Inside intelligence

Experienced in computational intelligence, **Professor Hussein Abbass** talks to *International Innovation* about his work developing models to improve decision-making processes for humans and machines

Can you outline your research on human and artificial intelligence?

For a long time, intelligence has been seen as a feature of a single individual – often measured by an IQ test, or a single machine – exemplified by the imitation test, wherein a machine aims to deceive a human through imitation. In new discussions about intelligence, such as social intelligence, the focus is still on the individual – but the emphasis is on that person's ability to interact. This is much closer to what I'm interested in than the classic individual-centric forms of intelligence. My work started by looking at strategists who were attempting to imitate an opponent's thinking. Strategists don't just imagine what an opponent is. Instead, they heavily rely on stories, dialogues and interactions among themselves to form a mental picture of an opponent. By studying the nature of these dialogues and interactions, I deduced that human intelligence is based on two fundamental concepts: the ability to understand context in order to assess risk; and the ability to challenge others. These notions have provided a framework for my research.

What function does challenge to intelligence play in interaction?

'Challenge' as a concept is very intriguing; I had to delve into this topic for many years to see how it could be computed. If I give my students a question they can solve using their existing knowledge and experience, they will generally consider it an easy one. If I give them a question requiring knowledge and skills that are far from what they know, they will become demotivated, and see the question as too hard, or perhaps impossible, to solve. To push the boundary of their knowledge further, the question I give them needs to sit just outside the boundary of what they know. Our ability to estimate this boundary and sustain this challenge is the fundamental building block to generate learning and innovation. From here, I concluded that intelligence can best be judged by 'interaction' through challenges and risks.

Could you introduce your current research goals?

I have one main goal in my research that branches in multidisciplinary directions. Alan Turing deciphered the Enigma machine by creating a machine; I would like to decipher intelligence without ending up with another machine. Instead, my aim is to achieve two outcomes: define in computable terms the basic operations for intelligence; and connect the human – including the brain – and the machine to design a next-generation form of intelligence. Put these together and my overall goal is achieved – we get true, trusted autonomy.

Why do you view trust as so central to your understanding of intelligent interaction?

Trust is a fascinating concept. Socially, it is the glue of our society. On a psychological level, trust is possibly the key mechanism we use to manage complexity. Imagine you do not trust anyone; you will simply be overwhelmed with information in every situation. Trust comes in to ease this complexity. When I trust my students, I can rely on indicators to see if they deserve my trust or not, but I do not need to sit with them all the time to see what they are doing. Similarly, any type of interaction, be it human-human, human-machine, or machine-machine, involves an exchange of information. Every time an exchange of information occurs, trust needs to be invoked or the interacting entities will go nowhere.

In what way is your work attempting to illuminate the role trust plays?

