A STRATEGY ROADMAP FOR HUMAN-LIKE COMPUTING

WHAT IS HUMAN-LIKE COMPUTING?

Human-like Computing (HLC) research aims to endow machines with human-like perceptual, reasoning and learning abilities, which support collaboration and communication with human beings.

INTRODUCTION

The idea of intelligent machines that can interact with humans has been around for many years, and if you count automatons, centuries. In recent times, studies in AI and machine learning have led to advances in computing allowing the deployment of self-driving cars and machines that can take on humans at Chess and Go and beat them.

The impressive performance of these machine learning systems relies on computer power to enable them to learn from huge numbers of examples. For example, Tesla's Autopilot program is built on 780 million miles of travel data (https://www.technologyreview.com/s/601567/tesla-tests-self-driving-functions-with-secret-updates-to-its-customers-cars/) and AlphaGo probably played hundreds of millions of games prior to beating Go professional Lee Sedol (http://www.nature.com/news/what-google-s-winning-go-algorithm-will-do-next-1.19573). Humans on the other hand can extrapolate from a small number of examples.

In addition, these systems cannot explain their reasoning. The inscrutability of statistically-based machine reasoning limits its applications in many areas, particularly when the outcomes have impact on humans. This is the case, for instance, in safety critical areas, law cases, medical diagnosis, job interviews, credit assessment and building or vehicle design. A system that could produce explainable human-like reasoning would have applications in many areas.

With this background the idea of creating systems that could emulate human performance was seen as a potentially important new direction for the Engineering & Physical Sciences Research Council's (EPSRC) Information and Communication Technologies (ICT) theme. The topic was incorporated into its strategy and discussions started about how an initiative could be developed.

The Engineering and Physical Sciences Research Council (EPSRC) recognised that this was a new approach and would require research from a range of disciplines within cognitive science and computer science. A workshop to explore the topic was organised in February 2016 with participants drawn from a variety of backgrounds, including the community supported by the Economic and Social Research Council (ESRC). This meeting reinforced the idea that the development of HLC systems was a timely and worthwhile goal. However, it was not clear what a research programme would look like although a number of challenge areas were identified. The workshop report is available here: https://www.epsrc.ac.uk/newsevents/pubs/humanlikecomputing/

Following the meeting, Professor Alan Bundy (School of Informatics, University of Edinburgh) produced a manifesto to describe the aims and objectives of research in HLC. Building on the February meeting and this manifesto and to attempt to clarify how to initiate a coherent research programme in HLC a second workshop on the topic was organised by Professor Stephen Muggleton (Department of Computing, Imperial College London) under the Machine Intelligence Series http://mi20-hlc.doc.ic.ac.uk/.

This roadmap has been synthesised from the output of the 2 meetings and the discussions of the manifesto.

