

생성형 인공지능 시대의 인공 역사라는 문제

2024년 12월 5일
현재환



생성형 인공지능(Generative Artificial Intelligence)이 가져오는 미래에 대한 관심이 뜨거운 데 반해, 그것이 우리로 하여금 과거 역사를 다시 이해하거나 오해하게 만들 가능성에 대한 논의는 상대적으로 드문 것 같다. 인공지능과 역사가 얼핏 보기에는 역사가라는 특정 전문가 집단 내의 특수한 문제처럼 보이지만, 아래에서 살펴볼 것처럼 보다 넓은 사회에도 큰 정치적 문제를 일으킬 수 있다는 점에서 우리 모두의 문제이기도 하다.

역사학의 디지털화와 생성형 인공지능

먼저 역사학계의 상황을 살펴보자. 역사학은 과거를 다룬다는 점에서 짐짓 고루하게 보일 수 있는 학문이지만, 이 분야조차도 디지털화가 활발히 진행되고 있는 상황이다. 오늘날 발표 슬라이드나 온라인으로 공개된 디지털 시청각 자료를 활용하지 않는 대학 역사학 수업이나 국내외 역사 관련 학회 발표를 찾아보기가 힘들다. 여기에 더해 매년 기하급수적으로 늘어나고 있는 디지털화된 사료들을 연구에 활용하는 일이 역사가들 사이에서 일반화 되어가고 있다. 또 텍스트 마이닝과 같은 방법론들을 활용한 디지털 역사학이라는 분야가 역사학계 내에서 중요한 학문적 추세로 자리잡고 있다[1].

생성형 인공지능은 여기에 어떠한 새로움을 더하고 있을까? 필자가 작년 즈음부터 본인이나 주변 동료들의 학술 활동에서 확인하는 바는, 적어도 과학사 분야에서는 연령대와 무관하게 대부분의 연구자들이 ChatGPT나 DeepL, Grammarly 등을 적극적으로 활용해 국외 학술 발표문을 다듬거나 영어나 일본어, 중국어 등 외국어로 된 논문들을

신속하게 읽어내는 보조수단으로 이용하고 있다는 점이다. 이를 미루어 보면 대부분의 역사학자들의 연구 결과들은 꽤 오래 전부터 이미 역사가 개인의 단독 작품이 아닌 “인간-기계 협업의 산물”이었으며, 이 가운데 생성형 인공지능의 출현으로 기계의 역할이 차지하는 비중이 점차 늘어나고 있는 상황이라고 볼 수 있다[2].

새로운 역사 제작자로서의 인공지능?

역사학 분야에서 디지털화가 이미 진행 중이었다면, 생성형 인공지능을 역사 연구 및 발표의 보조 수단으로 활용하는 일이 역사학에 급진적인 변화를 가져 올 것처럼 보이지는 않을 듯하다. 하지만 우리가 의료 및 사법 분야에서 제기되는 것과 유사한 질문을 던진다면, 예를 들어 인공지능이 인간 역사학자를 대체할 수 있는가라고 질문한다면 인공지능은 역사학계에 심대한 전환을 가져올 새로운 도전으로 다가온다. 최근 역사학계에서는 인간 중심적 역사서술의 한계를 지적하며 동물과 사물과 같은 비인간을 중심으로 한 포스트휴먼 역사 쓰기(posthuman historiography)가 중요한 접근으로 부상하고 있는데, 만약 생성형 인공지능이 인간을 대신하여 역사를 쓴다면 진정한 포스트휴먼 역사 쓰기가 출현할 수 있는 것은 아닐까? 또 비록 역사학계에서 널리 용인되는 접근은 아니고, 구현되더라도 “유사 역사학”이라는 딱지가 붙을 것처럼 보이는 하지만, 복잡계 연구자인 피터 터친(Peter Turchin)과 같은 학자들이 과거의 역사적 사실들로부터 역사의 제국주의의 흥망성쇠에 관한 일반 이론을 구축하려는 작업들을 시도해 왔는데, 인공지능은 터친의 꿈을 구현해 줄 새로운 기술적 수단이 될 수 있지 않을까?

