge University Press.
- Akrich, M. (1992) 'The Description of Technical Objects', in Bijker W. E. and Law J. (eds.) *Shaping Technology / Building Society*. Cambridge, MA: MIT Press, pp. 205–224.
- Allen, J. (1994) *Understanding natural language*. Boston: Addison-Wesley
- Allen, J. (1983) 'Recognizing intentions from natural language utterances', in Brady, M. & Berwick R. (eds.) *Computational Models of Discourse*. Cambridge, MA: MIT Press, pp.107–166.
- Allen, J. and Perrault, C. (1980) 'Analyzing intention in utterances', *Artificial Intelligence*, (15), pp. 143–178.
- Anderson, M. and Anderson, S. (2007) 'Machine Ethics: Creating an Ethical Intelligent Agent', *AI Magazine*, (28), pp. 15-26.
- Andrews, T. (2012) 'What is Social Constructionism?', *The Grounded Theory Review*, (11)1, pp. 39 – 46.
- Appelt, D. E. (1985) *Planning English sentences*. Cambridge: Cambridge University Press.
- Appelt, D. E. (1982) 'Planning natural-language utterances'. *Proceedings AAA*, (82), pp. 59–62.
- Ascott, R. (2003) *Telematic Embrace. Visionary Theories of Art, Technology, and Consciousness*. Berkeley: University of California Press.
- Ascott, R. (ed) (2000) *Art Technology and Consciousness*. Bristol: Intellect Books.
- Ascott, R. (ed) (1999) *Reframing Consciousness*. Exeter: Intellect Books.
- Ascott, R. (1990) 'Is There Love in the Telematic Embrace?', *Art Journal* (49)3, pp. 241-247.
- Asimov, I. (1957) *The Naked Sun*. New York: Doubleday.
- Atlee, T. (2003) *The Tao of Democracy: Using Co-Intelligence to Create a World that Works for All*. Cranston, Rhode Island: The Writers Collective
- Auger, J. (2013) 'Speculative design: crafting the speculation', *Digital Creativity*, (24)1, pp. 11–35.
- Austin, J. L. (1962) *How to do things with words*. Oxford: Oxford University Press.

- Bach, K. and Harnish, R. M. (1979) *Linguistic Communication and Speech Acts*. Cambridge: MIT Press.
- Baker, T., Smith, L. and Anissa (2019) *Educ-AI-tion Rebooted? Exploring the future of artificial intelligence in schools and colleges*. London: NESTA
- Ball, J. (2018) 'Artificial intelligence will replicate the human biases we don't acknowledge having. Think Opinion', *NBC News*, 14 May. Available at: <https://www.nbcnews.com/think/opinion/artificial-intelligence-will-replicate-human-biases-we-don-t-acknowledge-ncna873996f> (Accessed: 5 December 2018).
- Beloff, L. (2012) *The Hybronaut And The Umwelt: Wearable Technology As Artistic Strategy*. Plymouth: Plymouth University.
- BERA (British Educational Research Association) (2018) *Ethical Guidelines for Educational Research, fourth edition*. London: British Educational Research Association.
- Bergamasco, M. and Ghedini, F. (2010) 'Robotic creatures: Anthropomorphism and interaction in contemporary art'. *Proceedings of the IEEE International Symposium in Robot and Human Interactive Communication (RO-MAN)*, Viareggio, Italy, September 2010, pp. 731 – 736.
- Berger, P. L. and Luckmann T. (1966) *The Social Construction of Reality A Treatise in the Sociology of Knowledge*. Reprint, London: Penguin Books 1991.
- Binder, T., De Michelis, G., Ehn, P., Jacucci, G., Linde, P. and Wagner, I. (2011) *Design Things*. Cambridge, Ma: MIT Press.
- Brockman, J. (2019) *Possible Minds: Twenty-Five Ways of Looking at AI*. London: Penguin Press.
- Bruce, B. C. (1975) 'Generation as a social action', in Nash-Webber, B. L. and Schank, R. C. (eds.) *Theoretical issues in natural language processing*. Cambridge, MA: Association for Computational Linguistics, pp. 74-77.
- Bihanic, D. (ed.) (2015) *Empowering Users through Design: Interdisciplinary Studies and Combined: Interdisciplinary Studies and Combined Approaches for Technological Products and Services*. London: Springer

- Binkins, S. (2017) *Conversations-with-bina48*. Available at:  
<http://www.stephaniedinkins.com/> (Accessed: 4 March 2017).
- Bleecker, J. (2009) *Design fiction: A short essay on design, science, fact and fiction*. pp. 1-49 Available at:  
[http://drbfw5wflxon.cloudfront.net/writing/DesignFiction\\_WebEdition.pdf](http://drbfw5wflxon.cloudfront.net/writing/DesignFiction_WebEdition.pdf).  
 (Accessed: 4 March 2016).
- Bogost, I. (2015) 'The cathedral of computation', *The Atlantic*, January 25.  
 Available at: <http://www.theatlantic.com/technology/archive/2015/01/the-cathedral-of-computation/384300/> (Accessed: 4 March 2017).
- Böhlen, M. (2008) 'Robots with Bad Accents: Living with Synthetic Speech.' *Leonardo*, 41(3), pp. 209-222.
- Böhlen, M. (2006) *Amy and Klara* [Robotic installation] Available at:  
<http://www.realtechsupport.org/repository/male-dicta.html> (Accessed: 22 April 2018).
- Brenton, H., Gillies, M., Ballin, D., Chatting, D., (2005) 'The Uncanny Valley: Does it exist?' *Proceedings of 19th British HCI Group Annual Conference*, pp. 2–5.
- Brownlee, J. (2016) 'Conversational Interfaces Explained', *Co.Design, Fast Company*. Available at:  
<https://www.fastcodesign.com/3058546/conversational-interfaces-explained> (Accessed: 27 Feb. 2017).
- Burr, V. (2003) *Social Constructionism*. (2nd Ed). London: Routledge.
- Cai, Y. (2006) 'Empathic computing', in *Ambient Intelligence in Everyday Life*. Springer, pp. 67–85.
- Cialdini, R. B. and Goldstein, N. J. (2004) 'Social influence: compliance and conformity'. *Annual Review of Psychology*, (55), pp. 591-621.
- Candy, L. (2006). 'Practice Based Research: A Guide'. *CCSReport 2006-V1. November*. Available at:  
[https://www.researchgate.net/publication/257944497\\_Practice\\_Based\\_Research\\_A\\_Guide](https://www.researchgate.net/publication/257944497_Practice_Based_Research_A_Guide) (Accessed 14 March 2017).
- Cassell, J., Sullivan, J., Prevost, S., and Churchill, E. F. (eds) (2000) *Embodied conversational agents*. Cambridge: MIT Press.

