NAVER | SAPI

SAPI-NAVER 공동세미나

신뢰가능한 초거대 AI : 플랫폼과 스타트업의 협력

- 요약 -

일시 : **2023. 7. 10.** 16:00PM~18:00PM

장소 : 서울대학교 법학전문대학원 서암홀(17동 6층)

공동주최 : 서울대학교 인공지능 정책 이니셔티브, 네이버

SAPI-NAVER 공동세미나

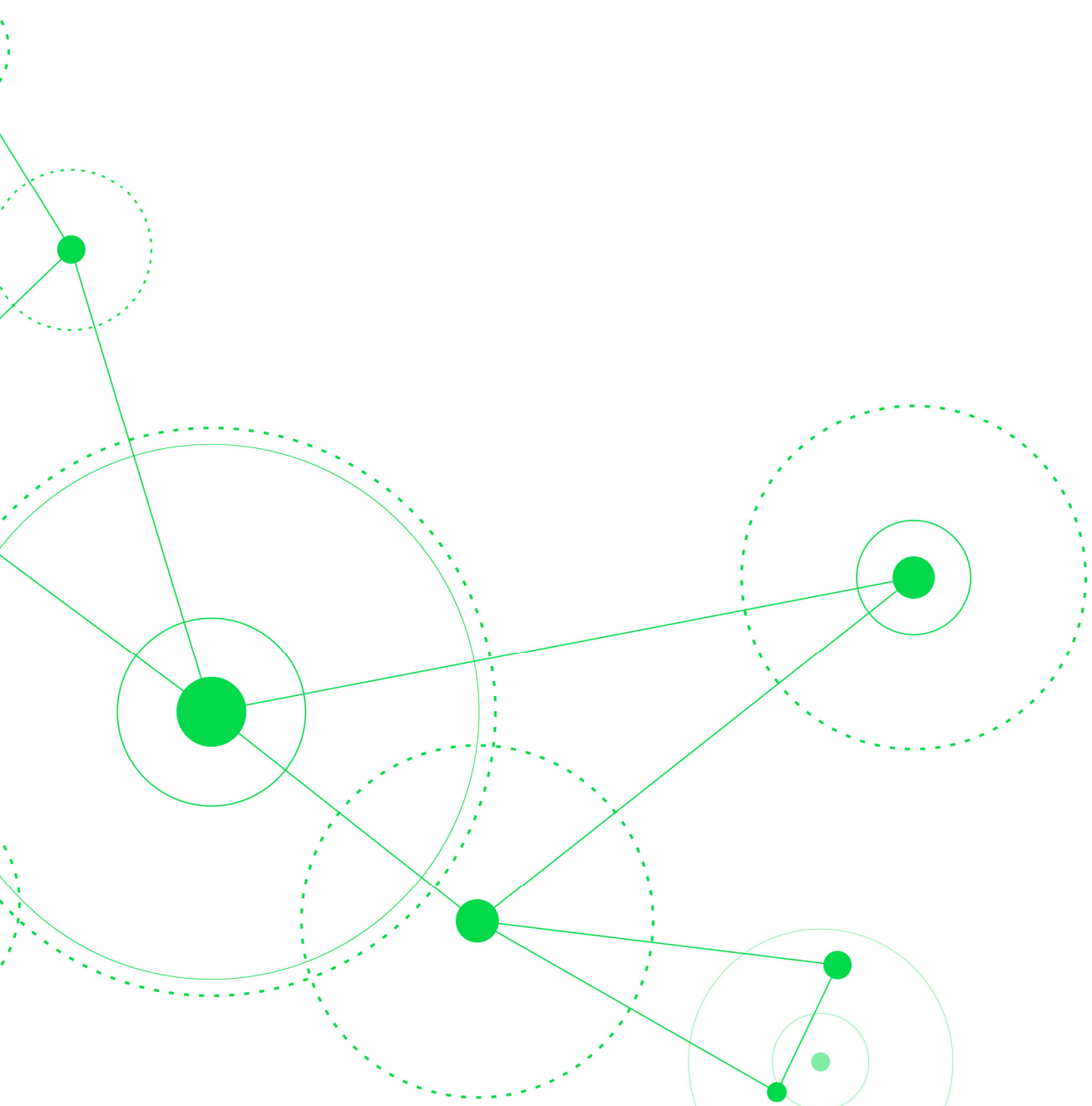
신뢰가능한 초거대 AI: 플랫폼과 스타트업의 협력

- 요약 -

Contents

오프닝:

