Tech Brief: Label your AI generated content privately using AON-PRISMA and the blockchain

AON-PRISMA, July 2023

AI generated content can do harm, but supporting fake news with generated images, text, sound and other artifacts that are not real. In response to this threat, the leading Tech companies made a pledge to label their AI generated content to the White House.

AON-PRISMA can help doing so privately, accurately and efficiently. Imagine, putting a copy of each of your AI generated content on a public blockchain. It can no longer be removed and everyone can see that it is AI generated. You have a reliable label. However, doing so is very privacy invasive since all content is now public.

In order to achieve privacy, one should only put a digest of the AI generated content on the blockchain. There are cryptographic hashes that do this. However, if someone changes even a single bit in the content, the cryptographic hash completely changes. Hence, it is easy to evade detection by comparing those cryptographic hashes. There are also perceptual hashes that are robust under small changes, i.e., small changes in the input result in small changes in the hash. However, those are invertible and publishing them is not much better than publishing the generated content itself.

AON-PRISMA allows publishing a perceptual hash in encrypted form. It remains comparable for small differences. A comparison will only reveal whether it is the (almost) same hash or not. However, inverting the perceptual hash becomes impossible, since it is encrypted. An adversary has to guess almost the hash before he can successfully compare it. AON-PRISMA is also efficient. Whereas finding the perceptual hash is exponentially difficult in the length of the hash, comparing AON-PRISMA protected hashes is easy (polynomial time). AON-PRISMA also allows you to reduce the expected search complexity to logarithmic in the database size while preserving (differential) privacy.

AON-PRISMA is a technology developed by researchers at the University of Waterloo and the National Research Council Canada. To find out more about AON-PRISMA, please visit https://aon-prisma.dev/ or email info@aon-prisma.dev