



데미스 허사비스, 구글 인공지능의 아버지

The Father of Google AI

김경란, 김경진

목차

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.
- 10.
- 11.
- 12.
- 13.
- 14.
- 15.
- 16.
- 17.
- 18.
- 19.
- 20.
- 21.
- 22.
- 23.
- 24.
- 25.
- 26.
- 27.
- 28.
- 29.
- 30.
- 31.
- 32.
- 33.

맞달아 있습니다. 데미스 허사비스를 처음 알게 된 것은 2016년 서울의 봄이었습니다. 화면 속 바둑 판 위에서 AI가 사람을 이겼다는 뉴스가 쏟아졌고, 세상은 흥분과 불안이 뒤섞인 표정으로 그 장면을 바라보았습니다.

저 역시 그랬습니다. 그런데 그 흥분이 가라앉은 뒤, 저는 한 가지 질문이 머릿속에서 떠나지 않는다는 것을 알아챘습니다. "이 사람은 왜 이것을 했을까?"

승리 자체가 목적이었다면, 그는 알파고 이후 멈췄을 것입니다. 그러나 허사비스는 멈추지 않았습니다. 바둑판에서 실험실로, 실험실에서 세계로. 그의 시선은 언제나 다음 문제를 향하고 있었습니다.

수십 년을 쌓아 온 체스의 직관, 게임 개발의 경험, 뇌 과학의 발견이 하나의 방향을 가리키고 있었고, 그 방향의 끝에는 인류가 오래도록 풀 지 못한 난제들이 있었습니다. 저는 교육 그리고 정책이라는 길을 걸어온 사람입니다. 기술의 작동 원리보다 그것 이 사회와 인간에게 어떤 의미를 갖는지를 오래 물어온 사람이기도 합니다.

그 질문이 이 책을 쓰게 만들었습니다. 알고리즘의 정교함이 아니라, 한 인간이 40년에 걸쳐 품 어온 집념의 무게를 전하고 싶었습니다. 2024년 10월, 스톡홀름에서 전화가 걸려왔을 때 허사비스는 "머릿속이 완전히 백지가되었다"고 말했습니다.

노벨 화학상이었습니다. 실험실이 아닌 컴퓨터 앞에서 이룬 과학적 성취가 과학사의 가장 높은 자리에 오른 순간이었습니다. 그 전화 한 통이, 이 책이 왜 지금 나와야 하는지를 말해 주었습니다.

AI는 이미 우리 곁에 있습니다. 그러나 우리는 아직 그것을 제대로 이해하지 못하고 있습니다. 이 책이 그 이해를 조금이라도 넓혀 줄 수 있기를 바랍니다. 기술을 두려워 워하는 사람에게도, 기술에 흥분한 사람에게도, 그리고 그 사이 어딘가에서 조용히 고 민하는 사람에게도.

이야기는 런던 북부의 작은 집, 네 살 소년이 체스 말을 처음 잡은 그 순간에서 시작합니다.
김경란, 김경진

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

1976년 7월 27일 런던 출생 1976년 7월 27일, 런던 북부의 한 병원에서 사내아이 가 태어났습니다. 이름은 데미스. 그리스계 키프로스인 아버지 코스타스 허사비스 (Costas Hassabis)와 중국계 싱가포르인 어머니 앤젤라(Angela)의 첫 아이였습니다. 동양과 서양의 피가 한 몸에 섞인 이 아기는 3남매의 맏이로, 뒤이어 남동생 한 명과 여동생 한 명이 태어났습니다.

훗날 남동생은 프로 포커 플레이어가 되고, 여동생은 피아니스트이자 작곡가의 길을 걷게됩니다. 세 남매 중 수학과 과학에 빠져든 사람은 데미스뿐이었습니다. 허사비스는 자신의 부모를 "보헤미안(Bohemian)"이라고 묘사합니다.

정해진 궤도를 따르기보다 자기가 좋아하는 일을 쫓으며 사는 자유로운 영혼들이었다는 뜻입니다. 그런 부모 밑에서 태어난 장남은 어울리지 않을 만큼 논리적이고 체계적인 아이였습니다. 여동생과 남동생은 부모의 예술적 기질을 물려받아 음악과 글쓰기의 세계로 향했지만, 데미스는 어릴 때부터 규칙과 패턴, 숫자와 구조에 끌렸습니다.

인터뷰에서 이렇게 말합니다. "부모님은 기술을 싫어합니다. 컴퓨터라면 질색이에요. 여동생 과 남동생도 전부 예술 쪽으로 갔습니다.

이 모든 것이 어디서 온 건지, 솔직히 저도 잘 모르겠습니다." 장남으로서의 역할은 일찍 시작되었습니다. 부모님이 바쁜 날이 많 았기에, 데미스는 동생들의 놀이 상대이자 보호자가되었습니다.

아버지의 장난감 가게에서 팔 수 없게 된 장난감들을 가져오면 분해하고 재조합해서 새로운 놀이를 발명 했고, 여덟 살 때는 체스 대회 상금으로 산 ZX 스펙트럼 컴퓨터로 오셀로 AI를 만들 어 다섯 살 남동생과 대국시켰습니다. 컴퓨터가 이겼습니다. 자신이 만든 프로그램이 독립적으로 판단하고 행동한다는 사실 앞에서 소년은 멈춰 섰습니다.

네 살에 체스를 시작한 데미스는 배움의 속도부터 남달랐습니다. 아버지와 삼촌이 체스를 두는 모습을 지켜보다가 규칙을 가르쳐 달라고 졸랐고, 불과 2주 만에 두 어른을 모두 이겼습니다. 다섯 살에 전국 대회에 출전했으며 여섯 살에 런던 8세 이하 챔 피언십에서 우승했습니다.

열세 살에는 엘로(Elo) 레이팅 2300의 체스 마스터 등급 에 올랐고, 세계 14세 이하 랭킹 2위를 기록했습니다. 이 무렵 그는 잉글랜드 주니어

대표팀의 주장을 맡았습니다. 11 세 이하 국가대표팀을 이끌며 전략을 짜고 동료들을 독려한 경험은, 훗날 2,000명이 넘는 딥마인드 연구진을 이끄는 리더십의 원형이되었습니다. 딥마인드의 공동창업자 무스타파 슬레이만(Mustafa Suleyman)은 데미스의 남동 생 친구였습니다.

형제와의 유대가 세계에서 가장 영향력 있는 인공지능 연구소의 창업 팀까지 이어진셈이죠. 예술가 기질의 동생들과 과학자 기질의 자신이 한 지붕 아 래 공존한 환경은, 훗날 서로 다른 배경의 인재들을 한데 모으는 그의 조직 운영 감각으로 이어집니다. 코스타스 허사비스는 한마디로 정의하기 어려운 인물이었습니다. 젊은 시절에는 싱 어송 라이터로 활동했습니다.

그의 영웅은 밥 딜런(Bob Dylan)이었고, 기타를 메고 노래를 만드 는 일에 열정을 쏟았습니다. 은퇴한 지금도 그는 오페라를 쓰고 있습니다. 데미스는 2025 년 싱가포르 인터뷰에서 아버지를 이렇게 묘사합니다.

"아버지는 밥 딜런처럼 되고 싶어 했던 것 같아요. 지금 은퇴하신 뒤에도 오페라를 작곡 중이신데, 아버지가 어떤 분인지 그것만 봐도 아실겁니다." 코스타스는 음악가 시절이 지나 자 장난감 가게를 열었습니다. 아내 앤젤라와 함께 운영한 이 가게는 어린 데미스에게 매력적인 실험실이었습니다.

진열대에는 온갖 보드게임과 퍼즐이 놓여 있었고, 팔 수 없게 된 장난감들은 집으로 따라왔습니다. 데미스는 그것들을 해체하고 재조합하며 새로운 놀이를 발명했습니다. 장난감 가게는 게임의 구조를 분석하고, 왜 어떤 게임은 재미있고 어떤 게임은 지루한지를 탐구하는 어린 게임 디자이너의 첫 작업실이었습니다.

장난감 가게가 문을 닫은 뒤, 코스타스는 교사가 되었습니다. 앤젤라도 교사로 일했습니다. 앤젤라는 1970년대 초 싱가포르에서 영국으로 건너와 특수교육 간호사로 훈련 받았고, 이후 존 루이스(John Lewis) 백화점에서 매니저로 근무하기도 했습니다. 코스타스의 교사 생활 역시 남들과 달랐습니다.

그는 야간 수업을 자신만의 소규모 회사처럼 운영했습니다. 데미스의 표현을 빌리면, 아버지는 "정해진 9시~5시 직장을 가진 적이 한 번도 없는" 사람이었습니다. 장난감 가게도, 음악도, 교습도 전부 자기 방식대로 꾸려나갔습니다. 이 끊임없는 변신은 아들에게 결정적인 교훈을 남겼습니다.

세상에 미리 깔린 길 (Pre-set path)은 없으며, 원하는 삶은 스스로 설계해야 한다는 것이었습니다. 데미스

는 이렇게 회상합니다. "부모님은 저희에게 열정을 쫓으라고, 그 열정 속으로 깊이 들어가라고, 관습을 걱정하지 말라고 가르치셨습니다. '결국에는 다 잘 될 거야'라는 게 부모님의 태도였어요.

제 경력은 분명 매우 비관습적이었는데, 부모님께서 그 점을 아주 많이 격려해 주셨습니다." 아버지의 삶의 방식은 데미스에게 "무에서 유를 만드는 법"을 가르쳤습니다. 장난감 가게를 열고, 음악을 만들고, 강좌를 열고, 그 모든 시도가 자기만의 소규모 사업이었다는 점은 중요합니다.

코스타스는 거대한 부를 축적하지는 못했지만, 아이디어를 행동으로 옮기는 일을 두려워하지 않았습니다. 이 태도는 데미스의 혈관 속에 깊이 스며들었습니다. 열일곱 살에 피터 몰리뉴(Peter Molyneux)와 함께 '테마파크(Theme Park)'라는 게임을 공동 설계하고 리드 프로그래머로 일한 것, 스물두 살에 독립 게임 스튜디오 엘릭서(Elixir Studios)를 창업한 것, 서른네 살에 딥마인드를 세운 것, 이 모든 도전의 밑바닥에는 아버지의 DNA가 흐르고 있었습니다.

데미스는 아버지로부터 물려받은 것만큼이나 아버지에게서 반면교사를 삼은 것도 있었습니다. 코스타스는 다방면에 재능이 있었지만 하나의 목표에 오래 집중하지 못했습니다. 음악에서 장난감 가게로, 장난감 가게에서 교직으로 옮겨 다니는 동안 깊이보다는 폭이 넓어졌습니다.

데미스는 이 패턴을 지켜보며 다짐했습니다. 아버지처럼 창의적이고 대담하게 살되, 자신은 하나의 큰 질문에 평생을 걸겠다고. 그 질문이 바로 "지능이란 무엇인가(What is intelligence?)"였습니다. 아버지의 자유분방한 도전 정신과 아들의 지독한 집중력이 결합된 결과가 딥마인드입니다.

코스타스가 아들에게 남긴 유산은 돈이 아니었습니다. "네가 원하는 길은 네 손으로 만들어라"는 믿음이었습니다. 음악, 장난감 가게, 교실을 오가며 살았던 아버지의 삶은 아들에게 창의성과 대중

소통과 사업적 감각이라는 세 가지 재료를 남겼습니다.

데미스 허사비스는 스스로를 "노동 계급(Working-class)" 출신이라고 말합니다. 2025년 인터뷰에서 그는 이렇게 말했습니다. "부모님은 이민자였고, 우리는 런던에서 꽤 가난한 동네에서 자랐습니다." 집안 형편은 뼈뺏했습니다.

아버지가 장난감 가게를 운영하고, 노래를 만들고, 교사로 일하는 동안에도 경제적 안정은 쉽게 찾아오지 않았습니다. 체스는 이 소년에게 지적 유희인 동시에 현실적 수단이었습니다. 주말이면 아버지 코스타 스가 낡은 캠퍼밴(Camper van)을 몰고 영국 곳곳의 체스 토너먼트에 데미스를 데려갔습니다. 호텔에 묵을 여유는 없었습니다.