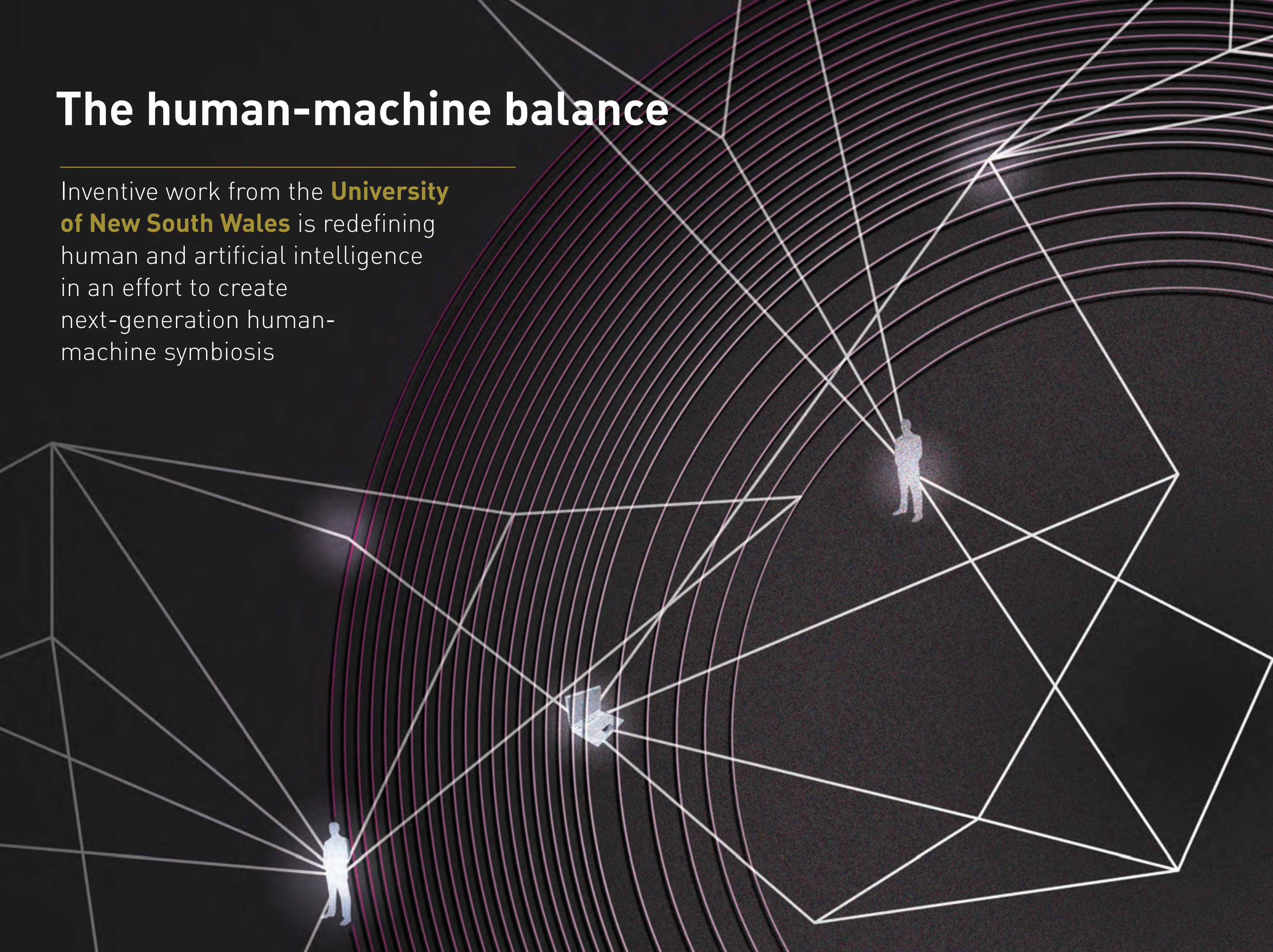
My current work attempts to understand trust from three perspectives. First, through an analysis of what social scientists and psychologists have found in their studies on trust. Second, by developing game-theoretic models to comprehend trust in interaction on a theoretical level. Third, through the design of computational trusting models to explicitly embed them in decision-making models in human-machine interaction.

What will be the wider impact of having a trusted system where humans and machines have different degrees of autonomy?

Imagine you have a trusted machine to cook for you when you're out, to look after the baby while it's sleeping, or to act on your behalf by negotiating a deal for your home loan with the bank and sign on your behalf. Do we have these machines today? Yes, we do. But why don't we use them? Because their autonomy is not trusted. If we could solve the relevant methodologies, symbiotic processes and trusted autonomy puzzles, we would crack the code for the concept of intelligence.

The human-machine balance

Inventive work from the **University of New South Wales** is redefining human and artificial intelligence in an effort to create next-generation human-machine symbiosis



ALTHOUGH SELF-GOVERNING machines capable of learning are frequently featured in futuristic films, many are not aware of the real abilities of artificial intelligence. Intelligent agents that can learn and make informed decisions are already used in numerous areas, such as traffic control. Indeed, the Institute of Electrical and Electronics Engineers (IEEE) has predicted widespread use of semi-autonomous vehicles within the next quarter century. As intelligent computer systems play an increasingly pivotal role in the modern world, it is important that the decision-making capabilities of both humans and computers are improved.

