Integrity: Generalized Artificial Image Classification With Noise Domain Localization

Westemeier, Huxley (School: Saint Paul Academy and Summit School)

AI image generation models can now create and augment photographs at scale. Recent reports predict that AI-assisted misinformation will rank above warfare and natural disasters as the greatest threat to global economic security by 2027. The AI-generated image detector space has struggled to keep pace with improved generated image quality, making it challenging for an untrained eye to detect AI content. Existing detectors fail when tested on unknown generators, are resource-intensive, rely heavily on machine learning, and are vulnerable to attacks. I created Integrity, a software tool without machine learning that detects images generated by any model by analyzing statistical deviations in the noise pattern, looking for indicators of authenticity rather than signs of artificialness. An original dataset of high-resolution authentic images was paired with artificial images from multiple models and run through Integrity's algorithm. Authentic image scores were empirically determined to fit within a threshold inconsistent with AI content. A working algorithm was achieved with an average classification accuracy of 98.6% on the custom dataset, reducing the computational cost and outperforming other detectors by more than 23%. Integrity also identifies small regions of manipulated authentic images, includes a built-in protection mechanism for detecting potential attacks, and is a comprehensive tool using a novel statistical approach for detecting AI-generated image content on a localized level. Integrity improves upon the accuracy, speed, security, and efficiency of ML-based detectors and has the promise of global reach, creating an opportunity for more people to have access to image authentication than ever before.

Awards Won: