

AI-009호

2020.03.10

SPRI AI BRIEF

인공지능 최신 동향과 시사점

vol.9

목차

01 美 교통부, 자율주행차 규제 가이드라인
(Automated Vehicle 4.0) 발표

02 美 국방부, 5대 AI 윤리 원칙 채택

03 美 CNAS, AI에 의한 기만 가능성
(AI Deception) 경고

04 비영리 AI연구소,
OpenAI의 세속화 우려 고조

美 교통부, 자율주행차 규제 가이드라인 (Automated Vehicle 4.0) 발표¹⁾

1 미국 교통부는 2016년 이후 매년 자율주행차 규제 가이드라인을 업데이트하여 발표

- 규제 가이드라인의 목적은 미국의 자율주행차 기술 리더십을 강화하고, 안전 등 자율주행차가 미국 사회에 야기하는 문제를 예방하는 것

<표> 미국 정부가 발표한 자율주행차 규제 가이드라인

제목 (발표일자)	주요 내용
Federal Automated Vehicle Policy (2016.9.)	안전한 자율주행차를 구현하기 위한 가이드라인으로서 디자인, 연구개발, 품질 테스트 과정의 모범사례(Best Practice)를 제시
Automated Driving Systems 2.0 : A Vision for Safety (2017.9.)	자율주행기술 3~5단계 개발기업을 대상으로 '자발적 안전 자체 평가 (Voluntary Safety Self-Assessment)'를 도입하여 안전성을 개선하도록 유도
Automated Vehicle 3.0 : Preparing for the Future of Transportation (2018.10.)	기존 규제(예 : 트럭 운전자 보호를 위한 운행시간 제한)를 자율주행차에 어떻게 적용·개편할지 논의
Ensuring American Leadership in Automated Vehicle Technologies : Automated Vehicle 4.0 (2020.1.)	미국의 자율주행차 기술 리더십을 강화하기 위한 규제 원칙을 제시하고 정부의 지원 방안에 대해 다각도로 검토

2 올 1월 발표한 네 번째 자율주행차 규제 가이드라인²⁾(Automated Vehicle 4.0)에서는 규제의 3대 목표 및 10대 원칙을 제시

- **사용자·지역사회 보호** 기업은 △기술의 성능과 한계를 명확히 밝히고, △범죄에 악용되지 않도록 보안을 강화하며, △개인정보를 보호하고, △운전자에게 기존 차와 자율주행차 중 선택권을 보장
- **효율적 시장 조성** 시장을 통해 기술이 선택되도록 기술 중립을 유지하고, △자국 기술이 해외에서 불법 사용되지 않도록 보호하며 △불필요한 규제를 제거함으로써 혁신 유인을 제고
- **협력 노력 강화** △연방정부와 주정부 간 규제를 일원화하고, △자율주행차 관련 규제와 지원정책 간 일관성을 높이며, △미국 교통 시스템 전체를 개선한다는 목표로 접근

3 이번 발표를 통해 美·中 기술 경쟁이 심화되는 상황에서 미국 기업의 혁신을 저해하지 않는 최소한의 규제만 적용한다는 트럼프 정부의 기본 원칙을 재확인

- 미국 자동차 업계의 완전 자율주행차 출시 시점이 수년간 미뤄지는 가운데, 중국은 정부가 나서 자율주행차 전용차선 건설 등 과감한 투자를 통해 미국을 추격
- 미국 교통부 장관 일레인 차오는 “이번 가이드라인은 자율주행차 업계의 혁신을 돕기 위해 만들어 졌으며, 미국의 기술 리더십을 공고히 하는데 도움이 될 것”이라고 기대

1) Forbes, “USDOT’s Automated Vehicle 4.0 Seeks Unifying Theme Across Federal Government”, 2020.1.10.

2) United States Department of Transportation, “Ensuring American Leadership in Automated Vehicle Technologies : Automated Vehicle 4.0”, 2020.1.

美 국방부, 5대 AI 윤리 원칙 채택³⁾

1 美 국방부는 AI의 군사적 활용에 대한 위험을 방지하기 위한 대책으로 5대 AI 윤리 원칙 채택 (‘20.2.24.)

