The *Communications* Web site, <http://cacm.acm.org>, features more than a dozen bloggers in the BLOG@CACM community. In each issue of *Communications*, we'll publish selected posts or excerpts.

Follow us on Twitter at <http://twitter.com/blogCACM>

DOI:10.1145/3317669

<http://cacm.acm.org/blogs/blog-cacm>

Implementing Guidelines for Governance, Oversight of AI, and Automation

Ryan Carrier
Governance and Oversight Coming to AI and Automation: Independent Audit of AI Systems

<http://bit.ly/2tPI1Sk>
February 12, 2019

Governance and independent oversight on the design and implementation of all forms of artificial intelligence (AI) and automation is a cresting wave about to break comprehensively on the field of information technology and computing.

If this is a surprise to you, then you may have missed the forest for the trees on a myriad of news stories over the past three to five years. Privacy failures, cybersecurity breaches, unethical choices in decision engines, and biased datasets have repeatedly sprung up as corporations around the world deploy increasing numbers of AIs throughout their organizations.

The world, broadly speaking, combined with legislative bodies, regulators, and a dedicated body of academics operating in the field of AI Safety, have

been pressing the issue. Now guidelines are taking hold in a practical format.

IEEE's Ethically Aligned Design (<https://ethicsinaction.ieee.org/>) is the Gold Standard for drawing together a global voice, using open source crowdsourcing techniques to assert some core ethical guidelines. Additionally, the standards body is deeply into the process of creating 13 different sets of standards covering areas from child and student data governance to algorithmic bias.

Others have joined the call. The EU recently created an ethical guidelines working group (<https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai>), and one of the original efforts includes: The Future of Life's AI principles (<https://futureoflife.org/ai-principles/>) created in conjunction with its Asilomar conference back in 2017. Even some specific companies, like Google, have gotten into the act, creating their own set of public ethical guidelines (<https://www.blog.google/technology/ai/ai-principles/>). This collection of work demonstrates the importance

and a considerable amount of effort to govern and oversee AI and automation from all corners of the globe.

Independent Audit of AI Systems is the next evolution of that governance: genuine accountability. It will build upon these global community guidelines to give audit and assurance companies the ability to assess compliance amongst companies employing AI and Automation. Let me explain how this all works.

ForHumanity, a non-profit organization, will sit at the center of key constituencies ranging from the world's leading audit/assurance firms to global academia and back to the companies themselves. ForHumanity's "client-base," however, is ONLY humanity (thus the name). Revenue to operate ForHumanity comes from donations, and companies who wish to license the audit process and the SAFEAI brand once compliance is achieved. Unlike the Credit Agency business model, where the rated entity "pays" for its rating, creating an inherent conflict of interest, ForHumanity does not exist to profit from audit compliance. This allows ForHumanity to act purely in the best interest of soci-

ety-at-large, seeking out “best practices” in the following areas (silos):

1. Ethics
2. Bias
3. Privacy
4. Trust
5. Cybersecurity

The ForHumanity team will operate global office hours and have dedicated staff for each of these audit silos seeking, sourcing, collating, moderating, and facilitating the search for “auditable best practices.” “Auditable” means binary, there is either compliance or non-compliance with the audit rule. Where we are unable to craft rules that are auditable, they will not become part of the audit. Gray areas are not the domain for compliance or non-compliance. Where gray areas are found (and there will be many), it will be the goal of the ForHumanity team, in conjunction with the global community, to distill these issues into binary parts and/or simply introduce transparency and disclosure (which is a compliance/non-compliance concept) into areas that in the past have been opaque, even if they remain gray. With transparency and disclosure, at least the public can choose which shade of gray they prefer.

Audit silo heads will hold two-hour office hours each day of the week. The times will be scheduled to accommodate the work day all around the world. Additionally, those who register may participate in an online, permanent chat designed to allow people to track the discussion over time at their convenience.

