미국

마이크로소프트社의 GitHub Copilot, 오픈소스 관련 첫 번째 저작권 분쟁 시작

이 나 라
University of Colorado 연구원

컴퓨터 소프트웨어 작동은 소스 코드를 읽어 내려가며 실행되는 방식임. 소스 코드 개발자의 의지에 따라 코드를 대중에게 공개하기도 하는데, 이런 소스 코드를 오픈소스라고 부름. GitHub는 세계적으로 유명한 오픈소스 공유 웹사이트를 일컬음.

1. 주요 용어

오픈소스 (혹은 오픈소스 소프트웨어, OSS)

- 흔히 컴퓨터 프로그램이라고도 불리는 '소프트웨어'는, 프로그램에 내장된 '소스 코드(source codes)'를 읽어 내려가면서 각 코드에 담긴 기능을 차례대로 구현하는 원리로 작동됨. 소스 코드란 소프트웨어를 실행시키는 치밀한 설계도와 같음.
- 대중에게 공개된 소스 코드를 오픈소스라고 부르며, '오픈소스 라이선스'의 허용 범위에 따라 변경 및 재배포도 가능함.

GitHub (깃허브)

- GitHub(깃허브)에는 git(깃)과 hub(허브)의 합성어임. git이란 컴퓨터 파일이 변경될 때마다 그 '버전'을 관리하는 명령어이자 프로그램을 칭함. 특히, 개발자들이 새로운 버전을 만들어 테스트하는 데 있어 제대로 작동되지 않으면, git 명령어를 이용해 옛 버전으로 되돌아갈 수 있음. '중심지'라는 뜻의 hub는 '중앙 집중'이라는 의미가 내포돼 있음.
- GitHub는 git 데이터를 온라인상에 저장해 주는 하나의 사이트(https://github.com)를 말함. 여타 다른 클라우드처럼 데이터 백업을 기본 기능으로 하지만, 여기서 더 나아가 개발자들이 서로의 소스 코드를 공유하는 장소이자, 공유된 코드를 수정하거나 자신의 코드를 첨가하면서 오픈소스 소프트웨어를 함께 제작하는 협업의 장소이기도 함. 그렇게 2008년에 설립된 GitHub은 세계 최대의 오픈소스 코드 공유의 플랫폼으로 성장함.

2. GitHub를 인수한 마이크로소프트, AI 코딩 서비스 Copilot 공개

지난 2018년 약 75억 달러에 GitHub를 인수한 MS(마이크로소프트)는 2021년에 OpenAI와 공동 개발한 GitHub의 Copilot(코파일럿)을 공개함.

- MS는 GitHub를 인수한 후에도 기본의 오픈소스 공유 플랫폼을 계속 유지하면서 많은 사용자를 보유하고 있음.
- GitHub Copilot은 GitHub에 올라온 수많은 오픈소스 코드로 학습된 AI 코딩 서비스임. 소프트웨어 개발자가 작성해야 할 코드에 대한 간단한 설명문이나 함수를 제시하면 Copilot이 해당 코드를 자동 완성하여 출력함.
- 보통 코드를 작성할 때 누군가가 만든 코드를 참고(reference 혹은 template)하여 써 내려가는 경우가 많음. 이러한 코드 초안으로서 Copilot의 코드가 활용되면서 사용자의 코딩 작성 시간 또한 획기적으로 단축시킴. 기존의 무료 서비스가 2022년 6월부터 유료로 전환됐고, 사용료는 한 달에 10달러나 일 년에 100달러로 책정됨.

GitHub 오픈소스 코드를 사용할 때는 오픈소스 라이선스 허용 범위에 맞게 활용해야 함.

