

Initial thoughts: Artificial Intelligence and Inclusion Symposium.
November 8-10, 2017, Rio de Janeiro, Brazil.

How privacy law and data on training sets might collide

Regarding privacy, data protection regulations enable individuals to request the withdrawal of data about them. Those strategies might have functioned well in the past, but would probably prove to be shortsighted for an automated future, where AI is widely deployed and smart devices proliferate. We are concerned about having accountable, unbiased algorithms that sort efficiently the data to perform a decision. To do so, they need robust training sets with appropriate representation. If we remove data due to privacy concerns, that would certainly have an impact on the final results. Some minorities would be underrepresented, and their particularities might be wiped out on the iterations to perform a classification. If the regulatory solution to avoid potential adverse automatized decisions is to remove the data, the remedy may be far worse than the disease, as decisions would be made without the proper data needed. Moreover, the so called "right of explanation" that data protection regulations are proposing (to inform the individual on the logic employed to achieve an automated decision), do not consider how AI models work. Machine learning uses several techniques that can neither be easily explained nor understood without proper technical background. Neural networks are complex and unpredictable schemes, where data is disambled and back propagated through several layers, so there is not a simple or linear correlation between a particular datapoint and the final result. Maybe the solution might not be to remove data, rather to policy how AI is deployed, and regulate those areas where it might not be advisable to allow AI implementations. On policy grounds, incentives should be created to detract for predictive business models that are secret or unaccountable.

AI awareness and literacy for the "smart" future

People are vaguely aware of what happens with their data behind the "smart" closed doors. Mostly they are focused and amused on what smart devices could do to improve their life management. Ridiculous devices collect all kinds of random data, ranging from smart bottles, egg holders and novelty appliances, to biometric monitoring devices, powered by strong marketing strategies of well being, coolness, and the promise of futuristic lifestyles. The data collection trend is going to intensify as smart devices are deployed all around us, on smart cities, public transportation and social interactions. AI awareness would be a must goal to policy and advocacy, to balance the "quantified self" rhetoric. Biased results created by inappropriate data or incorrect assumptions during processing, inadvertent bias that might slip during the data cleansing and model construction. Algorithmic models have steps where assumptions are made in order to prepare the data for usage. Dimensionality reduction implies to establish correlations between certain data points, to reduce the number of random variables under consideration. Some models deploy iterative strategies to improve their performance. In each loop, corrections are made that might erase underrepresented or isolated data points, leading to underfitting. This might have a potential impact on minorities, as their particularities could be wrongly clustered or erased. On the other hand, models are prone to overfit, trying to accommodate isolated data, and this affects their predictive power and the accuracy of the results, leading to wrong assumptions and correlations. Both phenomena should be carefully monitored. Biases are hidden in data sets, and might translate into unfair and inaccurate inferences. To tackle the issue we need statistics and AI literacy, to understand how black boxes work, the steps that are used to perform a decision, and the opportunities along this process where human assumptions are made, to detect where discrimination can nest (even if inadvertently or unintentionally), hoping to correct bias from inception.