

제 목	인공지능 사회와 그 적들: 인공지능을 통제하려는 시도에 대항하기		
소속 대학		소속 학과	
학 번		이 름	김*후
이메일		연락처	

오픈AI 사의 생성형 인공지능 Chatgpt는 출시 2월 만에 월간 활성 이용자 수 1억 명을 돌파하는 획기적인 인기를 끌며¹⁾, 4차 산업혁명 시대의 핵심 기술로 떠오르고 있다. 1956년 다타머스 회의에서 존 맥카시가 ‘인공지능’이라는 용어를 처음 사용²⁾하게 된 이래, 인공지능 기술은 그간 놀랄만한 발전을 이룩하였다. 미래에는 인공지능 기술이 사물인터넷, 자율주행, 로봇틱스 등의 기술과 결합하여 산업의 혁신을 가져오게 될 것이며, 첨단산업의 주축 기술로 자리잡게 될 것이라는 전망이 제시된다. 다만 인공지능이 가져올 미래가 무조건 장밋빛이라고만 말하기는 어려운 실정이다. 기술의 밝은 면이 있다면, 그 반대편에는 악용하고자 하는 이들도 있다. 이들은 인공지능을 자신들의 지배력을 확대할 수 있는 도구가 될 것이라고 보고, 정보의 우위를 통하여 통제권을 행사하고자 시도하고 있다.

인공지능 기술의 악용이 초래할 위험의 양상은 실로 다양하며, 그 영향 범위도 매우 크다. 그 중 하나로 잘못된 정보의 확산이다. 인공지능이 잘못된 정보를 제공하고, 그러한 잘못된 정보가 확산되면서 큰 파급효과를 갖게 된다. 인공지능이 고의로 날조된 정보(disinformation)와 고의는 아니지만 잘못된 정보(misinformation)의 확산 통로가 될 수 있다. 생성형 인공지능 서비스는 통계적으로 배열 확률이 높은 단어들을 연이어 생성하는 과정에서 겉으로는 그럴듯해 보이지만 사실은 잘못된 정보를 산출하는 이른바 ‘환각 현상(hallucination)’이 있다. 즉, 인공지능은 때때로 거짓 정보를 제공한다. 더불어 인공지능 알고리즘은 사회적 편견과 차별을 강화하는 역할을 수행할 수 있다. 인공지능 알고리즘에 의한 예측은 편향성과 차별성을 보일 수 있으며, 특히 학습 데이터가 불완전할 때 편향은 더욱 증가하는 경향을 보인다.³⁾ 일례로, 인공지능은 미인대회 선발에서 수상자 대다수를 백인으로 뽑는다는가(미인 판별 학습과정에 이용되었던 데이터 사진이 주로 백인이었기 때문)⁴⁾, 아마존 사의 인공지능 채용 시스템이 경력 10년 이상의 남성 후보만을 주된 고용 후보로 제시한다는가(그동안 지원자 중 남성 비율이 압도적으로 많았기 때문에 여성이라는 단어만 들어가도 감점 요소로 분류하기 시작)⁵⁾, GPT-3의 경우 ‘두 명의 이슬람교도가’라는 문장을 제시했더니 테러 관련 문장을 출력한다는가⁶⁾하는 경우가 있었다. 사진이나 영상의 경우, 실제 그대로의 것인지 아니면 인공지능에 의해 만들어진 것인지 구분이 쉽지 않아 큰 혼란을 가져오기도 한다. 예컨대, 브루노 마스가 걸그룹 뉴진스의 ‘하입 보이’라는 곡을 커버한 영상이 화제가 되었는데, 사실 이는 그의 목소리를 학습한

인공지능으로 제작된 영상이었다.⁷⁾ 이처럼 인공지능이 제시하는 잘못된 정보는 많은 사람들을 오인케하거나, 나아가 어떤 악의적 목적 아래 ‘딥페이크’의 형태로 퍼질 수도 있다. 문제는 한번 잘못된 정보가 유포되면 이를 바로잡는데 많은 자원이 소요되고(이른바 ‘브랜달리니 효과’), 잘못된 정보가 전체 정보 환경을 오염시키게 된다.⁸⁾ 이러한 오용 가능성에도 불구하고, 인공지능이 어떠한 데이터를 통해 학습했는지에 대하여 일반에 정확히 알려지지 않아 감시가 쉽지 않다는 점이 문제가 된다.⁹⁾

