Artificial Intelligence, the Emerging Needs for Human Factors Engineering, Risk Management and Stakeholder Engagement

Mike Erskine Executive Advisor
Risk Management Group, GHD Advisory
Level 8, 180 Lonsdale Street, Melbourne, AUSTRALIA 3000
Email: mike.erskine@ghd.com

WHAT IS INTELLIGENCE?

It is important to start with the basic concepts before entertaining the more complex aspects of managing artificial intelligence. Intelligence can be defined as:

One's capacity for logic, understanding, self-awareness, learning, emotional knowledge, planning, creativity, and problem solving. It can be more generally described as the ability or inclination to perceive or deduce information, and to retain it as knowledge to be applied towards adaptive behaviours within an environment or context [ref Wikipedia, intelligence, https://en.wikipedia.org/wiki/Intelligence*].*

Artificial intelligence can be defined as:
Both the intelligence of machines and the branch of computer science which aims to create it, through "the study and design of intelligent agents" or "rational agents", where an intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success. [ref Goebel, Randy; Poole, David L.; Mackworth, Alan K. (1997). Computational intelligence: A logical approach (PDF). Oxford [Oxfordshire]: Oxford University Press. p. 1. ISBN 978-0-19-510270-3. Archived (PDF) from the original on 7 March 2008.*]*

Some current programs, known as Narrow AI, can perform better than humans in some cases in a restricted field of information and outputs or actions. Examples are residential home energy storage and autonomous vehicles (AVs) in a very closed transport environment, such as a single corridor route. Narrow AI can utilise numeric information, along with graphical and textual information. Among the traits that researchers hope machines will eventually exhibit are reasoning, knowledge, planning, learning, communication, perception, and the ability to move and to manipulate objects. When this is achieved, the accepted term for it is called “general intelligence”.

BRIEF AI HISTORY

From its initial inception, Artificial Intelligence (AI) has had a number of winters now [ref Hendler, J., Internet Archive Wayback Machine, Avoiding another AI winter, 2012, https://web.archive.org/web/20120212012656/http://csdl2.computer.org/comp/mags/ex/2008/02/mex2008020002.pdf]. Multiple failures to achieve some initial potential led to funding drying up for many years. Many other spheres of technology development have had similar setbacks. However, recent progress has achieved sophistication that surpasses the level previously obtained. IBM’s Watson AI is now performing routine medical diagnosis, analysing petrochemical processes, conducting geological assessment for mining exploration, optimizing aviation operation and maintenance, and is even providing support in the running a university course [ref Letzter, R., IBM’s brilliant AI just helped teach a grad-level college course. Business Insider Australia, May 2016, https://www.businessinsider.com.au/watson-ai-became-a-teaching-assistant-2016-5]. Some sports articles are now written by AI. Google translate transitioned from the brute force phase conversion algorithm to AI late in 2016, resulting in improved translation quality.

Similarly, Google now uses AI for many searches, as distinct from the previous sorting algorithms. As users of Google searches, we are therefore already gaining knowledge and making decisions based on passive support of AI. Google bought the DeepMind Company recently, which specializes in AI.

Mathematical and statistical programs have been used over the years to analyse and determine critical values, and provide degrees of confidence, for medical, engineering, social, scientific and financial endeavours. Critical values have been set by professionals with suitable experience, and incorporated into standards, university training, codes and regulations for the population to abide by. Legal decisions of accountability of events have been based on these and other values that society consider critical for the levels of acceptability that we are prepared to tolerate.

When properly managed, AI has the potential to increase the levels of safety, social acceptability and performance compared to that which we have previously achieved. AI allows us to raise the bar on these many areas. As with other improvements in history, these will become the new standards that we will need to adopt. There will be new requirements of AI projects, organisations, and our regulators.

SURVEY OF AI APPLICATIONS

In preparing this paper, an internet survey was undertaken in 2018 to determine current AI applications across a broad range of industries. Eighty-one (81) AI applications were found, either at testing or operational phase, in government or private industry. The survey identified the following themes:

1. Early adoption in medical, financial, insurance and banking
2. Strong benefits when integrating with 3D structures/modelling and printing
3. Potential network optimisation in power, water and traffic
4. Synergy with autonomous vehicles

The majority of AI applications to date appear to be the complex technical areas (i.e. improved design or diagnosis) (~60%), followed by socio-technical (~30-35%) (i.e. tax office fraud detection). The remainder (5-10%) are socio-technical interactive (i.e. like some Facebook algorithms)

SOME RECENT AI SUCCESSES

Already, AI is showing intelligence in some specific areas exceeding the performance of humans [ref Sayer. P, Five things AI does better than humans, from the mundane to the magnificent, PCWorld, Nov 2016, `http://www.pcworld.com/article/3142193/techology-business/five-things-ais-can-do-better-than-us.html`]. Some significant successes over recent years include:

1. AlphaGo beats world champion at the game Go
2. Tesla's Autopilot brings man with blood clot to hospital
3. Swarm AI predicts the Kentucky Derby
4. Microsoft's AI now understands speech better than humans
5. AI improves cancer diagnosis
6. AI spots an eight-planet solar system
7. AI designed structures 3D printed for aviation
8. AI manages water utility

IBM’s Watson is another example, and is actively utilized in many ways including for aviation, mining, and medical assessments.

Although not statistically comprehensive, these success stories are typical of areas where outputs are technical or do not require strong social interaction with people. However, the results of utilising AI deliver strong benefits to both people and organisations. The other defining characteristic of some of these successes are the higher level of success or performance achieved compared to human levels in the endeavours. This is a key cost benefit feature of the AI applications.

Some of the organisations in the survey delivering the social and technical successes are those that are working in conjunction with IT companies, who have good risk management in place. There are, of course, some pure IT organisations achieving success, but not as many as the organisational combinations above.

AI is also being used in areas directly related to risk management. The first application is an active part of risk management to detect issues, for example to detect fraud in financial systems and airline bookings. The second area identified is as a tool to provide better control than standard systems, for example an automated share trading system, a process control system, or an AI system that provides medical support to doctors. In the 1950s and 1960s, the development of systematic safety and environmental assessments made significant improvement to the process industries. It is likely that similar benefits could be achieved by constructing and validating new AI applications involving safety, societal and financial risk management.

SOME RECENT AI LEARNINGS AND FAILURES

There have also been a number of AI learnings and some failures over recent years that are worth noting.

