Engineering Village™

1. **CoProtector: Protect Open-Source Code against Unauthorized Training Usage with Data Poisoning**
Sun, Z.; Du, X.; Song, F.; Ni, M.; Li, L. **Source:** *WWW '22: Proceedings of the ACM Web Conference 2022*, 652-60, 2022; **ISBN-13:** 978-1-4503-9096-5; **DOI:** 10.1145/3485447.3512225; **Conference:** WWW '22: ACM Web Conference 2022, 25-29 April 2022, Virtual Event, France; **Sponsor:** SIGWEB; **Publisher:** ACM, New York, NY, USA

Author affiliation:
Monash University, China
ShanghaiTech University, China
University of Technology Sydney, Australia

Abstract: Github Copilot, trained on billions of lines of public code, has recently become the buzzword in the computer science research and practice community. Although it is designed to help developers implement safe and effective code with powerful intelligence, practitioners and researchers raise concerns about its ethical and security problems, e.g., should the copyleft licensed code be freely leveraged or insecure code be considered for training in the first place? These problems pose a significant impact on Copilot and other similar products that aim to learn knowledge from large-scale open-source code through deep learning models, which are inevitably on the rise with the fast development of artificial intelligence. To mitigate such impacts, we argue that there is a need to invent effective mechanisms for protecting open-source code from being exploited by deep learning models. Here, we design and implement a prototype, CoProtector, which utilizes data poisoning techniques to arm source code repositories for defending against such exploits. Our large-scale experiments empirically show that CoProtector is effective in achieving its purpose, significantly reducing the performance of Copilot-like deep learning models while being able to stably reveal the secretly embedded watermark backdoors. (0 refs.) **Inspec controlled terms:** copyright - deep learning (artificial intelligence) - public domain software - security of data

Uncontrolled terms: CoProtector - open-source code - unauthorized training usage - Github Copilot - computer science research - deep learning models - source code repositories - ethical security problems - artificial intelligence - data poisoning techniques - embedded watermark backdoors

Classification Code: C6130S Data security - C0230B Legal aspects of computing - C6264

IPC Code: G06F21/00 - G06N3/02 - G06N20/00

Treatment: Practical (PRA)

Database: Inspec