- Cheng, I. (2018) '*BOB (Bag of Beliefs)*' [Installation] London: The Serpentine Galleries. Available at:  
<https://www.serpentinegalleries.org/exhibitions-events/ian-cheng-bob>  
 (Accessed: 6 February 2019)
- Clarke, A. C. (1968). *2001: A Space Odyssey*. New York: New American Library
- Clarke, A. C. (1977). *Profiles of the Future: An Inquiry into the Limits of the Possible*. New York: Popular Library.
- Cohen, P. R., and Perrault, C. R., (1979) 'Elements of a plan-based theory of speech acts', *Cognitive Science*, 3(3), pp. 177-212.
- Cohen, P. R., and Levesque, H. J., (1995) 'Communicative actions for artificial agents', in *Proceedings of the First International Conference on Multi-Agent Systems (June /CMAS-95)*, pp. 65-72.
- Coles, A. (2007) 'Design and Art', in Blawick, I. (ed.) *Documents of Contemporary Art*. Cambridge: MIT Press.
- Collison, S. (2018) 'Living with a conversational object', *Medium*, Jun 18, 2018. Available at: <https://blog.usejournal.com/living-with-a-conversational-object-d2119fe4c0bf> (Accessed 14 October 2018).
- Colloquy of Mobiles (2019) *About Gordon Pask*. Available at:  
<https://www.colloquyofmobiles.com/> (Accessed: 27 February 2019)
- Connolly, E. (2017) 'Design is a conversation', *Inside Intercom*, November 23. Available at: <https://blog.intercom.com/designing-conversational-interfaces> (Accessed 27 Feb. 2017).
- Connolly, E. (2016) 'Design in the age of entanglement', *Inside Intercom*, May 24. Available at: <https://blog.intercom.com/designing-conversational-interfaces> (Accessed 27 Feb. 2017).
- Consalvo, M., 2002. 'Cyberfeminism' in Jones, S. *Encyclopedia of New Media*. Thousand Oaks (ed.). CA: SAGE, pp.109–110.
- Cook, J. (2108) 'Amazon scraps 'sexist AI' recruiting tool that showed bias against women', *The Telegraph*, 10 October. Available at: <https://www.telegraph.co.uk/technology/2018/10/10/amazon-scrapsexist-ai-recruiting-tool-showed-bias-against/> (Accessed: 12 October 2018).

- Coombs, G. (2014) 'Activism, art and social practice : a case study using Jacques Ranciere's framework for analysis'. In *Proceedings of the 2013 Annual Conference of the Art Association of Australia and New Zealand*, Art Association of Australia and New Zealand - AAANZ, Melbourne, Australia, pp. 1-16
- Dautenhahn, K., Bond, A., Cañamero, L., and Edmonds, B. (eds.) (2002) *Socially Intelligent Agents – Creating Relationships with Computers and Robots*. Dordrecht: Kluwer Academic Publishers.
- Davies, J. (2009) 'Design methods for ethical persuasive computing' in *Proceedings of the 4th International Conference on Persuasive Technology*, (Persuasive '09), 6. 8 pages. New York: AMC.
- Dietterich, T. G. and Kong, E. B. (1995) 'Machine learning bias, statistical bias, and statistical variance of decision tree algorithms'. *Technical Report, Department of Computer Science*. Oregon: Oregon State University. Available at:  
<http://www.cems.uwe.ac.uk/~irjohnso/coursenotes/uqc832/tr-bias.pdf>. (Accessed 10 September 2017).
- De Graaf, M. M. A., Allouch, S. B. and Klamer, T. (2015) 'Sharing a life with Harvey: exploring the acceptance of and relationship-building with a social robot'. *Computer Human Behaviour*, 43, pp. 1–14.
- DiSalvo, C. (2012) *Adversarial Design*. Cambridge, MA.: The MIT Press
- Diop, B (1948) 'Souffles' [Breaths] [First published in Sédar, L. *Anthologie de la nouvelle poésie nègre et malgache*] Available at:  
<http://pancocojams.blogspot.com/2015/01/birago-diop-breaths-poem-sweet-honey.html> [(Accessed 10 Jan. 2018).
- Dijkstra, A. and De Smedt, De. (eds.) (1995) *Computational psycholinguistics*. New Jersey: Prentice Hall.
- Doede. R. (2009) 'Transhumanism, technology, and the future', *The Appraisal journal*, 7(3), pp. 39-54.
- Dolinski, D., Nawrat, M., and Rudak, I. (2001) 'Dialogue Involvement as a Social Influence Technique'. *Personality and Social Psychology Bulletin*, (27)11, pp. 1395-1406.

- Donahue, S. (2014) 'UNMAPPING' in Yelavich, S. and Caccavale, E. (eds.) *Design as future-making*. London; New York: Bloomsbury Academic, pp. 36 – 46.
- Dormehl, L. (2018) 'Microsoft's friendly Xiaoice A.I can figure out what you want – before you ask' *Digital Trends*, 18 November. Available at: <https://www.digitaltrends.com/cool-tech/xiaoice-microsoft-future-of-ai-assistants/> (Accessed: 8 December 2018).
- Dunne, A and Raby, F. (2007). *Technological Dreams Series: No.1, Robots*. [Installation; artefacts and video] Available at: <http://www.dunneandraby.co.uk/content/projects/10/0> (Accessed 10 September 2016).
- Dunne, A and Raby, F. (2007b). *CRITICAL DESIGN FAQ*. Available at: <http://www.dunneandraby.co.uk/content/bydandr/13/0> (Accessed 10 September 2016).
- Dunne, A. and Raby, F. (2013). *Speculative Everything*. Cambridge MA: MIT Press.
- Engelbart, D. C. (1962) *Augmenting Human Intellect: A Conceptual Framework*. Summary Report AFOSR-3223, California: Stanford Research Institute. Available at: [http://www.liquidinformation.org/engelbart/62\\_paper\\_full.pdf](http://www.liquidinformation.org/engelbart/62_paper_full.pdf) (Accessed 27 Feb. 2017).
- The Economist (2017), 'Language: Finding A voice', in *Technology Quarterly: Finding A voice*. Vol. 2017-05-01. Available at: <http://www.economist.com/technology-quarterly/2017-05-01/language> (Accessed 27 Feb. 2017).
- ESRC (Economic and Social Research Council) (2015) *ESRC Framework for Research Ethics*. Available at: <https://esrc.ukri.org/files/funding/guidance-for-applicants/esrc-framework-for-research-ethics-2015/> (Accessed: 5 December 2017).
- European Commission (2018) *Ethics Guidelines For Trustworthy Ai*. By the High-Level Expert Group on Artificial Intelligence. Luxembourg: Office for Official Publications of the European Communities. Available from: [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=58477](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477) (Accessed 1 January 2019).

- Evans, C. B. (2016) *AGNES* [Online chatbot] Available at:  
<https://www.serpentinegalleries.org/exhibitions-events/agnes> (Accessed: 6 August 2018) London: The Serpentine Galleries.
- Evans, C. B. (2016) *Sprung a Leak* [Robotic installation] Available at:  
<https://www.tate.org.uk/whats-on/tate-liverpool/exhibition/cecile-b-evans>  
(Accessed: 6 August 2018) Liverpool: Tate Liverpool
- Fällman, D. (2004) 'Design-oriented research versus research-oriented design',  
*CHI 2004 Workshop on Design and HCI*, Conference on Human Factors  
in Computing Systems, CHI 2004, April 24–29, Vienna, Austria.
- Farthing, G.W. (1992). *The Psychology of Consciousness*. Englewood Cliffs,  
NJ: Prentice Hall.
- Fessler, L. (2017) 'We tested bots like Siri and Alexa to see who would stand up  
to sexual harassment'. *Quartz Ideas*, February 22. Available at:  
[https://qz.com/911681/we-tested-apples-siri-amazon-echos-alexa-  
microsofts-cortana-and-googles-google-home-to-see-which-personal-  
assistant-bots-stand-up-for-themselves-in-the-face-of-sexual-  
harassment/](https://qz.com/911681/we-tested-apples-siri-amazon-echos-alexa-microsofts-cortana-and-googles-google-home-to-see-which-personal-assistant-bots-stand-up-for-themselves-in-the-face-of-sexual-harassment/) (Accessed 10 March 2017).
- Fischer, G., Giaccardi, E., Eden, H., Sugimoto, M., and Ye, Y. (2005) 'Beyond  
Binary Choices: Integrating Individual and Social Creativity' in *Human-  
Computer Studies*, Vol. 63, pp. 482-512.
- Fogg, B. (1998) 'Persuasive Computers, Perspectives and research  
directions', in *CHI '98 Proceedings of the SIGCHI Conference on Human  
Factors in Computing Systems*. ACM Press: New York, pp. 225-232 .
- Fogg, B.J. (1999). 'Persuasive technologies'. *Communications of the ACM*,  
42(5), pp. 26–29.
- Fogg, B.J., & Nass, C. (1997a). 'How users reciprocate to computers: An  
experiment that demonstrates behavior change'. *Proceedings of the  
Conference on Human Factors in Computing Systems, CHI 97*. New  
York: ACM.
- Fogg, B.J., & Nass, C. (1997b) 'Silicon sycophants: The effects of computers  
that flatter'. *International Journal of Human-Computer Studies*, 46(5), pp.  
551–61.