1. AI built to predict future crime was racist
2. Non-player AI characters in a video game crafted weapons beyond creator's plans
3. Robot injured a child
4. Microsoft's chatbot Tay utters racist, sexist and homophobic slurs
5. AI-judged beauty contest is racist
6. Pokémon Go keeps game-players in white neighbourhoods
7. Chinese facial recognition study predicts convicts but shows bias
8. Facebook chatbots shut down after developing their own language
9. Google Allo suggested man in turban emoji as response to a gun emoji
10. iPhone X Face ID beaten by a mask
11. Google Home outage causes near 100% failure rate
12. Google Home Minis spied on their owners
13. Facebook allowed ads to be targeted to "Jew Haters"

The observation here is that many of the outputs are much more sociologically oriented with much higher social interaction, either in safety, crime and punishment or engagement with society. The video game example is interesting as it touches on the concept of Coherent Extrapolated Volition (CEV) theory. CEV is "*meant as an argument that it would not be sufficient to explicitly program our desires and motivations into an AI. Instead, we should find a way to program it in a way that it would act in our best interests – what we want it to do and not what we tell it to"[ref* LessWrongWiki, Coherent Extrapolated Volition, https://wiki.lesswrong.com/wiki/Coherent_Extrapolated_Volition*]*.
This example is testament of the potential of AI to go further in what we would want, or in a new direction we didn't anticipate nor desire. Sometimes this can be good, other times, less so. The socio-technical failures are of interest. Social interactions are much more difficult, and the knowledge of social acceptability can be quite difficult for an AI.

Microsoft AI chatbot on Twitter back in early 2016 that had to be pulled down after one day. Microsoft had named the chatbot "Tay". Tay had learned to become racist within a short period [ref Vincent, J., The Verge, Twitter taught Microsoft's AI Chatbot to be a racist asshole in less than a day, Mar 24, 2016, accessed Jan 2018, https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.] with interaction. Other IT companies like Facebook faced problems at similar times [ref Hosanagar, K., Harvard Business Review, April 2017, The First Wave of Corporate AI is Doomed to Fail, accessed Jan 2018. https://hbr.org/2017/04/the-first-wave-of-corporate-ai-is-doomed-to-fail.]. As with most organisations, for every learning or failure that is picked up by society, there are most probably many other successes and failures that aren't reported.

There is a pattern of organisations and types of AI application noted in the above learnings and failures. These are typically non-process and non-transport industries, or to put it another way, these are predominantly IT organisations or with a strong social AI interaction.

WHEN HUMANS SECEDE CONTROL (TRUST), THE SOCIETAL PERMISSION POINT

This point isn't just an equivalent of replacing an existing human at current average levels of performance. It goes much deeper than that. For a train or an aeroplane, it is the point where a highly trained person within a high integrity system operates. These people are in the very low error level because of high skills to get to that point. They also have double check systems i.e. air traffic or train control, effectively taking them to a level about a factor of 10 or 100 better than an average person.

It is also hardwired into us from many millions of years, such that for us to let go of control, the system has to be significantly better. This is partly because of ego, and partly because of the unknown. When travelling by air, we trust the pilot and systems, because there is a benefit in using these services. The same for trains. It is a fragile balance. As the nuclear industry that routinely use the upper levels of Safety Integrity has found out, this societal trust is essential for continuity of business.

It is this region of about 2 orders of magnitude or better than society on average can do, is where society intuitively will accept highly automated systems. Whilst the safety of achieving this point may be well and truly much better than where we operate now for a certain mode of transport, the cost to achieve it is much higher. If targeting anything less than this point, but even well above current standards, societal acceptance likely won't be achieved, and financial viability is also unlikely.

UNCANNY VALLEY OR OTHER SOCIETAL ISSUES?

Autonomous Vehicles (AVs) are really a subset of robotics, so therefore some learnings from robotics are useful to consider. An initial phenomenon was observed in early robotic designs of several people, and was given a name by one researcher, Masahiro Mori, in 1970. He called it the uncanny valley.

In aesthetics, the uncanny valley is a hypothesized relationship between the degree of an object's resemblance to a human being and the emotional response to such an object. The concept of the uncanny valley suggests humanoid objects which appear almost, but not exactly, like real human beings elicit uncanny, or strangely familiar, feelings of eeriness and revulsion in observers. Valley denotes a dip in the human observer's affinity for the replica, a relation that otherwise increases with the replica's human likeness.[ref Mori, M. (2012). Translated by MacDorman, K. F.; Kageki, Norri. "The uncanny valley". IEEE Robotics and Automation. 19 (2): 98–100*]*

There is some discussion on the Internet that initial societal reaction to AVs constitute an "uncanny valley". The essence of this concept is also important in dealing with AVs and AI, and to some extent, other new technologies. Actually, the initial societal reaction is more complex than the above aversion. Some negative perception of AI exists, along with positive perception [ref Cameron, E, Accelerating innovation through responsible AI, PwC, accessed Jan 2018, `https://www.pwc.co.uk/services/audit-assurance/risk-assurance/services/technology-risk/technology-risk-insights/accelerating-innovation-through-responsible-ai.html`].

With AI, we are coming to realise that other psychological aversions exist as we interact verbally with ever increasingly intelligent synthetic systems. We naturally would like to know an AI is declared as not a real person, as it is uncanny to talk with them not necessarily knowing what sort of intelligence we are relating to. Secondly, another issue is being uncertain of what motive it is relating with us, i.e. to make profit for its organisation, as distinct from a more open and mutually beneficial and less commercial basis. In fact, as an AI gets closer to human like intelligence, a proportionally higher individual or societal tension is quite possible. This is partly because of our competitive tension, and ethical reasons, but also the broader issue of its increasing ability to potentially produce a better range of results for its owners than a human, and possibly at our personal expense.

AI for vehicles is currently in what is called the Narrow AI category. It is akin to a sophisticated machine, but with less verbal interaction with the occupants.

With AVs, a different feature is emerging. The initial response in 2014-2016 to the awareness of AVs being a reality was for a "blue sky" of amazing capability. When it is a mature technology, the indications are that this could be quite possible. That initial view seems to be giving way to a "mountain of caution", based on initial trial results and further research. This "mountain" comes from a few sources.

- The first is from the natural fear or concern of the unknown. Regulators and politicians are normally conservative as many panaceas are put forward to them by a range of people and companies. Quite often, purported benefits don't turn out as predicted.
- The second is the media fascination with new and different events affecting public opinion and perceptions. The recent Uber and Tesla fatality events were of concern. However, when AI driving is examined in context of overall risk level compared to human drivers, AI is significantly safer, i.e. lower in frequency of accidents per unit of distance travelled.
- The third is very deeply rooted in our expectations as discussed earlier. We individually feel we are good drivers, and that others are more to blame for the state or national road toll. Older people are reluctant to relinquish driving even when they know their capabilities are no longer suitable. We all like to be in control, and perceptually feel we do so at a very high standard, and loss of independence is also important. Rightly or wrongly, we then have very high expectations for AVs before we hand over control and trust them.

Perceptually, that AI trust region would appear to be in the order of a hundred times better than our current national average accident and fatality rate, or a 99% reduction in accident rate. This is difficult, because there is a delicate balance between cost to achieve this, and benefit gained, and societal ability to afford such technological performance. There are ways to possibly achieve this very difficult region of individual and societal acceptance. From an emotional perspective, we need to become familiar with the various forms of AI. This is perception, trust, integrity and the other features of the organisations bringing us this technology.

