

AI RFI Responses, October 26, 2018

Update to the 2016 National Artificial Intelligence Research and Development Strategic Plan RFI Responses

DISCLAIMER: The [RFI public responses](#) received and posted do not represent the views and/or opinions of the U.S. Government, National Science and Technology Council (NSTC) Select Committee on Artificial Intelligence (AI), NSTC Subcommittee on Machine Learning and AI, NSTC Subcommittee on Networking and Information Technology Research and Development (NITRD), NITRD National Coordination Office, and/or any other Federal agencies and/or government entities. We bear no responsibility for the accuracy, legality or content of all external links included in this document.



October 26, 2018

Re: Materials Germane to Revision of the 2016 National Artificial Intelligence R&D Strategic Plan

Dear Mr. D'Souza:

ACM, the Association for Computing Machinery, is the world's largest and longest-established association of computing professionals, representing approximately 50,000 individuals in the United States and 100,000 worldwide. Its US Technology Policy Committee is charged with providing policy and law makers throughout government with timely, substantive and apolitical input on computing technology and the legal and social issues to which it gives rise.

On behalf of the Committee, and in response to the National Science Foundation's Request for Information (RFI) of September 26, 2018 (Document 2018-20914), I am pleased to submit the attached *Statement on Algorithmic Transparency and Accountability* of May 2017 and its seven associated *Principles* for consideration in connection with revision of the nation's 2016 National Artificial Intelligence Research and Development Strategic Plan (see tinyurl.com/ydchmesh). The *Statement* is a joint product of ACM's Europe and US policy committees.

We respectfully and specifically commend the agency's attention to these materials in assessing strategies 3 and 4 as identified in the RFI: "understand and address the ethical, legal, and societal implications of AI" and "ensure the safety and security of AI systems," respectively. The Committee notes that building systems that achieve these aims is difficult. We believe, therefore, that emphasizing and enabling research to advance the field of accountable AI system design is especially important.

ACM U.S. Technology Policy Committee
1701 Pennsylvania Ave NW, Suite 200
Washington, DC 20006

www.acm.org/public-policy/ustpc

Statement on Algorithmic Transparency and Accountability

by ACM U.S. Public Policy Council, approved January 12, 2017

ACM Europe Policy Committee, approved May 25, 2017

Computer algorithms are widely employed throughout our economy and society to make decisions that have far-reaching impacts, including their applications for education, access to credit, healthcare, and employment. The ubiquity of algorithms in our everyday lives is an important reason to focus on addressing challenges associated with the design and technical aspects of algorithms and preventing bias from the onset.

An algorithm is a self-contained step-by-step set of operations that computers and other 'smart' devices carry out to perform calculation, data processing, and automated reasoning tasks. Increasingly, algorithms implement institutional decision-making based on analytics, which involves the discovery, interpretation, and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming, and operations research to quantify performance.

There is also growing evidence that some algorithms and analytics can be opaque, making it impossible to determine when their outputs may be biased or erroneous.

Computational models can be distorted as a result of biases contained in their input data and/or their algorithms. Decisions made by predictive algorithms can be opaque because of many factors, including technical (the algorithm may not lend itself to easy explanation), economic (the cost of providing transparency may be excessive, including the compromise of trade secrets), and social (revealing input may violate privacy expectations). Even well-engineered computer systems can result in unexplained outcomes or errors, either because they contain bugs or because the conditions of their use changes, invalidating assumptions on which the original analytics were based.

The use of algorithms for automated decision-making about individuals can result in harmful discrimination. Policymakers should hold institutions using analytics to the same standards as institutions where humans have traditionally made decisions and developers should plan and architect analytical systems to adhere to those standards when algorithms are used to make automated decisions or as input to decisions made by people.

This set of principles, consistent with the ACM Code of Ethics, is intended to support the benefits of algorithmic decision-making while addressing these concerns. These principles should be addressed during every phase of system development and deployment to the extent necessary to minimize potential harms while realizing the benefits of algorithmic decision-making.

Principles for Algorithmic Transparency and Accountability

- 1. Awareness:** Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.
- 2. Access and redress:** Regulators should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions.
- 3. Accountability:** Institutions should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results.
- 4. Explanation:** Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts.
- 5. Data Provenance:** A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals.
- 6. Auditability:** Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected.
- 7. Validation and Testing:** Institutions should use rigorous methods to validate their models and document those methods and results. In particular, they should routinely perform tests to assess and determine whether the model generates discriminatory harm. Institutions are encouraged to make the results of such tests public.