NAME : Hiroyuki Hattori
COUNTRY : Japan
REGISTRATION NUMBER : 7281
GROUP REF. : SC D2
PREF. SUBJECT : PS1
QUESTION N° : Q1-02

In Japan, there are no laws or regulations that restrict introductions of "black box" machine learning to critical infrastructure. Self-driving cars, however, are the only case of restriction to the Japanese industry by established safety standards for equipment.
The Japanese power industry is based on voluntary operational safety. It is a system in which each company establishes its own safety programs. The government authorizes them. Therefore, companies can select what algorithms are used for their own operational safety. The Japanese major companies share technical standards and install software and equipment based on these standards. In some areas represented by control system, companies can only use existing analytic algorithms in accordance with these standards. Currently, these standards do not take into account the existence of "black box" machine learning, which is practically unavailable.
In Japan, "black box" machine learning is restricted to existing technical standards in several domains. We suggest we should discuss on how to introduce machine learning in these standards.

While the overall the Japanese regulations are as such, we believe that we should not machine learning are used for critical infrastructures without some fail-safe system.
Whether it is a “black box” or not, machine learning based software is generally affected by input data. We call this uncertainty “black”. Most machine learning is good at finding optimal conditions from a statistical perspective. “Blackness” of machine learning varies greatly from one model to another. “Blackness” of each models depends on the training process of the machine learning, including the reliability of data labeling and the algorithm used, deciding process of parameters and so on.
On the other hand, humans make logical decisions based on calculations and empirical knowledge of physical characteristics, such as mechanical and electrical ones. There can be a system that replaces such human judgment, where the logic structure is obvious, and errors can be easily identified. There is an inherent risk in trusting the decisions that machine learning has made autonomously and to execute those decisions. When applying such machine learning based software for critical infrastructures, a checking mechanism is necessary.[1] The checking mechanism involves a human or is a system having the same functionality as that a human already performs. When we establish a checking mechanism, it is important to determine the robustness and accuracy of the whole system. The question is to what extent the decision mechanism as a whole will tolerate the degree of “blackness”.

If there is an existing decision-taking system, it may be possible to replace it by adding a human checking mechanism or using the existing decision-taking system as a checking mechanism. However, it is necessary to compare the accuracy of inputs such as past results, and to determine that the “blackness” is acceptable.
It is the same for decision-support systems to conduct accuracy verification. However, it is easy to introduce even “black box” machine learning by importing it into the existing decision-support systems. As these existing decision-support systems require a human as the final decision maker, machine learning can present the judgment to the human as a factor, regardless of the degree of “blackness”. In other words, it can coexist with the decision-making software of existing support systems, offering multiple possibilities. Alternatively, an existing support system can be used as a checking mechanism for the machine learning model. In the decision-taking process, we suggest that both statistical and physical computation-based decisions would be beneficial because they complement each other and may enable decision-making that is closer to the overall optimum. Therefore, in this case, machine learning is not limited by existing decision-support systems.

[1] Rudin, C. *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead.* Nat Mach Intell 1, 206–215 (2019).