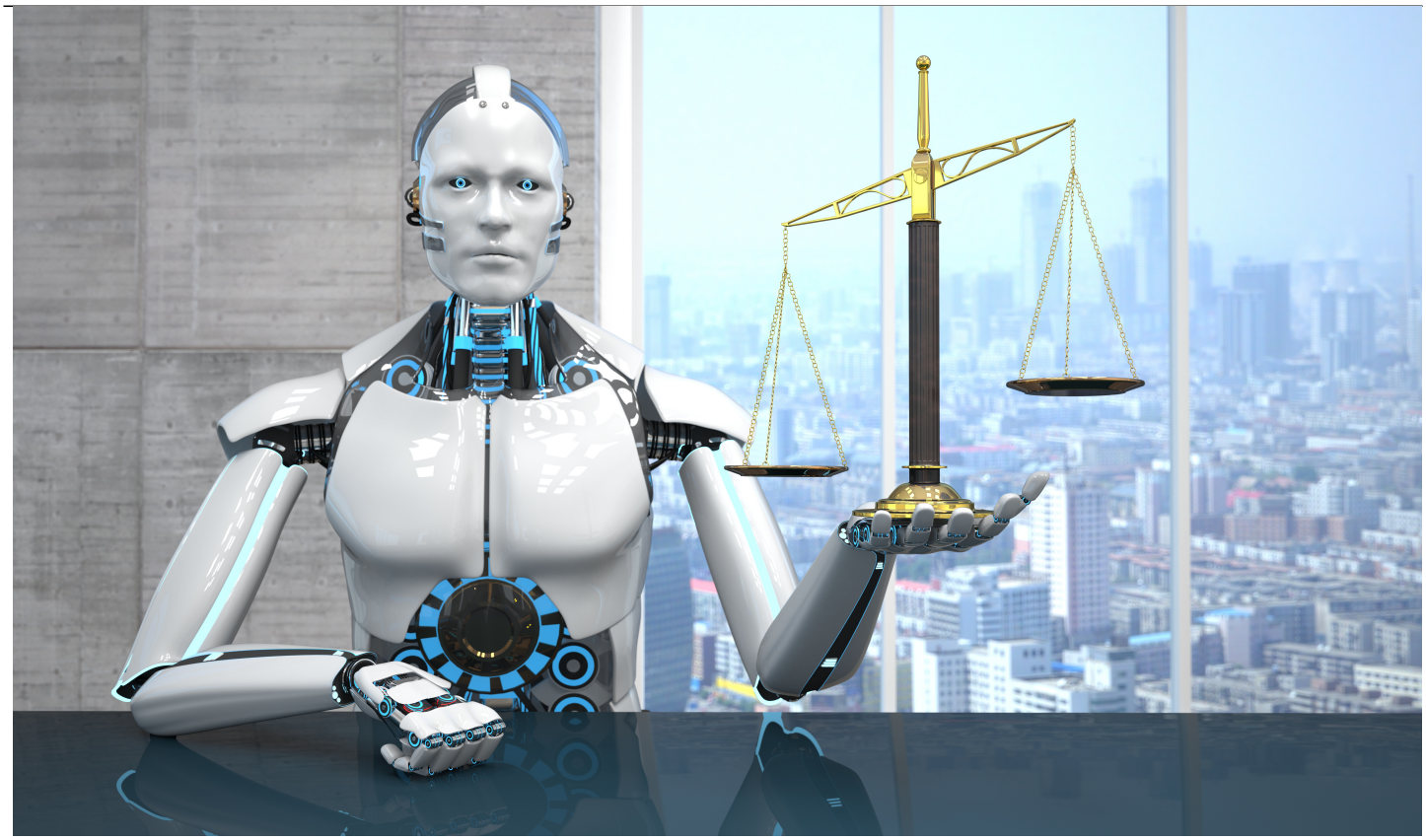
ACC DOCKET
INFORMED. INDISPENSABLE. IN-HOUSE.

