Panel Session: Can Artificial Intelligence Systems be Trusted?
12 December 2018

Author of these notes:

Prof. Adrian Hopgood, University of Portsmouth

Panel organiser:

Mr. Andrew Lea, Amplify Life

Panel chair:

Prof. Miltos Petridis, Middlesex University

Panel members:

Mr. Richard Ellis, RKE Consulting
Mr. Andrew Lea, Amplify Life
Prof. John McCall, Robert Gordon University Aberdeen
Dr. Simon Thompson, BT

Introduction:

Miltos
You trust that buildings work, that electrical switches work etc. AI is younger and is not always trusted. So, how trustworthy is AI?

Panel opening remarks:

Simon
You must be joking! Not trustworthy with anything like safety critical or nearly safety critical systems. We have not built the necessary control systems. Unlike airplanes, which are trusted because there are documents and test processes against predictive or real-world performance. AI test data are clean and artificial. Until we can evaluate performance and develop processes that document engineering processes, and management systems that document the systems, the answer is “no way”.

Andrew
Can human systems be trusted? Can AI be trusted as much as human systems? AI may be real, applied, or fake. Fake AI cannot be trusted. With applied AI it is hard to define the correct behaviour, so utility may be a better measure? In a medical system, reliability is hard to prove. Do we need a statutory framework, enabling the benefit to be exploited? Could real AI be trusted? It doesn’t exist yet, but it would have its own interests that may not match ours.

John
Relax – we have already lost control. The systems that we develop emerge from shared components. Mathematically we cannot prove correctness. So, trust should be about whether people are willing to use the software. Level 1 trust, i.e. the concept stage, describes what the AI is going to do for me. It is not about the technology. Does it do what I expect of it? – that is the useful definition of trust. Once the AI is there, how acceptable is it? Why did it do what it did? Trust can therefore be about explainability. For example, consider scheduling freight trucks in Aberdeen, which needed a real-time tool to assist. Over time, the truck drivers have accepted

the decisions without having to be educated about how the technology works. If it does reasonable things, then we trust it. We will evolve to live with AI. There may be some disasters along the way.

Richard
There are different sorts of trust, e.g. does it work as intended? A self-driving car should avoid a crash, which requires many things to be right. Even if it achieves that, is it biased? Biased examples include the Amazon recruitment system and bank loan systems. AI may not be biased by design, but may be subverted by its users, e.g. the chatbot that learned to be racist. Can we trust that the system is not designed to mislead you, e.g. search results that favour a product or political stance? Another form of trust is trust that the AI will not want to take over the world. Are there some decisions to reserve for humans? That is an ethical question. We need to understand which type of trust before we can answer the panel question.

Discussion:

Prof. Frans Coenen (audience)
Roborace presented ANNs for racing cars – it's OK to trust it in the safe situation of a race circuit. Frans uses AI for retinography, where clinicians are approx. 95% accurate and ANNs slightly better.

Simon
Image analysis systems are impressive, but they make different mistakes from humans. They don't make judgements on the same basis as humans. Things that could be picked up by humans but are missed by AI could lead to "car crash" for AI. Systems that make tools for humans may be a better approach.

Prof. Max Bramer (audience)
I like Simon's aircraft example. In other fields, e.g. medicine, mistakes are covered up. Consider the need to control autonomous weapons – it is not that the AI can't be trusted, but that it is cheap and may be used in an undesirable way. Do we trust AI for things that we are unaware of, e.g. AI to analyse surveillance devices? Should we assume it is safe and not misused? No, we should not.

Richard
That's about trust in authority, not trust in AI.

Simon
It is a debate like guns versus gun usage. Could we restrict the use of AI, like we restrict guns?

Richard
The issue is misuse of the data rather than of the technology. GDPR puts some bounds around data usage. AI is a tool for big data and its removal would restrict misuse, but governments say that they are applying it for societal benefit, e.g. to find the criminals.

Andrew
Regulation is not likely to be effective and would reduce the benefits e.g. medicine.

Simon
I am not convinced. AI may "lose the room", i.e. suffer a widespread collapse in confidence. His daughter wouldn't play against a computer Battleships opponent

trained by an ANN because it’s not fair. We expect fairness, a level playing field, with people in charge of machines. Tools are important.

Richard
When we place constraints on tools, we lose functionality and power. If we are happy with the downside, that’s OK, but we need to understand.

John
You are arguing that regulation can avoid the catastrophe, but regulation will stifle development. Technology will be developed to circumvent legal rules, so that’s not way the go. He is not arguing for no regulation, but we can learn from our mistakes, e.g. Cambridge Analytica.

Simon
This is like unbridled capitalism. Some accidents create zero learning opportunities. Climate change, for example, is a massive challenge that will benefit from the full unrestricted AI toolkit.

John
Consider an emergent intelligence example. There were floods in Aberdeenshire, with emergency services swamped and roads cut off. Groups popped up on Facebook – self-organisation spontaneously worked to provide help. Facebook is a free platform and constraints could have prevented this emergent behaviour.

Simon
Stated that he had been a Facebook fan but is now very suspicious of it. It had a good initial run, but then it was discovered that the system is flawed.

Andrew
It is better that look at all possibilities. Look at the substrate, i.e. who has access to the data.

Prof. Max Bramer (audience)
What would lead to loss of trust? Terrorists getting low-cost weapons. AI being used for surveillance to create a police state. We should think now to head it off rather than wait.

Gilbert Owusu (audience)
Those are specific AI evil uses. There are other positive AI uses. It is difficult to regulate the technology. YouTube is great, but it is also used to train in killing people. The debate will shift from the technology to other ways of using it.

Richard
Consider the autonomous drone example. Auto navigation and image recognition are good things, it is just their combination in that context that we don’t like.