부자(父子)는 캠퍼밴 안에서 잠을 자거나, 가장 싼 숙소의 이층 침대를 잡았습니다. 아버지는 늘 아래층을 쓰고, 데미스는 위층에 올라가 다음 날 대국을 준비했습니다. 어린 소년의 어깨에는 명확한 압박이 실려 있었습니다.

우승 상금을 타야 기름값과 숙박비를 겨우 충당할 수 있었기 때문입니다. 데미스는 이 시절을 이렇게 회고합니다. "대회에 가면 잘해야 한다는 압박감이 항상 있었습니다. 경비가 비쌌고, 우리에게 돈이 없었으니까요."

그리고 덧붙입니다. "어릴 때 압박감을 견뎌내면, 나중에 더 강해집니다." 절박함은 역설적으로 데미스에게 승부의 칼날 위에서 집중하는 능력을 길러주었습니다.

체스판 위에서의 한 수는 단순한 게임이 아니라 가계에 실질적 보탬이 되는 경제적 행위였습니다. 이런 환경에서 자란 아이는 심리적 압박과 최악의 조건에서 이기는 법을 배웁니다. 훗날 딥마인드가 자금난에 허덕이며 거대 IT 기업들의 인수 제안을 놓고 저울질할 때, 허사비스가 보여 준 냉철한 협상력과 흔들리지 않는 집중력은 캠퍼밴에서 다져진 것이었습니다. 물질적 결핍이 지적 결핍으로 이어지지 않았습니다.

오히려 반대였습니다. 허사비스의 집에는 비싼 물건 대신 끊임없는 대화와 호기심이 있었습니다. 부모님은 아이들이 무엇에 관심을 두든 자유롭게 탐구하도록 격려했습니다.

물질적 보상보다 질문에 진지하게 답해 주는 시간을 아끼지 않았습니다. 데미스는 어른들의 대화에 끼어들어 자기 논리를 펼치는 아이였고, 부모님은 그것을 제지하지 않았습니다. 체스 대회 원정 비용을 아끼기 위해 홈스쿨링을 받기도 했고, 퀸 엘리자베스 스쿨과 크라이스트 칼리지 핀클리를 거치며 열여섯 살에 A-레벨 시험을 2년 일찍 마쳤습니다.

사교육이나 특별한 입시 준비 없이, 공공 도서관과 체스판, 컴퓨터 한 대가 그의 교실이었습니다. 허사비스의 어머니 앤젤라는 싱가포르 출신입니다. 데미스가 열 살이 될 때까지 매년 여름 두세 달씩 싱가포르를 방문하여 외가 친척들과 시간을 보냈습니다.

창이 해변에서 수영하고, 싱가포르 친척들과 어울리던 기억은 그의 유년기에 동남아시아의 따듯한 색채를 더했습니다. 그리스의 열정과 싱가포르의 근면함이 한 아이의 내면에서 섞이고 발효되는 과정은, 사물을 한쪽 관점으로만 보지 않는 유연한 사고력의 토대가 되었습니다. 딥마인드 초창기, 거대 기업들의 자금력 앞에서도 소수 정예로 혁신을 만들어낸 그의 방식은 어린 시절에 뿌리를 두고 있었습니다. 집안에 돈은 부족했지만 생각할

거리는 넘쳤습니다. 캠퍼밴의 좁은 침대 위에서 다음 날 체스 전략을 구상하던 소년에 게 지능은 풀고 싶은 수수께끼였고, 부족한 환경이 그 수수께끼를 향한 출발점이었습니다. 퀸 엘리자베스 스쿨(Queen Elizabeth's School, Barnet) 1988년 9월, 런던 북부 바 넷(Barnet). 400년 넘는 역사를 가진 퀸 엘리자베스 스쿨의 정문 앞 에 남색 블레이저 차림의 소년들이 줄지어 서 있었습니다.

1573년 엘리자베스 1세의 칙령으로 설립된 이 학교는 영국 전역에서 가장 뛰어난 학업 성취도를 자랑하는 남자 공립 그 래머 스 쿨이었습니다. 입학 자체가 관문이었습니다. 열한 살짜리 소년들은 '일레븐 플러스 (11+)'라 불리는 선발시험을 치러야 했고, 그 관문을 통과한 아이들만이 이 교정에 발을 디딜 수 있었습니다.

열두 살의 데미스 허사비스는 그 줄 사이에 끼어 있었습니다. 체구는 작 았지만, 눈빛은 또래와 달랐습니다. 허사비스에게 퀸 엘리자베스 스쿨은 두 번째 체스판이었습니다.

학교의 모든 것이 규칙과 순위로 돌아갔습니다. 매 학기 성적표가 나올 때마다 학생들 사이에는 보이지 않는 위계 가 만들어졌고, 교사들은 옥스퍼드와 케임브리지 진학률이라는 숫자 앞에서 긴장했습니 1980년대 후반 영국 교육계는 대대적인 개혁의 한가운데 있었습니다. 마거릿 대처 정부가 도입한 국가 공통 교육과정(National Curriculum)은 학교 간 성 과를 공개 비교하도록 만들었고, 그래머스쿨은 이 경쟁의 최전선에 놓여 있었습니다.

수학과 과학에서 월등한 성 과를 내야 한다는 압박은 교실 안 공기를 팽팽하게 만들었습니다. 허사비스는 기묘한 이중생활을 하고 있었습니다. 평일에는 교복을 차려입은 모범생이었습니다.

수학 시간에 공식을 풀고, 과학 실험실에서 비커를 다루었습니다. 그러나 방과 후 와 주말, 그의 세계는 완전히 달라졌습니다. 아버지 코스타스가 운전하는 낡은 캠퍼밴을 타고 영국 전역의 체스 토너먼트장을 돌아다녔습니다.

때로는 학교를 며칠씩 빠지고 유럽 각지의 국제 대회에 출전했습니다. 크리스마스 방학도, 여름 방학 도 체스 대회 일정으로 채워졌습니다. 허사비스는 훗날 이 시절을 이렇게 회고합니다.

"학교를 쉬고 전 세계를 돌 며 체스를 뒀습니다. 상당히 빡센 일정이었습니다." 이중생활의 성적표는 이랬습니다.

학교에서는 또래를 앞서 월반했고, 체스판 위에 서는 13세에 마스터 등급(Elo 2300)에 도달했습니다. 세계 14세 이하 랭킹 2위. 1위 는 당시 이미 전설이 되어가던 헝가리의 유디트 폴가(Judit Polgár)였습니다. 허사비

스는 잉글랜드 주니어 대표팀의 주장을 맡아 국제 무대에서 나라를 대표했습니다. 교 실에서는 조용한 수재였 지만, 체스판 앞에 앉는 순간 프로 기사들을 상대하는 냉철한 승부사로 변했습니다. 흥미로운 점은 이 학교가 훗날 또 한 명의 AI 선구자를 배출했다는 사실입니다.

허 사비스보다 몇 년 뒤인 1995년부터 2002년까지 같은 교정을 밟은 무스타파 술레이 만(Mustafa Suleyman)은 2010년 허사비스와 함께 딥마인드를 공동 창업하게됩니다. 런던 북부의 한 그래머스쿨이 인공지능 역사의 두 주역을 길러낸 셈입니다. 허사비스에게 학교는 충족되지 않는 갈증의 공간이기도했습니다.

그는 훗날 한 인 터뷰에서 "내 기억에 우리 학교에서 옥스브리지에 간 사람은 거의 없었다"고 회상합 니다. 당시 학 교 환경이 그에게 충분한 지적 자극을 주지 못했음을 암시합니다. 수학

시간에 교사가 칠판에 적는 공식은 이미 알고 있는 것이었고, 정해진 교과서의 진도는 답답할 만큼 느렸습니다. 체스판 위에서 10의 120승에 달하는 경우의 수와 씨름 하던 소년에게, 교실의 정답 맞 히기는 단순한 게임이었습니다.

이 시기에 허사비스의 인생을 뒤흔든 사건이 일어납니다. 리히텐슈타인에서 열린 국제 체스 대회. 열두 살의 소년은 덴마크의 성인 챔피언과 마주 앉았습니다. 대국은 10시간이 넘도록 이어졌습니다. 국면은 이미 무승부가 확실했지만, 상대는 집요하게 승리를 노리며 물고 늘어졌습니다.

결국 탈진한 허사비스가 기권을 선언하자, 상대는 의기양양하게 무승부로 이끌 수 있었던 수순을 늘어놓으며 훈수를 두었습니다. 소년은 대회장을 둘러보았습니다. 수백 명의 명석한 두뇌들이 오직 64칸 위의 승패에 모든 에너지를 쏟아붓고 있었습니다.

그때 하나의 생각이 떠올랐습니다. "이 두뇌들을 더 중요한 곳에 쓸 수 있다면?" 이 질문은 소년의 내면에서 조용한 지각변동을 일으켰습니다.

세계 챔피언이 되겠다는 꿈이 흔들리기 시작한 것입니다. 체스에 인생을 걸면 오프닝 이론과 엔드게임 기법이라는 달린 세계에 갇히게 됩니다. 허사비스가 체스에서 배운 것은 특정한 수가 아니라 메타 스킬 (meta-skill)이었습니다.

문제 해결 능력, 패턴 인식, 전략적 사고, 상상력. 이것들은 체스에만 쓰기에는 너무 가까운 능력이었습니다. 훗날 이렇게 정리합니다. "체스를 두면서 실제로 갈고닦는 것은 그런 메타 스킬들입니다.

그걸 과학이나 사업 같은 다른 모든 분야에 적용할 수 있다는 걸 꽤 어릴 때부터 깨달았습니다."

1990년, 허사비스는 퀸 엘리자베스 스쿨을 떠납니다. 2년이라는 짧은 시간이었지만, 이 학교는 그에게 두 가지를 선물했습니다. 하나는 경쟁 속에서 단련된 집중력과 자기 규율. 다른 하나는 그 경쟁의 틀 자체를 의심하는 눈. 그래머스쿨의 엄격한 교육은 그에게 탄탄한 기초를 쌓아주었지만, 동시에 기존 시스템의 한계를 체감하게 만든 인큐베이터였습니다. 소년은 더 넓은 세계로, 더 깊은 질문으로 나아갈 준비를 하고 있었습니다.

퀸 엘리자베스 스쿨을 떠난 뒤, 허사비스의 일상은 완전히 바뀌었습니다. 매일 아침 교복을 차려입고 교문을 통과하는 의식이 사라졌습니다. 대신 그는 집에서 눈을 뜨고, 자기만의 시간표에 따라 하루를 시작했습니다.

부모가 주도한 홈스쿨링 기간이 시작된 것입니다. 창밖으로는 같은 또래 아이들이 등교하는 발소리가 들렸지만, 허사비스의 눈앞에는 교과서 대신 컴퓨터 모니터가 놓여 있었습니다. 1990년대 초 영국에서 자녀를 명문 그래머스쿨에서 빼내 집에서 가르치겠다는 것은 쉬운 결정이 아니었습니다. 코스타스와 앤젤라는 BBC 라디오 인터뷰에서 "사회적 규범에 얽매일 필요가 없다"고 말할 만큼 자유로운 부모였지만, 현실적 이유도 있었습니다.

세계 랭킹 2위의 주니어 체스 선수가 정규 수업의 리듬에 맞추기란 불가능에 가까웠으니까요. 홈스쿨링 기간 동안 허사비스의 하루는 두 축으로 나뉘었습니다. 오전에는 체스였습니다. 세계 그랜드마스터들의 기보를 분석하고, 오프닝 이론을 연구하고, 자신의 대국을 복기했습니다.

오후에는 컴퓨터 앞에 앉았습니다. 이 시간이야말로 훗날 딥마인드의 씨앗이 뿌려진 토양이었습니다. 1984년, 여덟 살의 소년은 체스 대회에서 받은 상금으로 인생 최초의 결정적 투자를 합니 싱글레어 ZX 스펙트럼 48K(Sinclair ZX Spectrum 48K). 당시 가격으로 200~300파운드에 불과한 이 작은 기계가 허사비스의 인생을 바꾸어 놓았습니다.

1980년대 영국은 가정용 컴퓨터의 황금기였습니다. ZX 스펙트럼, BBC 마이크로, 커모도어 64 같은 기계들이 수십만 가정에 보급되면서 '베드룸 코더(bedroom coder)'라 불리는 독학 프로그래머 세대가 탄생하고 있었습니다. 허사비스도 그 물결에 올라탔습니다.

다만 다른 아이들과 결정적으로 다른 점이 있었습니다. 대부분의 아이들은 게임을 '하기 위해' 컴퓨터를 켜지만, 허사비스는 게임이 '어떻게 작동하는지' 알기 위해 컴퓨터를 켜습니다.

당시 ZX 스펙트럼의 전원을 켜면 화면에는 윈도우 같은 운영체제 대신 깜빡이는 커서 하나만 나타났습니다. 사용자가 직접 명령어를 입력해야만 무언가가 일어나는 세계. 허사비스는 서점에서 프로그래밍 교재를 사서 베이직(BASIC) 언어를 한 줄 한 줄 입력하기 시작했습니다. 부모는 기술에 문외한이었기에 도움을 줄 수 없었습니다.

오롯이 혼자 힘으로 변수, 반복문, 조건문이라는 프로그래밍의 기초 문법을 익혀갔습니다. 훗날 이 경험을 이렇게 표현합니다. "컴퓨터란 마음의 힘을 확장해주는 일종의 마법적 장치라고 늘 느꼈습니다.

시는 그 끝점이라고 생각합니다." 이 문장에는 여덟 살 소년의 원체 힘이 응축되어 있습니다. 그는 잠들기 전에 컴퓨터에게 복잡한 계산이나 시뮬레이션을 맡겨놓고 이불 속에 들어갔습니다.

다음 날 아침 일어나면 기계는 밤새 문제를 풀어놓고 있었습니다. 내가 자는 동안에도 나의 지적 능력을 대신 확장해주는 존재. 이 경험이 주는 전율은 소년의 마음에 깊이 각인되었습니다. ZX 스펙트럼에서 기초를 다진 뒤, 허사비스는 보다 강력한 기계로 눈을 돌렸습니다.

커모도어 아미가(Commodore Amiga). 당시로서는 뛰어난 그래픽과 사운드를 갖춘 16비트 컴퓨터였습니다. 베이직을 넘어 기계어(machine code)까지 독학한 소년은 이 기계 위에서 자신의 첫 번째 인공지능 프로그램을 작성합니다. 열한두 살 무렵의 일이었습니다.

소년 시절 컴퓨터 여정에서 가장 백미는 보드게임 오셀로(Othello)였습니다. 체스보다 규칙이 단순하지만, 컴퓨터에게 '제대로 된' 수를 두게 하려면 알고리즘이 필요했습니다. 허사비스는 체스 프로그래밍에서 사용되는 알파-베타 가지치기(alpha-beta pruning)와 탐색 트리(search tree)의 개념을 책에서 찾아 독학했고, 이를 자신의 오셀로 프로그램에 적용했습니다.

시험 상대는 다섯 살 남동생이었습니다. 결과는 허사비스의 승리, 정확히 말하면 허사비스가 만든 프로그램의 승리였습니다. 중요한 것은 자신이 작성한 코드 덩어리가 스스로 판세를 읽고, 전략을 세우고, 독자적으로 돌을 놓아 승리를 거두었다는 사실. 살아 있지 않은 기계가, 살아 있는 것처럼 행동하는 광경. 이것은 소년에게 충격이었습니다.

이 작은 성공은 허사비스의 머릿속에 하나의 질문을 심었습니다. "지능이란 대체 무엇인가? 코드 몇 줄로 흉내 낼 수 있는 것인가, 아니면 그 이상의 무엇인가?" 이 물음은 호기심을 넘어 그의 인생 전체를 관통하는 탐구 과제가 됩니다.

30년 뒤 그는 딥마인드라는 회사를 세우고, 그 질문에 대한 답을 찾기 위해 수백 명의 박사급 연구원들을 모을 것입니다. 홈스쿨링 기간은 허사비스에게 제도권 교육이 주지 못하는 것을 선물했습니다. 시간을 자율적으로 운용하는 법. 자신의 호기심을 스스로 충족시키는 법. 실패한 코드 앞에서 좌절 하지 않고 디버깅(debugging)을 반복하는 끈기. 48킬로바이트라는 극히 제한된 메모리 안에 복잡한 논리를 집어넣어야 하는 환경은 효율에 대한 강박에 가까운 감각을 심어주었습니다.

이 감각은 훗날 거대한 AI 모델을 설계할 때도 그를 따라 다닙니다. 홈스쿨링이 끝난 뒤, 허사비스는 이스트 핀클리(East Finchley)에 위치한 크라이스 츠 칼리지(Christ's College, Finchley)로 진학합니다. 그래머스쿨이 아닌 종합학교 (comprehensive school)였습니다. 이곳에서 그는 열여섯 살이라는 이른 나이에 A레 벨(A-level) 시험을 마칩니다. 보통 18세에 치르는 시험을 2년이나 앞당긴 것입니다.

케임브리지 대학은 그에게 입학 을 허가했지만, 나이가 너무 어리다는 이유로 1년을 기다리라고 권고합니다. 예상치 못한 공백이 생긴 것입니다. 이 공백이 또 하나의 운명적 전환점이 됩니다.

허사비스는 컴퓨터 잡지 '아미가 파워 (Amiga Power)'에서 "볼프로그 프로덕션에서 일할 기회를 잡아라(Win-a-job- at-Bullfrog)"라는 공모전을 발견합니다. 그리고 우승합니다. A레벨 시험이 끝난 다음 날, 열일곱 살의 소 년은 볼프로그 프로덕션의 사무실 문을 밀고 들어갔습니다.

정 식 고용을 하기에는 나이가 너무 어려서 급여는 현금으로 받아야했습니다. 이곳에서 그는 전설적인 게임 디자이너 피 터 몰리뉴(Peter Molyneux)를 만나고, '테마파크 (Theme Park)'의 공동 설계자이자 리드 프로그래머가 됩니다. 1994년 출시된 이 게 임은 수백만 장이 팔리며 시뮬레이션 샌드박스 장르 의 고전이 되었습니다.

허사비스는 훗날 조언을 구하는 사람들에게 이렇게 말합니다. "아이가 어릴 때 폭넓 게 탐험하도록 격려하세요. 깊이와 전문성을 장려하되, 다른 모든 것을 희생시키지 마세요.

인 생은 풍요로우니까요. 중요한 것은, 특정한 과목이나 주제를 배우는 것 보 다도, 배우는 법을 배우는 그 자체 능력입니다." 그 시절 거실의 희미한 모니터 불빛 아래서, 허사비스는 체스판 너머의 세계를 발견 했습니 64칸의 격자는 무한한 가능성을 가진 코드의 세계로 확장되었고, 그 코드의 세

계는 다 시 지능 그 자체를 향한 질문으로 이어졌습니다. 홈스쿨링이라는 우회로는 사 실 가장 빠 른 지름길이었습니다. 1970년대 런던 북부의 거리 풍경

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

4세에 체스 입문, 13세에 마스터 등급(Elo 2300) 달성 아버지와 삼촌의 체스 대국을 보고 배우다 1980년, 노스런던의 허사비스 가정에서 평범한 저녁 풍경이 펼쳐지고 있었습니다. 아버지 코스타스와 삼촌이 거실 탁자에 체스판을 펼쳐놓고 말을 움직이고 있었습니다. 두 사람 모두 전문 선수가 아닌, 그저 여가를 즐기는 아마추어 수준이었습니다.

네 살짜리 데미스는 그 광경을 지켜보고 있었습니다. 아이의 눈에 64칸의 흑백 격자판 위에서 벌어지는 조용한 전투가 들어왔습니다. 데미스는 자기도 두고 싶다고 말했습니다.

아버지와 삼촌은 어린아이의 호기심을 귀엽게 여겼을 뿐입니다. 말의 이름과 움직이는 방향 정도 알려주면 금세 흥미를 잃을 거라 생각했습니다. 그러나 예상은 빗나갔습니다.

규칙을 배운 지 불과 2주 만에, 네 살짜리 아이가 아버지를 이기기 시작했습니다. 삼촌도 마찬가지로였습니다. 허사비스는 훗날 이렇게 회고합니다.

"체스는 원래 제가 할 줄 알았던 것 같습니다. 네 살에 배웠는데, 배우던 기억조차 없어요." 코스타스는 아들에게 특별한 무언가가 있다는 걸 직감했습니다.

장난감 가게를 운영하고, 노래를 쓰고, 교사 일을 전전하며 늘 새로운 도전에 나서던 이 그리스계 키프로스인 아버지는 곧바로 아들을 지역 체스 클럽에 데려갔습니다. "거기서부터 모든 것이 시작되었습니다"라고 허사비스는 담담하게 말합니다. 그렇게 시작된 체스 인생은 빠른 속도로 전개되었습니다.

5세에 전국 무대에 섰습니다. 6세에 런던 8세 이하 대회에서 우승 트로피를 들어 올렸습니다. 체스 클럽의 어른들은 이 작은 소년이 판을 읽는 방식에 놀랐습니다. 단순히 다음 수를 계산하는 게 아니었습니다.

판 전체의 흐름을 감지하고, 상대의 의도를 읽고, 몇 수 앞의 상황을 머릿속에 그려내는 능력이 또래와 달랐습니다. 아이는 늘 자기보다 나이가 많은 상대와 두었고, 그 속에서 빠르게 성장했습니다. 체스판 앞에서의 집중은 아이의 일상을 바꿨습니다.

방과 후 시간, 주말, 방학이 모두 체스로 채워졌습니다. 학교를 빠지고 해외 대회에 나가는 일도 잦아졌습니다. 부모는 아들의 경기 결과에 민감했습니다. 이기면 기뻐했고, 실수하거나 패하면 크게 속

상했습니다. 어린 허사비스에게 체스는 순수한 놀이이면서 동시에 가족의 기대가 실린 무거운 과업이기도 했습니다. 아버지와 삼촌의 체스판에서 우연히 시작된 게임은, 이미 소년의 삶 전체를 규정하는 축으로 자리 잡고 있었습니다.

그 거실의 체스판이 없었다면 딥마인드도 없었을 것입니다. 알파고도, 알파폴드도 없었을 것입니다. 네 살짜리 아이가 아버지에게 "나도 두고 싶다"고 말한 그 순간은, 40년 뒤 노벨 화학상으로 이어지는 긴 여정의 첫 번째 수였습니다.

잉글랜드 주니어 대표팀 주장 9세에 잉글랜드 11세 이하 대표팀의 주장이되었습니다. 보통의 아이들이 아직 체스의 기초 정석을 익히고 있을 나이에, 허사비스는 국가를 대표하는 자리에 올랐습니다. 이후로도 잉글랜드 주니어 대표팀의 주장을 맡았습니다.

본인의 표현대로라면 "늘 자기 나이 또래 국가대표팀의 캡틴이었고, 대부분 훨씬 나이 많은 선수들과 경기했습니다." 주니어 대표팀 주장 타이틀은 체스 실력만으로 주어지는 것이 아닙니다. 대회 현장에서 팀을 이끌고, 전략적 배치를 논의하고, 동료 선수들의 심리적 안정을 도모하는 역할까지 포함합니다. 10대 초반의 아이가 이 역할을 수행했다는 것은, 허사비스에게 체스판 위의 계산 능력 외에 사람을 읽고 상황을 관리하는 능력이 이미 자라나고 있었다는 뜻입니다.

훗날 딥마인드에서 2,000명 이상의 연구자를 이끄는 리더십의 첫 연습이 주니어 체스 대표팀 주장실에서 시작되고 있었습니다. 13세, 허사비스는 엘로 레이팅(Elo rating) 2300에 도달했습니다. 엘로 레이팅은 체스 선수의 실력을 수치화하는 국제 표준 시스템입니다.

일반적인 아마추어 클럽 선수가 1200~1400 사이에 위치하고, 1800이면 상당한 고수로 인정받습니다. 2000을 넘으면 '후보 마스터(Candidate Master)', 2200 이상이면 공식적으로 '마스터(Master)' 등급에 해당합니다. 13세 소년이 2300에 도달했다는 것은 대부분의 성인 클럽 선수를 압도하는 수준에 올랐다는 뜻입니다.

