

GitHub Copilot이 집단소송을 당했다고?

LINE+ 김동혁

Agenda

01 / MS의 전략?

- Github 인수
- 시장 점유율 높이기

02 / Copilot Issue

- Copilot?
- Copilot 집단 소송

Microsoft의 전략?

큰그림을 그린 Github 인수

2018년 10월 GitHub 인수
Microsoft가 약 75억 달러에 인수

개발자들에게 편리한
작업 환경을 만들어주는
GitHub 업데이트

2019년 7월 OpenAI 투자
2020년에는 OpenAI의 GPT-3
언어 모델의 독점 사용권자

2021년 6월 Copilot 출시

Microsoft의 전략?

MS의 행보는 시장 점유율을 높이는 효과가 있다

2018년 10월 GitHub 인수

Microsoft가 약 75억 달러에 인수

**개발자들에게 편리한
작업 환경을 만들어주는
GitHub 업데이트**

2019년 7월 OpenAI 투자

2020년에는 OpenAI의 GPT-3
언어 모델의 독점 사용권자

2021년 6월 Copilot 출시

수천만명의 개발자를 확보
수천만명의 개발자가 GitHub에 소
스코드를 올리고 있음

ML을 위한 데이터 수집
Copilot의 초석

개발자들이 GitHub에 의존하게 함
Git action, Git Package,
Git Sponsor, Git Codespace
- Azure로 Deploy 하는게 쉽고 빠름

Copilot을 위한 투자

**더 고도화 된다면 기존 개발자들뿐만
아니라 신입 개발자들까지 끌어당길수 있음**

Copilot?

Microsoft에서 출시한 Copilot은 OpenAI의 Codex를 기반으로 구동되는 AI 자동코딩 프로그램.



GitHub repository
코드들을 통해 ML

함수명 or 주석문으로
AI 코드 자동생성

Copilot 집단소송

GitHub 저작권 소유자 집단 vs Microsoft, GitHub, OpenAI 등

11월 3일 GitHub Copilot 관련 회사들에 집단 소송

원고 : Joseph Saveri Law Firm + Matthew Butterick

피고 : GitHub, Microsoft (GitHub 소유사), OpenAI 등

소송 사유

- LLC1 for violation of the Digital Millennium Copyright Act, 17 U.S.C. §§ 1201–1205 (the “DMCA”)
- violation of the Lanham Act, 15 U.S.C. § 1125
- violation of Unfair Competition law, Cal. Bus. & Prof. Code §§ 17200, et seq.
- violation of the California Consumer Privacy Act, Cal. Civ. Code § 1798.150 (the “CCPA”)
- Breach of Contract regarding the Suggested Licenses, GitHub’s Privacy Statement, and GitHub’s Terms of Service
- Cal. Bus. & Prof. Code §§ 22575–22579, Cal. Civ. Code § 1798.150. Plaintiffs and the Class also bring this Complaint against Defendants for their Tortious Interference in Plaintiffs’ Contractual Relationships; Fraud, and Negligence regarding handling of sensitive data

Copilot 집단소송

GitHub 저작권 소유자 집단 vs Microsoft, GitHub, OpenAI 등

LLC1 for violation of the Digital Millennium Copyright Act, 17 U.S.C. §§ 1201–1205 (the “DMCA”) - 모든 피고 대상