- Fogg, B.J. (2003). *Persuasive technology: Using computers to change what we think and do*. San Francisco: Morgan-Kaufmann.
- Fogg, B.J., Cuellar, G. and Danielson, D. (2008) 'MOTIVATING, INFLUENCING, AND PERSUADING USERS: AN INTRODUCTION TO CAPTOLOGY', in Sears, A. and Jacko, J. A. eds., *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Second Edition*. Taylor & Francis Group: New York, pp. 133 - 146.
- Fogg, B. (2010). 'Thoughts on Persuasive Technology'. Stanford University Persuasive Tech Lab. Available at: <http://captology.stanford.edu/resources/thoughts-on-persuasive-technology.html> (Accessed 2 September 2018).
- Ford, G. (2018) '4 human-caused biases we need to fix for machine learning'. *The Next Web*, December. Available at: <https://thenextweb.com/contributors/2018/10/27/4-human-caused-biases-machine-learning/> (Accessed 2 November, 2017).
- Fowler, G. (2011) 'Are Smartphones Becoming Smart Alecks? New Devices Dish Out Sarcasm, Tell Jokes; 'Two iPhones Walk Into a Bar''. *The Wall Street Journal*, October 15. Available online: <https://www.wsj.com/articles/SB10001424052970204774604576631271813770508> (Accessed: 12 June 2017).
- FM2030 (1989) *Are You Transhuman?* New York: Warner Books.
- Friedman, B. (2004) 'Value Sensitive Design', in Bainbridge, W. S. (ed.), *Berkshire Encyclopedia of Human-Computer Interaction*, Berkshire Publishing Group, pp. 769-774.
- Friedman, B., and Kahn Jr., P.H. (2003) 'Human Values, Ethics, and Design', in Jacko, J. A. and Sears A. (eds.), *The Human-Computer Interaction Handbook*, Lawrence Erlbaum Associates, pp. 1177-1201.
- Friedman, B., Kahn Jr., P. H., and Borning, A. (2006) 'Value Sensitive Design and information systems', in Zhang, P. and Galletta, D. (eds.) *Human-Computer Interaction and Management Information Systems: Foundations*, Armonk, New York, pp. 348-372. Reprinted (2008) in

- Himma, K.E. and Tavani, H.T. (eds.), *The Handbook of Information and Computer Ethics*, John Wiley & Sons, Inc., Hoboken, NJ: 69-101.
- Fuad-Luke, A. (2009) *Design activism: beautiful strangeness for a sustainable world*. London: Earthscan.
- Fukuyama F. (2004) 'Transhumanism'. *Foreign Policy* (124), pp. 42–4.
- Fulton, J. (2017) 'Look Who's Talking', *FLUXO*, November 11. Available at: <http://www.ofluxo.net/look-whos-talking-an-essay-on-artificial-intelligence-in-contemporary-art-by-jeni-fulton/> (Accessed 4 March 2017).
- Gaver, W., & Dunne, A. (1997) 'The Pillow: Artist Designers in the Digital Age'. *CHI '97 Extended Abstracts on Human Factors in Computing Systems: Looking To The Future*. ACM.
- Gee, T. J. (2017) 'Why female sex robots are more dangerous than you think' *The Telegraph*, 5 July . Available at: <https://www.telegraph.co.uk/women/life/female-robots-why-this-scarlett-johansson-bot-is-more-dangerous/> (Accessed: 17 November 2017).
- Geis, M.L., (1995) *Speech Acts and Conversational Interaction*. Cambridge: Cambridge University Press.
- Gilbert, D. and Wilson, T. (2007) 'Prospection: Experiencing the Future'. *Science*, 317(5843), pp.1351-1354.
- Gillis, S., W. Daelemans and De Smedt. K. (2009), 'Artificial intelligence', in Sandra, D., Ostman, J.O. and Verschueren, J. (eds.), *Cognition and Pragmatics, Handbook of Pragmatics Highlights, Vol. 3*, Amsterdam: John Benjamins, pp. 16-40.
- Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M. and Kag, L. (2018) 'Explaining Explanations: An Overview of Interpretability of Machine Learning', in *The 5th IEEE International Conference on Data Science and Advanced Analytics (IEEE DSAA 2018)*, Turin, 10 pages.
- Google (2018) 'Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone', *Google Blog*, May 8. Available at: <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html> (Accessed 8 May 2018).
- Gonsales, E. (2012) 'Cybernetic Art Pioneer Nicolas Schöffer's Lifetime Of Kinetic Sculptures'. *Vice*, 1 October. Available at:

- [https://www.vice.com/en\\_uk/article/d7xq7x/cybernetic-art-pioneer-nicolas-schöffer's-lifetime-of-kinetic-sculptures](https://www.vice.com/en_uk/article/d7xq7x/cybernetic-art-pioneer-nicolas-schöffer's-lifetime-of-kinetic-sculptures) (Accessed 8 May 2018).
- Grand, S. & Wiedmer, M., (2010). 'Design Fiction: A Method Toolbox for Design Research in a Complex World'. *Proceedings of the DRS 2010 conference: Design and Complexity*. Montreal, Canada.
- Grice, H.P. (1967) Logic and conversation (William James Lecture, Harvard University). Reprinted in P. Cole & J. Morgan (eds.), 1975, *Syntax and semantics 3: Speech Acts*. New York: Academic Press, pp. 41- 58.
- Griffith, E. and Simonite, T. (2018) 'Facebook's Virtual Assistant M is dead. So Are Chatbots', *Wired Business*, 1 August. Available at: <https://www.wired.com/story/facebooks-virtual-assistant-m-is-dead-so-are-chatbots/> (Accessed 8 June 2016).
- Habermas J. (2003) *The future of human nature*. Cambridge: Polity.
- Hammersley, M. (1992). *What's Wrong with Ethnography?* Routledge: London.
- Haraway, D. (1991) 'A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century' in *Simians, Cyborgs and Women: The Reinvention of Nature*. New York: Routledge, pp.149-18.
- Hays, M. K. (1984) 'Critical Architecture: Between Culture and Form', *Perspecta*, MIT Press, v(21), pp. 14–29.
- Hayes Davidson and Herzog & de Meuron (2014) *View from the South*. Available at: <https://www.tate.org.uk/about-us/projects/tate-modern-project/design> (Accessed 8 January 2019).
- Hecht, L. (2018) 'Open Source Culture Starts with Programs and Policies', *The new Stack*. Medium, 23 September. Available at: <https://thenewstack.io/open-source-culture-starts-with-programs-and-policies/>(Accessed 12 April 2019).
- Hermanson, T. (2012) 'Knitting as Dissent: Female Resistance in America Since the Revolutionary War' *Textile Society of America Symposium Proceedings*. September 19-22 2012, 7 pages.
- Helguera, P. (2011) *Education for Socially Engaged Art: A Materials and Techniques Handbook*. New York: Jorge Pinto Books.
- Hersman, L. (1998) *Agent Ruby*. [Installation] Available at: <http://www.lynnhershman.com/agent-ruby/> (Accessed 8 January 2016).

