Debating Artificial Intelligence: The Fox versus the Hedgehog

By Donald K. Emmerson

SYNOPSIS

Singapore in Southeast Asia and Stanford University in the United States are focal points for discussions of AI and how it can be made to help not hurt human beings. A recent panel at Stanford illustrates the difficulty and necessity of bringing both generalist and specialist perspectives to bear on the problem.

COMMENTARY

SINGAPORE HAS been described as "a thriving hub for artificial intelligence" (https://www.businesstimes.com.sg/opinion/artificial-intelligence-in-singapore-pervasive-powerful-and-present). In May 2019, Singapore's Personal Data Protection Commission (PDPC) released the first edition of "A Proposed Model AI Governance Framework (https://www.pdpc.gov.sg/Resources/Model-AI-Gov).

That "accountability-based" document would "frame the discussions around harnessing AI in a responsible way" by "translat[ing] ethical principles into practical measures that can be implemented by organisations deploying AI solutions". The guiding principles it proposes to operationalise are that AI systems should be "human-centric" and that decisions made by using them should be "explainable, transparent, and fair".

Ethical Principles in AI

Ethical principles are crucial in AI. They are philosophical compared with the technical character of practical measures. While Singaporeans discuss how to put which principles into practice, variations on that conversation are underway in Silicon Valley. A case in point is a recent discussion of AI at Stanford University, whose Artificial Intelligence Lab was established in 1962.

This comment focuses on how differently scholars in the humanities may approach the challenge of making AI "human-centric" compared with their colleagues in computer science.

At Stanford in April 2019, before an audience of nearly 1,700 people, a panel on AI (https://www.youtube.com/watch?v=d4rBh6DBHyw) brought together a fox and a hedgehog. The "fox" was a historian, Hebrew University of Jerusalem professor Yuval Noah Harari. The "hedgehog" was an engineer, Stanford professor Fei-Fei Li.

A poet in ancient Greece is said to have coined these metaphors by remarking: "The fox knows many things, but the hedgehog knows one big thing." The contrast is often used in academic discourse to distinguish generalists from specialisers. Viewed in that light, Yuval Harari's latest book, *21 Lessons for the 21st Century* (https://www.theguardian.com/books/2018/aug/15/21-lessons-for-the-21st-century-by-yuval-noah-harari-review), is an eclectic read worthy of a fox. The titles of its chapters include "God," "War," "Humility," and "Science Fiction". The subject of AI crops up as well.

The Hedgehog

As an undergraduate at Princeton, Fei-fei Li co-edited a book, *Nanking 1937: Memory and Healing* (2002), that delved hedgehog-style into "one big thing"— the Nanking Massacre. Since earning her doctorate in electrical engineering, Li has understandably concentrated on working and publishing in her discipline, computer science. Her specialty is AI, whose importance surely qualifies it as "one big thing," if only as shown by the huge turnout for the panel.

The conversation between Harari and Li was intriguing but incomplete. Prof. Li co-directs Stanford's Human-Centered AI Institute. "Human-Centered AI" activity sounds foxy — interdisciplinary. It was Harari, however, who played the boundary-crossing fox by linking infotech with biotech to suggest that their overlapping could gestate an ability and a proclivity to "hack human beings".

Linking AI to psychology, he wondered whether personal decisions could someday be "outsourced to algorithms". Could neuroscientific AI be used to "hack love" by causing an infatuation that would not otherwise have occurred? Harari brought illness in as well: "In a battle between privacy and health," he predicted, "health will win."

Shifting into political science, he worried that AI could become a "21st century technology of domination". Others share his anxiety. On biotech, for instance, there is Jamie Metzl's just-published *Hacking Darwin: Genetic Engineering and the Future of Humanity* (https://www.npr.org/2019/05/02/718250111/hacking-darwin-explores-genetic-engineering-and-what-it-means-to-be-human).

Hedgehogs & Foxes: Collaboration Needed

Harari's concerns almost made "human-centered AI" sound oxymoronic. But as a fox untrained in computer science, he lacked the knowledge that a hedgehog with digital depth would have brought to bear on the topic. Li had the necessary expertise on AI. But she did not respond to Hariri's worries and speculations beyond assuring him and

the audience that interdisciplinarity and ethics were definitely on her institute's agenda.

Without hedgehogs to keep them realistic, foxes can get carried away. Without foxes to keep them contextual, hedgehogs can silo themselves. Helpful in this context — forgive the foxy term — is a vigorous recent defence of foxiness as a career choice: David Epstein's *Range: Why Generalists Triumph in a Specialised World* (https://www.npr.org/2019/05/02/718250111/hacking-darwin-explores-genetic-engineering-and-what-it-means-to-be-human).

Already someone somewhere may be drafting an antithesis to the foxiness of *Range*. Perhaps its title will be *Depth: Why Specialists are Necessary in a Generalist World*.

In any case, to this author's shallow knowledge, foxes and hedgehogs are not sworn enemies, either on paper or in nature. So here's to deep range and wide-ranging depth, unlikely in the work of a single scholar, but possible through animalian collaboration.

Donald K. Emmerson, a confessed fox, heads the Southeast Asia Program in the Shorenstein Asia-Pacific Research Center at Stanford University, where he is also affiliated with the Abbasi Program in Islamic Studies and the Center on Development, Democracy, and the Rule of Law. He contributed this article specially to RSIS Commentary. His edited book, The Deer and the Dragon: Southeast Asia and China in the 21st Century, is forthcoming in 2019.