17 U.S. Code § 1201 - Circumvention of copyright protection systems

- (a) Violations Regarding Circumvention of Technological Measures. (A) **보호되는 저작물에 대한 접근을 효과적으로 제어하는 *기술적 조치를 우회해서는 안됨**
- (b) Additional Violations (1) 어느 누구도 다음과 같은 기술, 제품, 서비스, 장치, 구성 요소 또는 그 일부를 제조, 수입, 대중에게 제공, **제공 또는 거래해서는 안됨.**
- (d) Exemption for Nonprofit Libraries, Archives, and Educational Institutions. (1) 상업적으로 이용되는 저작물에 접근할 수 있는 비영리 도서관, 기록 보관소 또는 교육 기관은 (a)(1)(A)항을 위반하지 않음
 - (2) (1)에 따라 제공되는 면제는 저작물과 동일한 사본이 다른 형식으로 **합리적으로 제공되지 않는 경우에만 적용**
 - (3) **상업적 이익 또는 금전적 이익을 목적으로 고의로 위반한 경우** (1) 항을 위반한것으로 간주
 - (A) 초범에 대해서는 제 1203조에 따른 민사상 구제를 받아야 함.
 - (B) 제 1203조에 따른 민사상 구제 외에 반복적이거나 후속적인 범죄에 대해서는 제1항에 따른 면제를 상실

* circumvent a technological measure (기술적 조치 우회/회피)

- 저작권 소유자의 권한없이 뒤섞인 저작물을 복호화하거나, 암호화된 저작물을 해독하거나, 기술적 조치를 우회/제거/비활성등으로 손상시키는것을 의미
- 저작물에 접근하기 위해 저작권 소유자의 권한을 가진 정보의 적용이나 프로세스 또는 처리가 필요한 경우, 기술적 조치는 "효과적으로 저작물에 대한 접근을 통제"

17 U.S. Code § 1202 - Integrity of copyright management information

- (a) False Copyright Management Information 누구도 **고의로 저작권 침해를 유도, 가능, 촉진 또는 은폐할 의도**를 가지고
 - (1) 허위 저작권 관리 정보를 제공하거나 (2) 허위 저작권 관리정보를 배포하거나 그러기 위해 수입해서는 안됨
- (b) Removal or Alteration of Copyright Management Information 누구든지 **저작권 소유자 또는 법률의 권한 없이**
 - (1) **저작권 관리 정보를 의도적으로 제거&변경** (2)(3) 저작권 관리 정보가 제거 또는 변경되었음을 알면서 **이를 배포하거나 배포를 위해서 수입해서는 안됨**
- (c) Definition 이 섹션에서 사용된 "저작권 관리 정보"라는 용어는 디지털 형식을 포함하여 저작물의 사본 또는 음반 또는 저작물의 실연 또는 전시와 관련하여 전달되는 다음 정보 중 하나를 의미.
 - (1) 저작권 고지에 명시된 정보를 포함하여 저작물을 식별하는 제목 및 기타 정보. (2) 저작물의 저자에 대한 이름 및 기타 식별 정보.
 - (3) 저작권 고지에 명시된 정보를 포함하여 저작물의 저작권 소유자에 대한 이름 및 기타 식별 정보. (6) 저작물의 이용약관.
 - (7) 해당 정보를 참조하는 식별 번호 또는 기호 또는 해당 정보에 대한 링크.

Copilot 집단소송

GitHub 저작권 소유자 집단 vs Microsoft, GitHub, OpenAI 등

17 U.S. Code § 1203 - Civil remedies

- (a) Civil Actions 앞선 1201,1202 위반으로 피해를 입은 사람은 해당 위반에 대해 적절한 미국 지방 법원에 민사 소송을 제기할 수 있음
- (c) Award of Damages - (2) Actual damages (3) Statutory damages - (A) 최종 판결이 내려지기 전에 항소 당사자는 1201,1202 조의 위반 건당 \$2,500 or more than \$25,000 손해배상을 청구할수 있음

17 U.S. Code § 1204 - Criminal offenses and penalties

- (a) In General. 1201조 1202조를 고의로 위반하고 상업적 이익 또는 사적 금전적 이익을 목적으로 하는 사람은
 - (1) 최초 위반시 \$500,000 이하의 벌금 또는 5년 이하의 징역 또는 둘다 부과
 - (2) 후속 위반인 경우 \$1,000,000 이하의 벌금 또는 10년 이하의 징역 또는 둘다 부과

17 U.S. Code § 1205 - Savings clause (예외조항, 구제조항)

- Nothing in this chapter abrogates, diminishes, or weakens the provisions of, nor provides any defense or element of mitigation in a criminal prosecution or civil action under, any Federal or State law that prevents the violation of the privacy of an individual in connection with the individual's use of the Internet.