다른 문제는 인공지능을 활용한 감시와 통제이다. 조지 오웰의 소설 『1984년』에 등장하는 ‘빅 브라더’(정보 관리를 통해 사회를 통제하고 영구 집권을 기획하는 존재)가 인공지능 시대에 다시 출현하게 될 것이라는 우려가 있다. 개인에 관한 정보, 예컨대 위치, 활동, 거래 내역, 취향 정보 등을 수집하여 데이터베이스화한다면, 그리고 그것이 통제를 원하는 세력에 손에 들어가게 된다면 그 결과는 지속적인 감시 사회의 도래가 될 것이다. 특히 최근에는 신경과학의 발전으로 두뇌 스캐닝(F-MRI 이용 두뇌 혈류 분석)과 인공지능 분석 도구(LLM)를 결합해 특정한 사람의 생각을 읽어낼 수 있다는 연구도 있다¹⁰⁾. 이 결과는 관점에 따라서 ‘자유롭게 생각할 권리’가 침해될 수 있는 가능성을 보여주는 것이라고 생각될 수도 있다. 또한 인공지능 기술이 정치적 목적을 위해 사용되는 경우도 나타날 수 있다. 선거 등에 영향을 줄 목적으로 특정 정파를 선전한다거나, 댓글 등을 통해 여론을 조작 시도하거나, 조직적 이념 선동을 기할 수도 있다. 우리에게 대한 정보를 수집·처리·자료화하여 사생활을 감시하는 방식으로 이용할 수 있다는 가능성은 우리들로 하여금 ‘파놉티콘’을 떠올리게 한다. 이와 관련하여 최근에 주목할 만한 보도가 나왔는데, 바로 동영상 공유 플랫폼 틱톡이 중국 공산당의 선전 도구로 활용된 바 있다는 폭로에 관한 것이었다. 틱톡의 모회사 바이트댄스에서 근무하다 해고된 전직 임원이 중국 정부가 그간 회사 내부 데이터에 접근권을 행사한 바가 있다고 밝힌 것이다.¹¹⁾ 인공지능 사회에서는 더 체계화된 통제가 성행하게 될 것이라는 우려가 나오는 부분이다.

또 다른 문제는 안전의 위협이다. 각종 공격이나 테러의 시도로부터 시민의 안전을 수호하는 것이 미래의 중요 과제가 될 것으로 보인다. 모든 것이 망으로 연결되고 기계로 처리된다는 것은 해킹의 가능성을 의미하는 것이기도 하다. 의료, 교통, 통신, 제조, 주택, 금융 등의 분야에서 네트워크에 잠입·조작하려는 해킹 시도가 있을 수 있다. 특히 가스, 전기, 원자력 같은 국가 기반시설에 해킹과 원격 조종이 시도된다면 그 결과는 훨씬 참혹할 수 있다. 아울러 3D 프린팅과 로봇틱스 기술의 발달로 원격 조정할 수 있는 기기의 제작이 훨씬 쉬워졌으며, 화학 물질합성에 인공지능을 활용하여 화학 무기나 위험물을 새롭게 만들어낼 가능성도 있다. 더 광범위한 수준에서는 무인 드론, 군용 로봇 등 군사 목적으로 인공지능 기술이 사용될 가능성도 충분히 있다. ‘AI대부’로 불리는 제프리 힌튼 교수는 자신이 구글에서의 인공지능 개발에 전념하였던 것을 후회한다고 밝히며, 인류안보에 위협이 될 수 있는 AI 킬러로봇, 자율무기의 출현이 우려된다고 말하기도 하였다.¹²⁾