- GitHub에 공개된 모든 오픈소스 코드에는 원작자가 활용 범위를 지정할 수 있음. 이에

COPYRIGHT
TREND

지난 2018년 GitHub를 인수한 MS는 2021년에 AI 코딩 서비스인 Copilot을 공개했음. Copilot은 GitHub에 올라온 수많은 오픈소스 코드로 학습된 AI임. 처음엔 무료였던 Copilot이 올해 6월부터는 유료로 전환됨에 따라, 오픈소스 라이선스 침해 문제가 발견됨.

매튜 버터릭 변호사와 조셉 사버리 로펌은 Copilot에서 오픈소스 라이선스 위반 정황을 포착하고, AI 코딩 서비스에 대한 오픈소스 저작권 침해 소송을 대대적으로 준비하고 있음.

MS의 GitHub에서 만든 인공지능 Copilot이 작성한 출력문에서 오픈소스 라이선스 문제가 수면 위로 떠오르자, 다시금 AI 기반 서비스의 잠재적인 저작권 침해가 주목됨. 이번 사건으로 인해 AI가 작성한 오픈소스나 그 소스 코드로 개발된 소프트웨어 등에서의 저작권이 누구에게 있는지를 따지는 대규모 소송이 될 가능성이 큼.

따라 무료로 자유롭게 사용할 수 있는 오픈소스 코드가 있는가 하면, 반대로 영리 목적으로는 사용할 수 없는 코드도 있음. 잠재적인 저작권 문제를 피하기 위해서는 반드시 오픈소스 라이선스가 허용하는 범위에 맞게 오픈소스 코드를 활용해야 함. 사실 이는 GitHub의 사용자라면 모두 알고 있는 규칙임.

3. 사건 개요

지난 10월 17일(현지 시각), 작가이자 디자이너, 개발자 및 변호사인 매튜 버터릭(Matthew Butterick)과 조셉 사버리(Joseph Saveri) 로펌은 Copilot의 완성 코드에서 저작권이 있는 GitHub 오픈소스 코드와 유사성을 포착함. 이에, MS를 상대로 오픈소스 라이선스 위반에 대한 소송을 준비한다고 발표함.

- 버터릭 변호사 측의 증거 코드를 보면, 돈을 내고 산 Copilot 출력 코드에서 '비영리 목적'의 오픈소스 코드와 유사성이 발견됨. 버터릭 변호사는 이것이 코드 원작자의 오픈소스 라이선스를 위반하는 행위라고 봄.
- 또한, 해당 사건으로 인해, 라이선스 허용 범위에 상관없이 GitHub의 오픈소스가 Copilot의 학습용 메타 데이터로 사용되었음이 간접적으로 확인됨.
- 버터릭 변호사는 MS가 AI 학습용 메타 데이터의 저작권 문제를 해결하지 않는다면, 향후 Copilot 코드로 만든 영리 목적의 소프트웨어가 개발될 때 더 큰 저작권 분쟁이 일어날 것으로 봄.

4. 시사점

최근 AI 창작물과 관련된 저작권 분쟁이 AI 코딩 서비스인 MS GitHub Copilot에까지 번진 양상.

- 이번 건은 오픈소스와 관련하여, ▲AI가 작성한 소스 코드의 저작권, ▲해당 소스 코드로 만들어진 소프트웨어의 저작권, 나아가 ▲AI 학습용 메타 데이터의 저작권 침해에 대해 최초로 법적 책임을 제기했다는 점에서 의미가 있음.
- 이번 소송이 실제 법정 소송으로 이어질 경우, 그 결과가 오픈소스 커뮤니티와 AI 모델링 및 학습 방법에도 큰 영향을 끼칠 것으로 보임.

| 참고자료 |

- https://githubcopilotinvestigation.com/
- https://www.prnewswire.com/news-releases/joseph-saveri-law-firm-and-matthew-butterick-file-class-action-lawsuit-against-github-microsoft-and-openai-over-violations-of-open-source-licenses-arising-from-github-copilot-an-ai-based-product-301668255.html
- https://www.theverge.com/2022/11/8/23446821/microsoft-openai-github-copilot-class-action-lawsuit-ai-copyright-violation-training-data