Why We Should Care About Artificial Intelligence Ethics Frameworks

Technology, Privacy, and eCommerce



Cheat Sheet

- **Trust is critical**. In order for organizations to expand the use of AI, they need to be comfortable using it.
- **More regulation, more protection.** To assure users that AI is safe, more oversight is needed.
- **Regulation isn't a cure-all.** Even when regulated, AI will not always be used in a trustworthy manner.
- **Addressing distrust.** While regulation will likely increase trust, organizations need to develop their own policies to create a more trustworthy relationship with users.

What are the criticisms levied against artificial intelligence (AI) ethics frameworks? And do we really care about the role for law in the regulation of AI? After all, we can always play catch up after the fact as we usually do.

"You can be unethical and still be legal — that's the way I live my life," as Mark Zuckerberg said in 2004 while still a student at Harvard. With that in mind, let's look at the difference between ethics and legality.

Ethics versus law

Criticisms of ethical frameworks (or principles) intended to drive the design, adoption, and use of AI focus on three main points:

1. Ethics frameworks are optional. Therefore, they can be readily disregarded for other value-optimizing approaches.

2. Ethics frameworks don't guide organizations on how to implement aspirational standards, rather they rely on regulated standards — of which there are few.

3. Ethics frameworks do not work on a global level.

The inconsistent approaches taken by the AI ethics frameworks illuminate the problem of using the trust lens.

Trustworthy AI needs to be defined by regulatory compliance alone. And that, in turn, will build trust, which in turn will further inform regulation.

The growth and value of AI

The results of the 2020 McKinsey global survey on AI support the notion that AI is increasingly designed and adopted by companies to generate value. Increasingly, that value is coming in the form of revenue and cost savings. A variety of industries attributed up to 20 percent or more of their organizations pre-tax earnings to AI.

AI is “increasingly prolific in modern society and most every organization,” the report concluded. McKinsey anticipated that customer demand for AI would increase, driven by organizational owners’ demands for revenue increases and/or cost decreases. Organizational investment in AI attributed to the COVID-19 pandemic focused on revenue stability during the changed business environment. As organizations expand their AI, many publish or endorse principles and frameworks.

> A variety of industries attributed up to 20 percent or more of their organizations pre-tax earnings to AI.

The 2020 McKinsey report summarizes the analysis of 36 prominent AI documents consisting of principles and/or frameworks and discusses trends across them. The trends closely mirror themes identified by a similar study conducted the same time by the Berkman Klein Centre for Internet & Society Research, Harvard University.

Eight common themes appear across all AI principle frameworks:

1. Privacy;

2. Accountability;

3. Safety and security;

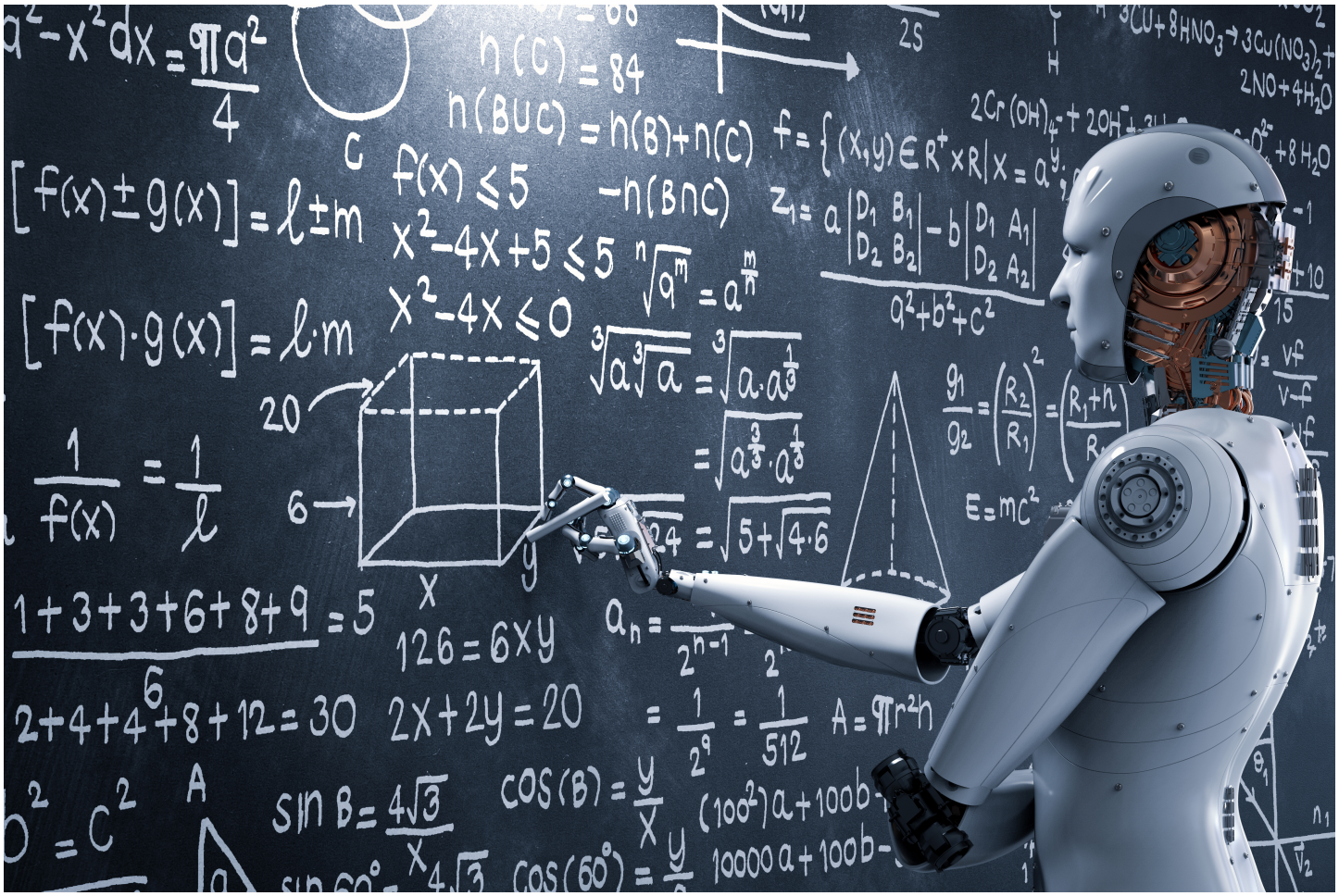
4. Transparency and explainability;
5. Fairness and non-discrimination;
6. Human control of the technology;
7. Professional responsibility; and
8. Promotion of human values.

Trust in AI and trustworthy AI

A decline of four percent in trust of technology globally and a decline of six percent in trust of technology in Australia was assessed by the 2020 Edelman Trust Barometer.

The 2021 Edelman Trust Barometer showed further declines in trust of technology over the 12-month period between assessments with another six percent decline globally and another five percent decline in Australia. The cumulative decline in trust of technology measured 2011-2021 was assessed at nine percent globally and 13 percent in Australia.

The continued decline in trust in technology, including AI, highlights concerns regarding the rapid pace at which technology develops, its augmentation of reality, and the inverse relationship between adoption and use of technologies and laws that regulate design and development.



Emerging technologies such as AI are among the most feared and, given this is the direction in which considerable investment is being made, trust is only set to decline further. This is consistent with the increasing impact that technologies such as AI are having on many aspects of life. The increasing activity to develop, adopt, and use AI has led the World Economic Forum to brand the current era the "Fourth Industrial Revolution," with much change still to come.

Trust in the AI context is the enabler of decisions between organizations (and people) as it reflects the level of confidence each has in the other. Digital trust is a concept based on each organization's digital reputation as well as the assurance levels provided by each organization's people, processes, and technology to build a secure digital world.