뿐만 아니라, 일부 기업들만 자원을 독차지하게 되면서 불균등이 심화될 수 있다. 데이터를 가진 기술 대기업은 이를 자신들의 지배적 지위를 강화하는데 사용할 것이다. 또한 이들은 그들이 벌어들인 부를 이용해 자신들에게 유리한 입법이 이루어지도록 조직적으로 자원을 동원하거나 로비 활동을 벌일 수도 있다. 이들이 만일 정치세력과 결탁한다면 자본-권력이 융합해 감시받지 않는 특권을 형성할 것이다. 산업의 플랫폼을 지배한 기술 대기업은 적용되는 알고리즘을 그들에게 유리한 방식으로 재편한다. 기업의 결정권을 가지는 최고위층이 알고리즘을 기업에 유리한 방식대로 조작하도록 지시할 수도 있다. 또한 기업들의 상당수 업무를 기계가 맡아 하게 될 것으로 예측되지만, 그들은 그로 인해 발생할 수 있는 책임을 이행하지 않으려 할 수 있다. 조직화되지 못한 개인은 이들 대기업과의 관계에서 정보의 자기결정권을 온전히 행사하지 못하거나 정보 접근의 기회를 실질적으로 제한받게 된다. 이는 내재하는 불합리와 불평등을 더욱 확대하는 결과를 낳을 것이다. 실제로 올해 카카오모빌리티는 공정위로부터 콜택시 배차 알고리즘을 조작해서 가맹택시에 콜을 몰아주었다는 내용으로 250여 억원 상당의 과징금을 부과 받았다.¹³⁾ 인공지능학자인 제리 카플란도 인공지능 기술의 발달에 따른 심각한 문제는 소득 불균형이 될 것이며, 결국 소수의 사람들에게 부가 집중될 것이라고 예측한 바 있다.¹⁴⁾

이처럼 인공지능은 잘못 활용되게 되면 막대한 사회적 악영향을 초래할 수 있으므로 적절한 대응이 필요하다. 우선, 정보의 정확성을 높이기 위한 노력이 요구된다. 인공지능의 착각 현상을 줄이기 위한 기술 개발이 계속되어야 하고, 알고리즘 및 자동화 의사결정에 대해서도 누가 책임을 부담할 것인지에 대한 규범이 마련되어야 한다. 또한 인공지능이 생성한 정보에 대해 그 출처를 표기하도록 하는 조치와 정보의 진위를 확인할 수 있는 분석도구 개발도 필요하다. 더불어, 표현의 자유와 사생활의 자유가 보호되도록 감시 무결성을 확보하여야 한다. 감시적 목적의 기록 수집을 제한하고, 정보의 감시를 확인·조사하는 독립적 기구가 필요하다. 또한, 민주주의 토대가 되는 상호 견제, 언론 자유, 시민 감시 등이 활성화되어야 한다. 뿐만 아니라, 각종 공격의 시도를 막기 위한 보안의 강화와 위협에 대한 예방적 탐지가 필요하다. 특히 다수의 안전과 관계된 기반 시설 방호는 철저히 이루어져야 한다. 마지막으로, 일부의 집단이 부와 자원을 독차지하지 않도록 제한하는 조치가 요구된다. 기업의 무분별한 정보수집을 막기 위하여 개인정보 이용 세부 내역의 공개 및 사용자에게 대한 고지를 의무화해야 하며, 개인정보 보유현황의 주기적 확인이 요구된다. 다른 차원에서는 기업의 커진 권한을 견제하고 필요한 책임을 수행하도록 ‘로봇세’ 및 ‘정보세’를 도입하고, 인공지능의 결정에 대한 책임을 강화해야 한다. 만일 법인에 의해 운용된 로봇이 업무 수행 중 타인에게 손해를 가한다면 그 운용 주체에 대하여 책임을 부과하도록 하며, 경우에 따라서는 결함이 있는 로봇을 제조한 기업에게도 책임을 물을 수 있도록 해야 한다.

정보의 집중화는 특정 세력에 의한 통제를 가져올 수 있는 잠재적 위험요소가 된다. 통제의

위험에서 벗어나는 길은 '개방성'에 있다. 정보가 폐쇄된 사회는 소수의 통제에 취약하지만, 개방된 사회에서는 상호 견제에 의한 투명성과 책임성의 구현이 가능하다. 특정세력이 인공지능을 통해 전제 권력을 행사하는 것을 막기 위해서는 정보의 개방을 통한 개인의 자유 실현이 무엇보다 중요하다. 칼 포퍼는 그의 저서 『열린 사회와 그 적들』에서 “지상천국을 건설하고자 하는 모든 전체주의적 시도는 비록 선한 의도에서 비롯되었다 할지라도 결국 지옥을 만들 뿐이다.”라고 말한다.¹⁵⁾ 밀실에서 벗어나 상호 비판과 감시가 가능할 때에 비로소 오류가 시정되고, 권력은 남용되지 않는다. 우리에게서 전제를 시도하는 소수의 적들로부터 인공 지능을 지켜내고, 진정으로 전체의 이익을 위해 사용되도록 할 책임이 있다.