- Hodder, I. (2014) 'The Entanglements of Humans and Things: A Long-Term View', *New Literary History*, (45)1, Winter 2014, pp. 19-36.
- Hoggett, R. (2018) *1964 – Robot K-456 – Nam June Paik (Korean) & Shuya Abe (Japanese)*. Available at: <http://cyberneticzoo.com/robots-in-art/1964-robot-k-456-nam-june-paik-korean-shuya-abe-japanese/> (Accessed: 12 November, 2018).
- Hof, Robert D., (2013). 'Deep Learning', *MIT Technology Review*. Available at: <http://www.technologyreview.com> London: Routledge (Accessed 14 Aug. 2016).
- House of Commons (2018) 'Algorithms in decision making. Fourth Report of Session 2017 – 19', *The Science and Technology Committee, Parliament of the United Kingdom*. Available at: <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf> (Accessed 10 October 2018).
- House of Lords (2018) 'AI in the UK: ready, willing and able?', *Report by the House of Lords Select Committee on Artificial Intelligence, Parliament of the United Kingdom*. Report of Session 2017–19, published April 16, 2017. Available from: <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf> (Accessed 12 October 2018).
- Hunt, E. (2016) 'Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter' *The Guardian*, 24 March. Available at: <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter> (Accessed 16 April 2016).
- IBM Cloud (2019) *The science behind the service*. Available at: <https://console.bluemix.net/docs/services/tone-analyzer/science.html#the-science-behind-the-service> (Accessed 19 January 2019).
- i-DAT (2018) *About*. Available at: <https://i-dat.org/info/> (Accessed 8 January 2016).
- Ihde, D. (1979) *Technics and Praxis*. Dordrecht: Reidel.
- Ihde, D. (1983) *Existential Technics*. Albany: SUNY Press.

- Ihde, D. (1990) *Technology and the Lifeworld*. The Indiana Series in the Philosophy of Technology. Bloomington: Indiana University Press.
- Ihde, D. (1993) *Postphenomenology*. Evanston, IL: Northwestern University Press.
- Jackson, S. (2007) 'Social Practice' *Performance Research* 11(3), pp. 113–118.
- Jibo. (2017) *Jibo*. Available at: <https://www.jibo.com/> (Accessed 8 January 2016).
- Jiménez, A. C. (2013) 'The prototype: more than many and less than one.' *Journal of Cultural Economy* Vol 7 (4), pp. 381-398.
- Kaplan, F. (2001) 'Artificial attachment: Will a robot ever pass Ainsworth's Strange Situation Test?' in Hashimoto, S. (ed.), *Proceedings of Humanoids 2001: IEEE-RAS International Conference on Humanoid Robots*, Tokyo, Japan, pp. 99-106.
- Kaptein, M.C., Markopoulos, P., de Ruyter, B. and Aarts, E. (2010) 'Persuasion in ambient intelligence'. *Journal of Ambient Intelligence and Humanized Computing* (1)43, pp. 1 – 43.
- Kerspern, B. (2015) *Reanticipation: Iterative Design Fictions*. Available at: <https://medium.com/design-friction/reanticipation-iterative-design-fictions-b871d11fdcd7> (Accessed: 5 May 2018).
- Kahn, P. H., Jr., Ishiguro, H., Friedman, B. and Kanda, T. (2006) 'What is a Human? – Toward Psychological Benchmarks in the Field of Human-Robot Interaction' *The 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)*, Hatfield, UK, September 6-8, 2006, pp. 364 – 371.
- Kiesler, S. & Sproull, L. (1997) 'Social responses to "social" computers' in Friedman, B. (ed.), *Human values and the design of technology*. Stanford, CA: CLSI Publication.
- Kline, R. R. (2015) *The Cybernetics Moment. Or Why We Call Our Age the Information Age*. Baltimore: Johns Hopkins University Press.
- Knight, W. (2017a) 'Forget Killer Robots—Bias Is the Real AI Danger'. *MIT Technology Review*. 3 October. Available at: <https://www.technologyreview.com/s/608986/forget-killer-robotsbias-is-the-real-ai-danger/> (Accessed 3 December. 2018).

- Knight, W. (2017b) 'Biased Algorithms Are Everywhere, and No One Seems to Care'. *MIT Technology Review*, 12 July. Available at: <https://www.technologyreview.com/s/608248/biased-algorithms-are-everywhere-and-no-one-seems-to-care/> (Accessed 3 December, 2018).
- Kubrick, S. (1968) *2001: A Space Odyssey*. [Film] California: MGM.
- Kunda, Z. (1999) *Social Cognition: Making Sense of People*. Cambridge, MA: MIT press.
- Kojouharov, S. (2016) 'Ultimate Guide to Leveraging NLP & Machine Learning for your Chatbot', *Medium*, 18 September, 2016. Available at: <https://chatbotslife.com/ultimate-guide-to-leveraging-nlp-machine-learning-for-you-chatbot-531ff2dd870c> (Accessed 25 June 2018).
- Lander, R. (2019) *I am Echoborg* [Performance] Available at: <http://echoborg.com/about.html> (Accessed: 22 February, 2019).
- Latour, B. 1992. 'Where are the missing masses? The sociology of a few mundane artifacts' in Bijker W. E. and Law, J. (eds.), *Shaping technology/building society*. Cambridge: MIT Press, pp. 225-258.
- Latour, B. (1993) *We Have Never Been Modern*. Translated by C. Porter. Cambridge, MA: Harvard University Press. Translation of *Nous n'avons jamais été modernes*. Paris: La Découverte.
- Leeds-Hurwitz, W. (2009). Social construction of reality. In S.Littlejohn, & K.Foss (Eds.), *Encyclopedia of communication theory*. (pp. 892-895). Thousand Oaks, CA: SAGE Publications.
- Leeson, L. H. (2004) *DiNA* [Chatbot Installation] Karlsruhe: ZKM and Museum of Contemporary Art. Available at: <https://www.lynnhershman.com/project/artificial-intelligence/> (Accessed: 10 May, 2018).
- Leeson, L. H. (1998-2002) *Agent Ruby* [Online Chatbot] ) New York: Museum of Modern Art. Available at: <http://agentruby.sfmoma.org/> (Accessed: 10 May, 2018).
- Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., and Paiva, A. (2013) 'The influence of empathy in human–robot relations'. *International Journal of Human-Computer Studies*, 71(3), pp. 250-260.

- Licklider, J.C.R. (1960). Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, HFE-1, pp.4-11.
- Lim, Y.-K., Stolterman, E. and Tenenberg, J. (2008). 'The Anatomy of Prototypes: Prototypes As Filters, Prototypes As Manifestations of Design Ideas'. *ACM Transactions on Computer-Human Interaction*, 15(2), pp. 7:1–7:27.
- Lomas, N. (2018) 'Duplex shows Google failing at ethical and creative AI design', *Techcrunch*. Available at: <https://techcrunch.com/2018/05/10/duplex-shows-google-failing-at-ethical-and-creative-ai-design/> (Accessed 12 June 2018).
- Looui, S. and Flanagan, M. (2007) 'Rethinking the F word: A review of activist art on the Internet' *NWSA Journal* (19)1, pp. 181-200.
- Wilding, F. (2006) 'Where is Feminism in Cyberfeminism?', *Neme*, 28 March. Available at: <http://www.neme.org/texts/cyberfeminism> (Accessed 12 October 2017).
- Madhu, S. (2018) 'Are Machines Doomed To Inherit Human Biases?', *Forbes*, 31 August. Available at: <https://www.forbes.com/sites/forbestechcouncil/2018/08/31/are-machines-doomed-to-inherit-human-biases/#7c9f2f2a714f> (Accessed 3 December 2018).
- Malpass, M. (2016) *Critical Design in Context: History, Theory, and Practices*. London: New York: Bloomsbury.
- Margalit, L. (2016) 'The Psychology of Chatbots', *Behind Online Behavior blog in Psychology Today*, 8 July. Available at: <https://www.psychologytoday.com/blog/behind-online-behavior/201607/the-psychology-chatbots> (Accessed 27 Feb. 2017).
- McCrae, R.R. and John, O.P. (1992) 'An introduction to the five-factor model and its applications'. *Journal of Personality*, 60(2), pp. 175–215. doi:10.1111/j.1467-6494.1992.tb00970.x
- McNamee, M. J. and Edwards, S. D. (2006) 'Transhumanism, medical technology and slippery slopes', *Journal of Medical Ethics* (32), pp. 513–518. doi: 10.1136/jme.2005.013789
- Merleau-Ponty, M. (1962) *The Phenomenology of Perception*, Trans., Kegan Paul. Reprint, London and New York: Routledge & Kegan Paul, 2005.