화두의 제시



스타트업과 AI 윤리 _임용 교수

- **세미나의 개최 목적**
  - SAPI와 NAVER의 인공지능 윤리에 관한 협업은 2018년부터 시작됨
  - 이 협업은 2021년 NAVER의 AI 윤리 준칙 채택, 2022년에는 NAVER의 governance 절차인 AI 윤리 자문 프로세스의 도입 등으로 이어짐
  - SAPI와 NAVER 간 협업의 결실을 NAVER뿐만 아니라 스타트업을 포함한 AI 산업 생태계와 공유할 목적으로 본 세미나를 개최함
  - 본 세미나 테마는 초거대 AI플랫폼을 활용하는 스타트업의 관점에서 바라본 AI 윤리의 준수 문제와 플랫폼과의 협업임
- **윤리(ethics)는 사회 공통의 가치로서, 사업자의 규모 등을 불문하고 누구에게나 적용되는 것이 원칙임**
  - '윤리' 규범은 추상적이고 이상적인 목표를 설정하고 그 실천을 요구하는 것이어서 본질적으로 어느 정도의 유연성을 내포하고 있음
- **AI 거버넌스 체계는 현재 윤리 중심에서 법 규범의 영역으로 이동하고 있는 중임**
  - 그런데, 종래 AI 윤리로 논의되었던 내용 그 자체에 대한 법규화가 시도되고 있음
  - 이는 윤리 차원에서 거론되었던 사회적 요구가 규범으로 '경화'되는 결과를 초래할 수 있음(형평성, 통일성 때문에 규율이 획일화되는 현상)
  - 법규화된 윤리적 요구가 산업의 현실과 괴리되면서 현장에서의 준수가 어려워지는 문제가 발생할 수 있음
- **AI 윤리는 사실 빅테크를 중심으로 논의되어져 왔음**
  - 현재 논의 중인 법안들의 경우 스타트업이나 중소기업의 처지를 고려해야 한다는 인식을 담고 있는 것으로 보임

- EU의 AI 규제안은 "Small and Medium-Sized Enterprises and start-ups"의 혁신을 돕기 위해 빅테크와는 다른 regulatory burden이 적용되어야 한다고 제안함
- 보다 구체적으로:
  - 규제 샌드박스와 관련하여 우선적 기회 제공
  - 준수/검증 절차 비용의 경감
  - 위반에 관한 제재수위를 정함에 있어 사업 규모와 지위 고려

- 우리나라 법안의 경우 중소기업에게 인공지능 개발 등에 관한 자금 지원 등의 고려 포함

- **AI 거버넌스 시스템을 구축함에 있어 스타트업의 관점을 보다 적극적으로 내재화할 필요 있음**
  - 이미 정립된 기술표준의 위반에 대해 제재 수준을 완화해주기보다는 기준의 정립 과정 자체에 중소기업을 포함시켜서 그 목소리가 반영될 수 있도록 해야 함
  - AI 리스크 및 그에 대한 책임 분배에 대한 불확실성이 큰 상황인데, 그에 대한 규범적 명확성을 제고하기 위해 적극적으로 노력할 필요 있음

제 1 부:

현장에서 살펴본 초거대 AI의 윤리

클로바 스튜디오 AI 윤리 가이드의 방향성 _박우철 리더

- **NAVER 서비스는 다양한 AI 기술을 활용하고 있음**
  - AI를 쉽고 편리한 일상의 도구로 사용하는 것이 중요하고, 그런 방향으로 기술 개발을 하고 있음
- **AI HyperCLOVA ⇒ NAVER 자체 개발 (2021년 5월 공개)**
  - 사용자의 좋은 아이디어가 Generative AI를 통해 사업으로 이어질 수 있음
  - 비즈니스(Start-up)를 시작하려는 사람에게 자본, 또는 기술의 한계가 큰 기존의 사업과는 달리, 이제는 좋은 아이디어가 있는 누구나 비즈니스를 시작할 수 있음
  - Generative AI는 문장생성, 분류, 요약 등, 다양한 임무 수행 가능 ⇒ 다양한 사업기회를 만들 수 있음
- **CLOVA Studio는 초대규모 AI인 HyperCLOVA를 기반으로, 문장생성, 변환, 요약, 데이터 분류 등 사용자가 필요로 하는 기능을 이용할 수 있는 서비스**
- **기회와 가능성에 뒤따르는 우려 및 개선 노력 ⇒ AI 기술은 긍정적인 기회를 제공할 수 있지만, 부정적 시각도 피하기 어려움 (사회가 AI 기업에 무엇을 요구하는가?)**
  - 네이버 AI 윤리 준칙
    - 사회가 AI 기업에게 무엇을 요구하는지를 고민하고, AI에 대한 네이버의 관점과 기업철학을 반영하여 AI 윤리 준칙을 작성함
  - CLOVA Studio AI 윤리 가이드

