Algorithms and the Black Box Problem

By: Edmund Ang

“It’s all in the algorithm” - a common and popular phrase used when discussing technologies making key decisions. Many know an algorithm is deployed in software services such as search engines, intelligent assistants, and others. What is not obvious, is that increasingly an algorithm is deployed in safety critical equipment, e.g. an intruder detection system, multi-criteria smoke detectors and video fire detection system.

These smart technologies rely on algorithms to reduce false positive, e.g. spurious fire alarm due to dust, and false negative, i.e. not identifying a real incident. For example, most modern aspirating smoke detection system relies on an algorithm to process the signals from the LED and infrared sensors to determine a real or spurious alarm.

Whilst I am excited by the possibilities pairing algorithms with advanced hardware technologies; I am concerned these are becoming a black box to end-users, and their creators. My first concern is the increasing complexities. Using the aspirating detection system as an example, although the decision-making process to differentiate a true or false alarm appears straightforward, the reality is anything but. The system needs to be sufficiently robust to differentiate this in many environment under different scenarios. If machine learning based algorithm is used in the future, this will exponentially increase the complexities, and no one will fully understand the cause and effect of a scenario because of the evolving nature of machine learning. Case in point in another industry is the financial market’s flash crashes over the last few years attributed to the deployment of trading algorithms in high speed trading.

Secondly, there is no standardized testing method. Currently, each manufacturer maintains their own proprietary design, and there is no one standardized testing method for the robustness of the algorithm in dealing with various scenarios and edge cases. Therefore, there is no consistency in the expectation on the performance of such systems. Thirdly, the algorithms are mainly closed sourced. Acknowledging the need to maintain a company’s intellectual properties, most of the algorithms paired with the hardware technologies are proprietary or hidden. This means only the owners with access to the source code have a reasonable chance in understanding and stress testing the algorithm.

If the current situation is left unchecked, soon we will end up with the black box problem where no one fully understands the technology at hand; and we can only trust, albeit blindly with no verifiable assurance the system is sufficiently robust where a catastrophic failure like those seen in other industries will not occur.
That said, at this juncture, compared to other industries, the use of algorithm in fire safety critical technologies is still primitive. This is precisely the opportunity for the fire industry to set a strong foundation to ensure we have a full understanding and control over the algorithm we deploy today, and in the future.

I have two humble suggestions. Firstly, we need to adopt an open source mindset. Acknowledging the necessity to protect intellectual properties, this mindset can still be instilled when products are developed. At this point of development, it is still possible to open the algorithm deployed in safety critical technologies in sufficient details to ensure the wider engineering and research community can help examine and stress test these algorithms. Whilst a manufacturer can test the general cases, it is impossible to identify all the edge cases for these algorithms. Secondly, we need to develop an industry agreed testing and algorithm disclosure method to provide a level playing field to the manufacturers. The fire industry, in collaboration with the testing laboratories and standards setting body need to develop a testing method for safety critical technologies where an algorithm is a crucial part in ensuring the functionality of the system. This is to ensure there is consistency in the expectation on the performance of such systems. Further, the industry must collectively agree on a standardized algorithm disclosure method to ensure professionals utilising such systems can understand the decision-making process of the algorithms.

I appreciate the effort required to implement the suggestions above are enormous. Nonetheless, I fully believe if we all come together, the collective intelligence of the fire industry will prevail.

Edmund Ang
Fire & Risk Engineer, and PhD Researcher at Hazelab Imperial College London