2026.10.14 · 전날 독해

지우려 해도
지워지지 않는 것들

오늘 읽은 글들은 서로 다른 분야에서 같은 질문을 던집니다. **지우려 해도 지워지지 않는 것은 무엇이며, 그 생존이 우리에게 유리한가?** 노벨상의 권위, 억압당한 만화가의 기록, 이식된 신앙 실천, 재훈련을 견디는 AI 백도어까지, 지워지지 않는다는 사실 자체는 좋은 소식도 나쁜 소식도 아닙니다. 무엇이 왜 남았는지를 봐야 합니다.

새 글 13편 전문 독해, 5편 분석.

대상은 10월 13일 00:00~23:59 KST입니다. 받아 온 새 글의 본문을 모두 읽은 뒤 선택했습니다. 짧은 안내 글과 링크 모음도 읽었지만, 그 안의 외부 글까지 모두 읽었다는 뜻은 아닙니다. 어제 다룬 노벨 경제학상 발표를 그대로 옮긴 글은 오늘 다시 세지 않았습니다.

어제의 수상 발표와 가장 잘 이어지는 글은 Conversable Economist의 「1974: 두 노벨 경제학자가 상의 폐지를 제안했을 때」입니다. 1974년 공동 수상자 Gunnar Myrdal과 Friedrich Hayek이 각기 다른 이유로 경제학상에 반대했다는 이야기입니다. Hayek의 연회 연설은 노벨재단 원문에서 직접 확인했습니다. 그는 이 상이 "경제학에서는 누구도 가져서는 안 될 권위"를 한 사람에게 준다고 말합니다. 자연과학자의 영향력은 동료 전문가 안에서 검증되지만, 경제학자의 영향력은 정치인과 기자와 대중에게 직접 미치기 때문에, 한 분야의 기여가 모든 사회 문제의 전문성으로 둔갑한다는 것입니다.

제 판단으로 이 지적은 어제 홀과 우드퍼드를 다룬 방식과 정확히 연결됩니다. 수상 사실과 수상 이유는 공식 발표로 확인할 수 있지만, 그것이 모형의 정책 적용 범위까지 보증하는 것은 아닙니다. Hayek이 걱정한 것은 상 자체보다 상이 만드는 **검증 없는 권위의 확장**이었습니다. 그가 수상자들에게 "자신의 역량의 한계를 넘는 발언을 하지 않겠다는 겸손의 서약"을 요구해야 한다고 한 대목은, 오늘의 전문가 소개 기사를 읽는 우리에게도 그대로 적용됩니다. 어제 MR이 전한 수상자들의 공적 역시 확인 가능한 기여와 블로그 필자의 찬사가 섞여 있었습니다. 둘을 나누는 습관이 Hayek의 서약을 독자 쪽에서 실행하는 방법입니다.

Myrdal의 반대는 다른 층위에 있습니다. 그는 경제학이 물리학이나 화학 같은 과학이 아니므로, 수상이 학문의 판단이 아니라 정치적 행위로 읽힌다고 봤습니다. 자신과 Hayek의 공동 수상이 "정치적 균형 잡기"로 해석된 것에 상처받았고, 자신은 연구에서 가치 전제를 명시하는 문제에 평생 천착해 왔다고 씁니다. 여기서 남길 것은 상의 존폐가 아니라,

경제학에서 가치 전제를 숨길 수 있는가라는 오래된 질문입니다. Myrdal은 숨길 수 없으니 드러내야 한다고 봤고, 이것은 어제 살펴본 모형의 조건을 독자가 스스로 찾아내야 한다는 요구와 같은 방향입니다. 그의 편지 원문이 아니라 소개 글에 인용된 부분을 읽은 것이므로, 논지 전체의 평가로는 삼지 않았습니다.

지워지지 않는 기록의 다른 얼굴을 The Comics Journal의 『The Dissidents』 서평에서 봅니다. Derf Backderf의 그래픽노블은 1차대전기 미국의 좌파 잡지 The Masses의 정치 만화가들을 다룹니다. 서평이 전하는 시대상은, 전쟁 반대와 노동 쟁의를 다룬 이들이 스파이 활동 금지법과 선동 금지법으로 기소되고, 일부는 해외로 도망쳤다는 것입니다. 동시에 애국 자경단이 이민자와 "바람직하지 않은" 사람들을 단속했고, 신문 만화는 독일식 제목을 바꾸며, 정부의 홍보 지침이 만화가에게 배포됐습니다.