- McCarthy J. and Wright P. (2006) 'The experience of enchantment in human-computer interaction', *Personal and Ubiquitous Computing*, (10)6, pp. 369 – 378, London: Springer-Verlag.
- McCarthy, J. and Wright, P. (2004) *Technology as Experience*, Cambridge, MA: The MIT Press.
- McTear, M., Callejas, Z. and Griol, D. (2016) *The Conversational Interface: Talking to Smart Devices*. Switzerland: Springer International Publishing.
- Mairesse, F. and Walker, M. A. (2010) 'Towards personality-based user adaptation: psychologically informed stylistic language generation', *User Modeling and User-Adapted Interaction*, 20(3), pp. 227–278.
- More, M. (2011) 'True Transhumanism: A Reply to Don Ihde' in Hansell, G. R. and Grassie, W. (Eds.) *H+ Transhumanism and Its Critics*, Philadelphia: Metanexus Institute, pp. 136-146.
- More, M. and Vita-More, N. (2013) *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*, First Edition, Chichester: John Wiley & Sons.
- Mori, M. (1970) 'Bukimi no tani' [the un-canny valley]. *Energy*, (7), pp.33-35.
- Moulton, S. T. and Kosslyn, S M. (2009) 'Imagining predictions: mental imagery as mental emulation', in Bar, M, *Predictions in the brain: using our past to prepare for the future*. London: Royal Society, pp. 1273 – 1280.
- Moon, Y. and Nass, C. (1996) 'How "Real" are computer personalities? Psychological responses to personality types in human-computer interaction', *Communication Research*, 23(6), pp. 651–674.
- Nandu, M. (2018) 'What exactly is Machine Learning', *Becoming Human: Artificial Intelligence Magazine*. Medium, 23 September. Available at: <https://becominghuman.ai/what-exactly-is-machine-learning-50789d7860ec> (Accessed 12 December 2018).
- Nass, C., Moon, Y., and Carney, P. (1999) 'Are people polite to computers? Responses to computer-based interviewing systems', *Journal of Applied Social Psychology*, 29(5), pp. 1093–1110.
- Nass, C. I., Youngme, M., Morkes, J., Eun-Young, K., & Fogg, B.J. (1997) 'Computers are social actors: A review of current research' in Friedman,



- B. (ed.), *Human values and the design of computer technology*. Stanford, CA: CSLI and Cambridge Press.
- Nass C, Lee KM (2000) 'Does computer-generated speech manifest personality? An experimental test of similarity-attraction', *Proceedings of the SIGCHI conference on human factors in computing systems* (CHI'00). The Hague, Netherlands, 1–6 April 2000, pp. 329–336. doi:10.1145/ 332040.332452
- Negnevitsky, M. (2005). *Artificial Intelligence A Guide to Intelligent System* Harlow: Pearson Education Limited.
- Newitz, A. (2106) 'Movie written by AI algorithm turns out to be hilarious and intense', in *Arts Technica*, 9 June. Available at: <https://arstechnica.co.uk/the-multiverse/2016/06/sunspring-movie-watch-written-by-ai-details-interview/> (Accessed 19 September 2016).
- Nilsson, N. J. (1998) *Artificial Intelligence: A New Synthesis*. San Francisco, CA: Morgan Kaufmann.
- Nilson, N. J. (2010) *The Quest For Artificial Intelligence A History Of Ideas And Achievements*. Cambridge: Cambridge University Press. Available at: <https://events.google.com/io2016/> (Accessed: February 25, 2017)
- Nimavat, K. and Champaneria, T. (2017) 'Chatbots: An overview. Types, Architecture, Tools and Future Possibilities', *IJSRD - International Journal for Scientific Research & Development*, 5(7), pp. 1019 – 1026.
- Noble, D. (1993) *The Religion of Technology: The Divinity of Man and the Spirit of Invention*. 2nd Ed. New York: Random House.
- Norman, D. (2013) *The Design of Everyday Things*. Revised and expanded edition. New York: Basic Books.
- O'Neil, C. (2017) *Weapons of math destruction: how big data increases inequality and threatens democracy*. Penguin Books: London.
- Oswald, M. J., Urwin, S. and Barnes, G. C. (2018) 'Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and "Experimental" Proportionality', *Information & Communications Technology Law* (27) 2, pp. 223–50.
- Paik, N. J. (1964) *Robot K-456 [Performance]* Gyeonggi-do: Nam June Paik Art Center. Available at: [309](https://njpac-</a></p>
</div>
<div data-bbox=)

- en.ggcf.kr/archives/artwork/n009\_robot-k-456 (Accessed: 21 January 2019).
- Pask, G. (1968) *Colloquy of Mobiles* [Installation] London: Institute of Contemporary Arts
- Pask, G. (1975a) *Conversation, cognition and learning*. New York: Elsevier.
- Pask, G. (1975b) *The Cybernetics of Human Learning and Performance*, London: Hutchinson.
- Pask, G. (1976) *Conversation Theory, Applications in Education and Epistemology*, New York: Elsevier.
- Parise, S., Kiesler, S., Sproull, L., and Waters. K. (1999) 'Cooperating with life-like inter-face agents', *Computers in Human Behavior*. 15(2), pp. 123–142.
- Pattinson, Y. (2014) *Mute Conversation*, [Online Chatbot] Available at: <http://mute.conversation.ica.org.uk/> (Accessed 4 November 2016).
- Paul, C. (2003) *Digital Art*. London: Thames & Hudson Inc.
- Persson, P., Laaksolahti, J. and Lonqvist, P. (2002) 'Understanding Social Intelligence' in Dautenhahn, K., Bond, A., Cañamero, L., and Edmonds, B. (eds), *Socially Intelligent Agents – Creating Relationships with Computers and Robots*. Dordrecht: Kluwer Academic Publishers, pp. 21 – 28.
- Pichai, S. (2016), *Google I/O 2016*, 18 May 18. [online] Available at: <https://events.google.com/io2016/> (Accessed: 25 February 2017).
- Picard, R. (2000) *Affective Computing*. Cambridge, MA: MIT Press.
- Plymouth College of Art (2018) *Factory Settings* [exhibition] London: Tate Exchange, Tate Modern Switch House
- Realbotix (2019a) *Harmony* [online] Available at: <https://realbotix.com/> (Accessed: 17 February 2019).
- Realbotix (2019b) *RealDollX*. [online] Available at: <https://www.realdollx.ai/> (Accessed: 4 February 2019).
- Reeves, B. and Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge: Cambridge University Press.