The current market for technologies, including AI, presents customers with vast choice, resulting in an increased ability for them to set expectations and take their business elsewhere when their expectations are not met. Lost trust between customers and organizations results in lost business and revenue and, in the current marketplace, trust is a highly traded commodity that impacts benefit and value in seconds.

> Lost trust between customers and organizations results in lost business and revenue and, in the current marketplace, trust is a highly traded commodity that impacts benefit and value in seconds.

AI challenges the traditional ways of establishing and maintaining trust, which often involve lengthy contractual negotiations and agreements between two or more parties prior to commencement. AI demands for the establishment of trust to be instantaneous in a dynamic way that also scales for a multitude of possibilities over the duration of the relationship. Vast potential for interconnectivity also requires a much wider chain of trust to be established between organizations, governments, loosely connected social groups, and people not known to each other.

Trust is hard to define. Its understanding varies from person to person. It is defined differently across cultural and religious norms. Many definitions claim it can be broken down into quantifiable ethical measures such as: truthfulness, integrity, confidentiality, responsibility, and more. It also extends into tangible components such as cybersecurity, compliance, and responsible business. It is important to not only understand what trust means in a particular context but also to define what it means for participants in that context as it underpins every decision and interaction between participants.

Trust and transparency are consistent partners. Organizations that actively demonstrate what they are doing, and that they are doing the "right" thing, inspire trust. Such demonstrations are not limited to regulatory compliance, but rather extend beyond that to expressions that encapsulate the concept of doing the "right" thing because it is the "right" thing to do. The value or benefit derived by organizations that make such statements is not readily measured or documented given uncertainty about what is "right" in any given context.

> Trust and transparency are consistent partners.

Regulation is the predominant measure of "right" in a legal sense, however other indicators such as market share, revenue growth, stagnation or decline, and externally authored publicity can indicate how accurately an organization is able to assess what is "right" in any given context.

The theme of trust appears to have emerged in the context of AI principles and ethics frameworks following the publication of the European Union AI Trustworthy Guidelines in 2019. The Guidelines identify trust as being defined by all underlying principles or framework elements so that implementation of such principles and frameworks will likely result in trustworthy AI. Trust is identified as pivotal in the Guidelines:

> In a context of rapid technological change, we believe it is essential that trust remains the bedrock of societies, communities, economies and sustainable development. We therefore identify 'Trustworthy AI' as our foundational ambition, since human beings and communities will only be able to have confidence in the technology's development and its applications when a clear and comprehensive framework for achieving its trustworthiness is in place.

Trustworthiness is presented by the Guidelines as fundamental to the development, adoption, and use of AI. Unwanted consequences are identified as is low adoption and use for AI developed in the absence of consideration of trust. The Guidelines advocate that AI devoid of trust denies humanity of its "potentially vast social and economic benefits."

However, the Guidelines do not identify trust as a discrete element of principles or frameworks, rather the Guidelines only identify trust as a secondary element to the discrete elements of lawfulness, ethical, and technical robustness. The inconsistency between the advocacy in the Guidelines for trust being fundamental to the realization of benefits or other value, and its simultaneous demotion to a secondary element, is stark. This seems to have set the tone for principles and frameworks that followed.

Inconsistent approaches to trust and "trustworthy AI"

Several frameworks developed and adopted in the private sector emphasized the role of the safety and security in fostering trust in AI, according to the Harvard University study, which was published after the European Guidelines. That study further references the AI Policy Principles of the German-based Information Technology Industry Council statement that the success of AI depends on users "trust that their personal and sensitive data is protected and handled appropriately." This interpretation appears to apply trust as a secondary element to regulatory adherence, in this case adherence to privacy regulation. Other private sector AI frameworks or principles documents also identify trust as a secondary element to the achievement of value, or within the context of regulation, or both.