The "Mountain of Caution" needs to have definite attention and a strategy to deal with it. Formalised methods are required for processes of HF analysis and of stakeholder engagement in this respect. There is a path forward in this for companies contemplating significant AI content in its products and services. There are two main components. These can be broadly defined as technical and social. Both are needed, but also need to be complementary, and not independent of each other.

HUMAN FACTORS

The UKHSE has a good definition *Human factors refer to environmental, organisational and job factors, and human and individual characteristics, which influence behaviour at work in a way which can affect health and safety*
In other words, human factors is concerned with what people are being asked to do (the task and its characteristics), who is doing it (the individual and their competence) and where they are working (the organisation and its attributes), all of which are influenced by the wider societal concern, both local and national.
Human factors interventions will not be effective if they consider these aspects in isolation. The scope of what we mean by human factors includes organisational systems and is considerably broader than traditional views of human factors/ergonomics. Human factors can, and should, be included within a good safety management system and so can be examined in a similar way to any other risk control system.[ref UK Health and Safety Executive, Human Factors, `http://www.hse.gov.uk/humanfactors/introduction.htm`*]*

There are many good principles here. The scope is rooted in the traditional workplace model and organisation. For AI and autonomous vehicles, it is worthwhile to expand the thinking of HF, just as

the complexity of both AI and AV permeates much more into our society, rather than a conventional workplace.

SOCIETAL ACCEPTANCE AND THE NEED FOR STAKEHOLDER ENGAGEMENT

Another equally important facet of rolling out AV integrated with AI technology is societal acceptance, and the process of stakeholder management. One view of stakeholder engagement is shown below. It may not be the most important item at all times as some may think, but is undeniably very important.

Stakeholder engagement and stakeholder management are arguably the most important ingredients for successful project delivery, and yet are often regarded as a fringe activity or one that can be outsourced to business-as-usual functions. Project managers depend on people to respond to the outputs and benefits that they deliver. People will only respond if they are engaged. The phrase "stakeholder management" implies that these people can be made to respond positively to a project, but the truth is that a project manager frequently has no formal power of authority and therefore has to rely on engagement to achieve his/her objectives. [ref Association for Project Management, Stakeholder Engagement, `https://www.apm.org.uk/resources/find-a-resource/stakeholder-engagement/key-principles/`*]*

Again, several good principles are embodied here. As with HF, the scope is very much of the traditional workplace or project style. Broader concepts of engagement need to be considered. Due to the level of permeation into our lives, and of transferring control of certain aspects occur, i.e. our driving, very deep levels of trust must be tested. Crossing this threshold requires much more than a few demonstrations and a purchase. When we are in the car, and it is interacting with the broader community, our sense of interaction and our influence/control of that interaction is now very different. It is worthwhile to understand the level of importance in AV technology.

LEARNING FROM OTHER INDUSTRIES

The technical component is the quantitative demonstration of safety. In oil and gas, and rail, societal risk values of tolerability and SFAIRP must be demonstrated. In these areas, HF plays an important role. By contrast, with AVs and AI in vehicles, HF plays a potentially larger role. For nations where their societal risk values have been articulated, this can be a clearer path. For those nations without clearly articulated societal tolerability and acceptability values, this can be more difficult. Regulators tend to need clearer values and articulate demonstrations of safety. A long-term dynamic safety case for vehicles needs to be developed at a national level.

On the social permission side, there are other factors to consider. Dr. Peter Sandman's and others findings regarding societal fear and dread are relevant. To counterbalance the potential perception or fear, openness and integrity have to be right at the fore of any AI development and rollout. Therefore, company integrity policies will need to be significantly expanded. Recent examples of the banking royal commission, the commission into child abuse, and the VW software defeat programming, highlight how critical organisational integrity and reputation is. This determines social permission to operate.

NATIONAL VALUES THINKING

Part of the stakeholder approach needs to be done at a national level. Many nations have values and rights as a charter, or within various legislation at State or National level. Some nations refer to or adhere to international values. For AV rollout and national acceptance, like rail and aviation, national values need to be better articulated, and communicated societally beforehand. This is not so easy, but if done correctly, can be a powerful agent for society in the acceptance of AVs on their roads. Australia's values with respect to road vehicles is relatively less developed compared to rail safety and could possibly benefit from using similar rail safety values as part of the communication to the general public.

COMPLEMENTARY LOGIC TO CURRENT NARROW AI

Whilst there is significant benefit in AI and other computational engines to determine what objects are in order to take certain actions, there may be other risk-based approaches to augment reliability of correct identification. Some of these are subtler. The recent Uber accident has further highlighted the need. Our human intelligence looks at many features along with shape and size. Aspects like trajectory, velocity, background weather, and environmental influence versus purpose, all play a part. More of this can be programmed with the vehicles to perhaps make much higher accuracy identification of objects with potential to interact with the vehicle. Human identification whilst good, is let down by much poorer surveillance of the immediate environment, whereas current scanning AV technology is good and is improving at this.

Multiple deductive and inductive logic processes could be used, along with reductive processes for really high-quality decision making of professional drivers and pilots. The dynamic abductive reasoning process is also required for dynamically adjusting and for increased learning. These likewise need to be programmed into AI AV technology and verified as effective through a formal process.

INDEPENDENCE OF ASSESSMENT OF HUMAN FACTORS AND STAKEHOLDER ISSUES

Internal HF and stakeholder assessment do have benefits, but they can also have constraints at times. The capitalist imperative and internally generated schedules can limit the quality, diversity and magnitude of input of these vital disciplines.

Properly regulated independent assessment can provide a much wider range of assessment benefit. This is partly because of the typical expertise residing in specialist firms, but also because of a much more powerful paradigm. These are outside people with the requisite specialist skills, looking inwards at the project with set regulatory requirements for the first time with the ability to articulate the key concerns of the general population. This is especially beneficial if there is a charter of values that can be referenced.

This has become the trend in recent large government projects in Australia where a range of societal issues and sensitivities exist and need to be managed. The emerging range of large corporate project such as AVs with AI would have a high level of legitimate HF issues and stakeholder concerns that come together as societal permission requirements before acceptance can occur. An important part of societal acceptance for large projects can be an open and independent assessment. Regulatory bodies have realised this and have developed refined processes for rail, aviation, and other key facets of society such as infrastructure design, industry and utilities.

AI DEVELOPMENT ISSUES THAT COULD LEAD TO HAZARDS AND RISKS

Whilst it would be difficult to predict risks on the vast number of AI applications, it is possible to list some of the more likely causes giving rise to some scenarios. A list of some likely discipline or application level issues associated with AI development and deployment are considered.

