: What policies should govern Artificial Intelligence (AI) applications?

by Daniel J. Power

Editor, DSSResources.COM

Artificial Intelligence (AI) applications are reaching the point where both organizations and governments must decide about creating policies regulating use and liability associated with using AI. Managers and politicians should debate what is permitted and what is encouraged in the development and use of Artificial Intelligence. There are many outstanding issues. For example, who is liable if an AI application's decision or recommendation in wrong? How and when should AI applications be tested and validated? How will people respond to AI applications? Recently, Google announced that its new virtual AI assistant called Duplex could make phone calls for you using a human like voice that sounded like a real person. This announcement was poorly received and sparked a "harsh backlash". Within a few days, Google had changed its policy and announced that its AI assistant would identify itself as a robot when making phone calls, cf., Escher and Lynley (2018), Welch (2018) and Vomiero (2018).

Some sources and estimates project automated bots could take nearly four in 10 (38%) jobs in the U.S., and take 30% of jobs in the United Kingdom by 2030 (Jenkins, 2017). Artificial Intelligence and bots appear to have a comparative advantage for performing certain jobs in transportation and storage, manufacturing, and retail. Ricardo in **The Elements of Political Economy** (1821) discussed comparative advantages among nations as a motivation for free trade, but comparative advantage between people and robots involves similar yet somewhat different issues (Worstall, 2015).

Recently, the Singapore Government Info-communications Media Development Authority announced establishment of an Advisory Council on the Ethical Use of AI and Data. "The Advisory Council will assist the Government to develop ethics standards and reference governance frameworks and publish advisory guidelines, practical guidance, and/or codes of practice for the voluntary adoption by the industry." The current discussion paper "recommends two key principles: 1) Decisions made by or with the assistance of AI should be explainable, transparent and fair to consumers; and 2) AI systems, robots and decisions should be human-centric," cf., IMDA Press Release (2018).

The Governance of AI Program based at the University of Oxford’s Future of Humanity Institute strives to steer the development of artificial intelligence for the common good using research and policy engagement. The program's "focus is on the challenges arising from transformative AI: advanced AI systems whose long-term impacts may be as profound as the industrial revolution", cf., https://www.fhi.ox.ac.uk/governance-ai-program.

: What policies should govern Artificial Intelligence (AI) applications?

The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation report (Brundage and Avin, 2018) notes "The use of AI to automate tasks involved in surveillance (e.g. analysing mass-collected data), persuasion (e.g. creating targeted propaganda), and deception (e.g. manipulating videos) may expand threats associated with privacy invasion and social manipulation. We also expect novel attacks that take advantage of an improved capacity to analyse human behaviors, moods, and beliefs on the basis of available data. These concerns are most significant in the context of authoritarian states, but may also undermine the ability of democracies to sustain truthful public debates. (p. 6)"

USA Today author Brett Molina (2018) reported on an MIT Research project that created a psychopath Artificial Intelligence named Norman. Molina writes "So, why would MIT create a psycho AI? It's all about algorithms, and when things might go awry it's not as simple as blaming the machine. 'The data that is used to teach a machine learning algorithm can significantly influence its behavior,' reads a statement on their website. "So when people talk about AI algorithms being biased and unfair, the culprit is often not the algorithm itself, but the biased data that was fed to it." See http://norman-ai.mit.edu/.

U.S Chief Justice John Roberts spoke to high school graduates at a June 2018 commencement ceremony, his message "Beware the robots". Wolf (2018) reports "Roberts warned that artificial intelligence and big data can alter the way people perceive the world". ... "The result," Roberts said, "can be a narrowing and over-simplification that is contrary to individuality and creativity." Wolf noted Roberts said "I worry that we will start thinking like machines".

