TPM-like authenticity function is needed for securing AI products

Yoshiyasu Takefuji

Mariarosaria Taddeo et al. wrote an article entitled "How AI can be a force for good" (1). The Explainable Artificial Intelligence program of DARPA (Defense Advanced Research Project Agency) is introduced (1). However, many lay public users of the AI technology for autonomous systems do not need to understand the details. They simply want to know the result. The explanations do not help the lay public at all. Translational Ethics with delegation and responsibility approach are also introduced to force AI for good (1). Translational Ethics with delegation and responsibility approach are weak and naive against adversarial and malicious AI developers. Instead of relying on the translational ethics, TPM-like (trusted platform module) robust authenticity functions are needed for securing AI products. According to Wikipedia, TPM is an international standard for a secure cryptoprocessor, a dedicated microcontroller designed to secure hardware through integrated cryptographic keys (2). TPM has been used in PCs and smart phones. However, currently, we do not have any tools for investigating authenticity whether adversarial/malicious AI modules are integrated/included in the system or not. In order to remove adversarial/malicious AI products from our market, robust authenticity regulations are required for verifying and justifying trusted AI systems.

References:

1. Mariarosaria Taddeo et al., How AI can be a force for good, Science 24 Aug 2018: Vol. 361, Issue 6404, pp. 751-752
2. https://en.wikipedia.org/wiki/Trusted_Platform_Module