Annex

Good Practices for Countering Deepfake Attacks

To further enhance AIs' defence against common deepfake attacks that target their identity verification solutions, AIs should review the relevance of the following good practices and take appropriate action(s) to enhance deepfake detection capabilities:

1. **Strengthen device security controls –** implement robust controls to ensure that customers' devices are secure and that mobile banking applications can directly access device cameras for secure image capture without manipulation (e.g. video injection[1]).

2. **Ensure the robustness of identity verification solutions –** incorporate a diverse and randomised set of liveness action tests during identity verification, and take appropriate follow-up actions in response to abnormal verification attempts.

3. **Equip systems with deepfake detection capabilities –** as part of these deepfake detection capabilities, AIs are strongly encouraged to implement image analytics solutions, such as face similarity detection and background analysis, to further enhance their ability to detect fraudulent verification attempts.

4. **Enhance fraud monitoring capabilities –** collect and analyse abnormal digital footprints and transaction patterns that may signal potential deepfake attacks.

5. **Implement end-to-end controls to support deepfake detection –** establish comprehensive processes that integrate deepfake detection solutions with other security controls, such as device security checks, transaction monitoring, manual reviews, and quality assurance checks, to provide a multi-layered defence system.

6. **Conduct regular reviews –** conduct regular reviews to assess the effectiveness of facial recognition solutions, including deepfake detection capabilities.

7. **Equip and train staff –** provide comprehensive training to staff responsible for manual checks, equipping them with necessary skills to effectively identify and respond to deepfake attacks.

---

[1] This involves compromising device security to feed images and videos directly to a device, thereby substituting the device's genuine camera source.