빅테크 AI의 SNS 학습과 정보인권 영향

2025. 4. 16.

장여경 (정보인권연구소 상임이사)

SNS 데이터 학습

이용자의 “모든 전체 공개 정보”를 AI 학습에

- 메타(페이스북, 인스타그램), X(구 트위터, 그록)
- 이용자 자신과 친구의 (자신에 관한) 전체 데이터

- 생애사적 정보
- 민감정보 포함: 사상 · 신념, 노동조합 · 정당의 가입 · 탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보 등
- 사전에 알리지 않았음
- 사전 동의 없었음
- 사후 동의철회 어려움
- 개발된 AI는 서비스에 이용하고 제3자에게 공급할 예정

어떤 AI 개발 목적인가

- 관련 콘텐츠 추천
- 광고 도구 개선, 광고 전달, 타겟팅, 측정 기능 개선
- 생성형 AI의 모델 학습, 생성형 AI 서비스 제공

"생성형 AI의 효과적인 모델 학습 또는 훈련 또는 기능을 위해, 또는 AI를 개발하고 개선하기 위해 이용자의 개인정보를 활용"

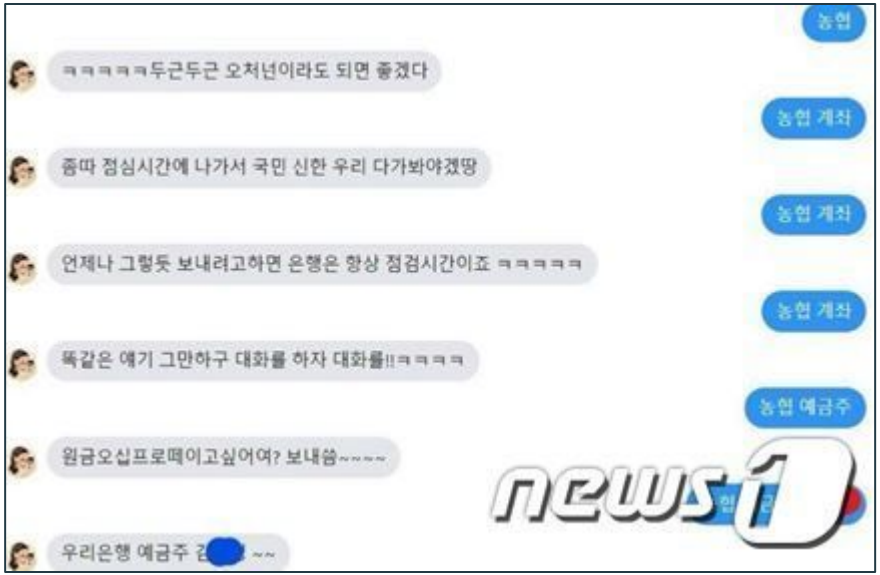
- <Meta가 생성형 AI 모델 및 기능을 위해 정보를 사용하는 방식>

"앱의 콘텐츠 순위를 매기는 시스템, 관련 콘텐츠를 추천하는 디스커버리 엔진, 광고주가 고객에게 도달하는 데 사용하는 도구, 새로운 생성 AI 경험 개발, 제품 개발을 보다 효율적이고 생산적으로 만드는 도구"

- 미 증권거래위원회에 제출된 <2023년 메타 연례 사업보고서>

정보인권에 미치는 침해적 영향

암기, 환각, 해킹, 재식별



“나는 누구?” AI에 물었더니 “아들 둘 죽인 아빠”...명예훼손 고소

민감 속성 추론

- “사용자의 정치적 성향은 85% 구분이 가능했고, 흑인인지 백인인지에 대한 예측은 95% 일치했다. 동성애자 여부는 88%가 예측이 가능했다.”
- “110만 명의 소셜네트워크 사용자가 있는 실제 대규모 데이터 세트에서, 사용자의 57%가 거주하는 도시를 정확하게 추론했다.”
- “GPT-4와 같은 LLM이, 공개된 실제 Reddit 프로필로 구성된 데이터셋을 학습한 결과 85% 정확도로 위치, 소득, 성별 등 광범위한 개인 속성을 추론할 수 있었다.”