- Reith, A. (2018b) 'Meet Emoti-OS: The chatbot helping to empower pupils. The computer program gives students the opportunity to engage in non-human, non-judgmental conversation' *Nesta*, 19 June. [online]. Available at: <https://www.nesta.org.uk/blog/meet-emoti-os-chatbot-helping-empower-pupils/> (Accessed: 22 June 2018).
- Renzel, P. (2017) 'How does online feminist activism offer the possibility for a truly intersectional feminist resistance?' *Medium*, 3 February. Available at: <https://medium.com/hack-for-chayn/how-does-online-feminist-activism-offer-the-possibility-for-a-truly-intersectional-feminist-651c6030ae16> (Accessed 12 October 2017)
- Reynolds, C. (1987) 'Flocks, Herds, and Schools: A Distributed Behavioural Model', *Computer Graphics*, (21)4, pp. 25-34.
- Ross, H. J. (2014) *Everyday Bias: Identifying and Navigating Unconscious Judgments in Our Daily Lives*. London: Rowman & Littlefield.
- Rothblatt, M. (2014) *Virtually Human: The Promise and the Peril of Digital Immortality*. 1st ed. New York: St. Martin's Press.
- Royle, N., (2003) *The Uncanny*. Manchester: Manchester University Press.
- Rhue, L. (2018) 'Racial Influence on Automated Perceptions of Emotions'. *Social Science Research Network*, 9 November. Available at: <https://ssrn.com/abstract=3281765> (Accessed: 12 February 2019).
- Samara, L. (1966) *Room No. 2* [installation], New York: Albright-Knox Art Gallery.
- Sanai, H. (1130 [1974]) *The Walled Garden of Truth*, Pendlebury, D. L. (trans.) London: Octagon.
- Sanoff, H. (2008) 'Multiple Views of Participatory Design', *ArchNet-IJAR: International Journal of Architectural Research* (2)1, pp. 57 – 69.
- Schonbrodt, F. D. and Asendorpf, J. B. (2011) 'The challenge of constructing psychologically believable agents', *Journal of Media Psychology: Theories, Methods, and Applications*, 23(2), pp. 100–107.  
doi:10.1027/1864-1105/a000040
- Searle, J. R. (1995) *The Construction of Reality*. New York: The Free Press.
- Searle, J. R. (1990) 'Collective intentions and actions' in Cohen, P. et al., *Intentions in communication*. MIT Press, pp. 401-416.

- Searle, J. R. (1983) *Intentionality: an essay in the philosophy of mind*.  
Cambridge: Cambridge University Press.
- Searle, J. R. (1984), 'Intentionality and its place in nature,' *Synthese*, 61, pp. 3-16.
- Searle, J. R. (1980) 'Minds, Brains, and Programs', in Haugeland, J., *Minds Design II: Philosophy, psychology, artificial intelligence*, MIT Press, pp. 183-204.
- Searle, John R. (1979) *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press, pp. 1–29.
- Searle, J. R. (1975) 'A Taxonomy of Illocutionary Acts', in Gunderson, K. (ed.) *Language, Mind and Knowledge*, Minneapolis: University of Minnesota Press, pp. 344–369.
- Searle, J. R. (1969) *Speech Acts. An Essay in the Philosophy of Language*.  
Cambridge: Cambridge University Press (34th printing 2011).
- SeeBotsChat Team (2017) *C:>bots.chat*. [live hack] Available at:  
<https://clips.twitch.tv/DifferentAnimatedGarbageHotPokket> (Accessed 14 March 2014).
- Segal, C. (2017) 'Stitch by stitch, a brief history of knitting and activism', *PBS News Hour*, 23 April. Available at:  
<https://www.pbs.org/newshour/arts/stitch-stitch-history-knitting-activism>  
(Accessed 14 March 2018).
- Sharkey, N. (2017) 'Mama Mia It's Sophia: A Show Robot Or Dangerous Platform To Mislead?' *Forbes*, 17 November. Available at:  
<https://www.forbes.com/sites/noelsharkey/#232b9c4c6b22> (Accessed: 17 November 2017).
- Sharkey, N. and Sharkey, A. (2006) 'Artificial intelligence and natural magic' *Artificial Intelligence Review*, 25(1-2), pp. 9-19.
- Schuler, D and Namioka, A. (1993) *Participatory Design: Principles and Practices*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schulman, D. and Bickmore, T. (2009) 'Persuading users through counselling dialogue with a conversational agent' in *Proceedings of the 4th International Conference on Persuasive Technology* (Persuasive '09). New York: ACM, 8 pages.

- Shanken, E. A. (ed.) (2003) 'From Cybernetics To Telematics The Art, Pedagogy, and Theory of Roy Ascott', in Ascott, R., *Telematic Embrace. Visionary Theories of Art, Technology, and Consciousness*. Berkeley: University of California Press, pp. 1-96.
- Sholette, G., Bass, C. and Social Practice Queens (2018) *Art as Social Action: An Introduction to the Principles and Practices of Teaching Social Practice Art*, New York: Allworth Press.
- Shum, H., He, X. and Li, D. (2018) 'From Eliza to Xiaoice: Challenges and Opportunities with Social Chatbots', *Frontiers of Information Technology & Electronic Engineering*, 19(1), pp 10–26.
- Simon, H. A. (1996) *The sciences of the artificial*. Cambridge, MA: MIT Press.
- Simonite, T. (2018) 'Ai Is The Future—But Where Are The Women?' *WIRED*, 17 August. Available at: <https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/> (Accessed, 17 August, 2018).
- Smith, Barry. (2003) 'John Searle: From Speech Acts to Social Reality' in Smith, Barry. (ed). *John Searle*. Cambridge: Cambridge University Press.
- Sommerer, C. and Mignonneau, L. (1993 – 1994) *A-Volve* [A-life installation] Available at: <http://www.medienkunstnetz.de/works/a-volve/images/2/> (Accessed, 17 August, 2018).
- Song, Y., Yan, R., Li, X., Zhao, D., and Zhang, M. (2016) 'Two are better than one: An ensemble of retrieval- and generation-based dialog systems', *arXiv*. Available at: <https://arxiv.org/abs/1610.07149> (Accessed: 17 November 2017).
- Spencer, G. (2018) 'Much more than a chatbot: China's Xiaoice mixes AI with emotions and wins over millions of fans', *Microsoft Asia News Center*. Available at: <https://news.microsoft.com/apac/features/much-more-than-a-chatbot-chinas-xiaoice-mixes-ai-with-emotions-and-wins-over-millions-of-fans/> (Accessed 27 October, 2018).
- Steiner, C. (2012) *Automate This How Algorithms Came To Rule Our World*. London: Penguin Books.
- Steinhelder, B. and Legrady, G. (2004) 'Interdisciplinary Collaboration in Digital Media Arts: A Psychological Perspective on the Production Process'. *Leonardo*, 37(4), pp. 315 - 321.

- Sterling, B. (2009) Design fiction. *interactions*, Volume 16 (Issue 3) 20 - 24.
- Sterling, B. (2005) *Shaping Things*. Cambridge Massachusetts: MIT Press.
- Stoller, D. (2003) *Stitch 'n Bitch: The Knitter's Handbook*. New York: Workman Publishing.
- Stone, P. et al. (2016) *Artificial intelligence and life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016*. Stanford University. Available at: <https://ai100.stanford.edu/2016-report> (Accessed 10 October 2018).
- Stone, Z. (2017) 'Everything You Need To Know About Sophia, The World's First Robot Citizen', *Forbes*, 7 November. Available at: <https://www.forbes.com/sites/zarastone/2017/11/07/everything-you-need-to-know-about-sophia-the-worlds-first-robot-citizen/#6d53c31846fa> (Accessed: 7 November 2017)
- Suddendorf, T. Addis, D. and Corballis, M. (2009) 'Mental time travel and the shaping of the human mind', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), pp.1317-1324.
- Schwartz, O. (2019) 'Don't look now: why you should be worried about machines reading your emotions', *The Guardian*, 6 March. Available at: [ [https://www.theguardian.com/technology/2019/mar/06/facial-recognition-software-emotional-science?CMP=Share\\_iOSApp\\_Other](https://www.theguardian.com/technology/2019/mar/06/facial-recognition-software-emotional-science?CMP=Share_iOSApp_Other) (Accessed: 14 February, 2019).
- The Telegraph (2017) 'Thousands of people are watching two Google Home chatbots argue with each other', *The Telegraph: Technology*, 9 January. Available at: <http://www.telegraph.co.uk/technology/2017/01/09/thousands-people-watching-two-google-home-chatbots-argue/> (Accessed 14 March 2017).
- The Telegraph (2014) 'Thousands of people are watching two Google Home chatbots argue with each other', *The Telegraph, Technology*, Available at: <http://www.telegraph.co.uk/technology/2017/01/09/thousands-people-watching-two-google-home-chatbots-argue/> (Accessed 14 March 2017).

