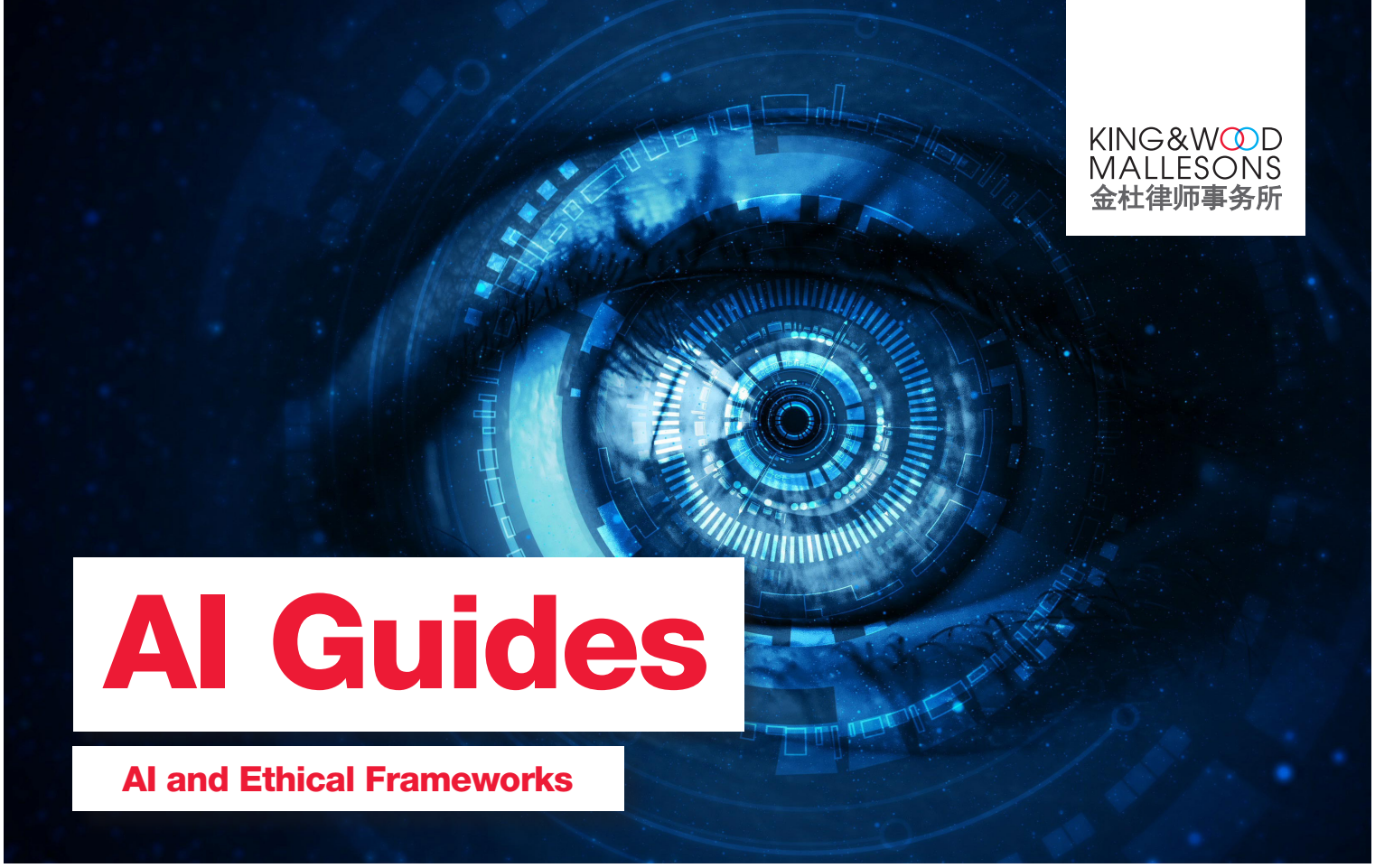

KING&WOOD MALLESONS
金杜律师事务所

AI Guides

AI and Ethical Frameworks

Is your AI system ethical? How do you know? Why should you care?

Ethical issues and technology are not new. As stated by Kranzberg in his first law of technology: *"Technology is neither good nor bad; nor is it neutral…."*

Although most companies do not set out to make their AI systems unethical or to use them in unethical ways, concerns are starting to be raised around the world about the ethical issues (and risks) that AI is increasingly presenting – this includes issues around bias, fairness, equality, transparency, empathy, dignity, privacy, human control/ oversight, sustainability, resiliency and reliability to name but a few.

So how can a non-sentient AI system be unethical?

Ethical issues arise not only from how AI systems are used but also as a result of how humans program it (i.e. the algorithms underlying the AI system) and as a result of how the AI system is trained (i.e. the initial 'training' datasets).

An example: Bias within an AI system

Bias can either be deliberately introduced into, or inherent within, an AI system.

The introduction of deliberate bias into an AI system is often easy to spot. For example, if an internet chatbot is released that learns from its interactions with people online and the chatbot is 'attacked' by internet trolls, it is likely (and in the past has quickly resulted in) the chatbot learning to respond in highly offensive and inflammatory ways.

Inherent bias may not be so obvious, and a degree of inherent bias is generally unavoidable (after all humans are likely to have collected the data underlying the initial datasets) but it can often be addressed if the developer, or the operator, actively considers whether an AI system may (or could) be biased.

For example, if you want to train your AI system to identify when a person is "in" an image but you train the AI system using pictures predominantly of Anglo-Saxon males, it is highly likely the AI system will learn to be biased towards Anglo-Saxon males and either not identify, or will incorrectly identify, non-Anglo-Saxon males (such as Asian men or women).

Similarly, if you have an AI system designed to automatically select the best candidates for a position based on their resumes and the system is trained on the traits of the top 100 employees over the last 20 years and those top 100 employees are almost all male, the AI system is likely to develop rules that preference male candidates at the expense of female candidates.

Consequences of using biased AI system (even when it is unintentional) include potential legal or regulatory issues (for example, a company may inadvertently breach anti-discrimination laws) and significant reputational issues (especially when an automated decision goes wrong (for example an individual is denied a service) or an AI powered object does not work as intended.

After all, it is significantly easier to lose the public's trust than it is to gain it.

A potential solution? Ethical frameworks

In response to growing concerns, the last 2 years has seen a proliferation of ethical frameworks, guidelines and principles (over 80 at last count) being developed

by governments, private companies, research institutions and not-for-profits.

Key examples include the European Commission's High-Level Expert Group on Artificial Intelligence (AI HELG) "Ethics Guidelines for Trustworthy Artificial Intelligence" and the OECD's "Principles on AI" (the latter having been endorsed by 42 countries including Australia and has now been adopted by the G20).

Australia has recently entered this space with the Department of Industry, Innovation and Science announcing Australia's eight "AI Ethics Principles" in November 2019.

Australia's eight principles focus on: Human, social and environmental wellbeing; Human-centred values; Fairness; Privacy protection and security; Reliability and safety; Transparency and explainability; Contestability; and Accountability.

The principles are currently being tested by a number of companies (including NAB, Telstra, CBA and Microsoft) and are currently voluntary for Australian companies who are using, or developing, AI systems.

However, whether they remain voluntary is yet to be seen…

The AI Guides are authored by:

John Swinson
Partner, Brisbane
T +61 7 3244 8050
john.swinson@au.kwm.com

Rebecca Slater
Senior Associate, Brisbane
T +61 7 3244 8147
rebecca.slater@au.kwm.com

Kendra Fouracre
Senior Associate, Melbourne
T +61 3 9643 4105
kendra.fouracre@au.kwm.com