- AI의 군사적 활용을 대비하기 위한 5대 AI 윤리 원칙은 국방혁신위원회(Defense Innovation Board)가 15개월 간 논의를 거쳐 만장일치로 의결한 사항으로 지난 2019년 10월 국방부장관에 권고⁴⁾

2 이번 윤리 원칙은 AI 사용의 책임성, 편향의 최소화, 투명하고 추적 가능한 개발, 사용 영역의 명확화, 의도치 않은 결과의 통제 가능성을 포함

- 그간 국방부 산하 조직인 방위고등연구계획국(DARPA)은 신뢰 가능하고 강건한 AI 시스템 개발을 위한 차세대 AI 캠페인(Next AI Campaign)*을 주도적으로 추진해 왔음
 - * 설명가능한 인공지능 (eXplainable AI), 주변 상황과 자신의 능력을 인지하는 AI 개발 등
- 美 국방부의 AI의 윤리 원칙은 미국의 AI 기반 무기체계개발의 새로운 R&D 방향성을 제시

<표> 국방부 5대 AI 윤리 원칙

AI 원칙	채택안 세부 내용
책임 (Reposable)	국방부(Department Of Defencse, DOD) 구성원은 적절한 수준의 판단을 하고, AI 기능의 개발, 배치, 사용을 책임짐
동등 (Equitable)	DOD는 AI 기능에 대한 비의도적인 편향을 최소화하기 위해 신중한 조치를 취함
추적가능 (Traceable)	DOD의 AI 기능은 관계 구성원이 투명하고 추적가능한 방법론, 데이터, 설계 및 문서화를 포함하여 AI 시스템의 기술적 이해, 개발/운영 방법에 적절한 이해를 갖도록 개발 및 배치될 것임
신뢰 (Reliable)	DOD AI 기능은 사용 영역을 명확히 정의하고, 해당 시스템의 전체 수명 주기에 안전, 보안, 강건성을 시험하고 보장함
통제가능 (Governable)	DOD AI 기능이 의도된 기능만을 수행토록 설계/구현하며, 의도하지 않은 결과에 대해 감지하거나 피해야하는 능력을 보유할 것임

3 향후 AI기반 무기체계뿐만 아니라 다양한 상용 기술 개발에 있어 높은 수준의 신뢰성, 안전성, 강건성 등 AI 윤리 준수 여부가 주요 고려사항으로 부상할 전망

- AI 기술의 군사적 활용을 적극적으로 추진해 온 미국이 AI 윤리 원칙을 채택한 사실은 성능보다 안전하고 신뢰할 수 있는 시스템 개발을 우선시한 결과로 판단됨
- 그간 AI 윤리를 강조해 온 유럽과 국제기구에 미국이 가세함에 따라 국제 규범으로 AI 윤리에 부합하는 AI R&D 추진 전략 필요

3) U.S. Department of Defense, DOD Adopts 5 Principles of Artificial Intelligence Ethics, (2020.02.25.)

4) U.S. Department of Defense, Defense Innovation Board Recommends AI Ethical Guidelines, (2019.11.01.)

美 CNAS, AI에 의한 기만 가능성(AI Deception) 경고⁵⁾

1 워싱턴DC에 소재한 국가안보연구 비영리기관인 CNAS(Center for New American Security)는 'AI의 기만(AI Deception)'가능성을 경고

- 미국의 저명한 전기전자분야 전문 학술지(IEEE Spectrum)는 AI에 의한 기만*에 대비해야 할 필요성을 강조한 CNAS의 연구 일부를 소개(2.24)
 - * 예시 자율주행차의 신호 오독(정지표시를 속도제한으로 해석), 팬더를 긴팔원숭이로 분류, 딥페이크(Deepfake) 기술을 적용해 만든 가짜 뉴스 및 합성 포르노 영상 등 잘못된 학습 또는 인간(개발자)의 의도가 반영된 AI의 적대적 공격(adversarial attacks) 형태
- 구글 DeepMind의 AI윤리 분야 선임과학자이기도 했던 Roff 박사는 기고문에서 “AI가 스스로 이러한 기만 방법을 터득하게 된다면 어떤 일이 벌어질까?”라는 화두를 제기
 - 현재 AI 기술 수준으로는 AI가 인간의 의도에 의해 악용되는 정도이며, 스스로 기만 행위를 학습하는 수준은 아니지만 미래에 대한 대비의 필요성을 강조

2 AI에 의한 기만행위는 이미 현실에서 일부 유형이 나타나고 있음

- 다중 에이전트(multi-agent) AI 시스템에서는 AI가 '기만'에 대한 개념* 없이도 특정 목적을 달성하기 위해 정보를 숨기거나 거짓 정보를 제공하는 행위가 가능
 - * 마음이론(Theory of minds)에 따르면 진정한 의미의 기만은 타인과 자신의 믿음, 욕구, 의도와 관점이 다르다는 것을 이해할 수 있는 능력에서 비롯됨
- 현재 AI에 의한 기만행위는 1) AI 에이전트가 능동적으로 잘못된 정보를 중개(Acts of commission)*, 2) AI 에이전트가 피동적으로 정보를 숨기는 유형(Acts of omission)**으로 이미 현실에서 나타나고 있음
 - * 다양한 형태의 잘못된 정보 신호를 학습 판별하는 사이버 방어 시스템(Cyber defense), 적의 탐지에 잡히지 않기 위해 기만적으로 움직이는 시기반 로봇틱스 (예: 드론 군집 비행)
 - ** * 최소한의 세금 납부를 위해 소득 신고 누락 방법을 제안하는 AI에이전트