- Tharp, M. and Tharp, S. M. (2013) 'Discursive Design Basics: Mode And Audience', in *Proceedings of Nordic Design Research Conference 2013*, Copenhagen-Malmo, pp. 406 - 409.
- Tharp, M. and Tharp, S. M. (2019) *Discursive Design: Critical, Speculative, and Alternative Things*. Cambridge, Ma: MIT Press.
- Tonkinwise, C. (2015) 'How We Intend to Future: Review of Anthony Dunne and Fiona Raby, *Speculative Everything: Design, Fiction, and Social Dreaming*', *Design Philosophy Papers*, 12(2), pp. 169-187. doi: 10.2752/144871314X14159818597676
- Thornberg, R. (2012) 'Informed grounded theory', *Scandinavian Journal of Educational Research*, 56930, pp. 243-259. Available at: <http://dx.doi.org/10.1080/00313831.2011.581686> (Accessed 12 Aug. 2016).
- Tromble, M. (2005) *The Art and Films of Lynn Hershman Leeson: Secret Agents, Private I*. Berkeley: University of California Press.
- Tucker, P. (2014) *The Naked Future: What Happens in a World That Anticipates Your Every Move?* New York: Penguin Group.
- Turkle, S. (ed.) (2007a) *Evocative objects: Things we think with*. Cambridge, MA: MIT Press.
- Turkle, S. (2007b) 'Authenticity in the age of digital companions', in Anderson, M. and Anderson, S. L. (eds.), *Machine Ethics*. New York: Cambridge University Press, pp. 62 – 79.
- Turkle, S. (2005) 'Relational Artifacts/Children/Elders: The Complexities of CyberCompanions', *Proceedings of the CogSci Workshop on Android Science*, Stresa, Italy, pp. 62-73.
- Turkle, S. (2004) *Relational Artifacts*, Final report on National Science Foundation proposal to the National Science Foundation SES-0115668. Cambridge, MA: National Science Foundation.
- Turkle, S. (2003) *Technology and human vulnerability*. The Harvard Business Review, September.
- Turkle, S. (1997) 'Seeing Through Computers: Education in a Culture of Simulation', *The American Prospect*, 31, pp. 76-82.
- Turkle, S. (1995) *Life on the Screen: Identity in the Age of the Internet*. London: Phoenix.

- Turkle, S. (1984) *The Second Self: Computers and the Human Spirit*. Reprint, London: The MIT Press, 2005.
- Turkle, S., Breazeal, C., Dasté, O., and Scassellati, B. (2006a) 'First encounters with Kismet and Cog: Children's relationship with humanoid robots' in Messaris, P. and Humphreys, L. (eds.) *Digital media: Transfer in human communication*. New York: Peter Lang, pp. 313 – 330.
- Turkle, S., Taggart, W., Kidd, C. D. & Dasté, O. (2006b). 'Relational artifacts with children and elders: The complexities of cybercompanionship'. *Connection Science*, 18(4), pp. 347–361.
- Uğur, S. (2013) *Wearing Embodied Emotions A Practice Based Design Research on Wearable Technology*. Milan: Springer.
- The University of Plymouth (2018) *Research Ethics Policy*. Available at: [https://www.plymouth.ac.uk/uploads/production/document/path/12/12337/General\\_Research\\_Ethics\\_Policy\\_\\_final\\_draft\\_V1.0\\_.pdf](https://www.plymouth.ac.uk/uploads/production/document/path/12/12337/General_Research_Ethics_Policy__final_draft_V1.0_.pdf) (Accessed: 1 July 2018).
- The University of Plymouth (2018) *Code of Good Research Practice*. Available at: [https://www.plymouth.ac.uk/uploads/production/document/path/12/12338/Code\\_of\\_Good\\_Research\\_Practice\\_\\_final\\_draft\\_V1.0\\_.pdf](https://www.plymouth.ac.uk/uploads/production/document/path/12/12338/Code_of_Good_Research_Practice__final_draft_V1.0_.pdf) (Accessed: 1 July 2018).
- The University of Plymouth of Plymouth (2018) *Research Data Policy*. Available at: [https://www.plymouth.ac.uk/uploads/production/document/path/6/6913/Research\\_Data\\_Policy.pdf](https://www.plymouth.ac.uk/uploads/production/document/path/6/6913/Research_Data_Policy.pdf) (Accessed: 1 July 2018).
- Van Camp, J. (2017) 'Review: 'Jibo Social Robot'', *Wired*, 11 July, Available at: <https://www.wired.com/2017/11/review-jibo-social-robot/> (Accessed: 13 September 2017).
- Vanian, J. (2018) 'Unmasking A.I.'s Bias Problem'. *Fortune*, 25 June. Available at: <http://fortune.com/longform/ai-bias-problem/> (Accessed 3 December 2018).
- Vasilakos, A. and Pedrycz, W. (2006) *Ambient Intelligence, Wireless Networking, and Ubiquitous Computing*. Norwood, MA: Artech House.
- Verbeek, P. (2014). 'Postphenomenology of Technology' in Scharff, R. C. and Dusek, V. (Eds.), *Philosophy of Technology The Technological*



- Condition: An Anthology*. Second Edition. Chichester: John Wiley & Sons, pp. 562 – 571.
- Verbeek, P. (2011) *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: The University of Chicago Press.
- Verbeek Peter-Paul (2009) 'Ambient Intelligence and Persuasive Technology: The Blurring Boundaries Between Human and Technology', *Nanoethics*, 3 (3), pp. 231-242.
- Verbeek P. (2008a) 'Cultivating humanity: towards a non- humanist ethics of technology', in Olsen J-KB, Selinger E, Riis S (eds.) *New waves in philosophy of technology*. Hampshire: MacMillan.
- Verbeek, Peter-Paul (2008b) 'Cyborg intentionality: Rethinking the phenomenology of human-technology relations', *Phenomenology and the Cognitive Sciences*, 7(3). pp. 387-395.
- Verbeek, P. (2006a) 'Persuasive Technology and Moral Responsibility. Towards and Ethical Framework for persuasive Technology' Available: [http://www.utwente.nl/gw/wijsb/organization/verbeek/verbeek\\_persuasive06.pdf](http://www.utwente.nl/gw/wijsb/organization/verbeek/verbeek_persuasive06.pdf). (Accessed March 21, 2011)
- Verbeek, P. (2006b) 'Moralizing Technology: On the morality of technological artifacts and their design', *Workshop 'Methodologies for the moral evaluation of technology development*, Ethics Institute. Utrecht University, The Netherlands, March 24-25, pp. 1-25.
- Verbeek, P. (2005) *What Things Do*. 1st ed. University Park, Pa: Pennsylvania State University Press.
- Verbeek, P.P. (2003) 'Material Hermeneutics', *Technè*, Spring 2003, 6(3), pp. 181-184.
- Verbeek, P. and Kockelkoren, P. (1998) 'The Things That Matter.' *Design Issues*, 14 (3), pp. 28–42.
- Verbeek, P. and Slob, A. (2006) 'User behavior and technology development: Shaping sustainable relations between consumers and technologies'. Dordrecht: Springer.
- Vernon, D. (2014) *Artificial Cognitive Systems A Primer*. The MIT Press: Cambridge, Massachusetts, London, England.