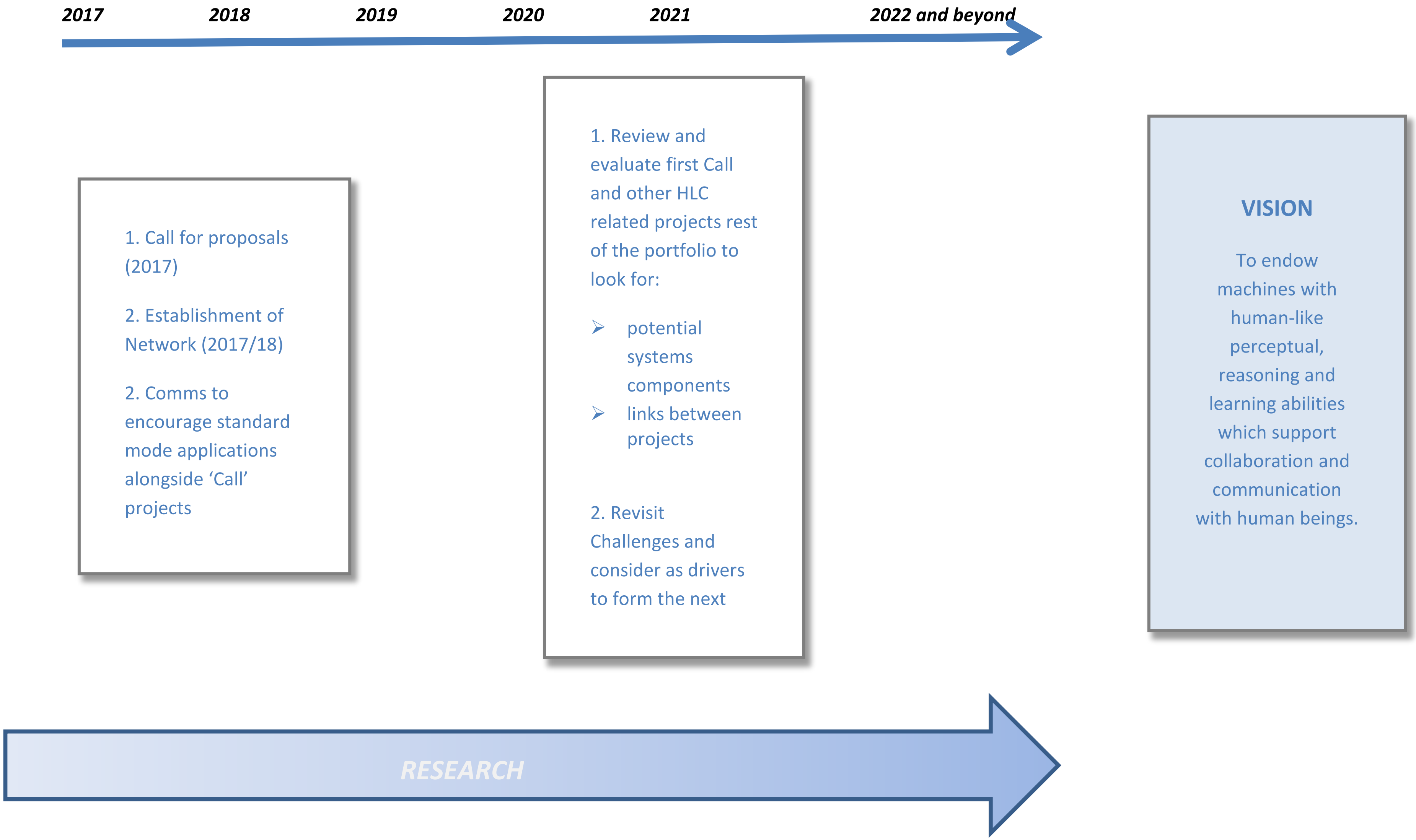


Fig. 1: HUMAN-LIKE COMPUTING – Key Strategic Milestones

THE ROADMAP

MOTIVATION

- To enable better communication and collaboration between humans and machines, especially in the context of hybrid teams in the workplace.
- To support the generation by Machine Learning of explicit and debuggable hypotheses and programs which incorporate and support reasoning, which can be understood by humans.
- To improve our understanding of human cognition via computational modelling.
- To inspire new forms of computation based on human cognition, especially on tasks where humans currently exhibit superior abilities.

REQUIREMENTS

(A) OF HUMAN-LIKE SYSTEMS

- Machines must provide explanations of their suggestions, decisions and actions that are intelligible to their human users. When the underlying processes are too large, complex and/or opaque, appropriate abstracts must be generated.
- Similarly, machines must be able to understand and act on instructions and information from the human users.
- Machines must be able to learn from a small number of examples, sometimes just one, using explicit and incrementally generated background knowledge to select appropriate hypotheses. They must be able to learn complex structures and to construct hypotheses that explain their observations.
- Machines must be able to construct their own world models, including models of humans and other machines, and to involve them as the environment, their tasks and their teams change. This includes changes to the language and logic of their models, as well as the beliefs stored in them. This may require exploration of the environment, including designing and executing experiments to distinguish between competing hypotheses.

(B) OF RESEARCH

- The research carried out in the name of Human-like Computing must be multidisciplinary and involve research that is at the cutting edge and represents the state-of-the-art in both the cognitive and computing science components.

SCOPE

Human-like Computing goes beyond designing improved AI or machine learning systems, and it is not about incorporating findings in neuroscience.

The systems envisaged will be operating in ways unlike current AI or machine learning systems. Those systems are not human-like although they exhibit impressive performance and carry out tasks that could be performed by humans. They rely on computationally-based solutions such as neural nets, they need large datasets and impressive computer power in order to operate.