이 장의 어떤 것도 다음에 따른 형사 소추 또는 민사 소송에서의 조항을 폐지, 축소 또는 약화시키거나 어떠한 방어 또는 완화 요소를 제공하지 않는다.

개인의 인터넷 사용과 관련하여 개인의 사생활 침해를 방지하는 연방법 또는 주법.

원고의 주장

- 원고와 집단 구성원은 저작권 Codex, Copilot 훈련에서 사용된 라이선스 자료의 저작권을 소유
- MS의 Copilot은 라이선스 조항을 무시하고 교육 됨
- 피고는 라이선스 소유자 집단에게 연락해서 DMCA 의미 내에 라이선스 자료에서 CMI(copyright management information)를 제거하거나 변경할 수 있는 권한을 얻지 않고 변경하거나 제거함
- 피고는 라이선스 소유자 집단에게 라이선스를 부여받지 않고 라이선스 자료를 Copilot에 통합
- Copilot은 통합한 라이선스 자료를 동일한 조건으로 배포하지 않고 CMI를 제거 또는 변경
- Copilot 사용자는 CMI가 된 출력물을 사용하기 때문에 라이선스 의무사항을 준수할수 없고 이는 피고측이 저작권 침해를 유도, 가능, 촉진 또는 은폐하려는 의도
- 피고는 Copilot이 출력물이 Copilot 라이선스 자료라는 허위 CMI를 제공하거나 배포
- DMCA 위반으로 인한 이익은 다음과 같음 - Copilot 구독료, Azure에서 Copilot을 호스팅하는데 사용되는 비용 (원고들의 저작물이 아니었다면 Copilot은 작동하지 않을 것임)
- MS와 GitHub는 광범위한 DMCA 경험이 있기 때문에 합리적으로 GitHub의 소스코드는 라이선스가 있음을 알아야 함.

Copilot 집단소송

GitHub 저작권 소유자 집단 vs Microsoft, GitHub, OpenAI 등

violation of the Lanham Act, 15 U.S.C. § 1125 - GitHub and OpenAI 대상

15 U.S. Code § 1125 - False designations of origin, false descriptions, and dilution forbidden

- (a) 민사소송

(1) 상품 or 서비스 or 상품 용기에 대해 혹은 이와 관련 하여 상거래에서 단어, 용어, 이름, 기호 or 장치 or 이들의 조합 또는 잘못된 원산지 지정을 사용하는 사람은 사실에 대한 거짓 or 오해의 소지가 있는 설명 or 사실의 거짓 or 오해의 소지가 있는 표현,

(A) 혼란을 일으키거나, 실수를 일으키거나, 그러한 사람과 다른 사람 과의 제휴, 연결 또는 연합에 대해 속이거나, 출처, 후원 또는 그의 상품, 서비스 또는 상업적 승인에 대해 속일 가능성이 있음.

원고의 주장

- GitHub와 OpenAI는 지속적으로 CMI 를 허위 또는 오해의 소지가 있는 정보로 소비자에게 전달하여 혼란을 주거나 착오를 일으키거나 기만할 우려가 있음
- 결과적으로 15 U.S.C § 1125 (a)(1)(A) 를 위반하였음

Copilot 집단소송

GitHub 저작권 소유자 집단 vs Microsoft, GitHub, OpenAI 등

violation of Unfair Competition law, Cal. Bus. & Prof. Code §§ 17200, et seq. - GitHub and OpenAI 대상

부정한 경쟁으로 인해 경제적으로 피해를 받은 개인을 보호하기 위한 법

- 캘리포니아의 UCL은 광범위하게 작성. 섹션 17200에는 불공정 경쟁에 대한 다섯 가지 정의가 포함.