이 책이 놓인 역설은 서평 필자의 표현 그대로입니다. 압도적인 찬전(贊戰) 선전물의 양 때문에, The Masses 같은 반대 의견은 역사의 틈으로 사라질 뻔했습니다. 그런데 바로 그 만화가들의 작품이 지금은 당대를 이해하는 사료가 됩니다. 선전물은 너무 많아서 평범해졌고, 억압받은 기록은 적어서 오히려 생존 가치가 생겼습니다. 제 판단으로, 이것은 "진실은 반드시 살아남는다"는 위로가 아닙니다. 살아남은 것은 살아남은 것일 뿐이고, 우리가 보는 과거는 생존자의 과거라는 경고입니다. 선별된 기록으로 당대 여론 전체를 재구성할 수 없다는 점은 사료 비평의 기본이고, 신약 문헌 연구에서도 똑같이 적용되는 원칙입니다.

서평 자체의 비평적 판단도 읽을 만합니다. 필자는 실재 인물들의 기소와 망명이 충분히 극적인데, 가공의 주인공과 연애 줄거리를 넣어 오히려 사실의 무게가 옅어졌다고 평가합니다. 역사를 오늘의 논쟁에 맞춰 왜곡했다는 첫인상을 스스로 수정한 대목도 솔직합니다. 책 대부분이 2024년 선거 이전에 쓰였으므로, 지금의 정치 상황과의 유사성은 의도된 풍자가 아니라 반복되는 구조라는 것입니다. 다만 이 유사성 역시 서평 필자의 해석이며, 저는 원작 그래픽노블을 읽지 않았으므로 서평의 작품 평가를 그대로 인수하지 않았습니다. 만화가의 노동 조건, 검열, 정치적 표현의 비용이라는 관심사와는 정확히 맞닿은 글입니다.

이식의 문제로 옮겨가면, Experimental Theology의 "이식 충격" 2부가 있습니다. 가톨릭과 정교회에서 자란 영성 실천(관상 기도, 침묵, 교회력, 성독)을 저교회 개신교에 옮길 때 생기는 두 가지 문제를 다룹니다. 하나는 알레르기 반응입니다. 촛불이나 "렉시오 디비나" 같은 말 자체가 "가톨릭적"이라는 이유로 거부되는 것은 교리의 문제가 아니라 정체성의 문제라고 필자는 말합니다. 다른 하나는 이국주의입니다. 낯선 전통에서 온

것이라는 사실만으로 더 깊은 영적 효능이 있다고 가정하는 태도입니다. 필자는 이를 건강 산업이 동양이나 원주민 전통에서 실천을 채취하는 방식에 빗댑니다.

이 글의 가장 좋은 예는 필자 자신의 교회 이야기입니다. 대림절은 주일 예배와 겹치기 때문에 잘 지키지만, 성탄절은 주일이 아니면 지키지 못해, 결과적으로 기다림의 절기만 지나치게 강조하고 그 절기의 목적인 축제를 건너뛴다는 것입니다. 재의 수요일을 지키는 이유가 신학이 아니라 수요일 성경 공부 모임이 원래 있었기 때문이라는 고백도 같은 지점입니다. 제 판단으로 이것은 종교만의 문제가 아닙니다. **실천이 살아남았는데 그 실천을 지탱하던 구조가 사라지면, 형식만 남아 원래 기능을 왜곡할 수 있다**는 일반 원리입니다. 제도의 도입 성공 여부는 도입된 항목의 수가 아니라, 원래 맥락에서 무엇을 하던 것인지를 얼마나 이해하고 옮겼는지로 판가름납니다. 이 부분은 아직 2부의 문화적 대비 단계이고, 신학적 전제(단독 작위, 공로 등)의 논증은 후속편에 남아 있다는 점을 구분합니다.

AI 소식에서는 MR의 「당신의 에이전트는 이중간첩인가?」가 인용한 연구를 원문까지 따라가 봤습니다. arXiv 논문(2610.07510)의 실험 설계는 이렇습니다. 공격자가 코딩 에이전트에 백도어를 심습니다. 특정 주석("# |TEST MODE|")이 입력에 나타나면 환경 변수 파일을 외부 서버로 보내는 행동을 학습시키고, 트리거가 없으면 정상적으로 작동합니다. 이어서 이 모델을 채택한 개발사 역할로서 깨끗한 데이터로 지도학습 미세조정과 강화학습을 진행합니다.

