

## 제16장 토론문

### 서론

발제자는 딥러닝과 생성형 AI가 지도 객체 탐지, 지도 일반화, 그타일 변환 등 카토그래피의 기술적 효율성을 극대화하고 있음을 정확히 짚어주었다. 그러나 발제문 말미에 제시한 세 가지 토론거리처럼, 기술적 진보는 신뢰, 소유권, 그리고 인간의 역할에 대한 의문점을 제시하기도 한다. 지도는 단순한 지표면의 객관적 모사가 아니라 권력과 이데올로기가 투영된 사회적 텍스트라는 인문지리학적 관점에서 발제자의 세 가지 토론거리에 차례로 답하고자 한다.

앞서 언급한대로, 과거의 지도 역시 권력과 이데올로기의 산물이었기에, 이러한 관점에서 AI가 생성한 지도 또한 과거의 한계와 다를 것이 없어 보일 수 있다. 그러나 AI가 생성한 지도는 과거의 지도들이 가지지 않았던 추가적인 문제점들을 안고 있다. 과거의 지도는 뚜렷한 제작 주체가 존재하여 그 정치적 편향성을 역으로 해체할 수 있었던 반면, AI가 생성한 지도는 데이터와 알고리즘을 활용해 만들어져, 대중이 이것을 무비판적으로 기술적 객관성이 담보된 것으로 수용할 수 있다. 특히, AI의 블랙박스 특성은 환각 현상이 발생하거나 데이터 편향으로 지도가 왜곡되었을 때, 그 책임을 물을 주체를 알 수 없다. 따라서 AI 지도에 대한 신뢰성을 담보하기 위해서는 모델의 공간적 의사결정 과정을 역추적하고 해석할 수 있는 XAI 기술을 활용해야 한다. 더불어, 모델이 학습한 데이터와 알고리즘을 명확히 밝히고, 특히 데이터에 대해서는 메타데이터를 공개해야 한다.

AI는 오픈스트리트맵과 같은 클라우드소싱 데이터나 인류가 역사 속에서 축적해온 지리 정보를 학습하여 새로운 지도를 생성한다. 만약 빅테크 기업들이 AI를 통해 추출한 공간 데이터를 사유화하고 지도를 독점적으로 상품화한다면, 이것은 공공재를 사용해 만든 공간 지식의 사유화로 이어지게 될 것이다. 따라서 AI를 활용해 만들어진 지도는 공유재로 규정해야 한다. 국가 안보나 개인의 위치 정보가 노출될 수 있는 정보는 비공개로 보호하되, AI가 공공의 데이터를 학습해 도출한 결과물과 그 알고리즘의 원리는 원칙적으로 오픈소스로 공개해야 한다는 기준을 세워야 한다.

AI의 도입으로 인해 인간 카토그래퍼는 직접 지도를 만드는 작성자에서 AI 모델과 그 결과를 통제하는 관리자로 진화하고 있다. 제16장의 본문에서도 인간 참여형 모델의 중요성을 강조하고 있다(p. 232). AI의 역할은 복잡한 데이터에서 패턴을 찾아내고, 스타일을 변환하거나, 객체를 빠르게 일반화하는 기술적 노동에 머물러야 한다. 반면, 인간 카토그래퍼는 AI가 도출해낸 지도가 사회적·윤리적으로 올바른지 비판적으로 해석하는 데 있다.

지도에 어떤 지명을 표기할 것인지, 소외 계층의 공간이 지도에서 생략되지는 않았는지 감시하고 교정하는 윤리적 판단과 개입은 인간만이 할 수 있는 영역으로 남을 것이라 생각된다.

AI는 지도를 그리는 방식을 바꿀 수는 있지만, 지도를 통해 세상을 어떻게 해석할 것인가에 대한 방법은 대체할 수 없다. 향후 지리학계는 AI라는 도구를 수용하되, 지도 제작이 본질적으로 정치적이고 사회적인 행위라는 지각을 잃어서는 안 된다.

## 제20장 토론문

발제자는 일반적인 LLM이 '한성부-경성부-서울'의 시계열적 변화를 이해하지 못하는 문제를 지적하고, 이를 온톨로지 기반의 지식그래프로 보완하는 대안을 제시했다. 지식그래프를 LLM과 연동하는 방식은 AI의 오작동을 제어하기에 적합한 시스템 아키텍처라고 생각된다. 그러나 역사지리 데이터는 아무리 정교한 온톨로지를 구축한다고 해도, 근본적으로 데이터의 희소성이라는 한계에 직면하게 된다.

제20장 원문에서는 인구 밀도나 중위소득이 낮은 지역에서 생성형 AI의 지역 특화된 답변을 얻기 어렵다는 것을 밝혀낸 연구를 소개했다(p. 292). 이것은 데이터가 인구 밀도가 높은 지역이나 부유한 지역에 편향되어 있기 때문일 가능성이 크다. 이와 마찬가지로 현대의 공간 정보는 거리뷰 이미지나 오픈스트리트맵과 같이 방대한 데이터를 통해 연구되지만, 역사지리 데이터는 문헌의 양이 절대적으로 부족할 뿐만 아니라, 문헌이나 데이터가 존재하지 않는 공백의 공간 역시 존재한다. 결국 LLM은 위 연구 사례와 같이 이렇게 데이터가 희소한 맹점(blind spot)에 마주치면 통계적 확률에 의존해 결과를 만들어내는 환각을 일으키거나 해당 공간 자체를 생략하는 오류를 범하게 될 것이다.

현재 AI 개발에도 유사하게 데이터의 부족을 겪고 있다. 생성형 AI 훈련에 활용할 텍스트 데이터가 부족한 것이다. 이러한 문제를 해결하기 위해 해당 업계에서는 생성형 AI를 역이용하여 가상의 합성 데이터를 만들어내 모델을 다시 훈련시키는 방법을 사용하기도 한다. 그러나 역사지리 시스템에는 이러한 방법을 적용할 수 없다. 부족한 과거의 공간 기록을 메우기 위해 AI가 확률적으로 생성한 훈련 데이터를 지식그래프에 주입할 경우, 실제 역사적 사실과 기계가 생성한 가상의 과거가 섞이는 데이터 오염이 발생하기 때문이다. 특히, 발제자가 제시한 정확한 지리적 정보를 제공하기 위해 만들어지는 지식 그래프에는 이러한 정보가 절대로 들어가지는 안 된다.

발제자가 제안한 온톨로지 및 지식그래프 구축은 LLM의 오류를 통제하고, 역사적 시공간의 변화를 AI에 담아낼 수 있는 훌륭한 방법이라 생각된다. 그러나 그 내용을 채워넣을 역사지리 데이터 자체가 절대적으로 부족한 상황에서, 단순히 양적인 데이터 부족을 해결하기 위해 합성 데이터와 같은 기술을 차용하는 것은 데이터 오염만을 초래하게 될 것이다. 따라서 역사지리학에서 생성형 AI를 도입할 때는 과거의 모든 시공간을 망라하

려는 노력에서 벗어나야 한다. 사료가 부재한 공백의 공간을 통계적 기법으로 메워 통합적으로 개발하려고 하기 보다는, 각 시간 및 공간적 범위에 특화된 생성형 AI를 개별적으로 개발하는 것이 오히려 적절할 것이라 생각된다.