Submission to Victorian Law Reform Commission – Artificial Intelligence

Chapter 2: What is artificial intelligence?

1. Should courts and tribunals adopt a definition of AI? If so, what definition?
    a. A tool that utilises mathematical and statistical approximations to analyses data and can replicate some primitive, human cognitive functioning ability for gathering and processing of data to derive meaning from patterns, relationships and trends in data, and its context. For example, 'learning' optimisation is the function of error reduction or likelihood maximisation techniques and adjustment processes. Ai can embody a process of mentalising homeostasis - which is a developmental process to optimise one's own model of the world through constant prediction, as to the cause and weighting of sensory stimuli, and corrective action required to regain an equilibrium.
2. Are there specific AI technologies that should be considered within or out of the scope of this review?
    a. In scope - All Ai models that utilise human data, Seed models and their mutations, that were trained on datasets created through illegal or deceptive methods – (theft, copy right infringements, intellectual property, personal data used without prior permission etc) - Fair Use principles apply in USA but must be reviewed as Australian legislators view these data acquisition techniques as theft consequently, output from these models is tainted and should be legally treated accordingly.

Chapter 3: Benefits and risks of AI

3. What are the most significant benefits and risks for the use of AI by:
    a. Victorian courts and tribunals – Ai benefits can include veracity assessments and support judicial decision making, case management, research, summaries, interpreting, voice to text? Threats from poor data and unethical data collection methods and poor quality assurance processes in model development
    b. Legal professionals and prosecutorial bodies? As above.
    c. The public including court users, self-represented litigants and witnesses?
4. Are there additional risks and benefits that have not been raised in this issues paper? What are they and why are they important?
    a. Benefits not considered in the issues paper include the ability of Ai to improve veracity assessments and support judicial decision-making. Ai tools can provide context and technical advice, augment or improve anachronistic judicial decision-making practices, and save many future appeals. Some AI models can do what humans can't do or have trouble doing. Judicial offices for example, still rely on their body language interpretation abilities for veracity assessments in spite of the fact that there is no global theory that explains body movements, language, emotional display, facial expression or micro-expressions, and certainly no global theory linking any of this to lie and deception detection. Polygraphs only measure biological responses to stress it doesn't detect lies, and a smile for example doesn't mean that person is happy, averted gaze doesn't mean they are lying. Yet judicial officers continue to interpret and weigh body language and emotions expression in their veracity assessments of witnesses and defendants and, as key determinants in their judgements. Lie detection is only 54% accurate

and much less for deception detection (as there are few, if any reliable and generalisable indicators). This outcome is consistent even amongst trained law enforcement and experienced judicial officers. Judgements are made on stereotypical behaviours that are not strongly correlated to lie or deception, and are affected by personal bias and experiences, and trauma. Decisions based on false assumptions and beliefs, stereotypes and pseudo-science are obviously problematic, particularly as they relate to assessments of the neurodiverse community and those who have social disorders, who mask, camouflage or display stereotypical nervousness behaviours….that are relied upon in judicial decisions.

b. For example Human action and interactions all have meaning that we use to predict the internal states of others - intentions, thoughts, feelings, beliefs etc. The ability to perceive these states in others is referred to as Theory of Mind. Deceptive behaviour however, masks/hides the hostile and malicious intentions of others, particularly in high-stakes environments. The task of deception detection based on behaviour analysis, over time, is simply too difficult for humans, therefore these mental states remain invisible, with no agreed expression and stereotypical cues are not correlated. However, deception can be revealed through the very acts and behaviour that causes someone to accept as true or valid, that which is false. Are there specific markers of behaviour in high-stakes environments that can be used to *infer* hostile and deceptive mindsets?

   For this model, a qualitative review and analysis of material was used to inform quantitative analysis using a computational approach. This process identified features in behaviour that can be used in a machine learning algorithm to *predict* hostile and deceptive intentions based on behaviours and their high-stakes context. Further, statistical analysis has shown how the observability of other minds could be measured through key features in action, the strength of observability and perception in attribution (not body language).

   Risks - Deep learning models can be trained to deliver whatever outcome is required. For example, several deep learning models were approved by the USA's FDA to diagnose autism at an early age, a phone application would make this process cheap and widely distributed. However, both models were trained on data primarily from boys diagnosed with ASD, and focused on stereotypical behaviours of ASD boys. ASD is actually diagnosed from a spectrum of observable behaviours; girls generally mask their behaviours. These applications therefore could only 'diagnose' a small segment of boys who displayed stereotypical ASD behaviours – a greater proportion of children, primarily girls, who probably have ASD, ADHD, OCD etc would therefore go undiagnosed with these applications.

   Another application was developed for teachers to help identify students who weren't paying attention (assessed by face direction) during class. The teacher would be able to identify these students and 're-engage' them by asking direct questions. Neurodiverse children attend in non-ableist ways and public ‘humiliation’ by directly asking questions of them would result in trauma for these children. There are many more examples.