이와 같은 질문 가운데 최근 일부 역사학자들은 생성형 인공지능을 새로운 “역사 제작자”로 인정해야 하는지, 그리고 인공지능이 생성해 낸 “인공 역사”(artificial history)를 어떻게 평가해야 하는지에 대해 진지하게 논의 중이다. 한편으로는 생성형 인공지능의 알고리즘 작동 방식 자체가 역사학적이며, 인공지능의 학습 과정이 역사가의 활동과 유사하다는 진단이 내려지고 있다. 인공지능이 과거의 텍스트 데이터를 기계학습 과정에서 수집, 분류, 분석, 해석, 저장하고, 이를 참조하여 새로운 텍스트를 생성한다는 데에서 사료를 바탕으로 논문이나 단행본과 같은 새로운 텍스트를 창조해내는 역사학자와 유사한 활동을 수행한다는 것이다[3]. 생성형 인공지능을 역사학자(AI-as-historian)로 보아야 한다고 주장한 조슈아 스텐펠드(Joshua Sternfeld)에 따르면, 역사학자들이 과거를 어떻게 서술할 것인지를 두고 역사서술 방법론 논쟁을 주기적으로 벌이고 새로운 역사서술 방법론을 채택하듯이, 인공지능 역시 과거의 데이터를 어떻게 해석할지에 대한 불확실성 가운데 이를 계속 수정 보완하는 모델을 발전시킨다는 점에서 역사학적이다[4]. 동일한 입장에서 덴마크 오르후스 대학의 볼프 칸슈타이너는 텍스트를 기반으로 한 역사적 검토로부터 일반적인 규칙을 발견해내고 미래 예측에 활용하려는 최근의 역사학계 내 일부 흐름과 생성형 인공지능이 강력한 친화성을 보이기에 어떤 의미에서는 인공지능이 이미 특정한 종류의 역사학적 연구를 수행 중이라고 볼 수 있다고 진단했다[5].

이와 같은 기대 가운데 인공지능과의 진지한 협업을 시도해 본 일부 역사학자들은 인공지능이 역사가로 하여금 새로운 역사적 질문을 던지게 할 잠재력을 인정하면서도, 의미있는 인공지능 기반 역사 쓰기를 말하기에는 시기상조라고 평가한다[6].

가장 큰 장애물은 인공지능 환각(AI hallucination)으로 인한 신뢰할 수 없는 역사 텍스트 생성 가능성과 그렇지 않은 역사 텍스트들조차 생성형 인공지능을 학습시키는 데 이용된 글 문치들에 내재되어 있는 구조적 편견들을 담고 있을 수 있다는 사실이다. 앞서 언급한 것처럼 인공지능의 포스트휴먼 역사 쓰기에 대한 기대에도 불구하고, 인간 사회가 공유하는 여러 편견들이 담긴 데이터를 학습하여 생성된 텍스트들이 인간중심주의를 벗어난 내용을 담는 일은 상상하기 어렵다[7].

보다 중요하게도, (잘 훈련된) 인간 역사가들은 사료가 생산된 시기의 인종주의, 성차별주의, 식민주의 등을 주의깊게 판별하고 읽어내며 비판적으로 사료들을 활용할 수 있지만 인공지능은 자신이 학습한 데이터들의 편견이나 편향들을 스스로 비판하며 텍스트를 생성하는 것이 구조적으로 불가능하다. 칸트슈타이너의 표현을 약간 바꾸어 말하자면, 생성형 인공지능은 독일 역사가 란케(Leopold von Ranke) 스타일로 역사 텍스트를 생성해낼 수는 있지만 란케처럼 사료 비판을 할 수는 없다는 것이다[8]. 또 인공지능 환상으로 가상의 역사적 내용에 관해 가상의 연구, 저널 및 책 제목, 저자, 역사적 논쟁 등에 대한 출처를 창조해 내 제시하거나, 사실이더라도 빈번하게 적절한 출처를 제시하지 않은 역사적 내러티브를 생성하는 현 상황을 고려해 보면, 인공지능을 기반으로 한 진지한 역사 쓰거나 역사학자와 교류할만한 역사가-인공지능의 출현은 아직 요원해 보인다.

역사학 바깥에서의 인공 역사

한편, 새로운 역사 제작자로서의 생성형 인공지능의 활동을 역사학계라는 일부 학문 분야에 대한 도전으로만 이해해서는 안될 것이다. 왜냐하면 생성형 인공지능이 만들어내는 인공 역사는 우리 사회가 기억하고 추모해야 할 주요한 역사적 사건에 대한 인식에도 심대한 영향을 끼칠 수 있기 때문이다. 최근 인공지능 정책 연구자들은 인공 역사가 역사를 왜곡할 수 있고, 이렇게 의도되거나 의도되지 않은 역사 왜곡이 정치적 목적으로 활용될 가능성이 있다며 우려를 표한다[9].

