역사학자가 前代未聞을 다룰 때

유발 하라리, 『넥서스(Nexus)』(2024, 김영사) 독서 토론

조병식

01 방법론적 한계

이 책의 가장 핵심적인 긴장은, 역사학자가 역사적 경험이 전무한 사태를 역사적 방법으로 경고하려 한다는 데 있다. 이를 이해하는 데 유용한 분석틀은 AI의 두 단계를 구분하는 것이다. 1단계 AI는 인간이 목표를 설정하고 AI가 그것을 실행하는 구조, 즉 도구로서의 AI이며, 2단계는 AI가 목표 자체를 스스로 설정하는 단계다. 하라리가 실제로 분석하는 사례들 — 페이스북 알고리즘, 이란의 히잡 단속 AI, NSA 스카이넷, 사회 신용 시스템 — 은 1단계에 해당한다. 그리고 이러한 비유기적 네트워크를 이해하기 위한 전제로서 1부에서 유기적 네트워크 ,이야기·신화·문서·관료제 등의 계보가 제시되었다.

반면 하라리가 수사적으로 경고하는 것은 2단계다. 페이퍼클립 최대화 하기, "독립적인 행위자(independent agents)", "새로운 종류의 신(new kinds of gods)" 같은 표현들이 그러한 의도를 단적으로 나타낸다. 2단계가 현실화되면 인간의 자정 장치 자체가 붕괴한다. 역사학의 방법은 인간이 과거의 오류를 자정할 수 있다는 전제 위에서 유효한데, 정작 경고가 가장 시의성을 가질 수 있는 순간은 역사학적 방법이 가장 무력한 순간이다.

저자는 이 긴장을 책 안에서 충분히 명시하지 않았고, 또 실제 내용은 1단계의 상황에 대한 분석과 예측이 주를 이루고 있지만, 필요에 따라 두 가지 단계를 혼용하고 있다. 다만 저자의 입장에서 생각해 본다면 2단계에 대한 경고는 ai의 위험을 강조하기 위한 수사적 장치로서, 실제로는 역사학자로서 다룰 수 있는 범위의 경계까지만 다룬 것으로 읽힌다.

02 비역사적 접근과 현재중심주의

『넥서스』의 전반부가 역사를 다루는 방식에는 방법론적으로 더 근본적인 문제가 있다. 하라리는 현재의 개념과 가치를 보편적인 것으로 전제하고 과거 전체에 소급 적용한다. 가장 두드러진 사례가 전체주의와 전제주의의 혼용이다. 파시즘, 나치즘, 스탈린주의 등으로 대표되는 전체주의는 대중정당, 대중매체, 산업사회 등을 전제조건으로 하는 근대적 개념이다. 반면 전제주의는 군주나 소수의 자의적 권력으로, 통제 범위가 정치·군사·조세에 한정되는 고대부터 적용 가능한 개념이다. 하라리는, 비록 두 가지를 구분하고 있으면서도, 정작 이 둘을 뒤섞어 로마, 진과 같은 고대 제국에서부터 AI 시대까지 연속적인 "전체주의의 정도"를 이야기한다.

이 문제는 단순한 개념 혼용을 넘어선다. 현재의 기준으로 보면 고대 제국의 통제는 느슨해서 낮은 점수를 받는다. 그러나 당시의 관점에서 보면 그것은 당대가 동원할 수 있는 수단과 자원의 최대치였다. "얼마나 전체주의적인가"는 절대적 척도가 아니라 그 사회의 기술적·제도적 조건을 따져 평가되어야 한다. 역사학은 현재의 기준과 가치, 개념을 가지고 과거를 재단하는 것이 아니라, 당대의 상황과 조건 속에서의 선택과 맥락을 이해하고 의미를 파악하는 학문이다. 하라리의 이러한 접근법은 역사학의 본령에서 벗어난다.