Human-like systems will be informed by biological models, but not mimicking the neuroscience. They will operate in ways similar to humans in that they will not require the level of data resources needed by AI and machine learning systems. They will learn in a way that is compatible with communication of the learned knowledge to human beings – they will possess an ability to communicate with humans to explain and describe the processes they have taken that have led to the decisions and actions they take

The research that needs to be supported to derive human-like systems will need to be collaborative between cognitive and computer scientists. The research will need to be inspired by cognitive science. Models of perception, cognition and development will inspire the technology such that the development of systems is informed by the way humans learn and behave.

Trends and drivers

A. Limitations of current AI.
B. Opportunities for systems that are more human-like.
C. The benefits to the computer and cognitive sciences of joint working: the research carried out has the potential to contribute to solutions in both domains.
D. The UK is in a good position internationally to capitalise on the potential and the opportunities.

These are described in more detail in the following sections.

(A) The limitations of current AI and statistical machine learning

Currently AI can do things that used to be done by humans, but not always in a human-like way. Existing AI systems cannot emulate general intelligence, for example like that of a child. Many existing approaches are either unable to deal with uncertainty or do so in a non-explicit fashion.. The success of AI systems today is based on large datasets, improved computing power and improved techniques such as pattern recognition. Its successes are also in narrow domains. Current AI systems lack the flexibility, generality and creativity exhibited by humans.

Success is also measured in terms of data with predictive accuracy being based on increasing data and statistical methods being used to measure how well machine learning performs.

The systems are also often "black boxes": it is not possible to see how they are working and how they are reaching the decisions they make. This limits their usefulness in many domains such as legal and medical reasoning and engineering design where knowing how the decision has been made is important.

There are many other ways in which current AI systems do not perform in a way that is humanlike and these are summarised in the table below.

Characteristics	Human	Statistical
Examples per concept	Few (around 1)	Many (> 10k)
Concepts	Many	Few
Background knowledge	Large	Small
Structure	Modular, reusable	Monolithic

Table 1: Comparison of Human and AI characteristics

Humans find it easy to transfer knowledge across domains and to recognise concepts. We are also good at using background constraints to identify what is relevant in a particular context – restructuring the problem to make it easier to solve. These are tasks that are beyond current AI systems.

Donald Michie (https://en.wikipedia.org/wiki/Donald_Michie) one of the founders of the field of artificial intelligence described 3 kinds of learning[1]:

> **Weak**: where a system improves its performance on unseen data based on learning from a sample of data
>
> **Strong**: where the system is able to communicate the learned hypotheses in explicit symbolic form
>
> **Ultra-strong**: where the user is able to comprehend the system's output and its possible consequences.

Current machine learning systems are weak. Human-like systems offer the opportunity for novel routes to develop strong and ultra-strong learning systems.

Such systems offer the possibility of using machine learning to explain how to improve human skills rather than replace human beings in the workplace. For instance, by analysing facial expressions of actors, such a machine could explain to a trainee how to portray emotion in a better fashion. Also by analysing the leg movements of a patient recovering from an accident, observations of the patient could be used to debug errors in a machine learned model of the patient's walking motion, and these errors could be explained in explicit form to the patient as well as to feed into the medical process, by helping a clinician to understand the nature of the remaining injury and the trajectory of the recovery.

(B) Opportunities

Opportunities for human-like systems abound in those domains where data are sparse, background knowledge is key, where complex structures must be learnt and explanation is essential. These are areas that current AI finds challenging or is unable to deal with. There are increasing opportunities where humans and systems could work together such as autonomous systems and in hybrid teams involving humans working collaboratively with robots, for example in manufacturing or tutoring contexts.

For these opportunities to be realised new research needs to be put in place that takes a different direction from traditional and current AI. We need research that exploits advances in cognitive science in understanding how humans and animals learn and reason, and link these to computer science to begin to build the components that could form systems that perform in a humanlike manner.

(C) Benefits

Research that we envisage in developing humanlike systems will have broad benefits. These will include not only in improving our ability to use machines to tackle problems that are currently challenging for computers, but also in providing a stimulus to cognitive science, as the collaboration will result in an increased understanding of human behaviour in the context in question.