Simon
It’s like assault rifles.

Richard
A drone with a computer on board is completely different from an assault rifle because it is easy to put together. We can’t ban the individual technologies.

John
People already predict explosives purchases through buying patterns, so maybe the same could apply to someone building a killer drone.

Miltos
Social media are not about AI, but we have a new generation of digital natives who accept all sorts of uses of data. We are oldies. The new generation accepts Google's access to their data. How can we trust AI in the world of fake news?

Simon
Put young people in the position where they can make appropriate choices. If we are not careful, we will create bad social norms – e.g. the way people interact with Alexa. We must challenge young people to think critically about their world.

Audience member:
Consider AI as tools that have to be used in accordance with a licence. You are responsible for it and its decisions. We can't validate the AI so we make sure that people using it are trained appropriately. It is like young pilots who cannot handle handover from an autopilot. Regulation is needed to ensure this licence model.

Richard
There are counterarguments; you would lose some of the essential benefits. Consider self-driving cars. It should be her way around: when the human is in trouble the machine should take over rather than vice-versa.

Audience member
I come from the marketing side, where there is not one unique solution. So there is a concern over whether the AI is really making a better decision. What do you think about trust in social science?

Andrew
You can't do the comparison with the alternative unselected choice, so you just have to try out the AI systems.

John
People do make decisions on the scantiest of evidence. The tools cannot give complete predictability, so we cannot expect too much of AI. People need to understand how the AI works and the basis of its decisions. If you pick up any tool without understanding it, you will have a bad decision. You need to know its limitations. Education is important.

Simon
If you can break down the solution into small steps then you can see the advantages and how you could have got there yourself, so then you will underwrite it.

Miltos
Consider other domains. Business intelligence shows a difference between scientists and softer disciplines. What are the challenges for different sectors in the use of AI?

Andrew
Yes, education is key. A lot of people are less able to think critically. Gullibility leads to acceptance of fake news and the lack of challenge.

Simon
It is important to understand the limitations of the technology, e.g. facial recognition is just associations. If we are looking for criminals in a database, we mustn't use it in other settings. A machine can give an appearance of spotting criminals, but that's not what it actually does.

John
We need the technical insight.

Miltos
He has witnessed people not getting a biometric recognition system to work as they misunderstood the underlying processes. Some parts of society could be excluded.

Richard
It shouldn't be beyond our wit to spread the word that we shouldn't trust everything that comes from the machine. It is just like the message not to trust strangers. Teach our kids not to place all their faith in machines.

Audience member
The issue is not the technology but people and institutions. The BBC iPlayer now demands my details, but they provided the service. You can increasingly access your bank only electronically. Institutions start to impose without consultation. The more we accept it, the greater the danger. Abuse of power is dangerous.

Audience member
Humans have cognitive bias. AI has availability bias. AI gives better decisions, but should we discuss those biases.

John
Progress is made by mistakes. AI just does what it is programmed to do. Human decision-making and AI – we are increasingly using AI as prosthetic ways of making decisions. We need to consider psychological use of AI.

Audience member
We could demand disclosure about which tools were used. Published papers could include a disclaimer, e.g. if this tool is used somewhere, these could be the consequences.

Andrew
AI is a horizonal technique, so it can be used in different ways. It's not so much what it is but what it is used for. Self-driving cars could provide a great improvement in quality of life for the elderly. AI that recognises animal pain would be a step forward for animal welfare.

Richard
That's true. There may be benefit in looking at similar things across society, e.g. you can't just go and buy explosives; there are licences and rules.

Simon
Be careful about the category. AI is horizontal, like blacksmithing in the past was a great boon for plough operators. Someone skilled has to make metal with which to plough. It is just like AI for decisions.

Audience member
Are we being too harsh on AI? AI doesn’t work like humans. Humans make errors, but we seem to expect more from the machine.

Simon
Humans are accountable, tools are not. Liability and due care need to be taken forward. A human must be accountable if they use an AI tool. If a self-driving car has my kids in it, who is accountable? The accountability is not clear. The systems need to have accountability built in, as we have for trains on the network. Self-driving cars can’t be excluded.

John
What you want is for an accident not to be repeated. Autonomous vehicles will get better from every accident, just like air accident investigations improve air safety.

Simon
Watch the video of the Tesla self-driving car image recognition learning system – it is terrifying and irresponsible. They should be accountable. Human error in an accident is understandable but driving when drunk is reckless.

John
We are in the pioneering cowboy stage, but the engineering will improve.

Andrew
Proper standards are needed.

John
Don’t put in barriers, otherwise things won’t develop. It’s like early disasters in aircraft design.

Simon/Richard
Early aircraft pilots took risks knowingly, but airships were different. The Comet disasters led to a royal commission; now we don’t make planes with square windows. We need institutions to respond.

John
We need to be more open and publish ideas of what could go wrong.

Simon
We can use the lessons of the past to short-circuit processes.

Miltos
Do we need the equivalent of a person with red flag in front of early vehicles?

John
Community can develop the systems.

Richard
What do we mean by AI that will be regulated? What’s in scope? A food regulation tool for managing use-by dates? A deeply embedded fuel sensor? There is a trade-off. Money makes things work. We need rules of legal liability and we need some international consistency.

Final comments:

Richard
The theme is that any control to improve trust will have consequences, but we may not know what they are. It is not simple.

John
We should not blindly trust AI, but we can go forward with it.

Andrew
As AI is more prevalent, trust will become more important. AI should have a health warning – don’t use it uncritically.

Simon
It’s a bag of snakes. We need to clean up our act.

Audience vote:
The ‘no’ votes win. We should not trust AI as it stands now.
Thanks to the panel.