240719

Dara Norman's thoughts on the grand challenges of the CCC Computational and Citizen Science Research workshop

The most human thing about Artificial Intelligence systems are that they make mistakes, they are biased, they mis-interpret and sometimes they just plain get things wrong. The problem is that unlike with people who exhibit these traits, we tend to not expect this behavior from the AI systems… that we humans have created! How so we get our culture to a place where we view and recognize that AI is a simply a tool and since humans programed that tool, we should have a healthy skepticism of its results.

As astronomers, we are accustomed to dealing with data sets and data inputs that are not perfect. This is, in part, because unlike with other experimental sciences, we must design our research around the light or gravitational waves that we can capture; we cannot design experiments to control the conditions of how those waves behave. Thus, our science is tied to making physical inferences about the universe and its content from complex data. We know that our data are likely to include issues like selection effects, sampling bias, obfuscating outliers, statistical and intrinsic errors and we are careful to attempt to identify and eliminate these concerns throughout the process of our scientific work from start to finish. We design our initial experiments with these issues in mind, we develop analysis methods to take these into account and we engage in peer review so that our colleagues can verify that we have thought of everything. The application of artificial intelligence (AI) and machine learning (ML) techniques to our data bring the opportunity to sort through datasets more quickly and with an eye to correlations and connections that we might not have otherwise recognized. However, like with science done by humans, we must also develop the similar verifications to be sure that conclusions reached through the use of these AI techniques are accurate, trustworthy, explainable and traceable.

Leveraging large groups of humans through citizen science programs could be a key way to develop these key verification strategies. Identifying the properties that make one person see a random cloud as looking like a dog, while someone else sees a horse, could give us insight into where AI codes might go awry, leading to biases, misinterpretations and errors. 'Real world' data is messy and AI algorithms designed to sort through this data in unsupervised (or under-supervised) ways have demonstrated a number of problems (Joyce, 2021 and references within). There are many examples of how training sets for AI fail to adequately capture the diversity of a sample set because of sampling bias (e.g., Buolamwini and Gebru 2018); how AI bots trained on public data sets, like the internet or social media, rapidly begin to reflect some of the worst traits of human behavior, e.g., sexism, racism, etc., through selection effects (Rodriguez, 2016); AI algorithms used in facial recognition, criminal sentencing or medical diagnosis disproportionately produce poorer outcomes for Black and Brown people (Angwin, et al., 2016). Recently, another big concern has been the tendency of AI to 'hallucinate' facts, i.e. make things up. (Vincent, 2023). Identifying how to recognize that these errors have occurred can be difficult if direct human expertise is not an integral part of the verification process, but this often does not happen in the 'wilds' of the 'real world' use of AI. Building algorithms that are more accurate and trustworthy without constant human 'hand holding' are needed to fully realize any of the gains we hope to achieve from AI's use. Therefore, the need for accurate, trustworthy, explainable and traceable AI in the 'real world' faces the same challenges that we must solve for astrophysical research. There is an opportunity, with this work, to bridge the fields of astronomy, computer science and sociology for the betterment of both science and society, at the same time. Identifying how the interaction of citizen science experiments to test AI could be an important part of this workshop.