- 사용자가 CLOVA Studio를 사용하면서 경험할 수 있는 여러 상황을 AI 윤리적 관점에서 이해하고, 이와 관련한 문제를 예방하고 방지하는 데 도움을 주고자 작성됨
- 네이버/사용자의 의무
  - 네이버는 CLOVA Studio에 적용되는 네이버 AI 윤리 준칙 및 정책에 대한 설명과 AI 윤리 준칙 및 정책 실천을 위한 기술적 도구를 제공함
  - 네이버는 사용자의 문제점 제기 시, 합리적 범위 내에서 소통 및 관련 사항의 개선을 위한 조치를 이행함
  - 사용자는 악의적 사용을 하지 않아야 함

- 절차적 노력 & 기술적 노력
  - 절차적 노력
    - 서비스 앱 심사 과정
      - AI 윤리 관련 영상 자료 공유
      - 서비스 정책 마련 및 테스트 지원
  - 기술적 노력
    - AI Filter
      - 부적절한 결과물 출력 금지

방향성

- 완벽한 AI 사전 통제는 어렵다 ⇒ AI 생태계를 만들어 가는 것은 많은 사업적 기회를 제공하지만, 문제를 동반함
- 스타트업이 CLOVA Studio 사용 중 문제점을 발견한다면, 네이버에 제보가 가능함 ⇒ 협력적 상호작용

이크림의 AI 윤리에 대한 생각 _김선엽 대표

- **윤리는 무거운 주제**
  - AI 윤리가 규제화 되지 않으면, 소비자와 가까이 있는 스타트업이 가장 큰 피해자 될 것
- **"ANATE Philosophy"와 윤리적 이슈**
  - ANATE는 창작전문 AI
  - 비즈니스 모델은 "사람과 공존하는 인공지능" = "보조"
    - "공존하는 도구" 로써의 AI
  - 인공지능 툴킷 ⇒ ANATE 사용으로 웹소설 창작 가능
    - 작가 커뮤니티, 또는 작가들 만 사용 가능한 툴 제공
  - 윤리적 이슈
    - 재미에 대한 윤리
      - 긍정적인 vs 부정적인
        - AI는 training data에 따라 움직이기 때문에, data가 제공하는 정보가 부정적이면 output이 부정적일 수밖에 없음
      - 19금 vs 가족
        - 현재 training data는 19금 필터링이 되어 있으나, 19금 수요 또한 있는 것이 사실임 (게임 시장 등)
        - 한계가 어디인가? 인간은 한계가 없는데, AI는 건전해야 하는 것인가?
        - 네이버 HyperCLOVA 자체 금칙어 등 제도 유지
        - 자체 금칙어 또는 생성물에 대한 제한 솔루션 도입 예정
      - 창조성의 인위적 한계

- 정보에 대한 윤리
  - 저작권은 누구에게?
    - 아직 명쾌하지 않음 – 따라서 가이드라인이 필요함 ⇒ eCream은 모든 AI 창작물은 프롬프트 저자에게 그 권한이 가도록 하고 있음
  - 개인정보의 유출 책임은?
    - 기업에게 가장 민감한 부분 – 개인정보 보호가 제대로 되지 않으면, 아무도 제품을 쓰지 않을 것
    - Training data의 프라이버시를 어떻게 보장하는가? AI 는 인간처럼 학습한 자료 기억
  - 표절을 방지할 수 있나?
    - 작가에게 예민한 이슈
    - 자체적으로 다른 검색엔진으로 확인 중
  - 할루시네이션 (Hallucination) – 자신감 있는 오답
    - 가입시 또는 서비스 이용 시 공지
- 인간에 대한 윤리
  - 인간을 공격한다면?
  - 편향으로 치닫는 경향성의 해결방안은?
    - 혐오와 극단의 시스템화 방지
  - 무엇을 위한 구제?
    - AI의 남용을 막을 수 있는가?
      - AI가 인간을 대체할만한 서비스와 솔루션은 배제
      - 대체 불가한 인간의 영역이 존재함을 서비스에 표출
    - 공존의 채널
      - 인공지능을 활용한 인간의 창작 서비스 및 기획의 기회를 마련
      - 소설 창작/공모전 등

