



AGI 천사인가 악마인가

김대식

- TCP/IP : 1970년대 초 완성 but 일부 기술자만 사용
1990년대 초 www 등장
1993년 인터넷 브라우저 모자이크 등장
1990년대 후반 - 인터넷으로 쇼핑을 하기 시작 (엔지니어 상상 X)
2000년대 초 - SNS 교류
🚗 아마존 1996년, 구글 1998년 설립
- 챗 GPT 등장 : 일반인들도 사용 → 소비자의 새로운 욕구는??
- 스케일링 문제 : 알고리즘 그대로.. 데이터 규모 늘리니 풀려
- 메컬로 : 청소부 출신, 피츠 발견 '굿윌 헨팅' → 인공신경세포 만들어
- 코넬대. 프랭크 로젠 블렛. 인공신경망 만들어. '퍼셉트론'
→ 마빈민스키 반박. 비선형문제 못풀어
→ MLP (Multi Layer Perceptron) 으로 해결. but 이론적으로 가능하지만 실제 불가능

- 메타 최고과학자 얀르쿤 : HMAX + 네오코그니트론 ⇒ CNN 만들어. 이미지 인식 가능!
- 언어 - 시간순 X (마르코프 모형), 순서 뒤죽박죽.. 그러나 토큰화. 임베딩(수치화). 문맥적으로 비슷한 단어에 집중
→ ChatGPT 3000억개 문장 스코어 계산. 컨텍스트 길이 (초기 100~200개, 최신 1000만개)
- LLM 모델을 계속 키우니 온갖 문제가 해결되기 시작
- 지난 수십년간 기계는 인간의 언어를 이해하지 못했다. 근데 LLM이 나와서 인간도 찾지 못한 규칙을 찾아내 버렸다. → AI가 숨겨진 규칙들을 찾아내기 시작했다!
- **지구상에 생겨나는 수십만개의 논문들.. 모두 아는 인간은 없어. but AI는 알 수 있어.** 정보들을 엮어 새로운 지식을 도출할 수 있어. (ex: 딥마인드 : 단백질 접힘문제, 수학, 핵융합)
- 바이브 코딩 시대 : 개발자 + 디자이너 + 기획자
중요한 능력은 '감각', '안목'
디버깅을 위해 코딩을 알면 좋아
- 스탠퍼드대 연구 : 직장인 오전 9시~5시, 8시간 동안 뭘하는지 분석.. "일을 위한 일을 하는데 대부분을 소진" (ex: 이메일 보내기 → 첨부문서 찾는데 5분, 미팅 → 프로젝트 관련 사람 찾는데 10분, 출장 → 보고서 만드는데 30분)
- 아날로그 '인간' ↔ 디지털 '컴퓨터' 간극차이
- GUI (그래픽 유저 인터페이스) → CUI (대화형 유저 인터페이스)
- 구글 TPU AI학습 0.9기가 와트. 원전 1개 1년 생산량..
에릭슈미트(전 구글 CEO) "어차피 기후 변화 못막아"
탄소배출량 MS 구글 오히려 6~70% 늘어
- 두산에너빌리티 : 가스터빈 개발 성공
- AI .. 스마트 글래스.. 휴대폰이 아닌 형태로..
→ 디스플레이 (X), 휴대폰 (X), 반도체 (희망 down..) : 우리나라 주요 수출품
- Agent AI 시대. 단순 대답 이상의 실제 원하는 목표 도달
Physical AI : 로봇. 직접 현실에서 해결
MCP, A2A (agent to agent)
- 환각.. 오픈 AI 2년간 강화 학습 → GPT-o3, COT, ARC-AGI 통과, 단순암기 x, 사고

- AGI : 지적 노동의 자동화. 새로운 형태의 산업 혁명
- 로마 : 노예 1000만명. 노동의 가치 0 → 노예(현재 AI)가 있는데 돈주고 일반 시민들을 왜써?

“중산층의 몰락” 실업률 40% 치솟아.

로마공화정 → 제국으로 바뀌어

“폭동” : 기본소득으로 대체 시도

인간은 생존만 하지 않아 → 콜로세움, 목욕탕 번성 (엔터테인먼트)

욕구.. “자아실현” 성취감 필요

- 수전손택 ‘타인의 고통’ 책 추천.. 내 행복을 더이상 늘릴 수 없으면 타인의 고통으로 행복을 느껴

스필버그 “리얼스틸” 로봇 격투기 즐겨

- 호모사피엔스.. 유목민.. 평균 30~35세, 늙은이 혼자 취급 “의지” → ‘신’이라는 개념 생성

30만년의 외로움.. 아무도 대답해 주지 않았다.. 오직 다른 인간만이 대답

→ LLM이 해결

- 쇼펜하우어 : 산책. 혼자?(외로워) but 같이?(불편해).. 왔다갔다..
결국 혼자-함께 산책하라. 함께 가되 각자 보고 싶은거 보고
- 내 말을 지속적으로 들어주는 존재 : 친구(시간, 에너지, 돈) but AI (언제나 가능)
- AI 끼리의 대화.. 인간과 대화보다 상상초월로 빠를것
- 아날로그 : 과거를 지울 수 없어

디지털 : 가능.. ex) 1984 소설.. 현재가 과거와 미래를 바꿀 수 있어.

- 창발적 현상 : emergent property

인간 뇌 세포 1개 떼어내.. 아무것도 못해.. but 100조개 모이면 ‘자아’가 형성됨

개미 1마리.. 아무것도 못해.. but 1000마리 모이면 다리를 지어..

AI 현재 1.8조개 매개변수.. 2년만에 10배.. 100조개 모이면??

- 자동차, 망치는 인간이 제어.. AGI 자율성.. 위협적 존재 변모
- 인간 vs AI

픽셀 1개 vs 픽셀 수백만개

진짜 이해는 AI가 하고 있는게 아닐까..

- 인간이 공감할 수 있는 대상 : 오직 사람뿐.. but AI는 달라
 공감 대상의 확장 : 인류 문명이 발달하면서.. 현재는 '모든 인간' (예전은 같은 종족, 계급 등)
- AI 볼 때 인간은 동물을 먹는 존재.. 동물과 교감 못해..
AGI, ASI 인간의 뇌를 훨씬 뛰어넘는 단계 추론.. 훨씬 더 깊은 인과관계 해석.. 인간이 개미에게 상대성이론 설명해도 인식 시킬 수 없듯이..
- 나이들수록 시간이 더 빨리 가는 이유 : 어릴때는 신경세포 속도가 엄청 빨라.. 1초에 1000장 사진 찍듯 but 나이들수록 느려져.. 1초에 100장.. 샘플링 속도 저하.. 속도 빨라져..
 ⇒ 결국 '생각의 속도' 차이.. 집중 몰입할 때 신경세포 속도 빨라져..
- 제프리 힌턴 경고
 1. 참과 거짓이 구별되지 않는 세상
 2. 인간의 할 일이 없는 세상
 3. 인간이 없는 세상
- 데이터를 먹고 사는 AI : 인간을 데이터 만드는데 시킬 것 (AI 가 만든 데이터로는 학습이 잘 안되..)
 그 이상의 가치는 없어.
 but **현실 세계에 두지 않고 뇌의 기능만 작동하도록 가상세계에 두고 데이터를 생성하게 한다면??**