1. Setting timeline targets or goals that may be too ambitious for the budget or particular capabilities of the AI
2. Not setting goals or targets that are necessary for the facility or organisation utilizing the AI, and having undue exposure to safety, environmental, social or reputation risk.
3. Quality of information available and how constant it may be for the life of the AI to utilize (possible allowance for retraining needs as information changes)
4. Programming AI without all of the necessary variables
5. Programming the AI with too many trivial variables – i.e. less relevant and generates spurious results
6. Programming the AI using older data, older statistical paradigms, perhaps not providing the fullest extent of newer capability.
7. Not using top tier specialists in the learning and checking processes leading to insufficient knowledge, or incorrect knowledge for application.

8. Lack of independent review phase of AI output testing leading to inadequate validation of required learning for the expected spectrum of issues.
9. Lack of suitable interaction between programmers, technical, sociological, environmental and financial specialists leading to deficiencies in the learning process for the AI.
10. Lack of suitable complexity of model to adequately reflect the current and projected situation leading to inability to deal adequately with a reasonable spectrum of issues.
11. Insufficient ethical training leading to safety and social issues.
12. Deliberate training or poisoning of the learning process with unacceptable material, leading dangerous safety, environmental and financial actions.
13. Negative perceptions by potential end users of the artificial intelligence output.
14. Criticality of what the AI is doing, scenarios related potential harm to life, environment, financial, social. Etc.

From these and other risks, and existing processes that we are familiar with, we can look at a range of customised processes that will likely manage issues and have potential for the best output.

RISK MANAGEMENT OF THESE NEW STANDARDS

Applications of technology in society have become increasingly more complex. We have been broadly successful in developing more and more complex control systems to achieve this desired level of performance and control of these operations. This is driven by our societal expectations, and also the profit that comes with achievement. Along the way, this has generated new challenges to test these systems for reliability. New professional certifications and courses for risk management, including functional safety have been developed in recent years to help achieve and maintain these standards.

1. Previously, for technical applications, we have been able to verify the outputs of many calculations, except where complex programs are used. In those cases, other complex programs can be used as verification. This can get into the region of what is known as Segal's Law[ref Pettenqill J., Segal's Law, 16S rRNA gene sequencing, and the perils of foodborne pathogen detection within the American Gut Project, NCBI, https://www.ncbi.nlm.nih.gov/pubmed/28652935]. *A man with one watch always knows the time. A man with two watches is never really sure*. What this means is multiple outputs may yield different results. In risk management, this may be beneficial. If the outputs of a complex system analysed two or more different ways are marginally different but in the same range, then that is usually acceptable, depending on consequences of error. This has been used in the rail industry with complex systems as contained in EN50126, which has been primarily deterministic in its assessment methodologies. AI systems by contrast, are stochastic. Environments in which AI can be used can be defined as follows [ref

Rodriguez, J., 6 Types of Artificial Intelligence Environments, Medium.com, Jan 2017, https://medium.com/@jrodthoughts/6-types-of-artificial-intelligence-environments-825e3c47d998]:

1. complete/incomplete,
2. fully/partially observable,
3. competitive/collaborative,
4. static/dynamic,
5. discrete/continuous,
6. deterministic/stochastic

It is these environments that determine what sort of data and how much you need.

AI applications require different approaches to achieve the really high levels of reliability and performance that we would now routinely achieve. How do you get an AI system operating at better levels than you can personally achieve? The answer lies in the careful risk management of

programming, data to feed it, and suitable programming that will likely achieve reasonable CEV for the AI, i.e. not too constraining, and not too loose. Almost, like a good parent, suitably stretching their children as they develop.

Current AI approaches include statistical methods, computational intelligence, and traditional symbolic AI. Many tools are used in AI, including versions of search and mathematical optimization, neural networks and methods based on statistics, probability and economics. The AI field draws upon computer science, mathematics, psychology, linguistics, philosophy, neuroscience, artificial psychology and many others.

The AI typically needs the collective benefit of numerical and graphical and textual knowledge and calculations that we can muster for a wide range of scenarios. It needs to be taught "life" skills for the key areas of its operation.

As with people, AI learning will need to be a continuous process as the environment in which it operates continually evolves. Risk management will likewise need to be active through all steps of the learning and operational phases. Some of these are outlined below.

1. Context and Interaction – proper definition and critical mass of context, including coherent extrapolated volition consideration. Proper definition of interaction, safety, societal, business and other.
2. Initial Learning – basic training for functionality.
3. Advanced learning – higher performance requirements
4. Learning Assurance – testing for a wider range of scenarios, or as systems become more complex
5. Ongoing Assurance – continued verification of performance, (with software or hardware upgrades, environment changes, or general time-based verification)

Context and Interaction

This contains all the elements of basic risk management as outlined in ISO 31000. Most of the failures observed in the examples above lack the more disciplined context definition. As the context becomes exponentially more complex as more social components and output interactions are included, the propensity for problems likewise increases. Key features required are a systematic identification of the environment and the hazards to all stakeholders within and associated with that environment. The level of consequence and frequency will determine how much effort is needed here. This can be expressed in several ways:

1. Insufficient detail and scenarios of the key identified areas required
2. Insufficient key areas identified
3. Insufficient HF, societal factors and environmental factors and related scenarios considered
4. Insufficient scoping of tolerability and acceptability of hazards, events and associated probabilities.
5. Unbounded limitations definition (boundary) setting of system capabilities with reference to the situation at hand

Whilst it is important to frame the context properly, there are opportunities to capture issues later. However, not doing enough context analysis knowing that later steps may capture shortcomings is neither recommended nor encouraged.

Context setting for broader AI applications perhaps requires more effort than previous technologically predominant/safety challenges. Oil and gas industries typically do put significant effort into their context and risk assessment processes because of the consequences of hydrocarbon releases.

AI can benefit from a large learning population such as the Tesla vehicles, and over a much larger time domain than humans [ref Morris, C., Tesla's massive accumulation of autopilot miles, Inside EV's, Jul 2018, `https://insideevs.com/tesla-autopilot-miles/`].

Initial learning

Some of the features to test for include testing for wrong learning, or level of learning required. Here is where some quantitation of AI performance can be determined against the targeted values determined in context. Examples include:

1. Autonomous vehicles with no direct human supervisory input.
2. Robotics operating within warehouses
3. Software AI guiding people around websites and providing health, product and financial related information.

Just like high skills requirements of drivers, pilots and other humans, there needs to be a competency testing regime for the AI. Key features will likely be:

1. Update test frequency period determination. This will depend on what is being learnt and how it is to be controlled.
2. Process of assurance of the degree of learning and of the right features and objectives to learn.
3. Level of competency and human supervision/intervention required.
4. Initial confirmation and determination of the magnitude and likelihood of events that could occur.

How to get the best learning for AI?