원인과 결과의 도치도 같은 맥락에서 발생한다. 하라리는 고대 제국에 민주주의가 없었던 사실을 설명해야 할 문제처럼 다룬다. 그러나 대중매체 없이 대규모 민주주의가 불가능했다면, 그런 수단이 없었던 고대 제국에서 민주주의가 등장하지 않은 것은 당연한 결과다. "왜 민주주의가 없었나"라고 묻는 것은 원인과 결과를 도치한 것이다. 올바른 질문은 "어떤 조건이 민주적 정보 네트워크를 불가능하게 했는가", "왜 민주주의가 불완전했는가"이다. 역사가 분석의 대상이 아니라 현재 주장의 근거 자료로 동원될 때 발생하는 전형적인 오류다.

03 민주주의 이상화

하라리는 현실의 민주주의가 아니라 이상적 민주주의를 분석의 준거점으로 삼는다. 더 구체적으로는 미국의 자유민주주의 체제가 그 이상형으로 기능한다. 이것은 그가 동원하는 구체적 사례들에서 드러난다. 알고리즘의 불투명성을 논하기 위해 등장하는 것은 COMPAS, 자정 장치의 붕괴를 보여주는 2021년 1월 6일 의회 습격을 제시한다. 무엇보다 하라리가 민주주의의 자정 메커니즘의 원형이자 준거로서 반복해서 제시하는 것은 미국 헌법이었다.

하라리가 그리는 민주주의는 이상적 민주주의다. 이상적 민주주의에서 포퓰리즘은 내부에서 생성될 수 없으므로, 하라리는 포퓰리즘을 민주주의 외부에서 침투한 위협으로 그린다. 그러나 포퓰리즘을 외부화하면 민주주의는 피해자가 되고, 민주주의 자체의 구조적 실패는 면책되고 포퓰리즘이 탄생하는 민주주의의 내재적 문제들이 간과된다. 이상화된 민주주의를 기준으로 삼는 것이, 현실 민주주의의 내생적 문제를 외생적 위협으로 치환하는 구조를 만들어낸다.

또한 이 논리 전체가 민주주의 자정 능력이라는 전제 위에 서 있다. 하라리가 민주주의의 핵심 장점이자 AI에 대한 중요한 대응책으로 내세우는 자정 능력이 실제로 작동했다면 포퓰리즘이 탄생할 수 있었을까? 포퓰리즘은 시민의 지지와 선거라는 제도 속에서 성장했다. 자정 장치가 제대로 작동했다면 포퓰리즘이 등장하는 어느 단계에서 교정되었어야 했다. 현실은 그렇지 않았다. 결국 하라리가 기대하는 민주주의 자정 장치란 근거 없는 낙관 혹은 이상에 가까워 보인다.

그리고 민주주의 이상화는 세계를 자유민주주의와 전체주의 두 범주로만 분할하는 이분법과 결합한다. 사회민주주의, 유럽식 의회 민주주의, 자유민주주의 외형 속의 일당 지배, 일본의 55년 체제나 싱가포르 인민행동당의 독점 지배는 선거와 법치의 외형을 갖추면서도 실질적 정치 경쟁을 구조적으로 제한한다. 이 대안 논리는 책에서 실질적으로 다루어지지 않는다. 민주주의의 위기와 극복을 이야기하면서 민주주의 체제의 다양성을 고려하지 않는 분석은, 그 처방이 가장 필요한 사회들의 실제 조건을 포착하지 못한다.

04 중국 실종과 이중 잣대

10장에서 AI가 전체주의 정권에 딜레마를 안길 것이라고 주장한다. AI를 신뢰하면 기술의 꼭두각시가 되고, 인간의 기관을 만들어 AI를 감독하게 되면 자신의 권력이 제한될 것이라는 논리이다. 그런데 이 예상을 반박하는 현실 사례의 분석이 빠져 있다. 바로 중국이다. 중국은 사회신용체제, 안면인식 인프라가 이미 작동 중이고 AI는 당의 목표를 실행하는 도구로 제도적으로 종속되어 있다. 하라리의 예측과 달리 현실에서 전체주의에 가장 근접한 체제가 AI를 효과적으로 통제하며 운용하고 있다. 즉, 중국은 딜레마를 해소해 가는 케이스로서 하라리의 논리에 대한 정확한 반증이다.