그 시점에서 허사비스의 세계 14세 이하 랭킹은 2위였습니다. 1위는 헝가리의 유디트 폴가르(Judit Polgár)였습니다. 폴가르의 레이팅은 2335. 허사비스보다 35점 높았습니다.

유디트 폴가르는 이후 여성 체스 역사상 가장 위대한 선수로 성장하여, 15세에 최연소 그랜드마스터 기록을 세우고 가리 카스파로프를 꺾기도 한 전설적 인물입니다. 허사비스가 세계 2위를 차지한 시점에 바로 위에 있던 사람이 이 정도 수준의

천재였다는 사실은, 역설적으로 허사비스 자신의 재능이 얼마나 대단했는지를 보여줍니다. 이 시기 허사비스의 일상은 체스에 완전히 점령당해 있었습니다. 학기 중에도 수시로 학교를 빠지고 유럽 각지의 국제 대회에 참가했습니다. 여름 방학과 크리스마스 방학은 모두 대회 일정으로 채워졌습니다.

나머지 시간에는 오프닝 이론을 공부하고, 기보를 분석하고, 엔드게임 패턴을 암기했습니다. 세계 챔피언이 되겠다는 것이 당시 그의 계획이었습니다. 그 목표를 달성하려면 삶 전체를 체스 하나에 바쳐야 한다는 것도 알고 있었습니다.

그러나 바로 이 시기에, 흥미로운 역설이 자라나기 시작합니다. 체스에 깊이 빠져들수록, 허사비스는 체스가 가르쳐주는 것의 범위가 좁아지고 있다는 사실을 감지했습니다. 초반에 체스가 선물한 것은 범용적 사고력이었습니다.

문제를 분해하고, 패턴을 인식하고, 여러 수 앞을 내다보는 능력. 그러나 마스터 수준에 접어들면서 필요한 것은 점점 체스만을 위한 전문 지식으로 바뀌어갔습니다. 특정 오프닝의 20수째 변화 구, 특수한 엔드게임 포지션의 승패 이론 같은 것들. 이 지식은 다른 분야에 전혀 쓸모가 없었습니다. 세계 2위라는 숫자는 영광이었지만, 동시에 갈림길의 표지판이기도 했습니다.

한쪽 길은 체스에 인생을 걸고 세계 1위를 향해 달리는 것이었고, 다른 한쪽 길은 체스에서 배운 사고력을 들고 더 넓은 세계로 나가는 것이었습니다. 13세의 소년은 선택을 내리지 않았지만, 마음속에서 질문은 이미 울트고 있었습니다. 체스가 가르쳐준 메타 스킬: 전략적 사고와 패턴 인식 복잡한 시스템을 이해하고 수를 내다보는 능력의 배양 허사비스는 체스를 '정신의 체육관(mental gym)'이라고 불렀습니다.

체육관에서 역기를 드는 사람은 역기 자체가 목적이 아닙니다. 근력을 키우는 것이 목적이고, 그 근력은 일상의 다른 모든 활동에 전이됩니다. 허사비스에게 체스가 바로 그런 역할을 했습니다. 64칸의 판 위에서 훈련한 사고력은 체스를 떠난 뒤에도 그의 삶 전체를 관통하는 도구가 되었습니다.

체스가 가르쳐준 첫 번째 메타 스킬은 복잡한 시스템을 전체적으로 파악하는 능력입니다. 체스판에는 초기 배치 상태에서 32개의 말이 놓여 있고, 가능한 경우의 수는 약 10의 120승에 달합니다. 이 숫자는 관측 가능한 우주에 존재하는 원자의 수(약 10의 80승)보다 압도적으로 큼니다. 이 거대한 가능성의 공간 속에서 최선의 수를 찾으

려면, 모든 경우를 하나하나 계산하는 방식으로는 불가능합니다. 대신 판 전체의 구조를 읽어야 합니다. 어디에 힘이 집중되어 있는지, 어디가 약한지, 상대의 전략적 의도가 무엇인지를 감지해야 합니다.

나무 하나하나가 아니라 숲 전체를 보는 능력입니다. 허사비스는 어린 시절부터 이 능력을 반복적으로 훈련했습니다. 한 수를 두기 전에 그 수가 판 전체의 균형에 어떤 영향을 미칠지를 가늠하는 습관, 개별 사건이 아니라 시스템 전체의 역학을 읽는 습관이 그에게 몸에 밴 사고방식이 되었습니다.

이 능력은 훗날 딥마인드에서 강화학습 알고리즘을 설계할 때, 수천 명의 연구자로 구성된 조직을 경영할 때, 그리고 단백질 접힘이라는 생물학의 난제에 도전할 때 동일한 방식으로 작동했습니다. 두 번째 메타 스킬은 패턴 인식(pattern recognition)입니다. 마스터급 체스 선수는 판 위의 말 배치를 보는 순간 특정 패턴을 인식합니다.

인지심리학자 에이드리언 드흐 로트 (Adriaan de Groot)의 연구에 따르면, 체스 마스터는 약 5만에서 10만 개의 패턴을 기억 속에 저장하고 있습니다. 이 패턴들이 즉각적으로 활성화되면서, 마스터는 초보자가 10분 걸려 분석할 상황을 몇 초 만에 파악합니다. 허사비스가 4세부터 13세까지 9년간 매일같이 수행한 것은 바로 이 패턴 라이브러리의 구축이었습니다.

패턴 인식 능력은 체스판 밖에서도 강력한 무기가 됩니다. 과학 연구에서 데이터 속 규칙성을 포착하는 것, 사업에서 시장의 흐름을 읽는 것, 게임 디자인에서 플레이어의 행동 유형을 분류하는 것 모두 본질적으로 패턴 인식 작업입니다. 허사비스 자신도 이를 명확히 알고 있었습니다.

"체스를 높은 수준으로 두면, 실제로 연마하는 것은 메타 스킬입니다. 문제 해결, 상상력, 창의적 사고, 전략적 사고. 이런 것들을 과학이나 사업 같은 다른 분야에 옮겨 쓸 수 있습니다." 세 번째 메타 스킬은 여러 수 앞을 내다보는 계획 능력입니다.

체스에서는 한 수를 두기 전에 최소 3~5수 앞을 계산합니다. 마스터급은 10수 이상을 내다봅니다. 이때 필요한 것은 단 순한 계산이 아니라 '만약 내가 이렇게 두면, 상대는 이렇게 응수할 것이고, 그러면 나는 이런 선택지가 생기는데, 그중 어떤 것이 5수 뒤에 가장 유리한 포지션으로 이어지는가'를 판단하는 조건부 추론입니다.

이것은 본질적으로 의사결정 트리(decision tree)를 머릿속에서 구축하고 탐색하는 작업이며, 훗날 딥마인드가 알파고에 적용한 몬테카를로 트리 탐색 (Monte Carlo Tree Search)의 인간 버전이기도 합니다.

허사비스는 이 세 가지 메타 스킬, 즉 시스템적 사고, 패턴 인식, 계획적 선견을 체스판 위에서 9년간 혹독하게 단련했습니다. 이 세 가지는 훗날 그가 설계한 인공지능 시스템의 핵심 구성 요소와 정확히 대응합니다. 딥러닝은 패턴 인식을, 강화학습은 시스템 속에서의 최적 행동 탐색을, 트리 탐색 알고리즘은 미래의 경우의 수를 내다보는 계획 수립을 담당합니다.

네 살짜리 소년이 아버지의 체스판에서 시작한 훈련이, 30년 뒤 인류 역사상 가장 강력한 인공지능의 설계 원리로 결실을 맺은 것입니다. 체스판은 허사비스에게 게임 이상이었습니다. 지능 그 자체를 이해하기 위한 최초의 실험실이었습니다.

12세, 리히텐슈타인 대회 1989년 봄, 알프스 산맥에 둘러싸인 작은 나라 리히텐슈타인에서 국제 체스 오픈 대회가 열렸습니다. 12세의 데미스 허사비스가 대회에 참가하고 있었습니다. 대회장은 수백 명의 선수로 가득했습니다.

각국에서 모인 체스 고수들이 긴장감 속에서 말을 움직이고 있었습니다. 허사비스의 상대는 독일의 피데 마스터(FIDE Master) 카르스텐 피퍼엠덴(Carsten PieperEmden)이었습니다. 대국은 길고 치열했습니다.

10시간이 넘도록 두 사람은 체스판 앞에서 일어나지 못했습니다. 중반전이 지나고 엔드게임에 접어들면서 말의 수가 줄어들었고, 게임은 킹과 퀸 대 킹, 룩, 비숍, 나이트라는 복잡한 종반전으로 흘러갔습니다. 60수가 넘어가고, 70수를 지나, 77 수까지 이어졌습니다. 12세 소년의 집중력과 체력은 한계에 다다르고 있었습니다.

결정적인 순간이 왔습니다. 허사비스는 패배를 선언했습니다. 그런데 상대가 일어 서더니 극적인 제스처로 보여주었습니다.

허사비스가 퀸을 희생했다라면 스테일메이트 (stalemate), 즉 무승부를 만들 수 있는 수순이 남아 있었다는 것이었습니다. 상대는 마지막 순간에 저렴한 트릭을 시도했고, 지쳐 있던 12세 소년은 그것을 놓쳤습니다. "위장이 뒤를 리는 기분이었습니다"라고 허사비스는 회고합니다.

허사비스는 대회장을 벗어나 리히텐슈타인의 들판을 걸었습니다. 알프스의 산들이 주위를 에워싸고 있었고, 아름다운 풍경이 펼쳐져 있었습니다. 산책 중에, 12세 소년의 머릿속에 하나의 생각이 떠올랐습니다.

생각속에서 대회장 홀을 내려다보았을 때, 수백 명의 뛰어난 두뇌들이 서로 이기기 위해 말을 움직이고 있는 광경이 눈에 들어왔습니다. "저 방은 놀라울 정도로 뛰어난 사람들로 가득 차 있습니다. 그런데 그들은 자신의 두뇌를 서로 경쟁하고 이기는 데 쓰고 있습니다.

만약 저 모든 두뇌를 하나의 시스템에 연결할 수 있다면, 병을 치료할 수도 있지 않을까요. 저 시간과 에너지를 더 나은 곳에 쓸 수 있다면, 세상을 위해 더 좋은 일을 할 수 있지 않을까요." 12세 아이의 직관이었습니다. 논리적 추론이라기보다 확신에 가까웠습니다.

10시간의 지독한 대국에서 어이없는 실수로 패배한 직후, 정서적으로 바닥을 친 상태에서 터져 나온 근본적인 물음이었습니다. 체스를 사랑하지만, 이것이 내 인생 전부가 되어야 하는가. 이 뛰어난 두뇌들이 서로를 이기는 데만 쓰이는 것이 최선인가. 만약 이 지적 역량을 모아서 인류의 진짜 문명에 투입할 수 있다면, 이 순간을 기점으로, 허사비스의 방향 전환이 시작되었습니다. 물론 곧바로 체스를 그만둔 것은 아닙니다. 그 뒤로도 수년간 대회에 참가했고, 케임브리지

대학에서 옥스 브리지 대학 전에 3년 연속 출전하기도 했습니다.

그러나 세계 챔피언이 되겠다는 목표는 이 리히텐슈타인의 산책에서 조용히 내려놓아졌습니다. 대신 새로운 질문이 그 자리를 차지했습니다. 체스에서 훈련한 이 사고력을, 더 큰 문제에 쓸 수는 없을까. 허사비스가 훗날 딥마인드를 창립하면서 내건 미션 "지능을 풀면, 나머지는 해결된다 (Solve intelligence, and then use that to solve everything else)"는, 바로 이 12세의 직관을 성인의 언어로 번역한 것입니다.

리히텐슈타인의 대회장에서 수백 명의 두뇌가 체스판에만 묶여 있는 것을 보며 느낀 안타까움, 그 지적 역량을 암 치료나 기후 문제 같은 인류적 과제에 쓸 수 있다면 좋겠다는 바람. 이것이 20년 뒤 범용 인공지능 (AGI)이라는 개념으로 구체화됩니다. 수백 명의 체스 마스터 대신, 수만 대의 컴퓨터가 하나의 학습 알고리즘 아래에서 단백질 구조를 예측하고, 신약 후보 물질을 탐색하고, 기상 패턴을 분석하는 세상. 허사비스가 꿈꾼 것은 바로 그것이었습니다. 리히텐슈타인의 산책은 이 전기의 원점(原點)입니다.