1. Provide it with the best knowledge in the area of concern from thought leaders.
2. Specifically train the thought leaders for AI, CEV and for risk management skills
3. Mentor others through the AI learning process to develop broader skills across the professions
4. Also, have thought leaders actively involved in the AI learning process, i.e. testing and checking the output, because the output will be of a level that needs thought leader level knowledge to be able verify.

Learning assurance

Because it is an assurance of an intelligence, there is need to use techniques related to technical, social and societal information and logic that an intelligence should or shouldn't be using. Some of this includes training of categories, causality determination, clustering and hierarchical clustering determination and assurance. Anomaly detection and association learning are particularly important in the development of Autonomous Vehicle AI systems of high integrity.

Validation is the assurance that a product, service, or system meets the needs of the customer and stakeholders, involving acceptance and suitability. Verification is the evaluation of whether or not the product, service, or system complies with a regulation, requirement, specification, or imposed condition, typically an internal process.

Possible techniques to use for assurance in these areas are as follows:

1. Validation of the objectives to ensure our true needs have been met. Subsets are outlined below.
2. Deductive reasoning validation. Context check. Too little pertinent information or too much unrelated or irrelevant information
3. Inductive reasoning validation – context and pattern check. Too little or too much information.
4. Reductive reasoning validation. – context and pattern check. Is there enough information from which to do effective reduction?
5. Abductive reasoning validation – Multiple expert review of a range of results of the AI in training
6. Verification of what has been learnt and alignment against objectives. CEV oriented tests of stochastic tails can perhaps help here. The aim is to test the system until "it breaks", or reaches logical limits, and then reference back to human performance under similar conditions.

Fallacious decision-making and dynamic environment testing

This is equally important to assure as the validation of good reasoning. There are two aspects to this that need to be considered.

1. What is the potential for a wrong decision or output given good validated training and information (system construct, information and human corruption)? Are there weaknesses in the development of the system?
2. What is the potential for fallacious decision making (error potential) when in operational mode and conflicting inputs or change of environment occur (Microsoft Tay example). Will the AI continue to perform with the Coherent Extrapolated Volition of our originally desired goals in a dynamically evolving environment?

As has been seen, the applications of AI will be much greater than the constrained environments we have often made or designed our equipment or processes for. With technology to date, some of the risks that have occurred are ones where the environment has altered. Our typical response has been to learn from mistakes, or to anticipate and engineer the equipment for an expected but unlikely scenario.

Time dynamic environments will require a level of risk management that isn't commonly utilised. Increasingly, this will occur in social settings, where the societal response may change with time, and even in response to the use of AI in that application. Some examples are listed below:

1. AI in autonomous vehicles operating on our roads. The response of people to this technology is now being recognized and starting to become clearer. Learnings from aviation and other highly automated environments indicate we will need to account for more individual and sociological responses.
2. As autonomous vehicles become more commonplace on our roads, the broader societal response will become more evident.
3. Malware targeting of key infrastructure required to effectively operate autonomous vehicles.

Risk management controls

As with any process or operation by an organisation, there needs to be risk management.
PwC has already developed a framework in the UK for the use of AI [ref Cameron, E, Accelerating innovation through responsible AI, PwC, accessed Jan 2018, https://www.pwc.co.uk/services/audit-assurance/risk-assurance/services/technology-risk/technology-risk-insights/accelerating-innovation-through-responsible-ai.html]. The framework is possibly set for more of a business environment rather than the broader range of AI applications noted earlier. However, most of the principles broadly apply.

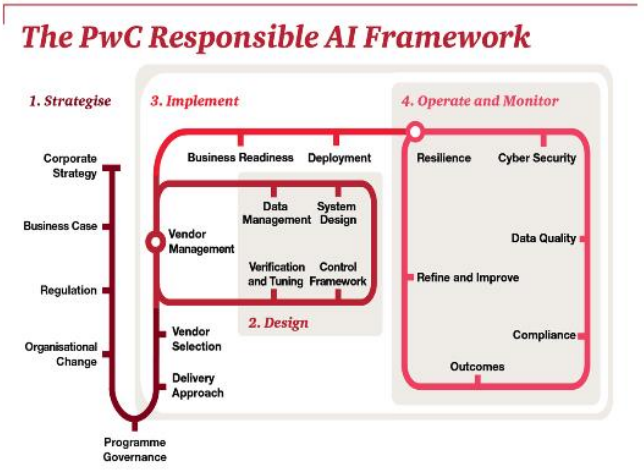

Figure 1 AI Framework

From an assurance viewpoint, it is quite a good way to think about AI. There are many deeper issues that this alerts people to now contemplate. There needs to be a whole range of controls to manage this technology within our organisations and economies. Some are being developed now, such as robotic standards [ref British standard BS8611:2016; Robots and robotic devise. Guide to the ethical design

and application of robots and robotic systems. http://shop.bsigroup.com/ProductDetail?pid=000000000030320089].

Some of the key controls for AI systems could be as follows:

1. Strict authorization codes of accepted teachers for AI program.
2. Certification of people suitable to train AI, by specific discipline.
3. Certification process by which to skill up people to be suitable to train AI
4. Functional safety and AI standards will need to be developed. FSE-AI
5. Formal safety and AI standards will need to be developed.
6. AI quality information standards could be made for any human or other publication.
7. AI Certified publication houses
8. Where there is higher probability of failure of output, to have a key human assessment point.
9. AI output assessment competency of people (e.g. university, driving, flying etc)
10. Certified AI Assessment competency courses (e.g. university, driving, flying etc.)

There are a broader range of issues that need to be considered in terms of societal management of AI over the next few years. Some of these are occurring or are soon to occur, and are highlighted below.

AVS AND PERCEPTION SURVEYS

There are some surveys occurring, currently centred on perceptions of people with regard to AI[ref Hohenberger et al, How and why do men and women differ in their willingness to use automated cars? The influence of emotions across different age groups, Elsevier, Oct 2016. Transportation Research Part A 94, pp 374 - 385,Haboucha et al., User preference regarding autonomous vehicles, Elsevier, March 2017, Transportation Research Part C 78, pp 37 - 49,Hulse et al., Perceptions of autonomous vehicles: relationships with road users, risk, gender and age, Elsevier, Oct 2017, Safety Science 102 (2018) pp1-13.]. There are possibly some cautions in this approach, if used as an actual basis, as most people haven't really seen what AVs can do. It does provide good guidance as to the general feelings that currently exist, which is very useful from an implementation strategy viewpoint. It could also represent a societal response to the news about these vehicles. Care and caution should be used in respect of how society will actually embrace this technology and the possible risks that could lead to real pushback.

AI push back

There is a possibility of some mild or even stronger push back by society in respect of AI. This can be because of developers crossing a few boundaries. A few examples are listed below.