Early AI was close to human development and cognition research, but as computing power has increased the fields have drifted apart as AI has had less need of insights from cognitive science using computer power instead. However, that link needs to be re-established. Research that is truly ground-breaking in all its component disciplines will lead to advances both in the computer and the cognitive sciences, leading to a better understanding of the human mind and with human abilities serving as a source of challenges and inspiration for computer science.

[1] Quoted in Besold, et al *Towards Ultra-Strong Machine Learning – Comprehensibility of Programs Learned with ILP* http://mi20-hlc.doc.ic.ac.uk/short_presentations/Schmid.pdf

To realise these benefits will require the formation of links between the component research projects as well as to the wider communities. This might require coordination by the community through, for example, a network or similar.

(D) THE UK'S STATUS

A humanlike computing initiative is timely. Although many groups worldwide are realising the limitations of current AI systems and the need to inform future developments from studies in cognitive science, the UK is well placed to capitalise on these opportunities. It has long had strengths in AI and cognitive science and those strengths remain. British researchers were among the founders of these disciplines. The US also has strengths in this area through, for example, programmes initiated by DARPA and initiatives in explainable AI, but UK researchers have good links with those groups.

Within Europe, EU programmes, such as Esprit, Framework and H2020, have helped build strong links between the UK and European researchers, with the UK often leading the way. For example, a recent series of Dagstuhl workshops held on Inductive Programming was jointly organised by leading UK and German researchers. In addition, the UK AISB series of conferences had meetings elsewhere in Europe, which led to the EU ECAI series – now one of the 3 top international AI conferences. AISB was the first AI society and the Machine Intelligence series was the first international AI conference series.

RESEARCH GOALS

1. COGNITIVE SCIENCE

Research in Cognitive Science and AI has drifted apart over the last 30 years, but the disciplines need to work together if the goal of human-like computing systems is to be realised. Human-like systems will need to be informed by an understanding of how humans and animals behave. In turn, the development of human-like computing can help drive cognitive science.

Recent AI developments seem to have been more about improvements in pattern recognition and motor skills. Machine Learning, game playing, and other applications have employed massive computer resources and data in a way that humans could not emulate, e.g., training on a million Go games.

Cognitive science, on the other hand emphasises the role of mental structures and processing in complex behaviours. There have been recent insights into psychology: learning, language, and memory. Human like systems will need to incorporate these and other aspects of being human and the challenge will be how to implement them into the design of such systems.

In the animal kingdom, human thinking is uniquely powerful, productive and abstract. We can reason, use natural language and solve novel complex problems. However, there are limitations to the way we think about the world. Understanding the way we learn about the world would be an important driver of the development of human-like systems. For example, human learning is exploratory, it is driven by external cultural influences. It is also characterised by prolonged development and a social scaffolding which supports that development. However, our capacity to learn can be limited (which can be an advantage in some situations), we are destructible (which makes us flexible) and we show increasing levels of social interactivity and cooperation.

Can we understand how humans develop and model this into systems? Could we, for example, develop a system driven by curiosity?

We could start by identifying the challenges and problems and breaking them down to understand how humans and animals tackle them. The lessons learned could then be applied to machine design.

The topics that need to be understood and built into computing models include:

- How do we create systems that create and revise their mental representations to fit the problem being addressed?
- What drives humans to act? The stimuli include emotions, rewards, et cetera. How can these be modelled and incorporated into human-like systems?
- How do we, as individuals, predict the behaviour of other humans? Human-like systems will need similar abilities.
- To what extent should humanlike computing include bad human traits? Current software can make decisions that are biased (see, for example criticism of Google's racial bias https://www.washingtonpost.com/news/morning-mix/wp/2016/06/10/google-faulted-for-racial-bias-in-image-search-results-for-black-teenagers/). Studying the bad aspects of human behaviour might inform the design of systems that are less biased.

2. Memory and forgetting: in people and AI systems

What is the value of forgetfulness? Humans forget, computers do not. Is there value in what humans do? Results in cognitive science indicate that human abilities in selective focussing on relevant background knowledge when reasoning about a novel task is intimately connected to highly selective mechanisms for retrieving memories. This begs the question as to whether human-like "forgetfulness" could have a beneficial effect when used by machines to learn in the context of overwhelming amounts of background knowledge. In this case, human background knowledge appears to be ordered in terms of ease of retrieval. More frequently used background knowledge is brought to the fore more rapidly, supporting efficient and pertinent reasoning. Such mechanisms could well be valuable if adapted for the purposes of machine reasoning, particularly in the context of making inferences which can be easily followed by human beings. Testing such an idea would require an AI system implementation together with human trials evaluated by Cognitive Scientists.

