SAND2021-10405C

# More is not better when Designing Ethical ML/AI

Finding Truth in a vast sea of unvetted data may be one of the greatest challenges of the next decade. Current approaches such as unsupervised clustering or association rule learning generate “insights” from these data pools with relatively little human involvement. Such techniques have real dangers. When designing or using AI, considering what *should* be done is at least as essential as considering what *can* be done. When decision makers act on ML/AI derived insights from unvetted data collections, they may act on the information received without a clear sense of the limits of the recommendations. In fact, determining these limits can be difficult when the characteristics of the collections are incompletely known. Adding to the potential problems, large data sets often embody poor decisions of the past. Basing future decisions on past data can amplify historical biases, inefficiencies, and injustice. Unvetted collections or, worse, incorrectly vetted collections lead to costly inaccuracies, including those with legal and/or ethical implications. All scientific progress carries risk. Placing work on AI within a larger ethical frame provides a means to anticipate adverse consequences and build remedies before irrevocable harm occurs.

In this talk, Dr. Ruby Booth will provide examples of seemingly innocuous data elements can lead to biased, even discriminatory algorithms. She’ll discuss strategies for moving test cases earlier in the design process to mitigate this risk. Finally, she will review ethical design practices, show unintended (but likely) consequences of AI decision support, and discuss techniques for to help reframe the way participants approach AI/ML driven decision support. Because, it doesn’t matter how “true” your results are if they aren’t trusted.

*Any subjective views or opinions expressed in this presentation do not necessarily represent the views of the U.S. Department of Energy or the United States Government. SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525*

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.