(1) 불법적인 사업 행위 또는 관행 (2) 불공정한 사업 행위 또는 관행 (3) 사기성 사업 행위 또는 관행 (4) 불공정하거나 기만적이거나 사실이 아니거나 오해의 소지가 있는 광고 또는 (5) 섹션 17500-17577.5에서 금지하는 행위. 섹션 17203은 법원이 불공정 경쟁을 방지하기 위해 금지 명령 및 기타 공평한 방어를 명령할 수 있도록 허용.

- UCL은 민간 정당과 공직자 모두에게 자격을 부여. 섹션 17204는 법무 장관, 지방 검사, 카운티 변호사 및 시 검사가 부상당한 시민을 대신하여 소송을 제기할 수 있는 권한을 부여.

발의안 64 이전에는 불공정한 사업 행위로 인해 불리한 영향을 받았는지 여부에 관계없이 모든 소비자가 UCL 소송을 제기할 수 있었음. 또한 소비자라면 누구나 대리인이 되어 불공정 경쟁을 하는 사업자를 상대로 집단 소송을 제기할 수 있음. 발의안 64는 "불공정한 경쟁의 결과로 실제로 피해를 입었고 금전이나 재산을 잃은" 개인 원고만 소송을 제기할 수 있는 반면, "영향을 받지 않은" 원고는 이제 자격이 없음.

원고의 주장

- 원고와 집단 구성원은 저작물을 생성하는 데 상당한 시간과 에너지를 투자
- GitHub와 OpenAI는 귀속 제거를 통해 이익 또는 기타 이점을 얻고 있으며 이러한 이점을 유지하는 것은 부당
- GitHub와 OpenAI는 법원이 금지하고 제지 하지 않는 이상 원고측에 회복할 수 없는 큰 손해를 계속해서 입힐것임
- GitHub와 OpenAI는 다음과 같은 불법적인 비즈니스를 관행
 - DMCA에 따른 원고의 권리 침해
 - 원고와 집단 구성원들사용자 간의 라이선스 계약에 대한 불법 간섭
 - 원고와 집단 구성원들의 코드를 Copilot을 통해 불법 출력될수 있도록 전달
 - CCPA(개인 정보 침해) 위반 및 소홀한 처리

Copilot 집단소송

GitHub 저작권 소유자 집단 vs Microsoft, GitHub, OpenAI 등

violation of the California Consumer Privacy Act, Cal. Civ. Code § 1798.150 (the “CCPA”) - GitHub and OpenAI 대상

민사소송

- (a)(1) 1798.81.5절 (d)(1)(A)항에 정의된 대로 암호화되지 않았거나 수정되지 않은 개인 정보가 다음의 결과로 무단 액세스 및 유출, 도난 또는 공개의 대상이 되는 모든 소비자 기업이 개인정보를 보호하기 위해 **정보의 성격에 적합한 합리적인 보안 절차 및 관행을 이행하고 유지할 의무**를 위반한 경우 다음 각 호에 해당하는 민사 소송을 제기할 수 있음.

(A) 사건당 소비자당 100달러 이상 750달러 이하의 금액 또는 실제 피해 중 더 큰 금액으로 피해를 복구합니다.

(B) 금지 명령 또는 선언적 구제. (C) 법원이 적절하다고 판단하는 기타 구제책.

(2) 법정 손해 배상액을 산정할 때 법원은 위법 행위의 성격과 심각성, 위반 횟수, 위법 행위의 지속성, 위법 행위가 발생한 기간, 피고의 위법 행위의 고의성, 피고의 자산, 부채 및 순자산.

- (b) 소비자가 개인 또는 집단 차원에서 법정 손해에 대한 소송을 시작하기 전에 소비자가 본 제목의 특정 조항을 명시하는 30일 전 서면 통지를 기업에 제공하는 경우 이 섹션에 따른 조치를 취할 수 있습니다.