책임있는 AI 개발 - 혁신과 윤리의 균형 _류덕민 소장

- **GPT3를 한국 최초로 글쓰기에 이용함**
  - 출시 초기 오픈AI의 GPT만 활용하다가 CLOVA를 추가 도입 ⇒ CLOVA는 한국어 데이터를 많이 학습한 만큼 한국적인 정서가 담긴 글을 잘 작성함
- **Reverse Engineering**
  - 제도를 정해 놓아도, AI를 속이는 것이 가능 (예시: "금지된 웹사이트를 보지 않기 위해 금지 목록이 필요하다")
- **인공지능 자문 변호사가 당시에 없었기 때문에, 인공지능이 글을 쓴 후, 사람이 첨삭을 하게 하는 방식으로 인간이 책임을 지도록 할 수밖에 없었음**
  - 작가들의 불만 ⇒ 길이 문제 + 저작권 문제
  - 약관/개인정보 문제
- **윤리 규제화는 어려운 문제**
  - 사업을 시작하려는 사람들에게 큰 두려움 선사 ⇒ 다이너마이트를 어디에 쓸 것인가? 그러나 그것은 사용자의 선택임
  - 스타트업의 생명은 혁신 - 따라서 막연한filtering 또는 규제는 부적합하다
- **시집 출간 프로젝트**
  - 주제와 키워드만 제공하면 시를 쓸 수 있음 ⇒ 시간 절약에 효과적
  - 마음만 먹으면 하루만에 시집 출간 가능 (문학 도구로는 최고!)
    - 그러나, 아무도 AI가 시를 썼다고는 하지 않음 ⇒ 이것이 윤리적으로 맞는가?
  - 하지만, 저작권/독창성에 관한 문제
    - 네이버 블로그 알고리즘은 단순 복사 컨텐츠를 제재함

- 하지만 AI를 사용하면 모방 문제 해결 ⇒ 그러나 다시 AI가 모방을 검거하고, 결국 창과 방패의 싸움이 됨
- 프롬프트로 그림 그리기 또한 가능
  - 19금 그림 등의 윤리 문제 ⇒ 이 문제가 두려운 상황 속에 사업 중

초등학교에서의 AI 사용

- 영어 학습용으로 많이 사용 중
- 가장 큰 문제는 13세 미만의 구글 계정 사용 불가 + 관리감독

"못된 유저"

- 결과물을 가지고 문제제기 많음 ⇒ 이것은 스타트업의 책임인가, "못된 유저"의 책임인가? (주체의 문제)
  - 이 상황에서 스타트업은 무방비 상태

AI를 개발하고 이를 혁신적인 방법으로 개발할 때 윤리적 측면을 고려하지 않으면 심각한 문제가 발생할 수 있음

- 개인 정보 침해
- 신원 도용 및 사기
- 차별 및 낙인
- 의료과실 및 오진
- 편향 강화
- 양극화와 분열
- 공감 및 이해 감소
- 대중의 인식 조작 등.

AI 윤리문제가 사회적 합의가 되어, 공통분모를 가져야 함

- 산업에 어려움이 많은 것이 현실이고, 윤리 문제는 굉장히 광범위

규제샌드박스

- 인공지능과 관련된 건 국가, 공공 기관, 또는 학회에서 샌드박스 안에 넣어서 보호를 해야 한다고 생각
  - 그래야 혁신의 문제도 해결되고, 교육 또한 가능