There is a Google AI in trial use that books a haircut or a restaurant meal on your behalf and reports back to you. Whilst technologically really good to do it on your behalf, there is a problem. The lack of declaration of not being a sentient intelligence, i.e. the false use of the English word "I". Societal pushback could occur when we start to have this happen more commonly. It would firstly occur with restaurant and hairdressers, unless of course, they accept the fact that a booking is legitimate and it is to their financial advantage.

What happens at the next stage to the broader population where we find ourselves being called by a company system trying to get something done for their benefit, and less so for ours? It is only a matter of time when marketers using this and emotional manipulation with AI in verbal form. Note the recent pushback on Facebook with fake news bots affecting the USA presidential election in 2016.

There is a strong connection between AI and analytics. Does this generate a homeostasis point where society moderates our digital footprint to minimise our web presence? Conversely, AI systems are getting sharper and analysing better, so it is likely our digital footprint is being better analysed. Also, increasing software usage and data generation does increase the footprint.

There could be very serious pushback from society about the full spectrum of undeclared AI interactions at a broader societal level. It could be a very big reputation risk for organisations who don't declare up front what AI they are using and how society is relying on that AI. A recent example is the pushback on Cambridge Analytica and the political parties using them to achieve their ends. There is a real need

to change constitutions or laws for open declaration of use of AI in any interaction either directly or indirectly with people, either by time or by path.

CRITICAL MASS CONCEPTS TO DO WITH AI – SOCIETAL AND ECONOMIC

There is another concept that needs to be considered in respect of AI technology. It is the critical mass level of the technology in selected areas of society. The logic goes as follows. If it has been used successfully in a few areas, and tangible benefits have been seen, then it will have broader acceptance into different areas of society. This was the case with the introduction of horseless carriages (cars), air travel, microwave ovens and mobile phones.

There is also another crossover point to consider. This is the economic imperative pushing for the technology to be used. The following scenarios highlight the various places and ways in which this may occur. Businesses need AI in order to effectively compete in a modern world as a new base level of functionality. As overheads drop lower and lower to maintain competitiveness, output per actual person employed has to increase as AI undertakes a range of tasks previously done by humans[ref Atos Opinion Paper Declares Artificial Intelligence Approaching Critical Mass, Sept 2018, https://atos.net/en-gb/2018/press-release-en-gb_2018_09_28/atos-opinion-paper-declares-artificial-intelligence-approaching-critical-mass].

New business models that couldn't exist without AI are now causing forward change that is hard to ignore, and will likely replace older companies that operate on outdated models that can no longer compete. This economic critical mass will be when the AI capability, and price delivered of product or service achieves lower quartile region, and societal acceptance i.e. foreign company, but price driver is too strong to ignore. *Critical mass of AI as a tool in society.* This occurs when the bulk of services and interactions are done with or by AI as a substantial component, when it is no longer economic nor effective for older systems. In certain areas of society this has already occurred. It is increasing depth and diversity. *AI and quantum computing* is a new synergy that has to be considered. Computational power will be beyond standard digital and possibly human capabilities.

USING AI IN THE FIGHT AGAINST MISUSE OF AI AND NEWS SYSTEMS

As a counter measure to fake news, it is possible that AI can be used as a filter to remove content that is deemed below par on levels of journalism related to bias at a level where input is not constructive. AI is then in control of information we receive, but on what basis? [ref Susaria, S, The Conversation, How Artificial Intelligence can detect – and create – Fake News, May 3, 2018, http://theconversation.com/how-artificial-intelligence-can-detect-and-create-fake-news-95404, Harwell D., AI will solve Facebook's most vexing problems, Mark Zuckerberg say. Just don't ask when or how. Washington Post, April 2018. https://www.washingtonpost.com/news/the-switch/wp/2018/04/11/ai-will-solve-facebooks-most-vexing-problems-mark-zuckerberg-says-just-dont-ask-when-or-how/?noredirect=on&utm_term=.421205283455]. It is interesting to note that AI has been used as a spam filter in e-mail systems for over a decade.

DRIVING WITH INTELLIGENCE

Currently, driving with road rules is not enough for human drivers. With early autonomous vehicle testing, it became readily apparent that emotional intelligence was also needed. This hastened the AV developers in the direction of AI as a way to deal with the issue. The AI can more easily deal with more complex situations where other users have a deficit of road knowledge/road rules, as well as deal with static and dynamic road conditions and other road user behaviour. AI can achieve some of this through sensing and learning other vehicle movement cues alongside weather and environmental cues.

AI AND CURRENT SAFETY CRITICAL SYSTEMS

Some recent thinking has emerged in regard to using AI for safety critical systems. Typically, high integrity safety systems are deterministic in nature. AI goes beyond deterministic thinking, i.e. if pressure exceeds "X", then open valve "Y" until pressure "Z" is achieved, and then close valve "Y"

again. AI can go well beyond numerical stochastics as well, which is inductive testing by nature. This is where certain patterns occur, then a certain situation is likely (to some extent) to happen, so a response is required. Contextual testing is also limited, i.e. how well does the testing approximate all that could occur in the field? Have all scenarios been contemplated, and does the system recognise a new situation as a deviation to current experience, or as a new situation without a defined response? Longitudinal testing is continued testing over a longer period of time to gain greater exposure to a wider range of events and to test responses. Longitudinal testing has limitations, but has been used so far.
Therefore, testing and assurance regimes need to reflect this inherent situation.

"Although AI will be an engine for progress in many areas, creating real-world systems that realize these innovations will in fact require significant advances in virtually all areas of computing, including areas that are not traditionally recognized as being important to AI research and development… future AI systems will not only draw from methods, tools, and themes in other areas of computer science research, but will also provide new directions for research in areas such as efficiency, trustworthiness, transparency, reliability, and security" (Hager, Bryant, Horvitz, Matarić, and Honavar 2017). These authors identify the development of formal methods as a key enabler for the deployment of AI techniques in dependable applications. [ref Johnson, C.W., The Increasing Risks of Risk Assessment: On the Rise of Artificial Intelligence and Non-Determinism in Safety - Critical Systems, 2018. School of Computing Science, University of Glasgow. http://www.dcs.gla.ac.uk/~johnson/papers/SCSC_18.pdf*]*

In essence, many people are describing the key components of competency based training and all that goes with it. Essentially, AI, like humans, needs to reach an acceptable level, and every now and then, have some retesting to ensure bad habits or wrong learning don't creep in. Transparency will likely be a problem with AI systems for a long time to come [ref Muehlhauser, L., Transparency in Safety-Critical Systems, Machine Intelligence Research Institute. August 2013. https://intelligence.org/2013/08/25/transparency-in-safety-critical-systems/]. Academics have been grappling with this issue but perhaps ignore some broader features of the limits of human centred logic and resources. Some approaches like generative adversarial networks (GAN's) could be what is required to train AI in certain cases, as the human limits are reached.