Chapter 4: AI in courts and tribunals

5. How is AI being used by: a. b. c. Victorian courts and tribunals legal professionals in the way they interact with Victorian courts and tribunals the public including court users, self-represented litigants and witnesses?
6. Are there uses of AI that should be considered high-risk, including in: a. b. c. court and tribunal administration and pre-hearing processes civil claims criminal matters How can courts and tribunals manage those risks?
    a. “High-risk’,’ legitimate’, ‘safe’ and ‘relied upon’ should be more clearly defined as they relate to Ai. High-risk for example, can an unintended consequence of poor data, synthetic data, data acquisition, lack of underpinning theory and model development and testing processes. Higher-risk exists in the deployment of these models generally and those models that are based on spurious claims, pseudo-science and false assumptions.
7. Should some AI uses be prohibited at this stage?
    a. See above. Development and application of any of these models currently, without appropriate guardrails (mentioned above), should continue but within the context of EU’s Ai legislation, and ISO 42001 used as a guide until Australia has a firm position.

Chapter 5: Regulating AI: the big picture

8. Are there lessons from international approaches that we should consider in developing a regulatory response for Victorian courts and tribunals?
    a. Big tech cannot be trusted with self-regulation. ISO and EU Ai legislation does not address the fundamental issue of data – the legalities around how it is acquired, its relevance towards required output goals, efficacy of the practices and processes for data collection, balanced and synthetic data – this is a critical issue that must be reconciled first. For example, an Australian Senate committee found the chatbot developers have committed “unprecedented theft” against Australia’s creative workers, using copyrighted materials to train their models without permission or payment. Photos of Australian children used for generative art models, without consent. Australia’s privacy regulator ruled Clearview AI’s facial recognition tool had breached privacy laws by using images from social media sites to train its AI facial recognition tool without individual’s consent. Data acquired in contravention of privacy, copy right or intellectual property law for training Ai models used in legal environments must be considered tainted and biased unless proven otherwise, or are exempted. Data issues could be the basis of future challenges. The USA utilises *Fair Use* principles to mitigate legal challenges.

       Fair use defence (doctrine in US law) is one of the limitations to copyright intended to <u>balance</u> the interests of copyright holders with the public interest *in the wider distribution and use of creative works*, this defence to copyright infringement claims certain limited uses that might otherwise be considered infringement. Fair use permits limited use of copyrighted material without first having to get permission from the copyright holder, and balances the interests of copyright holders with the public interest in the *wider distribution* and use of creative works that might otherwise be considered infringement. "Fair Use" and "transformative" form the foundation of the argument against data theft, but is does not address the issues directly and are not the solution – the wider distribution of *new material* generated by Ai based on original material is practiced but the *original material* is not widely distributed as per Fair Use principles. Fair Use is being argued by Big Tech, but they do not share or make their datasets available to the public as required by the principles, just some of the Ai models that have been trained on this contentious material. These concepts and practices should be determined by the courts and regulators.

9. What would the best regulatory response to AI use in Victorian courts and tribunals look like?
    a. Consider: a. which regulatory tools would be most effective, including rules, regulations, principles, guidelines and risk management frameworks, in the context of rapidly changing technology.

    Self-regulation or poor regulation does not work as the opportunities inherent in Ai currently outweigh the consequences, primarily because there is little accountability in Seed models and development platforms. The goals of law reform for this domain should be to promote the use of human-centric and trustworthy AI in all aspects of the justice system, and protect health, safety, fundamental rights and values, democracy, and rule of law from harmful effects of AI systems. These reforms should be considered within both the evolutionary (developmental) and revolutionary (deployment) environments of this technology. Therefore, law reform is required that holds the development and application of Ai within the justice system to high standards, heavily penalises its misuse but gives due consideration to its technical complexity, evolutionary trajectory, risk and opportunity with the context of future social and justice systems.

    A framework is required to mitigate the high-risks of Ai, but still allow legitimate development and use of AI, while slowing the use of mutant models and the development of 'frontier' models. A framework approach: Introduce a new framework legislation including definitions, the guardrails and thresholds for when they apply, and amend existing legislation, to enable enforcement by existing regulators. A quality assurance (ISO & compliance audits) and regulatory framework (set standards, compliance audits trains educates and prosecutes)is recommended that tightly controls the application of this technology within the justice domain. Self-regulation or poor regulation does not work as the opportunities outweigh the consequences currently .

    Quality assurance framework for - performance testing against fundamental standards - Framework for managing risk, data, responsible use, traceability, transparency and reliability, system maintenance, monitoring and oversight, evaluating output and feedback into system –beyond those already prescribed by the EU Ai Act and ISO 42001. These will form the baseline against which ethical and responsible Ai audits can be made.

b. whether regulatory responses should be technologically neutral, **or do some aspects of AI require specific regulation**?
Data and foundational Ai models require specific regulation - However, the issues and risks of automated decision-making and AI are inter-related. They include bias, privacy, models developed in culture that ignores social justice, rule of law, justifies data theft, free speech, facilitates scams, illegal and other anti-social activities, seeks financial rewards over social consciousness, who produce models, that reduce transparency, security and contestability.

However, current regulatory frameworks do not fully address the risks of AI. A quality assurance (ISO & compliance audits) and regulatory framework (set standards, compliance audits trains educates and prosecutes) is recommended that tightly controls the development and application of this technology within the justice domain.

c. How should court and tribunal guidelines align with AI regulation by the Australian Government?

[redacted]