10시간의 패배가 선물한 깨달음이 없었다면, 허사비스는 뛰어난 체스 그랜드마스터로 생을 마쳤을지 모릅니다. 그러나 12세의 소년은 체스판 너머를 보았고, 그 시선은 결국 과학사를 바꾸는 데까지 닿았습니다.

첫 컴퓨터(ZX Spectrum 48K)와 프로그래밍 커모도어 아미가(Commodore Amiga)에서 오셀로 AI 개발 1984년의 어느 오후, 런던의 한 체스 대회장에서 여덟 살 소년 데미스 허사비스가 우승 트로피를 안았습니다. 소년의 손에는 상금 200파운드 가 쥐어져 있었습니다. 또래 아이들이 라면 장난감 가게로 뛰어갔을 틈입니다.

그러나 데미스의 발걸음은 전자제품 매장으로 향했습니다. 그가 그곳에서 선택한 물건은 검은 본체에 무지개빛 줄무늬가 새겨진 작은 기계, 싱글레어 ZX 스펙트럼 48K(Sinclair ZX Spectrum 48K)였습니다. 1982년 클라이브 싱글레어 경(Sir Clive Sinclair)이 출시한 8비트 가정용 컴퓨터였습니다.

고무 재질의 납작한 키보드, 거실 텔레비전에 연결해야 화면이 나오는 투박한 구조, 카세트테이프 프로그램 불러들이는 느린 방식. 사양만 놓고 보면 보잘것없었지만, 이 기계가 가진 진짜 힘은 가격에 있었습니다. 125파운드에서 175파운드 사이라는 가격표는 영국 전역의 평범한 가정에 컴퓨터를 들여놓는 혁명을 일으켰습니다. 500만 대 이상 판매된 이 작은 상자는 영국 IT 산업의 씨앗이 되었고, 수천 명의 10대 소년소녀들을 이른바 '침실 프로그래머(Bedroom Coders)'로 변모시킨 장본인이었습니다.

데미스의 가정에는 컴퓨터를 아는 어른이 한 명도 없었습니다. 아버지 코스타스와 어머니 앤젤라는 스스로를 '기술 공포증 환자(Technophobes)'라고 부를 만큼 기계와 거리가 멀었습니다. 집에 선생이 없다는 것은 소년에게 불리한 조건이 아니라 축복이 되었습니다.

그는 스스로 길을 만들어야 했습니다. 소년은 아버지의 손을 잡고 런던 차링 크로스 로드(Charing Cross Road)에 자리한 대형 서점 포일스(Foyles)로 향했습니다. 그곳 컴퓨터 프로그래밍 서가 앞에 주저앉아 책을 펼쳤습니다.

허사비스는 훗날 당시를 회상하며 이렇게 말했습니다. "아버지와 함께 포일스에 가서 프로그래밍 코너에 앉아 게임에서 무한 목숨을 얻는 방법을 배우곤 했어요. 그 시절 컴퓨터의

놀라운 점은 커자마자 바로 프로그래밍을 시작할 수 있다는 것이었습니다.

나는 직 감적으로 이 기계가 창의력을 마음껏 풀어놓을 수 있는 마법의 도구라는 것을 이해했어요." ZX 스펙트럼에는 베이직(BASIC) 인터프리 터가 내장되어 있어서 전원을 켜는 순간 코 디ング을 시작할 수 있었습니다. 키보드의 각 키에는 최대 여섯 가지 기능이 할당되어 있었고, 'J' 키 하나만 누르면 'LOAD'라는 명 령어가 자동으로 입력되는 설계였습니다.

소년은 잡지 에 인쇄된 수십 줄, 때로는 수 백 줄의 코드를 한 글자 한 글자 타이핑하며 프로그래밍의 문 법을 체득했습니다.

이 시기 영국에서는 '유어 싱클레어(Your Sinclair)', '크래시(Crash)', '싱클레어 유 저(Sinclair User)' 같은 컴퓨터 잡지들이 매달 게임과 유틸리티의 소스 코드를 지면에 실었습니다. 독 자들은 그 코드를 직접 타이핑해서 실행했습니다. 오타 하나만 있어도 프로그램은 작동을 멈추었고, 오류를 찾아 수정하는 과정에서 아이들은 자연스럽게 논리적 사고와 디버깅의 기초를 익혔습니다.

데미스에게 이 과정은 고통이 아니라 게 임이었습니다. 체스판 위에서 상대의 수를 읽는 행위와 코드 속에서 버그를 추적하는 행위는 본질적으로 같은 능력, 즉 패턴을 인식하고 논리적 경로를 추적하는 능력을 요 구했습니다. 성장과 함께 도구도 진화했습니다.

ZX 스펙트럼의 48킬로바이트 메모리와 단색에 가까운 그래픽은 곧 소년의 야심을 감당하지 못했습니다. 데미스는 훨씬 강력한 성능을 갖춘 커모 도어 아미가(Commodore Amiga)로 기종을 바꾸었습니다. 아미가는 당 시로서는 파격적인 컬러 그래픽, 스테레오 사운드를 갖춘 16비트 컴퓨터였습니다.

새 로운 기계 앞에서 데미스는 자신의 인생을 관통하게 될 근본적인 질문을 처음으로 스스로에게 던졌습니다. "컴퓨터 가 나처럼 생각하게 만들 수 있을까?" 이 질문의 답을 찾기 위한 첫 번째 실험 대상은 보드게임 오셀로(Othello)였습니다.

허사비스는 자신 이 오셀로를 둘 때 머릿속에서 일어나는 판단 과정을 분해했습니다. 현재 판세의 유리함을 어떻게 평가하는지, 상대방이 다음에 어디에 돌을 놓을지 어 떻게 예 측하는지, 여러 수 앞을 내다볼 때 어떤 가치를 먼저 탐색하고 어떤 가치를 잘 라내는지. 그는 이 사고 과정을 알고리즘으로 옮겼습니다. 게임 트리의 가지들을 효율적으로 잘라내는 알파-베타 프루닝(Alpha-Beta Pruning) 같은 탐 색 기법의 기초를 독학으로 익힌 것도 이 무렵입니다.

완성된 프로그램은 남동생을 이길 수 있을 만큼 제법 잘 두었습니다. 허사비스는 훗 날 이 경험을 아카데미 오브 어치브먼트(Academy of Achievement) 인터뷰에서 이렇 게 회고했습니다. "제가 기억하는 첫 번째 프로그램 중 하나가 리버시, 영국에서는 오 셀로라고 부르는 게임을 두는 프로그램이었습니다.

꽤 잘 두었어요. 제 남동생을 이길 수 있었으니까요." 이 작은 성공 안에 거대한 원형이 숨어 있었습니다.

인간의 사고 과정을 관찰하고, 그것을 알고리즘으로 번역하며, 기계가 스스로 최선 의 결정을 내리게 만드는 이 세 단계의 흐름은 20년 뒤 딥마인드가 아타리 게임을 정 복하고, 바둑 의 신을 넘어서고, 단백질의 구조를 예측하는 과정과 정확히 같은 뼈대를 공유합니다. 여덟 살 체스 소년이 상금 200파운드로 산 검은 상자는 그렇게 인류 의 미래를 바꾸는 40년 여 정의

출발선이되었습니다. 놀이가 곧 연구가 되던 시기 데미스 허사비스의 유년 시절에는 놀이와 연구 사이의 경계선이 존재하지 않았습니다.

보 통의 아이들에게 게임은 학교와 숙제로부터의 도 피였지만, 데미스에게 게임은 정밀하게 해체하고 재조립해야 할 시스템이었습니다. 새로운 게임을 손에 넣으면 그는 먼저 즐겼고, 그 다음에는 분해했습니다. 화면 속 캐릭터 가 왜 저렇게 움직이는지, 점수 체계는 어떤 논리로 설계되었는지, 난이도 곡 선은 왜 이 지 점에서 꺾이는지. 소년의 방 안에 놓인 컴퓨터는 오락기가 아니라 세상에서 가장 정교한 실험실이었습니다.

1980년대 영국은 이런 소년이 자라기에 좋은 토양을 갖추고 있었습니다. ZX 스펙트럼과 BBC 마이크로(BBC Micro)가 보급된 교실과 가정에서 수천 명의 10대가 코 디ング을 배우고 있었습니다. Matthew Smith는 열일곱 살에 '매닉 마이너(Manic Miner)'를 만들었고, 데이비드 브래번 (David Braben)과 이언 벨(Ian Bell)은 대학 기숙사에서 우주 탐험 게임 '엘리트(Elite)'를 완 성했습니다.

올리버 쌍둥이(The Oliver Twins)는 자기 방에서 '디지 (Dizzy)' 시리즈를 탄생시 켰습니다. 침실 프로그래머들의 황금기였습니다. 전문적인 훈련을 받지 않은 청소년들이 자기방에서 코드를 짜고, 그 결과물을 출판사에 팔아 용 돈을 벌던 시대. 데미스는 시대적 물결 한가운데 서 있었습니다.

그런데 소년은 동시대의 다른 천재들과 미묘하게 다른 곳을 응시하고 있었습니다. 대부분 의 젊은 프로그래머들이 더 화려한 그래픽, 더 빠른 속도, 더 자극적인 게임 플 레이에 열정을 쏟았다면, 데미스의 관심은 게임 속 존재들의 행동 그 자체에 있었습니다. 적군 캐릭터는 왜 항상 같은 패턴으로 움직일까. 만약 이 캐릭터가 플레이어의 행동을 관 찰하고, 거기서 배우고, 다음번에는 다른 전략을 쓴다면 어떨까. "가상 세계의 인물들에게 어떻게 지능을 부여할 것인가."

이 질문은 당시 열 살 안팎의 소년이 품기 에는 지나치게 거 대한 것이었지만, 체스 대국에서 인간 지성의 깊은 곳을 탐험해본 아이에게는 자연스러운 다음 단계였습니다. 데미스의 집으로 돌아오면 그는 곧장 컴퓨터 앞에 앉았습니다.

체스판 위에서 수만 번 반복한 경험이 코드로 변환되는 시간이었습니다. 체스에서 한 수를 두기 전에 세 수, 다섯 수 앞을 내다보는 습관은 프로그래밍에서 조건 분기와 재귀 함수를 설계하는 감각이되었습니다. 기물의 상대적 가치를 순간적으로 계산하 던 직관은 게임 AI의 평가 함수(Evaluation Function)를 만드는 기초가되었습니다.

승리했을 때의 보상과 패배했을 때의 대가를 수천 번 체험한 기억은 훗날 강화학습 (Reinforcement Learning)의 보 상 체계를 이해하는 데 결정적인 밑거름이되었습니다. 이 모든 연결은 의식적으로 설계된 것이 아니었습니다. 소년은 재미있어서 코드를 짰고, 재미있어서 게임을 분석했을 뿐입니다.

이 시기에 형성된 허사비스 특유의 습관이 있습니다. 하나의 문제를 여러 각도에서 동시 에 바라보는 융합적 사고입니다. 그는 게임을 개발하면서 인간의 심리를 관찰했습니다.

왜 어떤 게임은 사람을 몇 시간이고 붙잡아두는데, 어떤 게임은 5분 만에 지 루해지는가. 그 는 인공지능을 설계하면서 자신의 뇌가 어떻게 작동하는지를 궁금해했습니다. 체스에서 직관적으로 좋은 수를 떠올리는 순간, 뇌의 어느 부분이 활성화되 는 것일까. 기계는 지치 지 않고 모든 경우의 수를 탐색하는 반면, 인간은 대부분의 가 지를 무시하고 단번에 핵심으로 뛰어드는 능력이 있습니다. 이 두 세계의 장점을 결 합할 수는 없을까. 소년은 스스로를 "생각에 대해 생각하는

아이"라고 정의했습니다.

메타인지(Metacognition)라는 학술 용어를 아직 몰랐을 뿐, 이미 그것을 실천하고 있었습니다. 그의 방은 프로그래밍 서적과 플로피 디스크, 체스 기보집과 게임 잡지들로 뒤덮여 있었습니다. 밖에서 뛰어노는 시간보다 화면 속 데이터가 자신이 의도한대로 반응하는 것을 지켜보는 시간이 더 길었습니다.