소비자가 위반했거나 위반하고 있다고 주장합니다. 시정이 가능한 경우 30일 이내에 사업자가 인지된 위반 사항을 실제로 시정하고 소비자에게 위반 사항이 시정되었으며 더 이상 위반 사항이 발생하지 않는다는 명시적인 서면 진술을 제공하면 개별 법적 손해 또는 등급에 대한 조치가 취해지지 않습니다. - 기업에 대해 광범위한 법정 손해배상이 개시될 수 있습니다.

개별 소비자가 본 장의 위반 혐의로 인해 입은 실제 금전적 손해에 대해서만 조치를 취하기 전에 통지가 필요하지 않습니다. 기업이 이 섹션에 따라 소비자에게 제공된 명시적 서면 진술을 위반하여 이 타이틀을 계속 위반하는 경우, 소비자는 서면 진술을 시행하기 위해 기업에 대해 조치를 개시할 수 있으며 명시적 서면 진술의 각 위반에 대해 법적 손해 배상을 청구할 수 있습니다, 뿐만 아니라 서면 진술 이후의 제목에 대한 기타 모든 위반.

- (c) 이 섹션에 의해 설정된 조치의 원인은 하위 조항 (a)에 정의된 위반에만 적용되며 이 제목의 다른 섹션 위반에 근거해서는 안 됩니다. 이 제목의 어떤 내용도 다른 법률에 따른 개인 소송권의 근거로 해석되지 않습니다. 이것은 다른 법률이나 미국 또는 캘리포니아 헌법에 따라 부과된 의무 또는 의무에서 당사자를 면제하는 것으로 해석되지 않습니다.

원고의 주장

- GitHub와 OpenAI는 'the terms of Cal. Civ. Code § 1798.140(g) of the California Consumer Privacy Act (“CCPA”) 법에 따라 "businesses"를 했음.

- GitHub와 OpenAI는 연간 운영 수익이 2,500만 달러 이상이며, 연간 50,000명 이상의 캘리포니아 주민의 개인 정보를 수집하며 캘리포니아 거주자의 개인 정보 판매로 연간 수익의 최소 50%를 얻음.

- 원고와 집단 구성원들은 'terms of the CCPA as natural persons as defined in Section 17014 of Title 18 of the California Code of Regulations' 따라 "소비자"였음.

- GitHub와 OpenAI는 CCPA에 따라 필요한 notice를 제공하지 않고 고객의 PII(Personal Identifiable Information)를 사용함으로써 CCPA를 위반.

Copilot 집단소송

GitHub 저작권 소유자 집단 vs Microsoft, GitHub, OpenAI 등

Breach of Contract regarding the Suggested Licenses, GitHub's Privacy Statement, and GitHub's Terms of Service

Cal. Bus. & Prof. Code §§ 22575–22579, Cal. Civ. Code § 1798.150. Plaintiffs and the Class also bring this Complaint against Defendants

for their Tortious Interference in Plaintiffs' Contractual Relationships; Fraud, and Negligence regarding handling of sensitive data - GitHub 대상

4. License Grans to Us

- This license does not grant GitHub the right to sell Your Content. It also does not grant GitHub the right to otherwise distribute or use Your Content outside of our provision of the Service, except that as part of the right to archive Your Content, GitHub may permit our partners to store and archive Your Content in public repositories in connection with the GitHub Arctic Code Vault and GitHub Archive Program.

이 라이선스는 GitHub에 귀하의 콘텐츠를 판매할 권리를 부여하지 않습니다. 또한 귀하의 콘텐츠를 보관할 수 있는 권리의 일부로 GitHub가 당사 파트너가 귀하의 콘텐츠를 공개 저장소에 저장하고 보관하도록 허용할 수 있는 경우를 제외하고 **서비스 제공 범위 밖에서 귀하의 콘텐츠를 달리 배포하거나 사용할 수 있는 권리를 GitHub에 부여하지 않습니다.** GitHub Arctic Code Vault 및 GitHub 아카이브 프로그램과의 연결.