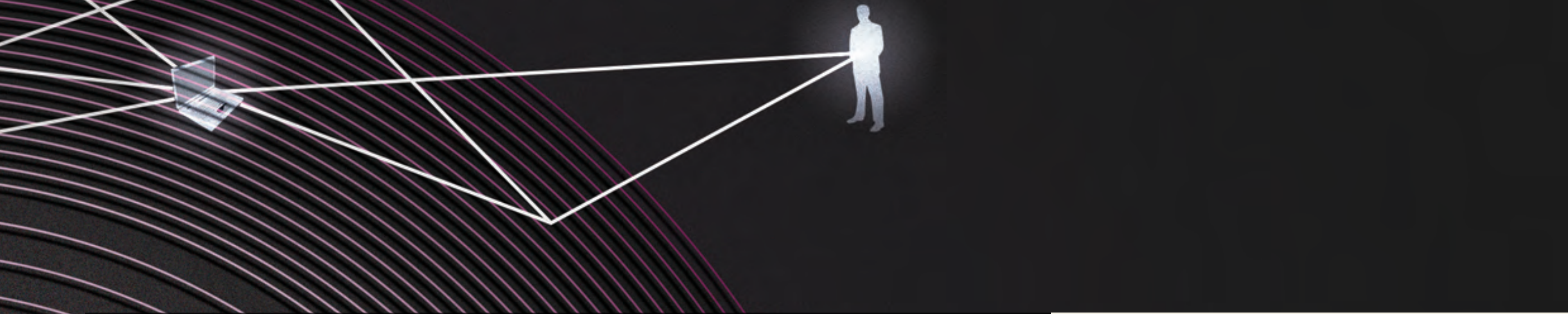
At the University of New South Wales (UNSW), Professor Hussein Abbass is endeavouring to understand how models can be developed that allow for improved decision-making processes by redefining intelligence and thus paving the way towards trusted autonomous systems.

DEVIL'S ADVOCATE

After more than a decade of researching the nature of competition and competitive interaction, Abbass' work has led to Computational Red Teaming (CRT), a state-of-the-art architecture to support decision making. The foundational concept of CRT is that human intelligence is derived from the ability to calculate risk and push boundaries by challenging an environment. "To have a true intelligent system, we need mechanisms to assess risks and to design and create challenges," explains Abbass.

In the strategic concept of Red Teaming (RT), individuals look at their own decisions through the eyes of direct competitors to make strategic

> It is hoped that this research, once distilled into computational models, can be embedded within decision-making models to design a trusted system for human-computer interaction



assessments. CRT employs multi-agent systems (MAS) and computational intelligence (CI) techniques in an attempt to transform this 'devil's advocate' approach into systemic, computable steps to improve decision-making processes. Whereas a classic computer program will be written to solve a problem, CRT involves programs that are developed with the ability to define their own programs, something Abbass likes to call 'meta-programming'. For example, in order to calculate risk, an agent can write its own objective while programmed with the mechanisms that allow it to define which uncertainties are most relevant and how these might impact the objective.

Implemented properly, CRT can be used to explore uncertainties, locate vulnerabilities, learn about other entities in the environment, understand biases, access information on other relevant decision cases and unlearn in order to learn. It is even able to explore ideas and scenarios that humans would not be able to process in the same timeframe.

As a proof of concept, CRT has been employed with success in an air traffic control (ATC) scenario. Electroencephalography has allowed Abbass to continuously measure and analyse the brain signals of air traffic controllers while simultaneously analysing air traffic information in real time. Using cues from either one or both data sources, a decision can then be made about a course of action, clearly showing the benefits of CRT as a decision support system in the context of an increasingly automated ATC environment.

HUMAN-MACHINE INTERFACE

The international use of CRT speaks volumes about the role of human mental processes within automation. "I have connected the human brain to the complex air traffic environment so that automation works in harmony with human cognitive abilities," Abbass elaborates. In working towards a next-generation form of intelligence, he envisions a team of humans and one of machines collaborating harmoniously to solve problems and make decisions. This vision is called Cognitive Cyber Symbiosis (CoCyS). Both human and machine 'thinking' are processes carried out within the electromagnetic spectrum, so why not blend cognitive space and cyber space together and transfer thoughts autonomously and seamlessly?