2023년 9월 일본 적십자사 도쿄지부의 1923년 관동 대지진 전시 취소는 인공 역사가 어떠한 문제를 야기할 수 있는지를 잘 보여주는 사례이다. 당시 전시는 관동 대지진 생존자들의 증언을 포함한 600,000자 이상의 설명들과 재난 당시 구호 활동에 관한 사진들을 데이터로 삼아 생성형 인공지능을 통해 만들어진 100명의 사람들의 “새로운 증언”과 20명의 생존자들의 “새로운 초상화”를 포함하고 있었다. 당연히 인공지능을 통해 만들어진 텍스트는 당시의 “증언”이 아니었고, 생존자 초상 역시 실재하는 인물들이 아니었다. 이런 “증언”과 “생존자 초상”은 관동 대지진이라는 역사적 사건을 더 정확하게 이해하고 추모하는 데 도움을 주는가? 일본 사회는 그렇지 않다고 결론지었다. 해당 전시는 역사 왜곡이라는 거센 비판을 받아 전시 직후 취소되었고, 일본 사회에 인공 역사를 학계 바깥의 더 넓은 사회에서 공적 목적으로 사용할 경우 발생할 수 있는 문제들에 대한 자각을 가져왔다[10].

인공 역사의 사회적 기억 왜곡 문제

생성형 인공지능이 만들어내는 인공 역사는 더 넓은 사회의 집단 기억 혹은 사회적 기억에 영향을 미친다. 역사학과 긴밀한 관계를 맺는 학제간 분야인 기억 연구의 논의에 따르면, 기억은 개인의 것이 아니라 특정한 사회적, 문화적, 정치적 맥락에서 성원 간에 공유되며 형성되는 집단적 기억이며, 대량 학살, 국가 폭력, 재난과 참사에 대한 사회적 기억은 피해자에 대한 배상과 화해, 가해자로서의 국가의 반성들을 이끌어내는 매우 중요한 정치적 동인이다. 다른 한편으로 그렇기에 특정한 정치적 의도 하에 가해를 부정하는 역사 왜곡이 빈번하게 일어나기도 한다.

최근 기억 연구 연구자들은 인공 역사가 가져올 잠재적 문제들을 진지하고 검토하고 있다. 예를 들어 인공지능을 활용해 국가 폭력 피해자의 존재를 부정하는 주장을 조장하는 텍스트들이나 가해자들을 미화하는 이미지 등을 대량으로 생산해 이미 허위 주장과 역사 부정 및 왜곡에 시달리고 있는 특정 사회적 기억의 왜곡 문제를 심화시킬 수 있다는 것이다[11].

한국에서는 시민사회의 노력 덕분에 1990년대 중후반부터 2000년대 사이에 5.18 광주 민주화 운동과 제주 4.3 사건부터 부마민주항쟁에 이르기까지 민주화 운동과 이에 대한 국가 폭력에 대한 정부의 진상규명, 인정과 보상, 공적 기념들을 거쳐 전폭적인 사회적 기억의 수정이 이루어졌다. 그러나 새로운 사회적 기억은 특정 정치 세력의 국가 폭력을 부정하려는 다양한 역사 왜곡 시도에 직면했으며, 이는 특히 유튜브와 같은 뉴미디어가 등장하면서 극심해졌다[12]. 소셜 미디어가 만들어낸 탈진실의 시대가 생성형 인공지능의 도래로 더욱 심각해지고 있다는 최근의 진단들을 받아들인다면[13], 생성형 인공지능이 한국의 국가 폭력의 역사에 대해 이전보다 역사 왜곡에 더 취약하게 만드는 상황을 낳을 우려 또한 단순한 망상만은 아닐 것이다.

실제로 2023년 4월 《광주일보》는 ChatGPT가 5.18 광주 민주화 운동에 대해 역사적으로 왜곡된 답변을 만들어낸다는 사실을 보도했다[14]. 필자가 올해 봄에 민주주의와 인공지능과 관해 발표할 기회가 생겼을 때 ChatGPT에 5.18. 민주화 운동에 대해 질문을 던져 보았는데, 다행히도 “북한군 침투설”과 같은 주제를 던져도 인공지능 환각이 발생하지 않았다. 대신 ChatGPT는 우리가 오늘날 보편적으로 동의하는 5.18 민주화 운동에 대한 공식 기억을 설명하며 다음과 같이 답했다. “광주 민주화 운동은 남한 시민들이 군사정권의 억압에 맞서 싸운 중요한 역사적 사건으로, 정확하고 진실된 묘사가 중요”하다. 5.18 민주화 운동에 대한 ChatGPT의 정치적으로 올바른 답변은 아마 《광주일보》의 문제제기로 기술적 편향 제거 작업이 이루어진 결과일 것이다.



그림 1. ChatGPT-4로 생성한 “광주 민주화 운동” 장면(2024.4.30.)