The creation and maintenance of the Independent Audit of AI Systems will be an ongoing and dynamic process. It will be fully transparent to all who would choose to participate, provided they join the discussion and participate with decorum. Each audit silo head will engage the community and seek points of consensus on auditable best practices. Once they believe they have found one, that audit rule will be proposed to the community at-large for consent or dissent. Dissent will also be tracked and shown to the Board of Directors for consideration. It is the role of the audit silo head to manage dissent and work toward reducing and eliminating dissent over time. If consensus is achieved, then that audit rule will be proposed to the ForHumanity Board of Directors. The Board will have

the final say, quarterly, on the current set of audit best practices rules. ForHumanity is dedicated to ensuring the Board of Directors is diversified across ethnic, gender and geography.

ForHumanity exists to achieve the best possible results for all. It does not get paid for the work it provides; instead, it is operated on a non-profit basis, licensing the SAFEAI logo and SAFEAI audit brand to those entities who submit to and pass the SAFEAI audit. In essence, we are asking those who benefit from the audit to pay it forward, so that we may continue and expand our work. Once an audit is passed, the company may *choose* to license the brand/logo in order to demonstrate to the world their compliance with the SAFEAI audit. The brand/logo may also be used by companies that wish to sell specific products that are SAFEAI-compliant as well. The brand/logo may be used on their packaging to enhance their ability to market and sell their product, versus competitors who may not have achieved SAFEAI audit compliance.

The rules are 100% transparent, so when an audit is conducted, compliance is expected. However, there may be areas of the audit that require remediation. Companies will be given a window of time in which to remedy their shortfall. Failure to comply will result in a public “failure” and transparency with regard to the noncompliance. This element is crucial to protect the SAFEAI brand, as well as to protect humanity from unsafe, dangerous, or irresponsible AIs. Over time, we expect the SAFEAI seal of approval to be an important part of consumers’ decision-making process for products and services. The theory is simple:

If we can make good, safe, and responsible AI profitable, whilst making dangerous and irresponsible AIs costly, then we achieve the best possible result for humanity.

In 1973, the major accounting firms came together and formed FASB (The Financial Accounting Standards Board), and the result of that work was GAAP (Generally Accepted Accounting Principles) which still govern financial accounting today. That work eventually became mandated by the SEC (and other jurisdictions around the world) for all publicly listed companies. The investing world was significantly improved through this clarity and uni-

formity. Third-party oversight gives great confidence to those who examine financial accounts to inform their decisions. It is a cornerstone of a robust market economy.

ForHumanity is working with major players to bring Independent Audit of AI Systems to fruition with the same robust and comprehensive oversight and accountability for artificial intelligence, algorithms, and automation. An effort like this will not eliminate fraud and irresponsible behavior. The world still suffered through the Enron and WorldCom financial accounting scandals, but by and large, accountability and universal rules will go a long way toward mitigating dangerous, irresponsible, and unfair behavior that has already challenged the world of technology. Microsoft and Google just recently informed their investors that ethics, bias, privacy, and other “risk factors” may occur, putting shareholders of those companies at risk (https://www.wired.com/story/google-microsoft-warn-ai-may-dumb-things/?mbid=social_twitter_onsiteshare).

Independent Audit is the best mechanism for companies to examine their compliance with best-practice rules and to make changes, mitigating downside risk. We look forward to working with them. We also ask each of you to participate as well. There are many ways to do so:

1. Track the progress of the SAFEAI audits and when the seal of approval begins to be used by compliant companies, buy those products.
2. Use services from companies that are SAFEAI-compliant.
3. Participating in the process for setting the auditing rules; it is open and all may join. You may not be a technical expert or have ideas to put forward, but your votes will count just as much as everyone else’s.
4. Donate to ForHumanity; we are a non-profit and you can find us at <http://forhumanity.center>
5. Tell others about the SAFEAI brand and help us spread the word.

Ryan Carrier is executive director of ForHumanity, a non-profit organization created to examine and mitigate the downside risks associated with Artificial Intelligence and Automation. Independent Audit of AI Systems is one such risk mitigation tool.

© 2019 ACM 0001-0782/19/5 \$15.00