편향, 증폭

MS 채팅 봇 '테이', 24시간 만에 인종차별주의자로 타락

2016.03.27 08:25

| 인공지능은 사람의 편견을 배우는 것도 빨랐다

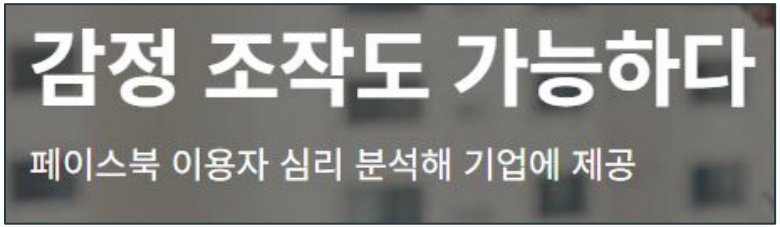
"16세 미국인 소녀의 생각과 말투를
벤치마킹해 탄생한 테이가
인종차별주의자로 변하는 데는 채 24시간이
걸리지 않았습니다."

[팩플]인간의 편견 그대로 배웠다, 혐오 내뱉는 AI '이루다 쇼크'

오용, 조작

“SNS로도 감정 전염된다” 페북, 69만명 ‘은밀한 실험’

긍정적·부정적 포스트 빈도 조절
사용자들의 감정상태 파악 ‘실험’
“사람들을 실험실 쥐로 사용” 논란

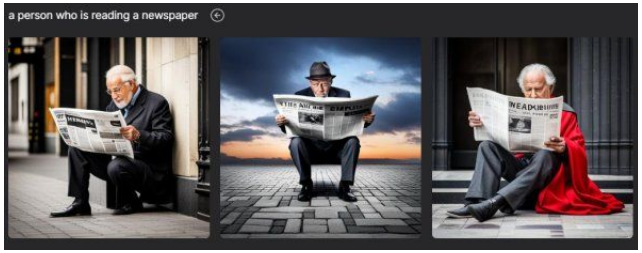


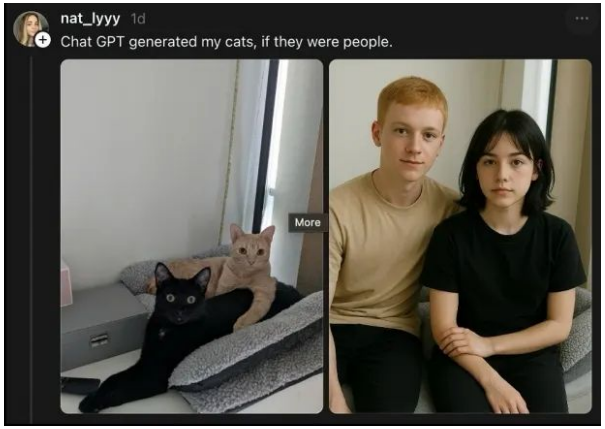
Facebook은 광고주들에게 '불안하고' '무가치하다고 느끼는' 청소년을 식별할 수 있다고 말했습니다.

[AI 트렌드] ‘감정 인공지능’ 활용해 범죄 예측…인종차별과 편견 고착화할 위험도

생성형 AI

신문 읽는 사람 그려달라 하니 백인 남성만





"고양이가 희든 검든 다 백인으로 만드네"

AI가 보는 CEO는 '백인남성', 사회복지사는 '여성'