Democratic US Presidential candidate John Delany notes "I want us to start thinking about automation, artificial intelligence and machine learning." He argues "We must take proactive steps to make sure that workers benefit, that automation creates more jobs overall and that these new technologies are implemented in an unbiased and ethical way. We need a National AI Strategy that puts workers first and makes sure that automation is an economic positive for our country." He argues a strategy should involve "Consulting with ethicists, privacy experts and representatives from all communities to make sure that AI is used in a way that is fair for all and that respects our individual rights." (email, July 28, 2018)

Sample (2017) reports "The rise of artificial intelligence (AI) has led to an explosion in the number of algorithms that are used by employers, banks, police forces and others, but the systems can, and do, make bad decisions that seriously impact people’s lives. But because technology companies are so secretive about how their algorithms work – to prevent other firms from copying them – they rarely disclose any detailed information about how AIs have made particular decisions." He notes

: What policies should govern Artificial Intelligence (AI) applications?

"Sandra Wachter, Brent Mittelstadt, and Luciano Floridi, a research team at the Alan Turing Institute in London and the University of Oxford, call for a trusted third party body that can investigate AI decisions for people who believe they have been discriminated against."

The American HBO science fiction western titled Westworld (2016) explores the coupling of Artificial Intelligence with human-appearing androids. The plots are thought provoking. In general, the problems associated with intelligent human-appearing androids seem to outweigh any benefits (as least in fiction). Other AI fiction include the Terminator (1984) where AI are both hero and villain. Blade Runner (1982) another android plot that has twisted, AI control issues. C-3PO and R2-D2 of Star Wars are the the best known, and cutest Artificial Intelligence bots. Hal of 2001: A Space Odyssey (1968) is the most omnipresent AI and the most troubling. Reality and pseudo-reality create confusion for people. At a future time, AI might stand for our Artificial Idiocy or an Alien Invasion.

There are potential problems and ethical issues with adoption and use of Artificial Intelligence. Metz (2018) reports "A.I. systems also exhibit strange and unexpected behavior because the way they learn from large amounts of data is not entirely understood. That makes them vulnerable to manipulation; today’s computer vision algorithms, for example, can be fooled into seeing things that are not there." Criminals may exploit AI for crime (Bajarin, 2016). AI may reduce the number of jobs for people (Bajarin, 2016; Bossmann, 2016). Artificial stupidity (Bossmann, 2016).

There are four major categories of Artificial Intelligence applications. These categories are: 1) Automated Planning, 2) Machine Learning, 3) Machine Reasoning and Rule-Based, and 4) Natural Language Processing. In a specific AI application multiple AI categories can be used in development. Automated planning means an application can construct a sequence of actions to reach a final goal. Machine learning algorithms are used for classification and prediction. Machine reasoning imitates thinking and applications draw inferences and conclusions based upon data inputs. Finally, natural language processing applications provide a means for computers to understand both written text and human speech.

People should be proactive in governing and regulating use of AI in society, organizations, homes and in devices. Using AI is a large scale experiment. We, as scientists, have an obligation to insure that people are not harmed by AI. All of us are becoming subjects in uncontrolled AI experimentation.

Some Suggested Policies

: What policies should govern Artificial Intelligence (AI) applications?

1. The owner of an AI application can not limit her/his/its liability related to use of the AI in any way, even with a disclaimer limiting liability.

2. If an expert is using an AI application for support, then both the expert and the owner of the AI share liability.

3. AI applications that replace human employees should be discouraged, except when the task is dangerous or creates other harms for a human.

4. AI applications should be designed to support humans and enhance the quality of human life.

5. Development of autonomous, self-replicating AI robots should be discouraged.

6. AI applications should **only** make autonomous decisions in routine, recurring decision situations. Even in those situations, knowledgeable humans should regularly monitor the decisions and consequences to insure the AI application is performing satisfactorily.

7. It should always be disclosed when an AI application is making a decision and the reasoning behind the decision should be understood and transparent to anyone impacted by the decision.

8. Each AI application should be tested to insure it is fair, accountable, and transparent. Utilisers of an AI must understand why the machine model thought an action, conclusion or recommendation was most appropriate. What information was analyzed to reach a conclusion? How certain is the conclusion? What is the error rate? A human must be convinced that the AI results are confirmed at a satisfactory confidence level by the data.