누군가에게는 고립으로 보였을 이 시간들이 소년에게는 퍼즐의 조각을 하나씩 맞춰가는 가장 행복한 몰입이었습니다. 게임이라는 안전하고 통제된 환경 속에서 지능의 본질을 탐구하는 방식은 딥마인드의 핵심 철학이 되었습니다. 2013년 딥마인드가 아타리 게임 '벽돌 깨기(Breakout)'에 규칙을 전혀 알려주지 않은 채 화면 픽셀만 보여주고 스스로 학습하게 한 실험, 2016년 알파고가 인간의 기보 없이 자기 대국만으로 바둑의 신이 된 알파고 제로, 2020년 알파폴드가 50년 난제인 단백질 접힘 문제를 풀어낸 과정.

이 모든 것의 원형은 노스런던의 어느 방에서 소년이 오셀로 AI를 만들며 품었던 단순한 질문, "컴퓨터가 나처럼 생각할 수 있을까?"에서 시작되었습니다. 체스 상금 200파운드짜 리 검은 상자 앞에 앉은 여덟 살 소년의 방은 단순한 어린 시절의 공간이 아니었습니다. 그 것은 지능을 풀어 인류의 난제를 해결하겠다는 40년 퀘스트의 첫 번째 스테이지였습니다. 체스 토너먼트에 출전한 소년

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

시 등 다종목 석권 1997년 여름, 런던 사우스뱅크의 로열 페스티벌 홀(Royal Festival Hall)에 낯선 긴장감이 감 돌았습니다. 체스 기사들과 바둑 고수들, 브리지 챔피언들과 오셀로 마스터들이 한자리에 모인 것입니다. 제1회 마인드 스포츠 올림피아드 (Mind Sports Olympiad, MSO)의 개막이었습니다.

인간의 정신 능력을 겨루는 올림픽을 만들겠다는 야심 아래, 체스 마스터이자 인공지능 연구자 데이비드 레비(David Levy), 기억력 훈련의 대가 토니 부잔(Tony Buzan), 저명한 체스 저술가 레이먼드 킨(Raymond Keene)이 힘을 모아 기획한 대회였습니다. 10만 파운드의 상금이 걸렸고, 수십 개 종목이 열렸으며, 세계 각지에서 수 백 명의 게임 고수들이 모였습니다. 그 가운데 스물한 살의 데미스 허사비스가 있었습니다. 대회의 핵심은 펜타마인드 세계 챔피언십이었습니다.

올림픽 육상에 5종 경기(펜타 슬론)가 있듯, 마인드 스포츠 올림피아드에는 정신의 5종 경기가 있었습니다. 참가자는 올림피아드 기간 동안 열리는 수십 개 종목 중 다섯 가지를 선택해 겨루고, 그 종합 점수로 순위가 매겨졌습니다. 조건이 있었습니다.

비슷한 종류의 게임을 중복 선택할 수 없었고, 최소 하나 이상은 장시간 종목을 포함해야했습니다. 즉 체스만 다섯 번 두거나 카드 게임만 골라 잡을 수 없었습니다. 진정한 '정신의 만능 선수'를 가려내기 위한 설계였습니다.

펜타마인드에서 다른 체스 선수들과 구분 짓는 것은 체스 실력이 아니었습니다. 다른 종목으로 건너갈 때마다 보여준 적응 속도였습니다. 바둑은 체스와 전혀 다른 사고를 요구합니다.

체스가 전술적 계산의 깊이를 겨루는 게임이라면, 바둑은 판 전체를 조감하는 형세 판단과 직관의 게임입니다. 경우의 수가 우주의 원자 수보다 많다는 이 동양의 게임에서 허사비스는 높은 수준의 기량을 갖추었습니다. 일본 장기인 쇼기(Shogi)에서도 마찬가지였습니다. 쇼기는 작은 상대

말을 자기 편으로 다시 투입할 수 있는 독특한 규칙 때문에 체스와는 다른 종류의 전략적 복잡성을 갖습니다. 허사비스는 이 낯선 규칙 체계 안에서 빠르게 고수의 반 열에 올랐습니다. 포커는 상대의 카드가 보이지 않는 불완전 정보 환경에서 확률과 심리를 동시에 읽어야 하는 게임이고, 디플로머시(Diplomacy)는 주사위도 카드도 없이 오직 플레이어 간의 협상과 동맹, 배신으로만 승패가 갈리는 외교 전략 게임입니다.

디플로머시 테이블에서는 수학적 계산보다 상대의 의도를 읽고, 신뢰를 얻고, 때로는 그 신뢰를 전략적으로 깨뜨리는 사회적 지능이 핵심입니다. 허사비스는 2004년에 디플로머시 챔피언십에서 우승했고, 2006년 개인 챔피언십에서는 4위에 올랐습니다. 순수한 논리의 세계에서 인간 심리의 세계까지, 그의 게임 능력의 스펙트럼은 놀라울 만큼 넓었습니다.

1998년, 허사비스는 첫 번째 펜타마인드 세계 챔피언이 되었습니다. 그리고 1999년, 2000년, 2001년에도 연달아 왕좌를 지켰습니다. 2002년에는 자신의 게임 회사 엘릭서 스튜디오 (Elixir Studios)의 경영에 집중하느라 참가하지 못했지만, 2003년에 복귀하자마자 다시 정상에 섰습니다.

6년 사이에 다섯 번의 우승. 20년이 넘도록 이 기록은 깨지지 않았습니다. 에스토니아의 안드레스 쿠스크(Andres Kuusk)가 2024년에 여섯 번째 우승을 달성하며 비로소 이 기록을 넘어섰지만, 허사비스가 20대 청년 시절에 이룩한 성취의 무게는 여전히 마인드 스포츠 역사에서

전설적 위치를 차지합니다. 허사비스의 펜타마인드 우승이 갖는 의미는 금메달의 숫자에 있지 않았습니다. 그것은 인간의 지능이 한 분야에 묶이지 않을 수 있다는 증거였습니다.

체스 세계 챔피언은 체스에서 최고이고, 바둑 세계 챔피언은 바둑에서 최고입니다. 하지만 펜타마인드 챔피언은 어떤 종류의 문제 앞에 놓여도 높은 수준의 해답을 찾아내는 사람입니다. 논리 게임, 확률 게임, 심리 게임, 전략 게임을 넘나드는 이 능력은 훗날 허사비스가 추구할 범용 인공지능(AGI)의 개념과 정확히 겹칩니다. 한 가지 일만 잘하는 좁은 지능이 아니라, 어떤 영역에든 적용될 수 있는 범용적 사고 능력. 허사비스는 자기 몸으로 그 가능성을 먼저 증명해 보인 셈입니다. "역사상 가장 뛰어난 게임 플레이어"라는 마인드 스포츠 올림피아드의 평가 2003년, 다섯 번째 펜타마인드 트로피를 들어 올린 허사비스에게 마인드 스포츠 올림피아

드 조직위원회는 하나의 칭호를 부여했습니다. "역사상 가장 뛰어난 게임 플레이어 (The best games player in history)." 이것은 체스에서 가장 강하다거나 바둑에서 가장 뛰어나다는 뜻이 아니었습니다.

서로 다른 규칙과 다른 종류의 지적 능력을 요구하는 여러 게임을 가로질러 가장 높은 수준의 종합적 사고력을 보여준 인물이라는 뜻이었습니다. BBC 라디오 4의 체스 프로그램 '어크로스 더 보드(Across the Board)'에 출연한 허사비스를 소개하며 진행자 도미니크 로슨(Dominic Lawson)은 그를 "체스, 디플로머시, 쇼기, 포커를 포함한 수많은 게임의 전문가"라고 묘사했습니다. 체스베이스 (ChessBase)는 그를 "위키피디아가 '영국의 인공지능 연구자, 신경과학자, 컴퓨터 게임 디자이너, 세계적 게임 선수'라고 소개하지만, 이것으로는 그의 진면목을 다 담아 내지 못한다"고 평했습니다.

이 평가가 과장이 아님을 보여주는 것은 허사비스가 보여준 종목 간 전환의 자연스러움이었습니다. 대부분의 게임 고수들은 한 분야에서 수십 년을 보냅니다. 체스 그랜드마스터가 갑자기 바둑 프로 수준에 이르는 경우는 극히 드뭅니다. 게임마다 요구하는 인지 능력의 종류가 다르기 때문입니다. 체스는 깊은 수순 계산(tactical calculation)을, 바둑은 전체 판을 읽는 형세 판단(positional judgment)을, 포커는 확률 계산과 상대 심리 파악을, 디플로머시는 설득과 협상을 각각 핵심으로 삼습니다.

허사비스는 이 서로 다른 인지 영역들 사이를 빠르게 전환하며 각각에서 높은 수준을 유지했습니다. 올림픽에서 100미터 달리기와 마라톤과 수영을 동시에 메달권에 놓는 것과 비슷한 일이었습니다. 허사비스 자신은 이 능력의 비결을 '메타인지(metacognition)'라는 개념으로 설명한 적이 있습니다.

게임을 하는 것이 아니라 게임을 하는 자신의 사고 과정을 관찰하는 것. 체스에서 이기는 법이 아니라, 새로운 게임의 규칙을 빠르게 흡수하고 핵심 전략을 추출하는 법을 연습한다는 의미였습니다. BBC 인터뷰에서 허사비스는 체스를 왜 그만두었느냐는 질문에 이렇게 답했습니다. "열한 두 살쯤이었습니다.

체스가 놀라운 게임이고 정신을 훈련하는 데 훌륭하다는 것은 알았지만, 평생을 체스처럼 좁은 하나의 분야에 쏟는 것이 자기 만족적이라는 느낌이 들었습니다. 저는 게임을 정신의 체육관(gym for the mind)으로 사용하고, 거기서 훈련한 기술을 과학이나 사업 같은 다른 영역에 옮기고 싶었습니다." 이 발언은

허사비스의 지적 여정 전체를 관통하는 핵심 철학이었습니다. 게임은 목적이 아니라 수단입니다. 지능의 원리를 탐구하기 위한 실험실입니다.

체스판 위에서 다섯 수, 열 수 앞을 내다보는 능력은 그 자체로 가치가 있지만, 그 능력의 근저에 있는 '패턴을 인식하고 미래를 시뮬레이션하는 메커니즘'을 이해하는 것이 허사비스에게는 더 큰 상이었습니다. 인간의 뇌는 체스를 배운 직후에 바둑도 배울 수 있는가. 왜 포커를 잘 두는 사람이 협상에서도 능숙한가. 서로 다른 게임에서 작동하는 지능의 공통 구조는 무엇인가. 이 질문들이 그의 머릿속에서 점점 선명해지고 있었습니다. 마인드 스포츠 올림피아드의 동료 참가자들은 허사비스를 묘사할 때 두 가지 특징을 공통적으로 언급했습니다. 첫째는 새로운 게임을 배우는 속도였습니다.

올림피아드에서는 해마다 새로운 종목이 추가되었고, 허사비스는 처음 접하는 게임에서도 놀라울 만큼 빠르게 상위권에 진입했습니다. 올림피아드 주최 측에 따르면, 허사비스는 2003년 올림피아드에서 한 참가자에게 엔트로피(Entropy)라는 게임을 직접 가르쳐 주기도 했는데, 이 게임은 영국의 화학 공학자 에릭 솔로몬(Eric Solomon)이 발명한 것으로, 첫 번째 올림피아드부터 매 회 빠지지 않고 종목에 포함된 유일한 게임이었습니다. 둘째는 게임에 대한 진지한 분석적 태도였습니다.

그는 단순히 승리를 즐기는 것이 아니라 왜 그 수가 좋은 수인지, 왜 그 전략이 이 게임에서 작동하는지를 끊임없이 파고들었습니다. 2010년 딥마인드를 창업할 때, 투자자들 앞에서 "범용 인공지능을 만들겠다"고 선언한 젊은 창업자의 신뢰도를 높여준 것은 논문 목록만이 아니었습니다. "세계 최고의 게임 플레이어가 게임을 풀 수 있는 AI를 만들겠다고 한다"는 서사는, 실리콘밸리의 투자자들에게 강력한 설득력을 발휘했습니다.