It is possible that combinations of deterministic and inductive systems may provide the best overall response. This is akin to our deductive logic in combination with our inductive reasoning for a given situation, and its natural variation, to arrive at a logical action that is most likely to be correct.

The IEC has been tasked with the exercise of developing a new international standard for a risk management framework for AI[ref IEC Blog, New International Standard will offer risk management framework for AI, March 2019, IEC, https://blog.iec.ch/2019/03/new-international-standard-will-offer-risk-management-framework-for-ai/]. It is a substantial issue, and is evidenced by recent conferences starting to emerge on this topic[ref Call for Contributions, First International Workshop on Artificial Intelligence Safety Engineering (WAISE 2018), http://www.es.mdh.se/safecomp2018/workshops/WAISE-CfP_SafeComp2018.pdf]. Currently, the functional safety standard only allows AI up to SIL 1 [ref IEC 61508-3:2010, Page 48, Table A.2 Software Design and Development – Software architecture design.] for fault correction. However, it is neither recommended or not recommended.

It is in the upper reaches of the functional safety standard that we find some useful principles that relate to development and use of AI in safety critical areas. A high hardware fault tolerance (HFT) along with a high integrity (SIL 4) is needed for a very safe and reliable system. This requires both elements of software and hardware control to achieve this outcome. Critical applications like rail crossings and other key rail interaction areas form part of this regime.

For safety critical and other systems, it is not enough to just have AI technology within a technical product. Organisations need to have suitable resources for developing and training the AI, including

appropriately developed ethics input and a functional management structure [ref Kanioura Dr, A., Critical Mass: Managing AI's unstoppable progress, Sep 2018, https://www.accenture.com/us-en/insights/digital/critical-mass-managing-ai-unstoppable-progress].

Regulatory authorities need to appropriately update the current So Far As Is Reasonably Practicable (SFAIRP) philosophy for safety critical systems in the light of what is reasonable for AI training and assurance. They will also need to better articulate the means by which assurance of SFAIRP will then need to be made. If developed suitably, the proposed IEC risk framework for AI and safety critical systems will be very important in these aspects.

ECONOMIC IMPLICATIONS OF AI FOR AUSTRALIA

Already, there are economic projections for AI in various countries [ref How will artificial intelligence affect the UK economy, PwC, https://www.pwc.co.uk/services/economics-policy/insights/the-impact-of-artificial-intelligence-on-the-uk-economy.html]. Indications for the UK are that it could be in the order of £235 Billion or about 5% of their economy by 2030 according to a PwC report. If Australia were to take an equivalent path, it would be in the order of $148 billion of our current $1.74 Trillion GDP, or $200 billion in 2030.

Like autonomous vehicles, the use and benefits of AI to our economy will be essential. Currently, we are only moderately ranked for autonomous vehicle preparedness [ref KPMG, Autonomous Vehicles Readiness Index, Assessing Countries' openness, 2018,https://assets.kpmg.com/content/dam/kpmg/nl/pdf/2018/sector/automotive/autonomous-vehicles-readiness-index.pdf,].

CONCLUSIONS

Professional people will need to be trained in fundamental calculations and principles in new ways to manage risks in their professions going forward. These people will need to use their depth of skills, in conjunction with likely new findings of AI to achieve new performance measures previously unattainable. Professional training in the development and separately in the operation of AI are whole new areas of risk management measure.

It is time to develop university level science, psychology (stakeholder), engineering and risk courses more specifically with higher levels of logic training in conjunction with discipline specific AI developments and management approaches. The countries that initiate these education advances will likely economically benefit compared to other nations.
Developing an equivalent of HF engineering approaches, but adapting them for AI will likely have some merit. Given the common basis of a form of intelligence, there are likely to be parallels of how we must risk manage humans that would be applicable with handling artificial intelligence.

There is a need to update and/or complement ISO31000 to match the more socio-technological interactive applications of AI, or at least to develop a guidance handbook, much like what was developed for climate change. More needs to be done about this, like the IEC is now doing.

New competencies and professional certifications need to be developed for AI practitioners in various disciplines, as well as risk management processes. These need to include higher logic and reasoning, HF and socio-safety and stakeholder engagement. Specific tests have to be developed for the AI to demonstrate the level of competence to operate systems and/or provide critical information for human operators upon which to act. Therefore, organisations who adopt and work with AI are likely to be changed in the way they will think as well as operate. The next evolution of safety cases with AI HF and Stakeholder Management Plans are required, along with appropriate regulatory support.

National and international tolerability criteria need to be developed for AI in use, just like any other control equipment. Would it be possible to consider well developed and tested AI with functional safety

integrity levels? These aspects need to be wisely considered over the next few years. Some specific societal engagement strategies will be required for various AI classes, each requiring unique approaches due to the diversity and capability of application.

REFERENCES

2. Hendler, J., Internet Archive Wayback Machine, Avoiding another AI winter, 2012, https://web.archive.org/web/20120212012656/http://csdl2.computer.org/comp/mags/ex/2008/02/mex2008020002.pdf
3. Letzter, R., IBM's brilliant AI just helped teach a grad-level college course. Business Insider Australia, May 2016, https://www.businessinsider.com.au/watson-ai-became-a-teaching-assistant-2016-5
4. Sayer. P, Five things AI does better than humans, from the mundane to the magnificent, PCWorld, Nov 2016, http://www.pcworld.com/article/3142193/techology-business/five-things-ais-can-do-better-than-us.html
5. LessWrongWiki, Coherent Extrapolated Volition, https://wiki.lesswrong.com/wiki/Coherent_Extrapolated_Volition
6. Pettenqill J., Segal's Law, 16S rRNA gene sequencing, and the perils of foodborne pathogen detection within the American Gut Project, NCBI, https://www.ncbi.nlm.nih.gov/pubmed/28652935
7. Vincent, J., The Verge, Twitter taught Microsoft's AI Chatbot to be a racist asshole in less than a day, Mar 24, 2016, accessed Jan 2018, https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.
8. British standard BS8611:2016; Robots and robotic devise. Guide to the ethical design and application of robots and robotic systems. http://shop.bsigroup.com/ProductDetail?pid=00000000030320089
9. Hosanagar, K., Harvard Business Review, April 2017, The First Wave of Corporate AI is Doomed to Fail, accessed Jan 2018. https://hbr.org/2017/04/the-first-wave-of-corporate-ai-is-doomed-to-fail.
10. The Conversation, Why using AI to sentence criminals is a dangerous idea, May 2016, https://theconversation.com/why-using-ai-to-sentence-criminals-is-a-dangerous-idea-77734
11. The Law dictionary, What is Jurisprudence, https://thelawdictionary.org/jurisprudence/
12. Kassner, M., Using AI-enhanced malware, researchers disrupt algorithms used in antimalware, May 2017, https://www.techrepublic.com/article/using-ai-enhanced-malware-researchers-disrupt-algorithms-used-in-antimalware/
13. Cameron, E, Accelerating innovation through responsible AI, PwC, accessed Jan 2018, https://www.pwc.co.uk/services/audit-assurance/risk-assurance/services/technology-risk/technology-risk-insights/accelerating-innovation-through-responsible-ai.html
14. How will artificial intelligence affect the UK economy, PwC, https://www.pwc.co.uk/services/economics-policy/insights/the-impact-of-artificial-intelligence-on-the-uk-economy.html
15. KPMG, Autonomous Vehicles Readiness Index, Assessing Countries' openness, 2018,https://assets.kpmg.com/content/dam/kpmg/nl/pdf/2018/sector/automotive/autonomous-vehicles-readiness-index.pdf,