A focus on value in evident in the 2019 Fujitsu global AI framework with trust a secondary element embedded as part of "fairness and safety to prevent discrimination and harm." The Fujitsu global AI framework identifies five aspirational elements: the provision of value to customers and society, striving for human-centric AI, striving for a sustainable society with AI, striving for AI that respects and supports people's decision making, and emphasizing transparency and accountability for AI. It is assumed that operationalizing these aspirations will result in "trustworthy AI," although the framework is heavily grounded in a regulatory measure of trustworthiness.

Let's use Facebook (now owned by Meta) as a different example of trust referencing, or lack thereof, in an AI ethics framework. Facebook's five key pillars for "responsible AI" disregard trust. The five pillars — privacy and security, fairness and inclusion, robustness and safety, transparency and control, and accountability and governance — rely heavily on existing regulatory frameworks for their definition and are devoid of any reference to also being trustworthy.

Absent from the five pillars is a clear statement of use of AI to drive value or benefit for customers, the company itself and its shareholders, or any specific party that could derive value or otherwise benefit. Instead, the five pillars advocate the use of AI for the benefit of "people, society, and the world."

Also missing is a statement explaining the company's goals in developing, adopting, and using AI to drive value or benefit. The absence of such a statement, coupled with inclusion of very broad classes of potential benefactors, is a lack of transparency, thereby sparking mistrust.

Government approach

Government-authored principles and frameworks also embody trust as a secondary theme. The Australian Government Artificial Intelligence Framework , for example, states an intention to guide business and governments to "responsibly design, develop, and implement AI. While the Australian Government refers to a need to "trust it [AI] is safe, secure, and reliable," the Framework is described as "aspirational and complementary, not to be viewed as a substitute to existing AI regulation and practices."

The Australian Government has developed an AI ethics framework and associated collateral to guide the operationalization of the framework, however, in doing so, it has failed to solidify the role of government in regulatory reform, by making the framework subservient to existing AI practices (with no reference to the extent to which existing AI practices adhere to current regulation).

> The Australian Government has developed an AI ethics framework and associated collateral to guide the operationalization of the framework, however, in doing so, it has failed to solidify the role of government in regulatory reform.

The requirement for trust in AI was also addressed in a subsequent Commonwealth Scientific and Industrial Research Organisation (CSIRO) study in which trust was identified as "essential to achieve widespread adoption of AI." However, the report's foreword CEO of CSIRO, Larry Marshall, Ph.D., does not acknowledge trust, and trust is again assigned as, at best, a secondary theme:

> AI is already a well-established technology, with applications across many industries starting to take shape … the success of our industries of the future will be determined by whether AI is simply used to cut costs, or whether we take full advantage of this powerful technology to grow new opportunities and create new value.

With no reference to trust as a primary element to be implemented in the design, adoption, and use of AI one is again left with the proposition that while trust is acknowledged as essential, it is defined by reference to regulation and considered secondary to value and benefits derived from the adoption and use of AI.

The European Commission approach further illustrates that trust, while an essential element to the design, adoption, and use of AI, is defined by reference to regulation and it is through regulation that trust in AI might increase.

> The European Commission approach further illustrates that trust, while an essential element to the design, adoption, and use of AI, is defined by reference to regulation and it is through regulation that trust in AI might increase.

The 2018 European AI strategy revolved around two focus areas, one of which was "trustworthy AI" (the other being excellence in AI). The General Data Protection Regulation (GDPR) introduced later that year, was foreshadowed by the European AI strategy as "a major step for building trust, essential in the long term for both people and companies." However, the 2020 Edelman Trust Barometer assessment of trust in technology does not incorporate the foreshadowed impact of the GDPR, assessing, instead, the decline of five percent in trust of technology in the United Kingdom during the 2019 calendar year. The 2021 Edelman Trust Barometer did not publish findings on trust in technology that are comparable with the earlier year assessments.

Indicative of a continued decline in trust in technology during the 2020 calendar year in the United Kingdom, is the European Commission action to propose an AI regulation (AI Act) 2022. The AI Act can be considered as an attempt to comprehensively regulate AI — it also supports the notion that trust in AI is grounded in regulation.