The precursors of what may sound like a fantastical proposition can already be glimpsed in today's world. One only needs to consider the brain-computer interfaces that allow disabled users to control motorised wheelchairs. What Abbass has in mind, however, is more accurately described as a next-generation human-machine cloud. Through the symbiosis of human-computer thinking, CoCyS aims to speed up the communication channel between humans and computers to bring about a superior real-time, evidence-based decision-making process.

Currently, CoCyS is a puzzle for Abbass – wherein all the pieces have been identified but do not yet fit together. Due to the sheer complexity of human brain signalling, responding constantly to innumerable sensory stimuli, a large ambiguity arises in the communication channels. Real-time, *in situ* data cleaning of brain signalling in complex situations is not yet possible, but instead of removing ambiguity, Abbass proposes to manage it. If person A tells person B something but the meaning is unclear, person B can ask questions that increasingly reduce the ambiguity, effectively cleaning up the signal. Interaction, therefore, is the key.

TRUSTED AUTONOMY

Before humans and computers can work together seamlessly, there are some trust issues that need to be addressed. In every domain that relies on reciprocal interaction between agents, trust plays a critical role – and yet a true understanding of the dynamics of trust remains elusive. Equally important to both human-machine and human-human interaction, Abbass' work attempts to understand how trust plays a pivotal role in designing an environment in which the CoCyS dream can become a reality.

To elucidate the dynamics of trust, Abbass wants to understand how it is reinforced in society and how it is transferred through the development of game theoretic models. In classical games, the decision-making process is carried out in a way that denies researchers the chance to study the role of influence, whereas the decision-making process in trust games allows this. There is little research regarding strategies to influence and transfer trust, but CRT can be used to provide important insights. Abbass' main goal is to discover whether a strategy can be employed by a truster to change an unreliable trustee into a reliable one and what those strategies are. It is hoped that this research, once distilled into computational models, can be embedded within decision-making models to design a trusted system for human-computer interaction.

RESHAPING INTELLIGENCE

OBJECTIVES

To redefine human and artificial intelligence and improve decision-making processes through:

- Computational Red Teaming (CRT): analysing challenge and risk in humans and machines
- Cognitive Cyber Symbiosis (CoCyS): connecting humans with cyber space
- Trusted autonomy: investigating the role of trust in a cooperative human-machine environment

KEY COLLABORATORS

Dr Sameer Alam; **Dr Michael Barlow**; **Dr George Leu**; **Dr Kathryn Merrick**; **Associate Professor Ruhul Sarker**; **Dr Kamran Shafi**; **Dr Jiangjun Tang**, University of New South Wales, Australia • **Dr Axel Bender**; **Dr Svetoslav Gaidow**, Defence Science and Technology Organisation, Australia • **Professor Anastasios Bezerianos**, SINAPSE, National University of Singapore • **Professor David G Green**, Monash University, Australia • **Professor Garrison Greenwood**, Portland State University, USA • **Dr Eleni Petraki**, University of Canberra, Australia • **Associate Professor Kay Chen Tan**, National University of Singapore • **Professor Xin Yao**, University of Birmingham, UK

FUNDING

Australian Research Council

CONTACT

Professor Hussein Abbass

University of New South Wales
School of Engineering and Information Technology
Northcott Drive, Canberra
ACT 2600, Australia

T +61 262 688 158
E h.abbass@unsw.edu.au
E hussein.abbass@gmail.com

www.husseinabbass.net

PROFESSOR HUSSEIN ABBASS has worked for the past 25 years on characterising and connecting human and artificial intelligence. His objective is to design dual-use models that can both be implemented as pencil-and-paper tools as well as sophisticated autonomous smart computer systems for organisations. The main driver for his research is to improve decision making on all levels, from individuals to government.