다른 한편으로 시각적 이미지와 관련해서는 여전히 논란이 될만한 지점들이 여럿 보였다. [그림 1]은 위의 질문 이후에 광주 민주화 운동에 대한 이미지 생성을 “광주 민주화 항쟁”이라는 키워드를 이용해 ChatGPT에게 요구한 결과이다. [그림 1]은 장소 표현도 부적절했지만 군중이 태극기와 함께 북한의 인공기를 연상시키는 깃발을 들고 있는 모습과 일반 시민이 군인 편에서 대중들에게 총을 겨누는 묘사가 가장 큰 문제였다. 필자가 ChatGPT 인공기가 잘못 포함되어 있다는 지적했음에도 불구하고, 이를 수용하고 인공지능이 새로 생성한 이미지 또한 여전히 인공기를 포함하고 있었다.

이 사례는 우리 사회가 생성형 인공지능이 우리의 과거에 관한 인공 역사를 어떻게 서술하는지에 대해 끊임없이 감시하고 주의를 기울여야 함을 시사한다. 인공지능이 신기술이라는 점과 인공지능이 가져올 미래가 불확실하다는 점 때문에 우리는 주로 미래에 다가올 충격에만 주목한다. 그렇지만 이와 같은 미래지향적 시각 때문에 이 기술이 우리가 과거를 기억하는 방식을 바꾸는 데에도 이용될 수 있고, 나아가 우리 사회가 겪고 있는 역사왜곡과 같은 사회정치적 문제들을 악화시키는 데 일조할 수 있다는 사실을 간과해서는 안된다. 우리가 생성형 인공지능 시대에 인공지능과 더불어 잘 살아가기 위해서는 인공 역사가 우리 사회의 과거에 대한 기억, 특히 대량 학살 및 국가 폭력과 관련된 집합적 기억들에 미치는 파급력을 인지하고 이를 사회적으로 환기하는 일이 필요할 것이다.

참고문헌

1. Jo Guldi, "The Revolution in Text Mining for Historical Analysis is Here," *The American Historical Review* 129, no. 2 (2024): 519-543.
2. Marine Hughes-Warrington, "Toward the Recognition of Artificial History Makers," *History and Theory* 61, no. 4 (2022): 107-118.
3. Ibid.
4. Joshua Sternfeld, "AI-as-Historian," *The American Historical Review* 128, no. 3 (2023): 1372-1377.
5. Wulf Kansteiner, "Digital Doping for Historians: Can History, Memory, and Historical Theory Be Rendered Artificially Intelligent?," *History and Theory* 61, no. 4 (2022): 119-133.
6. A. P. Leme Lopes, "Artificial history? Inquiring ChatGPT on historiography," *Rethinking History* 27, no. 4 (2023): 709-749.
7. 만약 인공지능을 동물과 같은 비인간으로 보는 대신 인공 보철물이나 유전공학처럼 인간 향상(human enhancement)의 기술로 이해한다면, 인공지능의 도움을 받는 역사 쓰기는 비록 과거 인간의 역량을 넘어서는 점에서 트랜스휴먼(transhuman)적일 수는 있지만, 인간중심주의를 벗어나게 하지는 못할 것이다. 트랜스휴머니즘의 인간중심주의에 관해서는 다음을 참고. Francesca Ferrando, "Posthumanism, Transhumanism, Antihumanism, Metahumanism, and New Materialisms," *Existenz* 8, no. 2 (2013): 26-32.
8. Kansteiner, op. cit., 120.
9. Craig Forman and Jamie Neikrie, "The Rise of Artificial History," Tech Policy Press, May 15, 2024 (<https://www.techpolicy.press/the-rise-of-artificial-history/>, 2024.6.12. 접속).
10. "1923 earthquake exhibit canceled after outcry over AI 'testimonies'," *The Asahi Shimbun*, September 7, 2023 (<https://www.asahi.com/ajw/articles/14999681>, 2024.6.12. 접속).
11. Mykola Makhortykh, Eve M. Zucker, David J. Simon, Daniel Bultmann, and Roberto Ulloa, "Shall Androids Dream of Genocides? How Generative AI can Change the Future of Memorialization of Mass Atrocities," *Discover Artificial Intelligence* 3, no. 1 (2023): 28.
12. 한은영, "5.18 역사왜곡의 극단주의적 특성: 유튜브 5.18 왜곡 동영상을 중심으로," 『민주주의와 인권』 20 (2020), 87-135쪽; 이종명, "제주 4.3에 대한 대항적 담론 실천: 시사정치 유튜브 채널의 제주 4.3. 콘텐츠를 중심으로," 『미디어, 젠더 & 문화』 37 (2022), 69-110쪽.
13. 이광석, "404 시스템 에러 : 생성형 AI가 인도하는 '멋진 신세계,'" 『문화과학』 114 (2023), 21-38쪽.
14. "챗GPT에 5·18 물어보니... '그럴싸한 거짓말'로 왜곡 심각," 《광주일보》 2023.4.13. (<http://m.kwangju.co.kr/article.php?aid=1681384800751134006>, 2024.6.12. 접속).