Title:
From methods to increase internal transparency to algorithm package leaflets for improving accountability towards (re-)users

Authors: Rob Heyman, Cora van Leeuwen, Myriam Smitt

Abstract:
The complexity of algorithms has risen to a level that is beyond human understanding. This is one of the reasons why there are limited means and opportunities for debate about algorithms, and tackling algorithmic biases. Algorithm as a concept remains black boxed in current research and debate about algorithms, machine learning and algorithmic bias. As a result, what we know about algorithms is general and anecdotal, the so-called tip of the iceberg. What is not vague, is the challenge algorithmic bias poses for society; unfairness, unequal treatment and consumer or citizen manipulation. A myriad of initiatives have expressed their concerns and attempted to raise awareness on the biases that algorithms could have and how they will influence our lives. Yet they often fail in their ability to valorize the red flags in applicable, practical or functional tools. This challenge calls for increased scrutiny and thus the need to define methods and tools to analyse algorithms, their design process, their potential biases, and their consequences for society.

At imec-SMIT we have researched improving transparency during algorithmic development (of health algorithms), which we call *internal transparency*: What are the moments in an algorithm's lifecycle that require (more) transparency and are prone to biases? How do we capture this human-decision making information from algorithm developers?
We also consider what should be disclosed externally: How and to whom should we be transparent? And how to share knowledge about biases without losing competitive advantages algorithm developers wish to safeguard? Are there biases we can prevent? Can we gauge and share information about the preconceived notions on health of the developers themselves, to the extent that they are relevant to the algorithm's outcomes?

We wish to elaborate on our methods to increase internal transparency in an ongoing imec project related to predictive health monitoring. We discuss mapping exercises, bias identification workshops, and the types of biases that are common in algorithm development. Secondly, we would like to receive input on our ongoing work to create an algorithm package leaflet. This is a transparency document similar to a medical package leaflet. It informs those who might be affected by the algorithm's biases and shortcomings; what should we include for whom in order to increase *external transparency*? What information would be considered relevant for the different end-users (there is the owner of an algorithmic solution and the one who is subject to the solution) and allow for informed decision making. We aim to extend our approach and tools towards other domains in society, like smart cities, (predictive) policing and migration.

What we have learned so far is that bias detection is not a structural part of algorithm development yet. This does not mean that bias is not addressed. Indeed, certain statistical biases are approached quite conscientiously. However, there are quick wins possible to address biases by adding transparency templates to the design process.