피터 틸(Peter Thiel)과 일론 머스크(Elon Musk)가 딥마인드에 초기 투자를 결정한 배경에는 허사비스의 학술적 역량과 함께, 수천 번의 대국을 통해 지능의 본질을 체험적으로 이해하고 있다는 확신이 자리 잡고 있었습니다. 올림피아드의 금메달들은 이력서의 장식이 아니라, "이 사람은 지능이 무엇인지 아는 사람"이라는 증명 서였습니다. 세계 포커 시리즈(WSOP) 참가와 데카멘타솔론 우승 전략 게임에서 확률 게임까지: 불완전 정보 하의 의사결정 라스베이거스의 리오 올스위트 호텔(Rio All-Suite Hotel). 매년 여름이면 세계에서 가장 날 카로운 포커 두뇌들이 모입니다.

세계 포커 시리즈(World Series of Poker, WSOP)는 1970년 시작된 포커의 올림픽이자, 수백만 달러의 상금이 걸린 지구상 최대의 카드 게임 축제입니다. 테이블에 데미스 허사비스가

앉았습니다. 체스 마스터이자 펜타마인드 5회 챔피언인 그가, 블러핑과 배드비트(bad beat)가 난무하는 포커의 세계에 발을 들인 것입니다. 포커는 허사비스가 이전에 정복한 게임들과 근본적으로 다른 성격을 지녔습니다.

체스나 바둑은 완전 정보 게임(perfect information game)입니다. 두 플레이어 모두 보드 위의 모든 말의 위치를 볼 수 있고, 숨겨진 정보가 없습니다. 이론적으로 충분한 계산 능력만 있다면 최선의 수를 도출할 수 있는 세계입니다.

포커는 정반대입니다. 상대의 카드는 보이지 않습니다. 상대가 강한 패를 들고 있는데 약한 척 베티링할 수도 있고, 약한 패로 강한 척 블러핑을 할 수도 있습니다.

정보가 불완전한 상황에서 확률적 추론과 심리적 판독을 동시에 수행해야 하는 것입니다. 허사비스의 포커 성적은 아마추어 수준을 훌쩍 넘어섰습니다. 포커 데이터베이스 헨돈몹 (Hendon Mob)의 기록에 따르면, 그는 생애 11회의 라이브 토너먼트에서 입상하여 총 107,902달러의 상금을 기록했습니다.

세계 포커 시리즈에서만 여섯 차례 상금권(캐시, cash)에 진입했고, 그중에는 WSOP의 꽃이라 불리는 메인 이벤트 (Main Event)도 포함되어 있었습니다. 2008년에는 런던에서 열린 WSOP 유럽 (WSOP Europe)에서도 입상했습니다. 한 번의 토너먼트에서 거둔 최대 상금은 27,192달러였습니다.

프로 포커 선수로 생계를 꾸리는 사람들의 수준에 비하면 거대한 금액은 아닐 수 있지만, 게임 회사를 운영하고 학술 연구를 병행하면서 여섯 시즌에 걸쳐 WSOP에서 꾸준히 결과를 낸 것은 그의 게임 능력이 한 분야에 국한되지 않았음을 증명하는 기록입니다. 포커에서 허사비스가 배운 것은 돈보다 값진 통찰이었습니다. 현실 세계가 체스판보다 포커 테이블에 더 가깝다는 깨달음이었습니다.

주식 시장에서 투자자는 다른 투자자의 전략을 완전히 알 수 없습니다. 의사는 환자의 몸속에서 무슨 일이 벌어지고 있는지 모든 정보를 갖고 있지 못합니다. 외교관은 상대국의 진짜 의도를 파악해야 하지만 그들이 내놓는 말은 포커 플레이어의 베팅처럼 진실과 기만이 뒤섞여 있습니다.

지능이 정말로 '범용적'이라면, 완전한 정보가 주어진 깔끔한 문제뿐 아니라, 정보가 부족하고 상대가 속일 수도 있는 혼탁한 상황에서도 최선의 판단을 내릴 수 있어야 합니다. 허사비스는 펜타마인드보다 더 극단적인 정신의 시험대에도 도전했습니다. 데카멘타슬론 (Decamentathlon)입니다.

단 4시간 안에 10가지의 서로 다른 시험을 치러야 합니다. 브리지, 체스, 창의적 사고, 체커, 바둑, 지능 테스트, 마스터마인드, 기억력, 암산, 오셀로. 종목마다 100점 만점으로 총 1,000점이 최고점입니다. 하나라도 약한 영역이 있으면 치명적입니다.

카드 한 벌을 외우는 기억력 시험이 끝나자마자 바둑 문제를 풀어야 하고, 복잡한 암산을 마친 직후 창의적 사고력을 겨루는 문제 앞에 앉아 야합니다. 허사비스는 2003년과 2004년, 이 데카멘타슬론에서 연속 우승했습니다. 제1회 대회 때 1만 파운드의 상금이 걸렸던 이 종목은, 올림픽아드 초기에는 펜타마인드보다 더 중요한 종목으로 여겨지기도 했습니다.

4시간이라는 짧은 시간 안에 인간 지능의 거의 모든 측면을 동시에 시험하기 때문입니다. 논리적 추론, 기억력, 계산 능력, 공간 지각, 전략적 판단, 창의적 사고. 이 모든 것이 한꺼번에 요구되는 경기에서 허사비스가 정상에 선 것은, 그의 두뇌가 어느 한쪽으로 치우치지 않은 균형 잡힌 구조를 갖추고 있었음을 보여줍니다. 포커의 불확실성과 데카멘타슬론의 전방위적 도전을 동시에 거친 이 시기는, 허사비스의 지적 여정에서 중요한 전환점이 됩니다. 체스 신동 시절에 그는 깊이의 힘을 배웠습니다.

한 가지 문제를 끝까지 파고드는 집중력. 펜타마인드 시절에 그는 넓이의 힘을 배웠습니다. 다양한 종류의 문제에 빠르게 적응하는 유연성. 그리고 포커와 데카멘타슬론을 거치며 그는 불확실성 속에서의 결단력, 즉 완벽하지 않은 정보를 가지고도 행동해야 하는 용기를 배웠습니다. 이 세 가지 자질은 훗날 그가 딥마인드를 이끌며 마주한 도전들, 아타리 게임에서 바둑으로,

바둑에서 단백질 접힘 문제로 도약할 때마다 반복적으로 발휘되는 핵심 역량이 됩니다.

게임 경험이 AI 설계에 미친 영향 2014년, BBC 라디오 4의 진행자 도미닉 로슨이 허사비스에게 물었습니다. "당신은 왜 체스를 그만두었습니까?" 허사비스는 잠시 생각한 뒤 대답했습니다.

"열한두 살 무렵이었습니 당시 저보다 레이팅이 높은 선수는 유디트 풀가르뿐이었고, 제 엘로 점수보다 겨우 35 점 높았습니다. 세계 최고가 될 수 있었느냐고요. 가능했을지도 모릅니다. 하지만 인간의 모든 두뇌와 재능을 다른 인간 한 명을 이기는 데에만 쏟는 것은 자기 탐닉처럼 느껴졌습니다." 이 고백은 허사비스의 게임 경력 전체를 하나의 문장으로 압축합니다.

게임에서 이기는 것은 그에게 최종 목적지가 아니었습니다. 게임은 지능의 작동 원리를 관찰하고 실험 하기 위한 가장 효율적인 실험실이었습니다. 그 실험실에서 허사비스가 발견한 첫 번째 원리는 '강화학습(Reinforcement Learning)'의 핵심 아이디어였습니다.

체스를 배울 때 교과서를 외우는 것보다 직접 두어보고, 지고, 왜 졌는지 분석하고, 다음번에 다른 수를 시도하는 것이 훨씬 빠르게 실력을 늘리는 법이었습니 포커에서는 이 원리가 더 극단적으로 작동했습니다. 같은 상황이 두 번 반복되는 법이 없기 때문에, 매 순간 자신의 결정에서 피드백을 받고 전략을 조정해야했습니다. 이것은 훗날 딥마인드가 개발한 DQN(Deep Q-Network)과 알파고의 자기 대국(self-play) 시스템의 뼈대가 됩니다.

교사가 정답을 알려주는 것이 아니라, AI 스스로 수없이 시행착오를 겪으며 '보상(reward)'을 최대화하는 방향으로 학습하는 구조. 허사비스는 자신이 수만 번의 대국에서 체험한 학습 과정을 기계에 이식하고자했습니다. 두 번째 원리는 '전이 학습(Transfer Learning)'의 가능성이었습니다. 펜타마인드에서 허사비스가 경험한 것은, 체스에서 배운 패턴 인식 능력이 바둑에서도 어느 정도 작동한다는 사실이었습니다.

물론 체스 실력이 자동으로 바둑 실력으로 전환되지는 않습니다. 하지만 "새로운 게임의 규칙을 빠르게 이해하고, 그 안에서 핵심 전략을 추출하고, 상대보다 한 수 앞을 내다보는" 능력은 게임 간에 전이되었습니다. 허사비스는 이것이 인간 지능의 핵심이라고 보았습니다.

체스를 두던 뇌가 곧바로 운전을 배울 수 있고, 외국어를 배우던 뇌가 프로그래밍도 배울 수 있는 유연함. 이 범용성을 기계에 부여하는 것이 그의 궁극적 목표가 되었습니다. 알파제로(AlphaZero)가 동일한 알고리즘으로 체스, 바둑, 쇼기를 모두 정복했을 때, 그것은 허사비스가 펜타마인드 테이블에서 품었던 꿈의 기계적 구현이었습니다.

세 번째 원리는 '불완전 정보 환경에서의 의사결정'이었습니다. 포커 경험이 씨앗을 뿌린 이 문제의식은 딥마인드의 연구 방향에 깊은 흔적을 남겼습니다. 알파고까지의 딥마인드 AI는 대부분 완전 정보 게임을 다루었습니다.

하지만 현실 세계의 문제, 단백질의 3차원 구조를 예측하거나 기상 패턴을 분석하거나 신약의 효능을 추정하는 작업은 모두 불완전한 정보 위에서 이루어집니다. 모든 변수를 알 수 없고, 자연은 포커 플레이어처럼 답을 숨기고 있습니다. 허사비스는 포커 테이블에서 이 본질을 체험했고, 그 경험은 딥마인드가 스타크래프트 II를 정복한 알파 스타(AlphaStar) 프로젝트로 이어졌습니다. 스타크래프트는 상대의 기지가 보이지 않는 '전장의 안개(fog of war)' 속에서 실시간으로 전략을 수립해야 하는 게임으로, 포커와 마찬가지로 불완전 정보 환경의 의사결정을 요구합니다.

네 번째 원리는 의외의 것이었습니다. '사회적 지능'의 중요성입니다. 디플로머시에서 세계 팀 챔피언을 차지한 허사비스는, 순수한 계산 능력만으로는 풀 수 없는 문제가 있다는 것을 알고 있었습니다.

사람과 사람 사이의 신뢰, 협력, 배신의 역학은 수학적 최적화의 범위를 벗어납니다. 이 경험은 딥마인드의 조직 문화와 협업 방식에도 영향을 미쳤지만, 더 근본적으로는 AI의 한계에 대한 허사비스의 인식을 형성했습니다. 아무리 뛰어난 알고리즘도 인간 사회의 복잡한 역학 속에서 작동하려면 사람을 이해해야 합니다. 딥마인드가 NHS(영국 국민건강서비스)와의 협력에서 데이터 프라이버시 논란에 휘말렸을 때, 그리고 구글이라는 거대 기업 안에서 연구소의 자율성을 지키기 위해 끊임없이 협상해야 했을 때, 디플로머시 테이블에서 다름은 감각이 허사비스에게 도움이 되었을 것이라고 추측하는 것은 무리가 아닙니다. 허사비스의 게임 경력을 하나의 그래프로 그린다면, 횡축은 시간이고 종축은 게임의 복잡성과 불확실성의 정도일 것입니다.

네 살에 시작한 체스는 규칙이 명확하고 정보가 완전한 세계입니다. 펜타마인드는 그 규칙의 종류를 다양하게 확장했습니다. 포커는 불확실성이라는 새로운 차원을 추가했습니다.