3. High to low level learning: bridging the gap

While low-level learning and cognition may be adequately handled by existing black-box approaches, a distinctive feature of human cognition lies in the ability to learn explicit and communicable knowledge. Explicit white-box approaches to Machine Learning exist with the areas of Logic-based and Statistical Relational Learning. In application areas involving closely coupled interactions between human and machines involving spatial and perceptual tasks it is vital to find ways of bridging the gap between low-level black-box approaches and high-level white-box approaches. Success in this area would bring the benefits of recent advances within both these forms of Machine Learning, leading to systems which reason in an explicit and communicable form while performing effective low-level signal analysis.

4. Comprehensibility: language meaning, learning, explanation and accountability

One of the concepts at the heart of human-like computing is the notion that to inform the human any system will have to produce something that is comprehensible and that can be demonstrated objectively. Meaning is important here, more so than simply performance. To concentrate on performance alone would exclude humans and inhibit their interaction with the system.

Learning is another important component of such systems. The main difference between current AI systems and humans is that human learning is often based on "small data" with humans requiring significantly fewer examples than machines (see Table 1) and their ability to use background knowledge to compensate for the lack of data.

Interaction between machine and humans will be an important part of any complete system. To achieve interaction, however, will require mutual understanding of the joint activity. Each partner, whether human or computer, will need to be able to infer what the other is trying to do. This is sometimes known as "plan recognition". Both machines and humans need an understanding of what each knows and does not know.

Related to this is the concept of trust. Interaction requires trust so we need to know the characteristics that enable us to know who to trust. How would a machine know this? What are the cues?

As systems become more human like and autonomous the relationship with humans becomes important. For example, if such systems do better than humans why not rely on the system alone to carry out the tasks without human intervention? However, if this is done, what happens when something goes wrong? Who is responsible? It will be important to incorporate an understanding of ethics into the system design and this includes addressing the question of what makes a human-like system different from other engineering. Other components (for example tyres) fail and cause fatalities. What would be unique about a human-like system?

5. Verbal versus non-verbal communication

Humans use a variety of ways to communicate effectively. These include not only verbal descriptions, but also diagrams, gestures, facial expressions and physical demonstrations of tasks. In order to support more effective human-machine communication an improved understanding of the cognitive and computational aspects of such modes of communication needs to be better understood.

6. Small-data learning: the AI challenge and connection to cognition

Statistical machine learning techniques typically require thousands of examples to reliably learn a single concept. By contrast human beings regularly learn concepts from a small number of instances. This human cognitive ability is hard to explain using the available theory of computational learning, since it implies a highly tuned and biased form of learning. However, unlike the standard machine learning setting, humans simultaneously learn many inter-related concepts at the same time, which may be critical to small-data learning ability. Better understanding of this phenomenon will not only provide an improved model of human learning, but also many engineering applications in which data is sparse and potentially expensive to obtain.

Some animals often play in order to learn. Those species that do tend to be the big-brained ones and so play seems to be evolutionarily important. Could this be a strategy to employ with, for example robots, developing systems that play to learn?

Unrewarded object play is widespread among nonhuman animals, though the amount and complexity of exploratory behaviour varies widely (reviewed in Power 2000). This propensity to explore objects is thought to confer adaptive benefits by allowing individuals to develop motor skills and engage in novel behaviour patterns that may lead to innovative foraging strategies (Bateson & Martin 2013; Auersperg 2015). Some of the most pronounced and complex (combinatorial) object play is found among large-brained birds and primates (Glickman & Sroges 1966; Parker 1974; Torigoe 1985; Auersperg *et al.* 2015).

Such play might provide opportunities to learn about objects and potentially develop abstract concepts about "invisible" physical properties such as weight and rigidity, which could provide a route to flexible problem solving and tool use (Call 2012). The exploratory play of young humans is thought to scaffold such learning, whereby infants construct knowledge about an object's functionality through repeated interaction (Piaget 1952; Gibson 1988; Greif & Needham 2011). However, whether any animals gain information about object properties through exploration is little studied.