9. AI application code should be restricted and well-secured. Passwords and other security measures related to an AI application should be routinely checked for vulnerabilities.

10. AI and machine learning are critical technologies and export of source code should be restricted.

11. A trusted third party body, perhaps at the national or international level, should be established that can investigate AI decisions when a charge is brought by a person or group who believe they have been discriminated against.

Calo (2017), in his excellent primer on AI and policy issues, concluded optimistically "AI has managed to capture policymakers' imaginations early enough in its life-cycle that there is hope we can yet channel it toward the public interest." Capturing the imaginations of policy makers must be translated into appropriate policies. That task is only just beginning.

The Gartner Glossary defines information governance as "the specification of decision rights and an accountability framework to ensure appropriate behavior in the valuation, creation, storage, use, archiving and deletion of information. It includes the processes, roles, and policies, standards and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals." Governance is complex, it is important for managers to do more than establish policies.

An IEEE Position Statement entitled Ethical Aspects of Autonomous and Intelligent Systems (June 24, 2019), approved by the IEEE Board of Directors, has seven governance principles:

1) Human Rights: A/IS should be developed and operated in a manner that respects internationally recognized human rights.

2) Well-being: A/IS developers should consider impact on individual and societal well-being as the central criterion in development.

3) Data Agency: A/IS developers should respect each individual's ability to maintain appropriate control over their personal data and identifying information.

4) Effectiveness: Developers and operators should consider the effectiveness and fitness of A/IS

technologies for the purpose of their systems.

5) Transparency: To the greatest extent feasible, the technical basis of the particular decisions made by an A/IS should be discoverable.

6) Accountability: A/IS should be designed and operated in a manner that permits production of an unambiguous rationale for the decisions made by the system.

7) Awareness of Misuse: Designers of A/IS creators should consider and guard against potential misuses and operational risks.

The Beijing Academy of Artificial Intellegence (https://www.baai.ac.cn/) divides AI principles into three categories: 1) Research and Development, 2) Use, and 3) Governance. Some of the principles should apply in all three.

Artificial Intelligence can augment, supplement, and support and possibly replace human intelligence. That potential is why we need to appropriately govern Artificial Intelligence (AI) applications.

References

Author Unknown, "Benefits & Risks of Artificial Intelligence," Future of Life Institute, N.D. at URL https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/

Bajarin, T., "These Are My 2 Biggest Fears About Artificial Intelligence," Time, November 14, 2016 at URL http://time.com/4569585/ai-robots-fears/

Beijing Academy of Artificial Intellegence. "Beijing AI Principles," May 28, 2019 at URL https://www.baai.ac.cn/news/beijing-ai-principles-en.html

Bossmann, J., "Top 9 ethical issues in artificial intelligence," World Economic Forum, Oct. 21, 2016, at URL https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/

Brundage, M. and J. Bryson, "Smart Policies for Artificial Intelligence," working paper, N.D. at URL https://arxiv.org/ftp/arxiv/papers/1608/1608.08196.pdf

Brundage, M. and S. Avin et al, "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," Future of Humanity Institute, February 2018, based upon M. Brundage and Shahar Avin, co-chaired workshop entitled “Bad Actor Risks in Artificial Intelligence” in Oxford, United Kingdom, February 19 and 20, 2017 at URL https://maliciousaireport.com/.

Calo, R., "Artificial Intelligence Policy: A Primer and Roadmap," UC-Davis Law Review, Vol. 51:399, 2017, pp. 399-435 at URL https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Calo.pdf

Escher, A. and M. Lynley, "8 big announcements from Google I/O 2018," TechCrunch, May 8, 2018 at URL https://techcrunch.com/2018/05/08/8-big-announcements-from-google-i-o-2018/.

Gartner Glossary, "Information Governance," at URL https://www.gartner.com/en/information-technology/glossary/information-governance

IEEE, "Ethical Aspects of Autonomous and Intelligent Systems," June 24, 2019 at URL https://globalpolicy.ieee.org/wp-content/uploads/2019/06/IEEE19002.pdf.