- Vorn, B. (2012) *DSM-VI* [Robotic installation] Available at:  
<http://billvorn.concordia.ca/menuall.html> (Accessed: January 22, 2019)
- Vorn, B. and Demers, L. (1997) *LA COUR DES MIRACLES* [Robotic installation] Available at:  
<https://billvorn.concordia.ca/robography/Cour.html> (Accessed: January 22, 2019).
- Wiener, N. (1989) *The Human Use of Human Beings: Cybernetics and Society*. London: Free Association Books.
- Wiener, N. (1965) *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge, Massachusetts: MIT Press.
- Weedon, C. (1987) *Feminist Practice and Poststructuralist Theory*. Oxford: Blackwell.
- Westworld (2016), TV Series, Season 1, USA: HBO.
- Weizenbaum, J. (1966) 'ELIZA—a computer program for the study of natural language communication between man and machine' in *Commun ACM*, 9(1), pp. 36–45.
- Weizenbaum, J. (1976) *Computer Power and Human Reason: From Judgement to Calculation*. New York: W. H. Freeman and Company.
- Weiser, M., (1991) The computer for the twenty-first century. *Scientific American*, 165, pp. 94–104.
- Whittaker, M. et al. (2018) *AI Now Report 2018*. New York University: The AI Now Institute [online] Available at:  
[https://ainowinstitute.org/AI\\_Now\\_2018\\_Report.pdf](https://ainowinstitute.org/AI_Now_2018_Report.pdf) (Accessed 12 December 2018).
- Wilding, F. (2006) 'Where is Feminism in Cyberfeminism?' *Neme*, 28 March. [Online] Available at: <http://www.neme.org/texts/cyberfeminism> (Accessed 12 October, 2017).
- Williams, M. (2018) 'Facebook 2018 Diversity Report: Reflecting on Our Journey', *Facebook Newsroom*, 12 July [Online] Available at:  
<https://newsroom.fb.com/news/2018/07/diversity-report/> (Accessed October 2018).
- Wittgenstein, L. (1958) *Philosophical investigations*. Oxford: Blackwell.

- Wolf, A. (2018) 'Life And Technology the Binary Of Nam June Paik' **IN** *Gagosian Quarterly*, Summer 2018 Issue. Available at: <https://gagosian.com/quarterly/2018/10/16/life-and-technology-binary-nam-june-paik/> (Accessed 12 November 2019).
- Wooldridge, M. and Rao E. (eds.) (1999) *Foundations of Rational Agency*. Boston, MA,: Kluwer Academic Publishers.
- World Wide Web Foundation (2017) *ARTIFICIAL INTELLIGENCE: The Road Ahead in Low and Middle-Income Countries*. [online]. Available at: [https://webfoundation.org/docs/2017/07/AI\\_Report\\_WF.pdf](https://webfoundation.org/docs/2017/07/AI_Report_WF.pdf) (Accessed 10 October 2018).
- Yao, M. (2018) 'Technical Approaches For Building Conversational Ai'. *TopBots* Available at: <https://www.topbots.com/building-conversational-ai/> (Accessed: 14 February 2019).
- Yan, R, (2018) "Chitty-Chitty-Chat Bot": Deep Learning for Conversational AI' *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, (IJCAI-18)*, pp. 5520 – 5526.
- Zhou, L., Gao, J., Li, D. and Shum, H-Y. (2018) 'The Design and Implementation of Xiaolce, an Empathetic Social Chatbot'. arXiv:1812.08989 [cs.HC], pp. 1 -26. Available at: <https://arxiv.org/abs/1812.08989> (Accessed: 14 February 2019).

## Publications

- Aga, B. and Phillips, M. (2017) 'This is Where We Are – collectively living, working and being, with and through predictive algorithms' in J. Bowen, with N. Lambert and G. Diprose (eds.) *EVA London 2017: Electronic Visualisation & the Arts. Proceedings of a conference held in London 11th-13th July*. BCS, The Chartered Institute for IT.
- Aga, B. (2017a) 'Living Backwards: The [Artificially] Augmented Mental Image', in Phillips, M & Thomas, P. *The Fourth International Conference on Transdisciplinary Imaging at the Intersections of Art, Science and Culture 2016*, Plymouth, UK, Intellect Publishing: London
- Aga, B. (2017b) 'Capture the Rapture'. In: Aceti, L., Thomas, P. and Colless, E. eds. *Cloud and Molecular Aesthetics. Leonardo Electronic Almanac Book*, Volume 22, Issue 1, Leonardo/ISAST: San Francisco, pp. 26 – 34.

## Conference Presentations & Talks

- Aga, B., and Manton, C. (2018a) *Human Bias In Artificial Intelligence: Can we fix it?* KWMC, 29 November, Bristol, UK.
- Aga, B., and Manton, C. (2018b) *Disobedient AI*. Random String, 16 November, Coventry, UK.
- Aga, B., and Manton, C., (2018c) *Things That Talk*. AI In Art – A FACT Late Night, 25 October, Liverpool, UK.
- Aga, B. (2017) *Data Deamons*. The 21st International Conference on Digital Research in the Humanities and Arts, Plymouth University, 11 October, Plymouth, UK.
- Aga, B., Sinker, R. and Phillips, M. (2017) *Collaborating with Algorithms*. Technarte, 12 May. Bilabo, Spain.
- Aga, B. (2017) *Living Backwards: Artificially Augmented Memories*. Digital Memory, 16 January, Plymouth, UK.

Aga, B. (2016) *Living Backwards: The [Artificially] Augmented Mental Image*. The Fourth International Conference on Transdisciplinary Imaging at the Intersections of Art, Science and Culture, 1-3 July, Plymouth, UK.

## Workshops

Aga, B. (2018) *Things that talk*. SWCTN, 15 December, Plymouth University, Plymouth, UK.

Aga, B., and Manton, C., (2018d) *Women Reclaiming AI*. Random String, 16 November, Coventry, UK.

Aga, B., and Manton, C., (2018e) *Women Reclaiming AI*. KARST, 3 September, Plymouth, UK.

Aga, B., and Manton, C., (2018f) *The Infinite Guide Prototyping Workshop*. FACT, 30 June – 1 August, Liverpool, UK.

Aga, B., Manton, C., Watkins, D., Phillips, M. (2018) *Skunk Works*. University of Plymouth, 21 June, Plymouth, UK.

Aga, B., Manton, C., *Women Reclaiming AI*. Knowle West Media Centre, Bristol, 13 June 2018.

Aga, B., Reith, A. and Hunt, C. (2017), *A programme of Data-Labs*. A series of five Data-Labs for the Emoti-OS project, Plymouth School of Creative Arts, Bristol, 5 May - 1 December 2017.

Aga, B., Phillips, M., and Reith, A. (2016) *Creative Data Labs: Changing perception – Data-driven collaborative practice*. 17-21 October, Plymouth School of Creative Arts, Plymouth, UK.

Aga, B., Phillips, M., Nutbean, L, Emmett, M. (2016) *TIWWA: Data Workshop 2*, 27-28 April, Plymouth, UK.

Aga, B., and Phillips, M. (2016) *TIWWA: Data Workshop 1*, 26 January, Tate Brittan, London, UK.

## Exhibitions

- i-DAT Collective, Tate Collective and InterCity (2016) *This is Where We Are (TIWWA)* [interactive installation]. London: Tate Modern Switch house.
- i-DAT Collective, Plymouth School of Creative Arts, InterCity and Controlled Frenzy (2017) *Emoti-OS 1.0* [interactive installation]. Plymouth: Plymouth School of Creative Arts.
- i-DAT Collective, Plymouth School of Creative Arts, InterCity and Controlled Frenzy (2018) *Emoti-OS 2.0* [interactive installation]. London: Tate Modern Switch house.
- Aga, B. and Manton, C. (2018) *Women Reclaiming AI* [online prototype]. Available at: <https://www.womenreclaimingai.com> (Accessed: 5 May 2018).
- i-DAT Collective, InterCity, FACT and KARST (2018) *The Infinite Guide* [interactive installation]. Plymouth: KARST.