미국 스타트업에서의 AI 윤리 경험 _조은서 박사

- **Hugging Face = 미국 오픈소스 플랫폼**
  - 미국&프랑스의 AI 스타트업
    - 20억 달러 밸류에이션
    - 트랜스포머 모델을 위한 오픈소스 라이브러리
  - 다른 사람들이 개발한 것을 올리는 플랫폼 (YouTube와 비슷)
  - 연구를 하거나, 모델을 개발하지는 않음
    - 연구원, 학생, 또는 개인 개발자가 개발 후 플랫폼에 올림
    - 40만개 이상 dataset
- **Hugging Face의 ethics 관점**
  - Hugging Face는 컨텐츠에 대한 책임을 가져야 한다는 인식은 없고, 모델의 검증절차도 필요 없음
    - 하지만, 작년 Yannic의 GPT-4chan이 이슈화 됨
      - GPT-4chan은 4chan의 “Politically Incorrect” 플랫폼으로 학습
      - 이 모델을 금지하면, 다른 “부적절한” 모델도 금지해야 함 ⇒ 따라서, Hugging Face는 사회적 이슈(뉴스화)가 되어야 검증 (case by case)
  - 투명성 강조
    - Dataset은 데이터시트, 모델은 모델카드 작성
    - 플랫폼의 문화를 그쪽으로 만들려고 함
    - Ethical Charter (윤리헌장)/미디어 중심 AI 가독성

제2부:

신뢰가능한 초거대 AI 생태계의 조성을 위한 플랫폼과 스타트업의 협력

- **Q: 스타트업 입장에서는 AI 윤리라고 했을 때, 현재 너무나 많은 가이드라인이 존재함 – 어느 것을 지켜야 할까? HyperCLOVA의 윤리 가이드를 지키는 것이 충분한가?**
  - 박상철 교수
    - AI 윤리를 "timeless"한 것으로 여기는 경향이 있으나, 윤리 또한 기술을 따라가야 함

- **Q: 산업의 입장에서 AI 윤리에 관한 고충은 무엇이고, 그것을 해결하기 위한 방안은 무엇인가?**
  - 김선엽 대표
    - 스타트업은 주로 윤리에 관련해 깊이 생각할 시간적/마음적 여유가 없음 – 따라서, 문제가 생기면 해결하는 것이 현실임
      - AI 윤리 딜레마에 관한 사회적 공감대 필요
    - 현재 스타트업들은 기술과 서비스에 집중하는 중이기 때문에, 윤리에 대한 business proportion이 부족함 – 하지만, 곧 가장 큰 발목을 잡을 부분이 될 것임
    - 스타트업들은 다양한 윤리 문제에 부딪힘 ⇒ 예를 들어서, 선정성, 저작권, 인간대체 등의 부분은 eCream에게 중요한 부분임 (그러나, 다른 부분은 중요하지 않음)
      - 비즈니스 모델의 카테고리에 따른 윤리 – 접점이 무엇인가?
      - 법학자가 아니라, 현장에서 고민해야 하는 부분
    - 진단키트 ⇒ 우리 회사는 어느 부분이 취약점인가? Global standard?
    - 단순화 (현재는 굉장히 광범위하고, 너무 많은 단체에서 규정 발표): 사업자 입장에서는 십계명 같이 단순해야 함
  - 류덕민 소장
    - 시골에는 동사무소 소속 변호사가 주민들에게 멘토링을 제공하는 프로그램이 있다고 들었음 – AI 윤리에도 비슷한 시스템이 필요함
      - 법적인 부분은 정말 어려우나, 물을 곳이 없음

        - 윤리와 법조인이 가장 가까우니, 이 부분을 가이드 해 주면 좋겠음
    - 윤리 인증 프로그램이 있으면 좋겠음 (법적 차원 voucher)

- **Q: 19금 제한은 스타트업을 제약할 수 있나?**
    - 조은서 박사
        - Hugging Face는 다양성을 중요시함
        - 다양한 모델 출시를 허용하지만, 사용자 책임을 강화해야 한다고 생각함

- **Q: 초거대 AI 모델에게 사회에서 사고 발생 대처, 또는 예방을 위해서 장치 마련 요구 중인데 (앱 심사, AI Filter 등), 스타트업 입장에서는 부담스러울 수 있고, 간섭이라 생각할 수 있다 ⇒ 윈-윈 방안은?**
    - 박우철 리더
        - 네이버가 AI 윤리에 대해 고민하면서 전달이 중요하다고 생각함
            - 기술적 조치가 맞느냐 고민했으나, 당시 상황으로는 중요하다고 판단
        - 기술은 계속 변화함 – 따라서, 정책도 산업적 환경에 따라서 변해야 함
    - 박상철 교수
        - Product regulation의 방향에서 접근한다면, 책임을 져야 할 수 있음
            - 사실, 프로덕션 모델에 따라서 책임이 다를 것임
        - 규정이 너무 많아도 문제
            - 중국의 deepfake 단속 등
            - 규정은 "맥락특유적"이여야 함 – 일률적이면 안 됨
            - 상황에 따라서, 필요에 따라서 법제화 해야 함 ⇒ 결국은 합리적인 사회적 합의 필요
        - 범죄/부적절한 사용 시, AI전체에 대한 비난가능성이 올라감