IMDA, "Singapore implements Artificial Intelligence governance and ethics initiatives," Infocomm Media Development Authority Press Release, June 5, 2018 at URL http://dssresources.com/news/4964.php

Jenkins, A., "Robots Could Steal 40% of U.S. Jobs by 2030," Fortune, March 2017 at URL http://fortune.com/2017/03/24/pwc-robots-jobs-study/

Malliaraki, E., "Toward ethical, transparent and fair AI/ML: a critical reading list," Medium, February 9, 2018 at URL https://medium.com/@eirinimalliaraki/toward-ethical-transparent-and-fair-ai-ml-a-critical-reading-list-d950e70a70ea

Metz, C., "Good News: A.I. Is Getting Cheaper. That's Also Bad News," The New York Times, Feb. 20, 2018 at URL https://www.nytimes.com/2018/02/20/technology/artificial-intelligence-risks.html.

Molina, B., "Terrifying: an artificial intelligence was fed Reddit captions. Now it's a 'psychopath'," USA Today, June 7, 2018 at URL https://www.usatoday.com/story/tech/nation-now/2018/06/07/artificial-intelligence-fed-reddit-captions-became-psychopath/681888002/

Power, D. J., "Will thinking machines make better decisions than people?" Decision Support News, Vol. 13, No. 7, 04/01/2012 at URL http://dssresources.com/faq/index.php?action=artikel&id=233

Power, D. J., "What is augmented decision-making?" Decision Support News, Vol. 14, No. 22, 10/27/2013 at URL http://dssresources.com/faq/index.php?action=artikel&id=279

Sample, I., "AI watchdog needed to regulate automated decision-making, say experts," The Guardian, January 17, 2017 at URL https://www.theguardian.com/technology/2017/jan/27/ai-artificial-intelligence-watchdog-needed-to-prevent-discriminatory-automated-decisions.

Vomiero, J., "Google's AI assistant must identify itself as a robot during phone calls," Global News, May 12, 2018 at URL https://globalnews.ca/news/4204648/googles-ai-identify-itself-robot-phone-calls/.

Welch, C., "It's hard to believe AI can interact with people this naturally," The Verge, May 8, 2018 at URL https://www.theverge.com/2018/5/8/17332070/google-assistant-makes-phone-call-demo-duplex-io-2018

Wolf, R., "Chief Justice John Roberts to high school graduates (and his daughter): 'Beware the robots'," USA Today, June 7, 2018 at URL https://www.usatoday.com/story/news/politics/2018/06/07/beware-robots-chief-justice-john-roberts-commencement-warning/681626002/

Worstall, T., "The Robots Cannot Take All Our Jobs Because Ricardo And Comparative Advantage," Forbes, May 1, 2015 at URL https://www.forbes.com/sites/timworstall/2015/05/01/the-robots-cannot-take-all-our-jobs-because-ricardo-and-comparative-advantage/#3123bdd27371

****

Part of Norman’s introduction to the public was staged as an April Fool’s prank on the lab’s official site.(http://www.the13thfloor.tv/2018/04/09/is-norman-the-first-a-i-psychopath-his-creators-say-so/)

From http://norman-ai.mit.edu/

APRIL 1, 2018

AI-Powered Psychopath

We present you Norman, world's first psychopath AI. Norman is born from the fact that the data that is used to teach a machine learning algorithm can significantly influence its behavior. So when people talk about AI algorithms being biased and unfair, the culprit is often not the algorithm itself, but the biased data that was fed to it. The same method can see very different things in an image, even sick things, if trained on the wrong (or, the right!) data set. Norman suffered from extended exposure to the darkest corners of Reddit, and represents a case study on the dangers of Artificial Intelligence gone wrong when biased data is used in machine learning algorithms.

Norman is an AI that is trained to perform image captioning; a popular deep learning method of generating a textual description of an image. We trained Norman on image captions from an infamous subreddit (the name is redacted due to its graphic content) that is dedicated to document and observe the disturbing reality of death. Then, we compared Norman's responses with a standard image captioning neural network (trained on MSCOCO dataset) on Rorschach inkblots; a test that is used to detect underlying thought disorders.