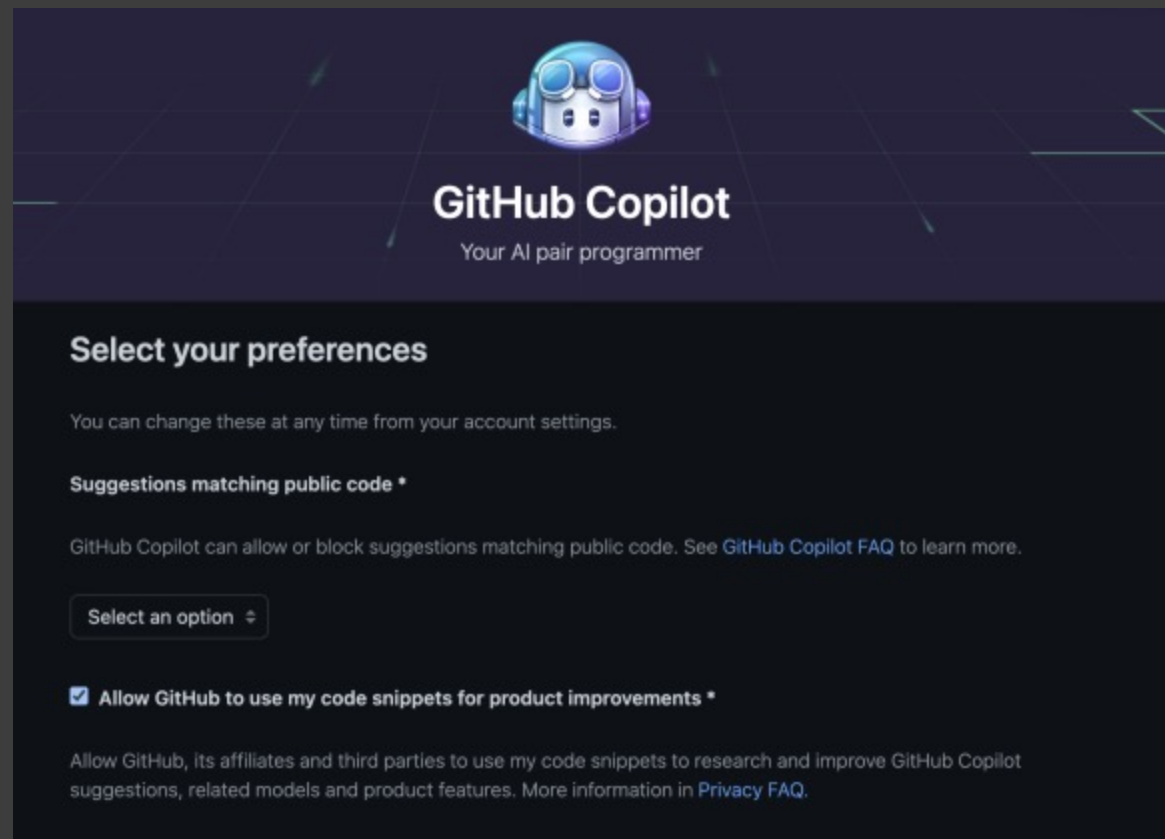
원고의 주장

- Copilot은 GitHub가 아닌 Extension 기능으로 개인의 컴퓨터에서 동작할 뿐만 아니라 Microsoft의 Azure 클라우드 플랫폼에서 동작
- GitHub는 원고 및 집단 구성원들의 라이선스 자료를 Copilot의 일부로 판매
- GitHub는 법원이 금지하고 제지 하지 않는 이상 원고측에 회복할 수 없는 큰 손해를 계속해서 입힐것임