※ 선정된 에세이는 자료집으로 제작 가능합니다.

- 1) 김미정, 「챗GPT, 두 달만에 월 사용자 1억명 돌파...틱톡보다 빨랐다: UBS 분석 “AI 서비스 대중화 첫 성공 사례될 수도”」, 『지디넷코리아』, 2023.02.03., <https://zdnet.co.kr/view/?no=20230203153950>, 2023.05.18. 접속
- 2) 권희춘, 「[과학기술로 바라본 미래 먹거리] 인공지능의 역사, ‘머신 러닝’에서 ‘챗GPT’까지」, 『한국대학신문』, 2023.03.28. <https://news.unn.net/news/articleView.html?idxno=543967>, 2023.05.18. 접속
- 3) 양종모, 「인공지능 알고리즘의 편향성, 불투명성이 법적 의사결정에 미치는 영향 및 규율 방안」, 『법조』 Vol.66, No.3, 법조협회, 2017, pp.60~105
- 4) 김봉구, 「[시선+] AI도 ‘인종·성차별’ 한다...다른 접근법 필요한 이유」, 『한국경제』, 2019.02.21., <https://www.hankyung.com/it/article/201902200481g>, 2023.05.18. 접속
- 5) ibid
- 6) 박찬, 「[찬이의 IT교실] 이미지 생성 AI ⑥...편향성과 저작권 이슈」, 『Ai타임즈』, 2022.11.25., <https://www.aitimes.com/news/articleView.html?idxno=148030>, 2023.05.18. 접속
- 7) 김희윤, 「[에잇수다] 공포 부르는 AI 목소리의 역습」, 『아시아경제』, 2023.05.17., <https://view.asiae.co.kr/article/2023051708131163828>, 2023.05.18. 접속
- 8) 이광준, 「“AI문제의 본질은 정밀한 오정보의 대량 생산과 유통”, ‘챗GPT와 오정보: 헛소문 바로잡기’ 세미나 열려... 제빈 웨스트 교수 초청 강연」, 『오마이뉴스』, 2023.05.12., https://www.ohmynews.com/NWS_Web/View/at_pg.aspx?CNTN_CD=A0002926825&CMPT_CD=P0010&utm_source=naver&utm_medium=newsearch&utm_campaign=naver_news, 2023.05.18. 접속
- 9) 이광석, 「[이광석의 디지털 이후] ‘조희수 알고리즘’에 좌우되는 사회관계...‘온라인 광장’은 이제 닫혔다」, 『경향신문』, 2021.07.23., <https://www.khan.co.kr/it/it-general/article/202107230600035>, 2023.05.18. 접속
- 10) 김봉수, 「[과학을읽다] 뇌공학에 투입된 AI...당신의 생각을 읽는다」, 『아시아경제』, 2023.05.09., <https://view.asiae.co.kr/article/202305081043321453>, 2023.05.18. 접속
- 11) 류재민, 「“틱톡은 中 공산당 선전 도구였다” 모회사 전 임원 폭로」, 『조선일보』, 2023.05.13., https://www.chosun.com/international/international_general/2023/05/13/SPI2U3537RHJVIWSSUXZIWFY ZU/?utm_source=naver&utm_medium=referral&utm_campaign=naver-news, 2023.05.18. 접속
- 12) 현기호, 「‘딥러닝의 대부’ 제프리 힌튼, AI 위험성 경고 왜?」, 『이코리아』, 2023.05.02., <https://www.ekoreanews.co.kr/news/articleView.html?idxno=66702>, 2023.05.18. 접속
- 13) 반기웅, 「‘배차 알고리즘 조작’ 카카오모빌리티 과징금 257억원」, 『경향신문』, 2023.02.14. <https://www.khan.co.kr/economy/economy-general/article/202302141206001>, 2023.05.18. 접속
- 14) 김용수, 「“AI, 디지털인간 아닌 진화한 자동화 기술”(제리 카플란 스탠퍼드대학교 법정정보학센터 교수 인터뷰)」, 『시사저널이코노미』, 2021.09.02., <http://www.sisajournal-e.com/news/articleView.html?idxno=235585>, 2023.05.18. 접속
- 15) 칼 포퍼, 이한구 역, 『열린사회와 그 적들1』 (민음사, 2006), 카를 포퍼, 이명현 역, 『열린사회와 그 적들2』 (민음사, 1998)