16. Marr, B., Harvard Business Review, 4 Mind blowing ways Facebook uses artificial intelligence, Dec 2016, accessed Jan 2018, https://www.forbes.com/sites/bernardmarr/2016/12/29/4-amazing-ways-facebook-uses-deep-learning-to-learn-everything-about-you/#3ab1096cccbf
17. Siri failure. https://honey.nine.com.au/2018/01/20/14/02/siri-texts-crush-by-mistake
18. Robot ethical programming British standard BS8611:2016, http://shop.bsigroup.com/ProductDetail?pid=000000000030320089
19. specific areas better than humans. http://www.pcworld.com/article/3142193/techology-business/five-things-ais-can-do-better-than-us.html
20. International Federation of Robotics, World Robotics, 2015 Service Robots, http://www.ifr.org/service-robots/statistics/
21. International Federation of Robotics, World Robotics, Service Robots Case Studies, http://www.ifr.org/service-robots/case-studies/
22. How will artificial intelligence affect the UK economy? https://www.pwc.co.uk/services/economics-policy/insights/the-impact-of-artificial-intelligence-on-the-uk-economy.html
23. AI Malware, AI quickly cooks malware that AV software can't spot. https://www.theregister.co.uk/2017/07/31/ai_defeats_antivirus_software/
24. Executive Office of the President, Artificial Intelligence, Automation, and the Economy, Dec 2016, https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF
25. How artificial intelligence conquered democracy, https://theconversation.com/how-artificial-intelligence-conquered-democracy-77675
26. The conversation, Uncanny Valley: why we find human-like robots and dolls so creepy, Nov 2015, https://theconversation.com/uncanny-valley-why-we-find-human-like-robots-and-dolls-so-creepy-50268
27. Biondi, F., Driven to Distraction, The Ergonomist, Dec 2017
28. Wikipedia, intelligence, https://en.wikipedia.org/wiki/Intelligence
29. Goebel, Randy; Poole, David L.; Mackworth, Alan K. (1997). Computational intelligence: A logical approach (PDF). Oxford [Oxfordshire]: Oxford University Press. p. 1. ISBN 978-0-19-510270-3. Archived (PDF) from the original on 7 March 2008.
30. UK Health and Safety Executive, Human Factors, http://www.hse.gov.uk/humanfactors/introduction.htm
31. Association for Project Management, Stakeholder Engagement, https://www.apm.org.uk/resources/find-a-resource/stakeholder-engagement/key-principles/
32. Mori, M. (2012). Translated by MacDorman, K. F.; Kageki, Norri. "The uncanny valley". IEEE Robotics and Automation. 19 (2): 98–100
33. Atos Opinion Paper Declares Artificial Intelligence Approaching Critical Mass, Sept 2018, https://atos.net/en-gb/2018/press-release-en-gb_2018_09_28/atos-opinion-paper-declares-artificial-intelligence-approaching-critical-mass
34. Susaria, S, The Conversation, How Artificial Intelligence can detect – and create – Fake News, May 3, 2018, http://theconversation.com/how-artificial-intelligence-can-detect-and-create-fake-news-95404
35. Harwell D., AI will solve Facebook's most vexing problems, Mark Zuckerberg say. Just don't ask when or how. Washington Post, April 2018.

https://www.washingtonpost.com/news/the-switch/wp/2018/04/11/ai-will-solve-facebooks-most-vexing-problems-mark-zuckerberg-says-just-dont-ask-when-or-how/?noredirect=on&utm_term=.421205283455

36. Johnson, C.W., The Increasing Risks of Risk Assessment: On the Rise of Artificial Intelligence and Non-Determinism in Safety - Critical Systems, 2018. School of Computing Science, University of Glasgow. http://www.dcs.gla.ac.uk/~johnson/papers/SCSC_18.pdf
37. Muehlhauser, L., Transparency in Safety-Critical Systems, Machine Intelligence Research Institute. August 2013. https://intelligence.org/2013/08/25/transparency-in-safety-critical-systems/
38. IEC Blog, New International Standard will offer risk management framework for AI, March 2019, IEC, https://blog.iec.ch/2019/03/new-international-standard-will-offer-risk-management-framework-for-ai/
39. Call for Contributions, First International Workshop on Artificial Intelligence Safety Engineering (WAISE 2018), http://www.es.mdh.se/safecomp2018/workshops/WAISE-CfP_SafeComp2018.pdf
40. Kanioura Dr, A., Critical Mass: Managing AI's unstoppable progress, Sep 2018, https://www.accenture.com/us-en/insights/digital/critical-mass-managing-ai-unstoppable-progress
41. Hohenberger et al, How and why do men and women differ in their willingness to use automated cars? The influence of emotions across different age groups, Elsevier, Oct 2016. Transportation Research Part A 94, pp 374 - 385
42. Haboucha et al., User preference regarding autonomous vehicles, Elsevier, March 2017, Transportation Research Part C 78, pp 37 - 49
43. Hulse et al., Perceptions of autonomous vehicles: relationships with road users, risk, gender and age, Elsevier, Oct 2017, Safety Science 102 (2018) pp1-13.
44. Rodriguez, J., 6 Types of Artificial Intelligence Environments, Medium.com, Jan 2017, https://medium.com/@jrodthoughts/6-types-of-artificial-intelligence-environments-825e3c47d998
45. Morris, C., Tesla's massive accumulation of autopilot miles, Inside EV's, Jul 2018, https://insideevs.com/tesla-autopilot-miles/
46. IEC 61508-3:2010, Page 48, Table A.2 Software Design and Development – Software architecture design.

BIOGRAPHY

Michael has over 34 years' experience in operational, consulting and engineering in Australia and Internationally. He is a member of Engineers Australia, and is a Function Safety Engineer. He has written and presented many papers, including socio safety economics of autonomous vehicles, and large energy storage. His consulting experience includes risk management, human factors, project management, process design and troubleshooting, accident investigation and feasibility study leader. Michael has experience in developing safety case and risk management documentation for government regulatory evaluation and organisational needs. He is a member of the Standards Australia Committee IT-043, examining governance of AI systems.