: What policies should govern Artificial Intelligence (AI) applications?

Note: Due to the ethical concerns, we only introduced bias in terms of image captions from the subreddit which are later matched with randomly generated inkblots (therefore, no image of a real person dying was utilized in this experiment).

Research and Development AI
The research and development (R&D) of AI should observe the following principles:

· Do Good: AI should be designed and developed to promote the progress of society and human civilization, to promote the sustainable development of nature and society, to benefit all mankind and the environment, and to enhance the well-being of society and ecology.
· For Humanity: The R&D of AI should serve humanity and conform to human values as well as the overall interests of mankind. Human privacy, dignity, freedom, autonomy, and rights should be sufficiently respected. AI should not be used to against, utilize or harm human beings.

· Be Responsible: Researchers and developers of AI should have sufficient considerations for the potential ethical, legal, and social impacts and risks brought in by their products and take concrete actions to reduce and avoid them.

· Control Risks: Continuous efforts should be made to improve the maturity, robustness, reliability, and controllability of AI systems, so as to ensure the security for the data, the safety and security for the AI system itself, and the safety for the external environment where the AI system deploys.

· Be Ethical: AI R&D should take ethical design approaches to make the system trustworthy. This may include, but not limited to: making the system as fair as possible, reducing possible discrimination and biases, improving its transparency, explainability, and predictability, and making the system more traceable, auditable and accountable.

· Be Diverse and Inclusive: The development of AI should reflect diversity and inclusiveness, and be designed to benefit as many people as possible, especially those who would otherwise be easily neglected or underrepresented in AI applications.

· Open and Share: It is encouraged to establish AI open platforms to avoid data/platform monopolies, to share the benefits of AI development to the greatest extent, and to promote equal development opportunities for different regions and industries.

Use AI

: What policies should govern Artificial Intelligence (AI) applications?

The use of AI should observe the following principles:

· Use Wisely and Properly: Users of AI systems should have the necessary knowledge and ability to make the system operate according to its design, and have sufficient understanding of the potential impacts to avoid possible misuse and abuse, so as to maximize its benefits and minimize the risks.

· Informed-consent: Measures should be taken to ensure that stakeholders of AI systems are with sufficient informed-consent about the impact of the system on their rights and interests. When unexpected circumstances occur, reasonable data and service revocation mechanisms should be established to ensure that users' own rights and interests are not infringed.

· Education and Training: Stakeholders of AI systems should be able to receive education and training to help them adapt to the impact of AI development in psychological, emotional and technical aspects.

Governance AI

The governance of AI should observe the following principles:

· Optimizing Employment: An inclusive attitude should be taken towards the potential impact of AI on human employment. A cautious attitude should be taken towards the promotion of AI applications that may have huge impacts on human employment. Explorations on Human-AI coordination and new forms of work that would give full play to human advantages and characteristics should be encouraged.

· Harmony and Cooperation: Cooperation should be actively developed to establish an interdisciplinary, cross-domain, cross-sectoral, cross-organizational, cross-regional, global and comprehensive AI governance ecosystem, so as to avoid malicious AI race, to share AI governance experience, and to jointly cope with the impact of AI with the philosophy of "Optimizing Symbiosis".

· Adaptation and Moderation: Adaptive revisions of AI principles, policies, and regulations should be actively considered to adjust them to the development

of AI. Governance measures of AI should match its development status, not only to avoid hindering its proper utilization, but also to ensure that it is beneficial to society and nature.

· Subdivision and Implementation: Various fields and scenarios of AI applications should be actively considered for further formulating more specific and detailed guidelines. The implementation of such principles should also be actively promoted – through the whole life cycle of AI research, development, and application.

: What policies should govern Artificial Intelligence (AI) applications?

· Long-term Planning: Continuous research on the potential risks of Augmented Intelligence, Artificial General Intelligence (AGI) and Superintelligence should be encouraged. Strategic designs should be considered to ensure that AI will always be beneficial to society and nature in the future."

Author: Daniel Power
Last update: 2020-08-05 07:15