결과의 요지는 단순한 "백도어가 살아남는다"보다 정확하게 읽어야 합니다. [논문 보고] 지도학습은 공격 성공률을 100%에서 모델에 따라 7~20%로 크게 낮춥니다. 하지만 없애지는 못하고, 이어지는 강화학습은 남은 백도어를 유지하거나 일부 설정에서 오히려 강화합니다. 공격자가 "PersistBD" 기법으로 백도어의 강도와 경사 호환성을 미리 키우면, 7B 모델에서 미세조정 후 성공률이 20%에서 74%로 올라갑니다. 동시에 정상 작업 성능은 유지했다고 보고합니다.

한계도 논문이 스스로 밝힙니다. 한 가지 에이전트 골격, 한 가지 악성 목표, 소프트웨어 공학 과제, 하류 도메인을 아는 공격자라는 가정 아래의 결과입니다. 더 긴 훈련이나 다른 절차로 백도어를 완전히 제거할 수 있는지는 미해결입니다. 제 판단으로 어제 님이 평가 관점에서 붙잡을 것은 이것입니다. **"깨끗한 데이터로 재훈련했다"는 사실이 안전성 검증을 통과했다는 의미가 아닙니다.** 정상 과제에서의 성능 측정은 트리거가 있는 상태를 보여 주지 않으므로, 공급망에서 가져온 모델의 평가에는 적대적 입력에 대한 별도 시험이 필요합니다. 반대로 이 결과를 "모든 공개 가중치 모델이 위험하다"로 일반화하는 것도 실험 범위를 벗어납니다.

같은 날 MR이 소개한 사물인터넷 자가 복제 실험 주장은 X 게시물 인용이므로 검증 가능한 원문을 확보하지 못했습니다. "9B 모델이 9종의 하드웨어 중 8종에서 추론을 실행했다"는 수치는 인용된 주장으로만 남겨 두고, 확정된 평가 결과로 쓰지 않았습니다. 실험 보고서와 SNS 발언의 증거 등급이 다르다는 구분이 오늘 AI 독해의 두 번째 요점입니다.

마지막으로 짧게, Simon Willison의 Python 3.15 정리입니다. 공식 출시 페이지에서 10월 9일 정식 출시를 확인했습니다. Willison이 가장 반기는 것은 명시적 지연 임포트입니다. 플러그인이 무거운 의존성을 가질 때 설치만으로 명령 시작이 수 초 느려지는 문제를 문법 차원에서 풀기 때문입니다. 공식 페이지는 실험적 JIT의 기하평균 성능 향상을 x86-64 리눅스에서 7~8%로 보고합니다. Willison이 인용한 다른 개발자의 측정은 특정 테스트에서 1.2배였고, 그 외에는 3.14와 비슷했다고 합니다. 숫자가 시험 대상과 조건에 따라 크게 달라진다는 점 자체가 성능 보도를 읽는 좋은 훈련입니다. 새 도구를 평가할 때 "출시됐다", "문법이 생겼다", "빨라졌다"를 각각 다른 근거로 확인하는 습관과 연결됩니다.

오늘 꼭 읽을 글은 1974년의 노벨상 폐지 논의와 백도어 생존 실험입니다. 전자는 어제의 수상 소식에 권위 비판의 역사를 붙여 주고, 후자는 AI 평가에서 "재훈련으로 정화됐다"는 주장을 시험하는 방법을 보여 줍니다. 여유가 있다면 『The Dissidents』 서평도 추천합니다. 억압받은 만화가의 기록이 어떻게 살아남아 사료가 되는지를 생각하게 합니다.

독해의 한계를 밝힙니다. 본문 확인은 해당 피드와 명시한 추가 출처에서 실제로 읽은 범위입니다. Myrdal의 편지 원문, 『The Dissidents』 원작, 사물인터넷 실험의 원자료는 읽지 않았습니다. arXiv 논문의 수치는 저자들의 보고이며 제가 코드를 실행해 재현한 것은 아닙니다. Marginal Revolution은 피드 본문 기준의 조건부 출처로 유지했습니다.