7. Social robotics: theory of mind, sense of self, context clues, spatial reasoning

Current developments are reaching the point where robots and softbots can work with and will be active in the same space as humans. To develop this further requires more research into the way humans behave (including trust) and using this to inform the way robots are designed. The development of complex social interactions between robots and humans will need a sense of self – including body image and perception.

These interactions may involve multiple human-like systems behaving as agents. These agents must have world models but when they interact the models must align. This means that the representations must be dynamic, evolving over time and be automated. Multiple agents must also be able to communicate and operate together.

The world as we perceive it is causally live and the causal potential goes beyond our own actions. That potential must be inferred.

8. INDUCTIVE PROGRAMMING: PSYCHOLOGY AND APPLICATION

The aim of automatically programming computers using examples has a long history. However recent implementations of such inductive programming techniques, such as Microsoft's FlashFill system have been successfully released as mass market products. These systems are still limited in their applications, and are still limited to construction of relatively simple and limited programs. Research in this area promises wide-scale application and high-value in human-like computing.

There is other research in this area that begins to point the way using of procedural representations of knowledge: Tenebaum's (MIT) character drawing programs as representations of the characters; Muggleton's (Imperial College) learning of a logic program; Chater's (University of Warwick) use of intervention of value passing by programs to model counterfactual reasoning. So, automated programming can be seen as a learning technique to construct such programs.

One aim could be to develop programs that are creative and that can produce, new characters, or new programs. There is no reason why everything should be problem-based.

GRAND CHALLENGES

Participants at the original workshop in February 2016 did identify a number of challenges or scenarios, which they felt could be used to drive research. On reflection the state-of-the-art is such that these are probably one step removed from the kind of research programme that we need to set in place in the immediate future. However, they do provide some potential application areas and, as such, might help to inspire the initial steps that will lead to the building blocks of human-like systems.

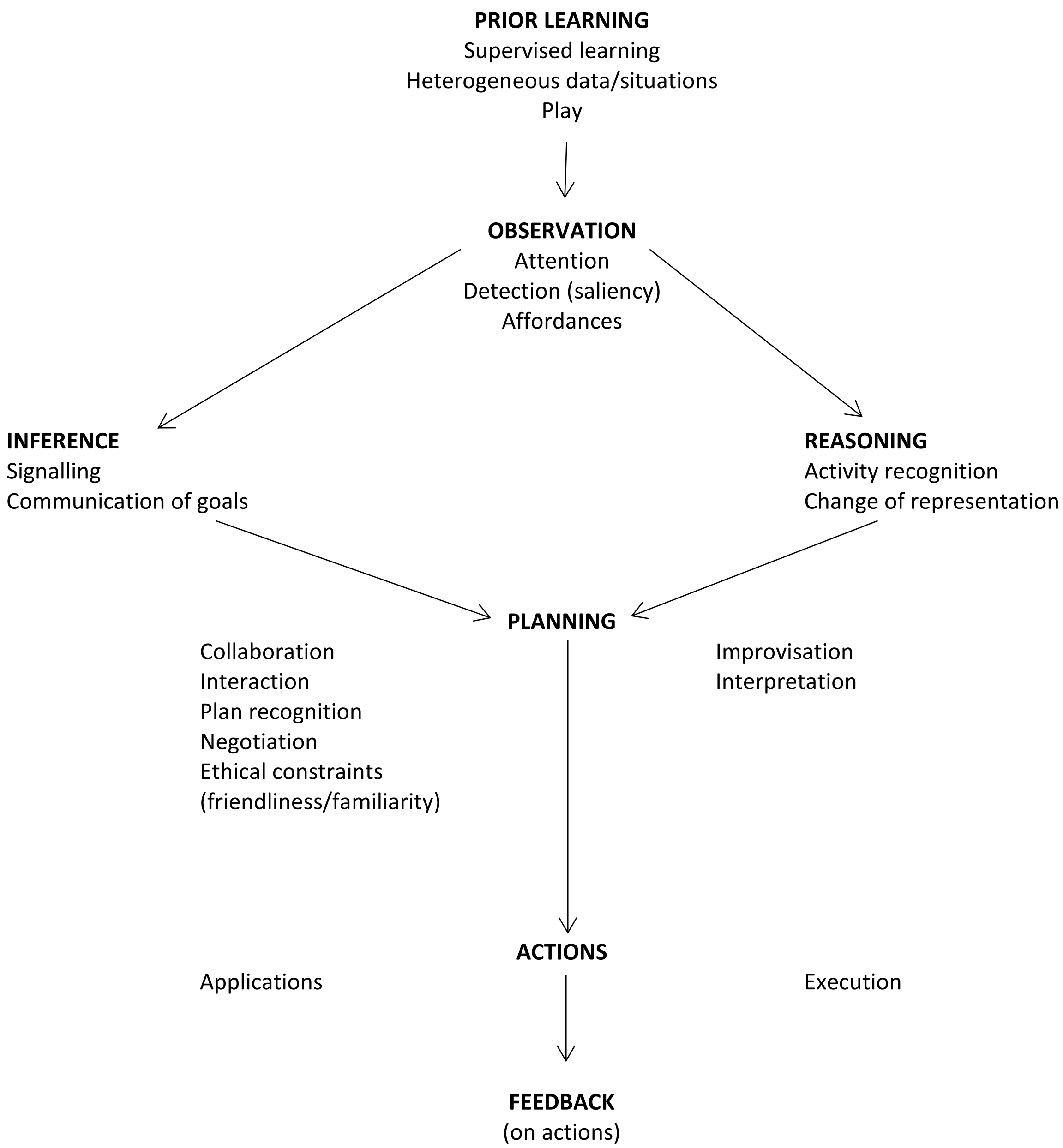
Details of the challenges can be found in the original workshop report:
https://www.epsrc.ac.uk/newsevents/pubs/humanlikecomputing/