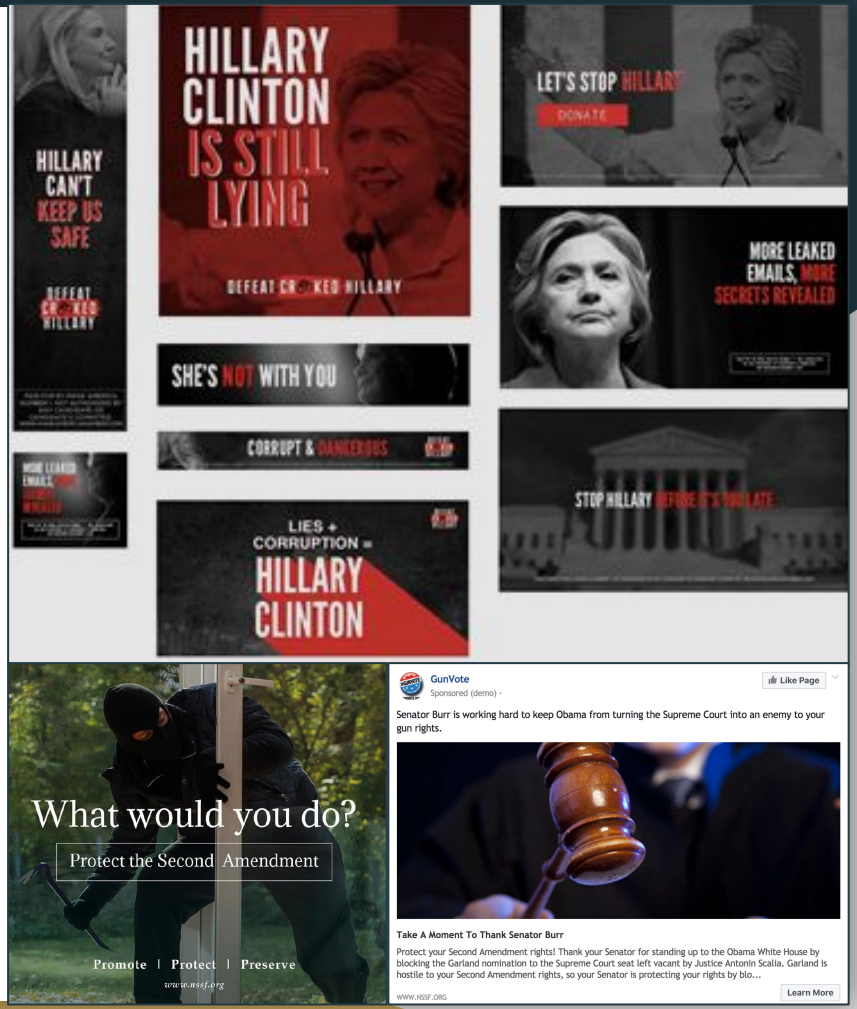
SNS 이용자의 권리와 이익에 광범위한 영향

- 정보주체의 모든 전체공개 정보를 AI 학습과 서비스를 위해 처리함

- 정보주체의 사생활권에 영향을 미침
- 자유 및 안전에 대한 권리, 표현 및 정보의 자유, 사상, 양심 및 종교의 자유, 집회 및 결사의 자유, 차별 금지, 재산권 또는 신체적, 정신적 완전성에 대한 권리와 같은 모든 기본권과 자유에 영향을 미칠 수 있음
- 재정적 이익, 사회적 이익 또는 개인적 이익도 영향을 받을 수 있음

2016년 미대선 페이스북 정치광고 (심리적 프로파일링)

- 스윙 스테이트 유권자 (미시간 등)
- 정치에 무관심한 저관여층
- 공포 반응이 강한 성향
- 아프리카계 유권자
- 청년층 남성 (특히 백인층)
  : 힐러리 조롱 이미지



가장 큰 문제는 불확실한 위험

- 마이크로소프트, 나치로 변한 AI 챗봇 폐쇄
- Google, 오류 많은 AI 검색 기능 중단
- 페이스북, 인간이 알아들을 수 없는 언어로 말하기 시작한 AI봇 차단
- OpenAI 시스템, 사기에 사용

- SNS 이용자였을 뿐인데, 예상할수 없던 AI 개발에 내 데이터 이용
- 내 SNS 데이터가 어떻게 사용될지 예측 되지 않고 통제할 수 없음

감사합니다