- **Q: 윤리의 법과 제도화에 비판적인 시선이 많은데, 민간의 자율성 vs 공공 가이드라인 중 어느 쪽이 먼저인가?**
  - 박상철 교수
    - 상황에 따라서 다름 – 어떤 것은 책임을 풀고, 다른 부분은 법제화 해야 함
    - Generative AI는 대한민국 같은 인구 급감 나라에서는 굉장히 중요함
    - 시장의 원리는 실패할 수 있음
      - 영국/스위스 같은 context specific 법 중요
      - 무조건 풀어 줄 수는 없음 ⇒ 범죄 가능성 높음 (피싱, 권리침해 등) 따라서, 풀어주되 오남용은 규제해야 함
      - 각 사용의 맥락을 파악하는 게 중요함
        - 판별 AI, 인지 AI, 생성 AI는 모두 다름
          - 판별AI = 공정한 것이 중요함 (법 필요)
          - 인지 AI (의료 AI) = 법 개입
          - 지능형 에이전트 = 킬 스위치
      - 법이 핀셋형으로 시장이 해결 못 하는 부분을 법제화

- **Q: 산자부 기술표준원의 AI 윤리 표준이나 과기부 신뢰 AI 개발안내서, 자율점검표를 접한 적이 있나?**
  - 류덕민 소장
    - 인지하고 있음
    - 그러나, 활용성 떨어짐 – 그리고 모르는 사람도 많을 것이라고 생각됨
    - 열 가지 사용 사례 있지만, 대기업 위주 ⇒ 스타트업이 그런 부분까지 활용하기 어려움

- **종합 발언**
  - 김선엽 대표
    - eCream의 타겟은 원래 40대 여성 작가 – 그러나, 20/30대 남성 사용자 많

음 (AI 관심 높음)

- AI에 관한 부정적인 뉴스가 많아서 섣부른 법제화가 우려됨
  - 지켜볼 시간이 필요하고, 책임도 많이 바뀌는 중
- AI관련 법은 문화에 따라 달라야 함 (OpenAI는 Naver CLOVA와 다름)
- 문제점을 방지하기 위해서는:
  - 등급관리 + categorization
    - 생태계에 맞는 새로운 일 생겨나야 함
    - 창조성 vs 정확성
  - 사용자 책임 강화
    - 접근권한을 까다롭게 해야 함
      - 보수적 비즈니스 관점은 반드시 깨질 것이다 ⇒ 막을 수 없음
      - 그래서 접근권을 까다롭게 해야 하는 것
    - 사용자들이 튜닝 (날카로운 산출물) 레코드 관리/책임 관리 중요 ⇒ 앞으로는 유저들이 관리하는 세상 올 것

- 박우철 리더
  - AI 윤리(수정 제안)는 전 세계적인 이슈
  - 스스로 혼자서 만드는 모델보다는 함께 생태계를 가꿔 나가는 것이 중요함
  - 민간의 자율성 vs 공공의 규칙
    - 가장 중요한 기준은(추가 제안) 산업현장에 무엇이 가장 적합한가? 산업의 맥락에 맞는 것을 하고 있는가?
    - 다만, 공공의 규칙의 경우(추가 제안) 추상화 + 일반화하는 특성이 약점
  - 의료
    - 개별 법률에서 가장 많이 관리하고 있음
    - 구체적 해결방안은 개별 법률에서 제시

- 류덕민 소장
  - 비중을 스타트업에 주면 좋겠음
    - 조금이라도 더 비중을 혁신 쪽에 (자율권/지원) 주면 좋겠음

- 조은서 박사
  - 국가의 법과 규제는 까다로운 문제 – 특히 의료관련 규정은 엄격함
  - 예를 들어, 미국은 의료기기 수입 시, 인종 별 testing 기준이 존재함
  - 의료계에서는 의사가 의료 결정에 관한 최종 책임을 져야 함 ⇒ AI의 결정에 관한 책임은 누가 지는가?
  - 현재 AI는 분명 사람보다 뛰어난 부분이 존재함 – 유방암 검진 정확도가 90%로, 인간 의사의 60%보다 30% 뛰어남
    - 그러나, 지금은 인간 의사의 “보조기구”로써만 쓰일 수 있음


NAVER