분석 기준의 이중성도 지적할 필요가 있다. 민주주의 위협을 다룰 때 하라리는 일관되게 1단계(도구로서의 AI)의 프레임을 사용한다. 봇을 배치하는 외국 정부, 알고리즘을 설계하는 기업, 그 인프라를 활용하는 포퓰리스트 — 언제나 인간 행위자가 배후에 있다. 반면 권위주의 위협을 다룰 때는 2단계에 가까운 서술이 등장한다. 반체제 봇이 "자율적으로" 정권을 비판하고, 알고리즘이 독재자로부터 권력을 "빼앗는다". 1단계와 2단계라는 구분이 분석 도구가 아니라 상황에 따라 편의적으로 사용되는 수사적 도구가 되는 것이다.

05 설득 텍스트의 한계

에필로그에서 하라리는 두 가지 서로 다른 AI 인식을 나란히 제시한다. 하나는 "새로운 정보 기술의 발명은 항상 중대한 역사적 변화의 촉매가 된다"는 것 — AI를 도구로 정의한다. 다른 하나는 "AI는 주체성을 지닌 우리 정보 네트워크의 정식 구성원이다"는 것 — AI를 행위자로 정의한다. 이 이중성은 수사적 장치로서 기능한다. 도구로서의 AI는 역사적 유비를 제공하여 기존의 이론 틀을 유지하고, 주체로서의 AI는 경고의 긴박성을 확보한다. 두 인식을 동시에 작동시키면서 역사적 근거와 실존적 공포 양쪽을 모두 동원한다.

책 전체의 구조는 순환 논리로 되어 있다. 호모 사피엔스는 허구적 이야기를 공유하

며 협력한다는 전제에서 출발하여, AI가 정보 네트워크를 왜곡하고 집중화한다는 진단을 거쳐, 자정 장치를 갖춘 민주적 제도를 구축해야 한다는 해법에 도달한다. 그리고 그 해법의 근거는 다시 호모 사피엔스의 협력 능력이다. 점토판에서 실리콘 제국까지의 방대한 여정이 출발점의 전제를 확인하고 강화하는 구조이며, 분석이 결론을 생산한 것이 아니라 결론이 처음부터 전제 안에 있었다.

또한 하라리가 말하는 협력에는 대상이 명확하지 않다. 협력을 이야기할 때 암묵적 주어는 자유민주주의 국가들일 것이다. 그러나 러시아, 중국과 같이 AI 기술 경쟁의 핵심 행위자이자 비 민주주의 행위자는 그 협력의 틀 안에서 찾아보기 어렵다. 이들을 배제하게 되면 협력이라는 처방 자체의 실효성이 흔들린다.

총평

『넥서스』는 분석 텍스트와 설득 텍스트의 두 얼굴을 가진 책이다. 설득 텍스트로서 이 책은 AI에 대한 대중적 불안을 역사적 경험과 언어로 풀어내는 데 성공했다. 분석 텍스트로서는 AI 개념의 이중성, 1단계와 2단계의 미분리, 비역사적 접근과 현재중심주의, 민주주의의 이상화, 분석 기준의 자의성, 자정 능력에 대한 근거없는 낙관이라는 한계가 드러난다. 그러나 가장 정직한 평가는, 이 책이 역사학자로서 할 수 있는 것 — 과거에서 패턴을 식별하고, 역사적 맥락에서 현재의 위험을 설명하고, 선택 가능성을 논증하는 것 — 을 했고 그 경계에서 멈춘 것이라는 독해일 것이다. 역사학이라는 방법론으로 AI 시대에 대해 말할 수 있는 것과 말할 수 없는 것을 이 책이 동시에 보여준다는 점에서, 『넥서스』는 그 자체로 하나의 방법론적 사례 연구이기도 하다.