The AI Act lists prohibited AI applications including:

- Manipulative online practices that produce physical or psychological harms to individuals or exploit their vulnerability on the basis of age or disability;
- Social scoring producing disproportionate or de-contextualized detrimental effects; and
- Biometric identification systems used by law enforcement authorities in public spaces (where

their use is not strictly necessary or when the risk of detrimental effects is too high).

Such a prohibition proposed within the AI regulation suggests that trust in AI is only borne out of regulatory compliance because of the need to prohibit uses and enforce such prohibitions. If trust in AI was in fact a primary element in the design, adoption, and use of AI, there would be no need for regulation to identify prohibitions.

What is the role of regulation in promoting trust in AI?

Three main elements were identified by CSIRO as encouraging AI adoption by increasing trust:

1. AI users need to feel trust that the AI will complete a task while achieving safety, efficiency, and overall quality measures that would not be achieved by a human.

2. The user must trust that personal information won't be inappropriately managed. This element relies on regulatory adherence to current privacy and secrecy/security laws, and to some extent any contractual requirements for a lesser or greater standard that regulation imposes.

3. The user needs to trust that the AI is doing what it is supposed to do, which may or may not be consistent with what a user thinks the AI is supposed to do.

While the first element might guide regulation to assist increased trust in AI, the second and third elements are grounded in existing regulation and the rights of independent users to adopt and adapt AI and the results of AI in any manner they see fit.

Regulation constraining the way users adopt and adapt technologies is not generally accepted for its perceived stifling of human creativity and development, especially when technologies can be used for both legal and illegal pursuits by users. For example, in Australia, the use of technologies that facilitate the illegal copying or distribution of music are not regulated, however it is a breach of copyright law when music is copied or distributed without the permission of all relevant copyright owners (except in limited circumstances). In a similar fashion, regulation of AI currently focuses on data owners' rights rather than the AI technology and algorithms.

A new monopoly right has been supported in the form of the European Commission plans for a new Data Act anticipated in 2022. The new Data Act is intended to "create a solid framework for digital trust, opening up public sector data, removing digital borders, encouraging trade in data, opening up competition and facilitating better security within the EU single market." And while there are calls for a globally consistent approach to data ownership protection, copyright, confidentiality, and database rights, and this may promote the use of AI, the likelihood of achieving such consistency globally is low. Nation states fight to retain their autonomy and the rights of citizens within their own jurisdictions (and sometimes extraterritorially).

> And while there are calls for a globally consistent approach to data ownership protection, copyright, confidentiality, and database rights, and this may promote the use of AI, the likelihood of achieving such consistency globally is low.

Aspects of the AI Act, such as different rules for different risk-levels of AI that are not prohibited, might operate to increase trust in AI, and indicate a "trustworthy AI" platform, but these are not

readily altered as to keep pace with new AI designs, adoptions, adaptations, and uses, and are insufficient to manage trust as the highly traded commodity that impacts benefit and value on organizations associated with AI. While regulation can shape user perception of trust in organizations that design, adopt, and use AI, as we have seen in the United Kingdom, regulation does not always operate to increase trust. Increasingly, it is regulatory compliance that is the benchmark for “trustworthy AI” and not the measure of trust itself.

Conclusion

A life lived by the rule of law alone is, in and of itself, a choice grounded in ethical considerations, albeit to reject unregulated ethical considerations such as trust. In a similar fashion, the increasing distrust in technology and associated rise in regulation of AI illustrates that reliance on trust as a primary element in an AI ethics framework, is increasingly required to be defined by regulatory compliance alone.

Sara Wedgwood

Corporate Counsel

Fujitsu Australia Limited

Sara Wedgwood is corporate counsel for Fujitsu Australia Limited and gives specialist advice to one of its subsidiaries as general counsel and company secretary to Fujitsu Australia Software Technology Pty Ltd.