The challenges are summarised below. They are listed as examples of where the research might go and are in no way to be regarded as definitive descriptions of the research needed. They are as yet by no means fully thought out and would require more work to clarify points such as requirements before they could form the basis of any activity.

1. The Physician's assistant

This would be a multi-agent system involving: robot, softbot, doctor, patient. It would need to be teachable. It would go beyond the current capabilities of telemedicine systems. A system would need to be able to hold a dialogue with the patient and the physician and this was the key feature. It would be co-located with the patient and the doctor would be distant so communication will be needed between all three. It would need to ensure that any plans were comprehensible for humans – how would a machine do this?

2. Childlike situational reactions



Overarching challenges	
Online/realtime	
Integration of techniques	
Tying together knowledge bases (multimodal)	
Communication	
Goal invention/detection	
AI Areas	**Psychology**
Robotics	Social psychology
Machine vision	Social signal processing
Speech	Observational learning (+ animal cognition)
BDI agents	Developmental psychology
Play	non-verbal communications
Machine learning	Situational awareness
Automated reasoning	Sensory motor perception
Planning	Attention
Multimodal communications	
Overarching behaviours	
Adaptability	
Intelligibility (understanding)	
Intentionality	
Feedback (continuous)	
Awareness	

3. Training system

The goal should be to allow training where there is no human expert, or a lack of time or resources, for example training for job interviews.

The research would need to develop tools and representational systems to capture the features that affect training such as culture, individual differences and to carry out observational learning.

It would need to have:

- Multicultural interaction systems
- Knowledge representation systems

- Ability to read the emotion of the interviewee through, for example biofeedback from wearables as input to the system (and be able to make sense of the data)
- Comprehensible inferences and tools to understand the training provided
- Credible, non-verbal communication that interacts at the right moment.

There should be a focus on the ability to carry out the short-term repair of representations and decisions under time pressure.

It should be an assistant working for the teacher and supporting the learners and it should learn from both of them. It would need to be able to deal with the diversity of teachers and students.

Application Areas

There were also a number of Application Areas identified:

- Creativity
- Robot companions
- "Find" engine (retrieves what you need, not what you ask for)
- Hybrid manufacturing (Industrie 4.0[2])

[2] See for example https://www.gtai.de/GTAI/Content/EN/Invest/_SharedDocs/Downloads/GTAI/Brochures/Industries/industrie4.0-smart-manufacturing-for-the-future-en.pdf