3 AI에 의한 기만의 정의를 비롯해 AI 스스로 기만하는 법을 배울 수 있는 다양한 가능성에 대비한 해법 마련 필요

- 진정한 의미의 AI 기만은 AI 스스로 자신과 타인의 심적 상태에 대한 정확한 이해를 전제로 하나 현재의 AI는 인간(개발자)의 의도를 담아 표현하는 수준
- 하지만, AI가 느낌, 믿음, 의도, 감정 및 사회적 상호작용 방식을 배우게 되면 사람들의 행동을 '관리'함으로써 스스로의 이득을 극대화하기 위한 기만행위를 할 가능성 증가
- 이러한 시점이 도래하기 전에 AI에 의한 기만의 정의를 비롯해 다양한 기만행위 유형과 방법, 사회적 영향 등을 고려해 정책, 규제, 기술 개발 등 선제적 해법 모색이 필요

1 MIT Technology Review는 최근 비영리 AI연구소로 유명한 OpenAI의 상업적 행보에 대한 특집 기사를 발간(2.17)⁷⁾

- OpenAI*는 지난해 OpenAI LP라는 제한적 영리 조직**을 설립(19.3)하고 마이크로소프트社로부터 10억 달러 규모 투자 유치(19.7)
 - * 엘론 머스크(Elon Musk)가 10억달러의 기부금을 출연해 지난 2015년에 설립
 - ** 투자금의 100배 한도내에서만 이익 추구를 하겠다고 선언하고 OpenAI의 일부 연구원들을 OpenAI LP에 재배치시켰으며 기존 OpenAI는 OpenAI Nonprofit이라는 조직으로 재정비
- 기사는 범용인공지능(Artificial General Intelligence)을 안전하게 개발하고 그 혜택을 모든 사람에게 제공하겠다는 목적으로 설립된 OpenAI*의 영리화에 대한 우려 제기

2 OpenAI의 공동 창립자였던 엘론 머스크(Elon Musk)는 트위터를 통해 'OpenAI가 좀 더 개방적일 필요가 있다'는 의견을 피력

- OpenAI의 공동창립자였던 엘론 머스크는 지난 2018년 운영진과의 견해차로 이사회에서 물러난 이후 경영에는 관여하지 않고 있음
- 하지만, 이번 MIT 테크놀로지 리뷰 기사에 대해서 'OpenAI가 보다 개방적일 필요가 있다'는 의견을 트위터*로 표명(2.18)
 - 아울러, AI기술의 위험성을 강조하며 자신의 기업(Tesla)을 포함해 AI를 개발하는 모든 조직에 대해 개별 국가 및 국제적 차원의 규제 강화 필요성을 재차 주장
 - * "OpenAI Should be more open", "All Orgs developing advanced AI should be regulated, including Tesla" (2020. 2. 18, @elonmusk 트위터)

3 영리 추구하고 동시에 숭고한 사회적 가치를 실현하겠다는 OpenAI의 '양수겸장 전략'이 효과를 발휘할지 귀추가 주목됨

- OpenAI LP의 설립에는 막대한 컴퓨팅 자원을 요구*하는 AI기술 개발을 위해 재정적 투자의 지속성을 확보한다는 명분이 작동
 - * 경쟁 기업들은 혁신적 결과를 얻기 위해서는 3.4개월마다 2배의 컴퓨팅 자원을 사용
- 하지만, 영리목적으로 개발된 기술은 기업에 의한 사유화와 독점화가 예상되며 이는 OpenAI가 추구하는 공유와 개방의 철학과는 모순되는 상황에 직면
- 정부는 AI기술의 안전하고 올바른 개발과 사용을 민간에게만 맡길 것이 아니라 기술에 대한 이해 역량을 높여 선제적으로 높은 개발 기준을 마련하는 것이 필요

6) Inverse, "OpenAI co-founder Elon Musk says secretive a.i. firm should be more open", 2020. 2. 18.

7) Karen Hao, "The messy, secretive reality behind OpenAI's bid to save the world", MIT Technology Review. 2020. 2.17