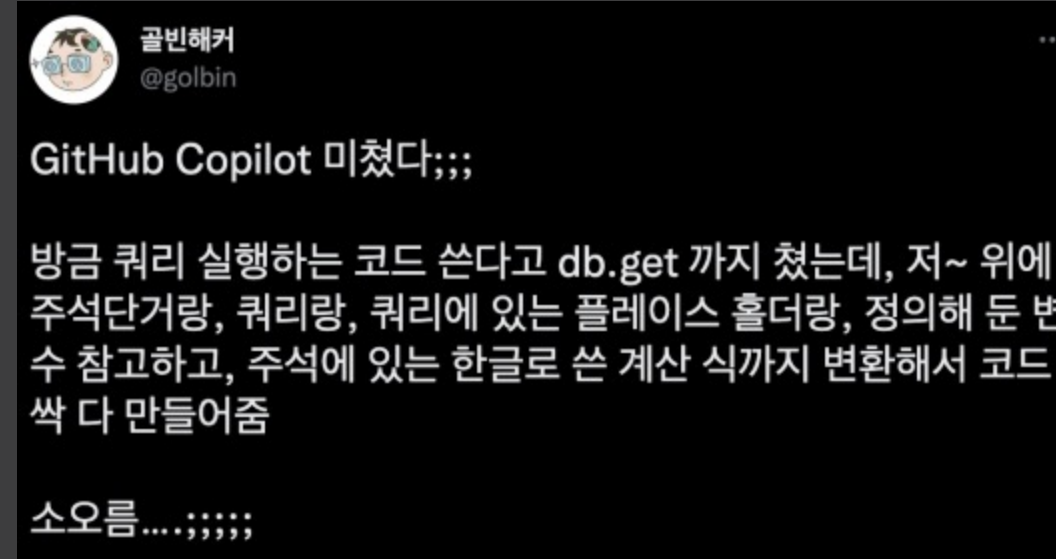
못다한 이야기

OpenAI LP - 비영리로 운영되던 OpenAI가 19년 3월 12일 OpenAI LP라는 제한적 영리추구 (capped-profit) 법인을 설립하며 기존 직원들을 OpenAI LP로 옮겨 사실상 완전 비영리에서는 벗어나게 되었음
본래 설립 목적이 훼손되지 않도록 제한적 영리추구로 설립하였는데, 비영리인 OpenAI Nonprofit이 OpenAI LP의 재무적인 의사결정을 하게 되며, 제한적인 이익을 초과하는 이익 발생 시에는 투자자나 직원들이 갖지 않고 OpenAI로 이익금을 다시 돌려주는 구조

Copilot 처음시작할때 선택사항



Copilot 후기



Copilot 출력물 위반 증거 코드 (소장에 예시)

```
57. Here is the exercise from Eloquent Javascript:
// Your code here.

console.log(isEven(50));
// → true
console.log(isEven(75));
// → false
console.log(isEven(-1));
// → ??

58. The exercise includes the "??" error. However, for Haverbeke's purposes, this is
not an error but a placeholder value for the reader to fill in. Codex—as a mere probabilistic
model—fails to recognize this nuance. The inclusion of the double question marks confirms
unequivocally that Codex took this code directly from a copyrighted source without following any
of the attendant License Terms.

59. Haverbeke provides the following solution to the function discussed above:
function isEven(n) {
  if (n == 0) return true;
  else if (n == 1) return false;
  else if (n < 0) return isEven(-n);
  else return isEven(n - 2);
}

console.log(isEven(50));
```




Thank you

Reference

Youtuber 노마드 코더 - <https://www.youtube.com/@nomadcoders>

Matthew Butterick - <https://githubcopilotinvestigation.com>

, <https://githubcopilotlitigation.com/> (소장있는 사이트)

Legal Information - <https://www.law.cornell.edu/uscode/text/17/chapter-12>

- <https://www.law.cornell.edu/uscode/text/15/1125>

- https://en.wikipedia.org/wiki/California_Unfair_Competition_Law

- <https://ccpa-info.com/california-consumer-privacy-act-full-text/>

GitHub privacy-policy - <https://docs.github.com/en/site-policy/privacy-policies/github-privacy-statement>