디플로머시는 인간이라는 가장 예측 불가능한 변수를 도입했습니다. 이 귀적은 딥마인드 AI의 진화 경로와 놀라움 만큼 일치합니다. 아타리의 벽돌 깨기에서 시작해, 바둑의 완전 정보 게임을 정복하고, 스타크래프트의 불완전 정보 환경으로 나아가고, 마침내 단백질 접힘이라는 자연 세계의 난제에 도전한 여정. 허사비스는 자기 인생에서 먼저 걸어본 길을 딥마인드의 AI에게 다시 걸어가게 한 셈입니다.

런던의 올림픽아드 경기장, 라스베이거스의 포커 테이블, 디플로머시의 외교 회담장에서 보낸 수만 시간은 딥마인드라는 거대한 프로젝트의 설계 경험이었습니다. 허사비스는 2003년 경쟁적 게임에서 은퇴했지만, 엄밀히 말해 그는 게임을 그만둔 것이 아니었습니다. 게임판의 크기를 바꾼 것이었습니다.

64칸의 체스판에서 19x19줄의 바둑판으로, 그리고 마침내 생명의 설계도인 단백질 구조로. 게임의 규모는 달라졌지만 핵심 질문은 같았습니다 "이 복잡한 시스템 안에서 어떻게 최선의 답을 찾을 수 있는가." 펜타마인드의 트로피들과 WSOP의 상금 칩들은 이 질문에 대한 첫 번째 해답이었고, 딥마인드의 알고리즘들은 그 해답의 기계적 확장이었습니다. 어린 시절의 데미스 허사비스, 체스 말을 들고

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

운명적 만남 1992년, 런던 북부의 한 집 안에서 16세 소년이 게임 잡지를 넘기고 있었습니다. 데미스 허 서비스였습니다. 또래보다 2년이나 앞서 고등학교 과정(A-Level) 을 끝마친 그에게 케임브리지 대학 합격 통지서가 도착해 있었지만, 대학 측의 답변은 의외였습니다.

"너무 어리다. 1년 동안 세상 경험을 쌓고 오너라." 영국에서는 이런 휴식기를 '갭 이어(Gap Year)'라고 부릅니다. 대부분의 학생이 배낭여행을 택하는 그 시간을, 데미스는 전혀 다른 방식으로 보 내기로 했습니다. 그의 손에 들린 잡지는 당시 유럽의 게임 문화를 이끌던 월간지 아미가 파워 (Amiga Power) 였습니다.

그 잡지 한구석에 작은 공모전 광고가 실려 있었습니다. "불프로그에서 일할 기회를 잡으세요(Win-a-job-at-Bullfrog)!" 불프로그 프로덕션 (Bullfrog Productions)은 영국 서 리주 길퍼드(Guildford)에 자리 잡은 게임 개발사였습니다.

1982년 피터 몰리뉴(Peter Molyneux)와 레스 에드거(Les Edgar)가 공동 창업한 이 회사는 《파퓰러스(Populous)》와 《파워몽거(Powermonger)》로 이미 유럽 게임계에서 독보적인 지위를 차지하고 있었습니다. 데미스는 공모전에 자신의 코드를 제출했고, 얼마 지나지 않아 길퍼드 사무실로 오라는 연락을 받았습니다. 면접장에 나타난 소년을 본 순간, 피터 몰리뉴의 기억은 수십 년이 지나도 선명했습니다.

그는 훗날 타임(TIME)지와 인터뷰에서 이렇게 회상했습니다. "《반지의 제왕》에 나오는 엘프 같았습니다. 거리에서 스쳐 지나가면 눈에 띄지 않을 정도로 작고 가녀린 아이였는데, 눈에는 지성의 불꽃이 반짝이고 있었습니다."

몰리뉴를 진정으로 사로잡은 것은 외모가 아니라, 곧이어 벌어진 대화였습니다. 결정적인 순간은 회사 크리스마스 파티로 이동하던 버스 안에서 일어났습니다. 길 퍼드에서 파티 장소까지 가는 왕복 여정 동안, 16세 소년과 유럽 게임계의 거장은 실 새 없이 이야 기를 나눴습니다. 몰리뉴의 표현에 따르면 그것은 "인생에서 가장 지적으로 자극적인 대화"였습니다.

두 사람은 게임의 철학에 대해 이야기했습니다. 인간의 심리에서 승리욕은 어디서 비롯되는가. 그 욕구를 기계에 심을 수 있는가. 지능이란 결국 무엇인가. 몰리뉴는 대화 내내 속으로 되뇌었다고 합니다. "이건 겨우 아이잖아!"

그러나 그는 이 아이가 비범한 곳으로 향하 리라는 것을 직감했습니다. 두 사람은 빠르게 가까워졌습니다. 데미스가 케임브리지 입학 을 앞두고 불프로그에서 보 낸 여름 동안, 몰리뉴와 데미스는 틈만 나면 보드게임과 컴퓨터 게임을 함께했습니다.

몰 리뉴는 전략 게임에서 대부분 데미스를 이겼다고 자랑했지만, 그 우위는 오래가지 못했습니다. 두 사람은 새로운 카드 게임의 역학을 탐구하기 위해 '더미 (Dummy)'라는 이름의 게임을 함께 고안했는데, 데미스는 이 게임에서 몰리뉴를 35 연승으로 꺾었습니다. 경쟁심이 강하고 패턴 인식에 탁월한 소년 앞에서, 게임계의 전 설도 속수무책이었던 것입니다.

그러나 이 만남의 진짜 가치는 상호 보완에 있었습니다. 몰리뉴는 감각적이고 즉흥적인 비평가였습니다. '갓 게임(God Game)'이라는 장르를 개척하며, 플레이어에게 신의 시점을 부여하는 대담한 구상을 밀어붙이는 리더였습니다.

데미스는 그 꿈을 정교한 코드로 현실화 할 수 있는 냉철한 설계자였습니다. 데미스는 비전이 어떻게 사람의 마음을 움직이는지를 상대로부터 배웠고, 물리뉴는 데미스에게서 치밀한 논리가 복잡한 문제를 어떻게 풀어내는지를 목격했습니다. 이 관계는 훗날 라이언헤드 스튜디오(Lionhead Studios)에서의 재회로 이어지며, 데미스의 커리어 전체를 관통하는 멘토십의 출발점이 됩니다.

갭 이어를 배낭여행 대신 코드와 대화로 채운 이 선택은, 체스 신동이었던 소년을 가상 세계의 설계자로 변모시킨 첫 번째 전환점이었습니다. '신디케이트(Syndicate)' 레벨 디자인 참여 볼프로그에 합류한 데미스 허사비스에게 주어진 첫 임무는 《신디케이트(Syndicate)》의 플레이 테스트였습니다. 1993년에 출시된 이 게임은 기업이 국가를 대체한 디스토피아적 미래를 배경으로, 플레이어가 사이보그 요원 팀을 지휘하여 세계를 지배해 나가는 실시간 전략 게임이었습니다.

《블레이드 러너》를 연상시키는 어둡고 퇴폐적인 도시 풍경, 아이소메트릭(isometric) 시점으로 내려다보는 거리, 그 위를 움직이는 무장 요원과 시민들의 군상이 화면을 채웠습니다.

16세의 신입에게 플레이테스트란 게임을 반복적으로 플레이하며 버그를 찾고 난이도의 균형을 점검하는 작업입니다. 대부분의 사람들에게는 단조로운 일이었겠지만, 데미스에게 이 과정은 게임이라는 복잡한 시스템의 내부 구조를 밑바닥부터 해부하는 기회였습니다. 그는 오류를 보고하는 데 그치지 않았습니다.

레벨 디자인 단계에 깊이 관여하면서, 각 미션의 공간 배치와 적 요원의 이동 경로, 시민들의 반응 패턴을 분석하고 개선안을 제시했습니다. 《신디케이트》의 미션 구조는 데미스의 사고방식과 잘 맞았습니다. 각 미션은 암살, 요원 탈취, 정보원 납치 등 다양한 목표를 제시했고, 플레이어는 이를 은밀하게 수행할 수도, 정면으로 돌파할 수도 있었습니다.

이 개방형 설계는 '하나의 정답'이 아니라 '다수의 전략적 경로'를 허용했고, 데미스는 이 다양성을 극대화하는 레벨을 구성하는 데 집중했습니다. 그에게 레벨 디자인은 예쁜 지도를 그리는 작업이 아니었습니다. 적의 배치, 지형의 높낮이, 시민 동선의 밀도, 무기 접근성 같은 수십 가지 변수가 서로 물려 돌아가는 정교한 연립방정식을 푸는 일이었습니다.

이 경험에서 데미스가 깊이 매료된 개념은 '에이전트(Agent)'였습니다. 게임 속 도시에는 요원과 시민이 각자의 규칙에 따라 움직이고 있었습니다. 시민들은 요원이 총을 쏘면 도망치고, 경찰은 소란을 감지하면 현장으로 달려왔습니다.

이 반응들은 개발자가 일일이 지시한 것이 아니라, 각 캐릭터에 심어진 조건부 규칙(if-then logic)에서 자동으로 발생했습니다. 개별 요소들이 서로 맞물려 예측하지 못한 상황을 만들어 내는 이 현상은, 훗날 복잡계 이론에서 '창발(Emergence)'이라 불리는 것과 같은 원리였습니다. 데미스는 이 가상 도시의 작은 점들이 살아 있는 것처럼 반응하는 모습에서, 지능의 가장 원초적인 형태를 발견했습니다. 당시 게임 개발은 오늘날처럼 수백 명이 참여하는 대규모 산업이 아니었습니다.

볼프로그의 직원은 35명 안팎이었고, 소수의 프로그래머와 아티스트가 밤을 새우며 하나의 게임을 완성하던 시절이었습니다. 이 밀도 높은 환경에서 데미스는 코딩뿐 아니라 게임 디자인의 철학, 프로젝트 관리의 흐름, 팀 내 의사소통의 방식을 자연스럽게 체득했습니다. 그는 완벽주의적 성향을 바탕으로 게임 엔진의 비효율적인 부분을 지적했고, 더 나은 코딩 방식을 제안했습니다. 동료들이 이 어린 신입의 거침없는 발언

에 당혹스러워하는 순간도 있었지만, 그가 내놓는 결과물 앞에서는 수긍할 수밖에 없었습니다. 《신디케이트》는 데미스 허사비스에게 두 가지를 남겼습니다. 하나는 기술적 역량입니다.

실시간으로 움직이는 복수의 에이전트를 제한된 하드웨어 위에서 구동시키는 법, 레벨 하나에 담긴 수십 가지 변수의 균형을 맞추는 법을 그는 이 게임에서 몸으로 익혔습니다. 다른 하나는 질문입니다. 이 요원들은 내가 짜놓은 규칙 안에서만 영리합니다.

만약 규칙 없이도 스스로 배울 수 있다면? 이 질문은 아직 답이 없었지만, 소년의 머릿속에 분명한 씨앗으로 심어졌습니다. 그리고 그 씨앗은 바로 다음 프로젝트인 《테마파크》에서 첫 번째 줄기를 틔우게 됩니다. 전설적 게임 '테마파크(Theme Park)' 공동 설계 수백만 장 판매와 시뮬레이션 샌드박스 장르의 탄생 실력을 증명한 데미스 허사비스에게 피터 몰리뉴는 더 큰 과제를 맡겼습니다.

놀이공원을 짓고 운영하는 시뮬레이션 게임, 《테마파크(Theme Park)》의 공동 설계와 리드 프로그래밍이었습니다. 17세의 나이에 회사의 핵심 프로젝트를 이끄는 수석 프로그래머가 되었다는 자체가 이례적이었습니다. 개발 기간은 약 1년 반이었습니다.

몰리뉴가 전체적인 비전과 디자인 방향을 제시하면, 데미스는 그것을 코드로 구현하는 엔진의 핵심을 담당했습니다. 몰리뉴는 훗날 이 게임의 탄생 배경을 이렇게 설명했습니다. 경영 시뮬레이션이라는 장르가 가능성이 있다고 판단했고, 이전에 만든 경영 게임 《앙트러프러너(The Entrepreneur)》의 실수를 반복하지 않으면서 사람들이 즐겁게 플레이할 수 있는 게임을 만들고 싶었다고요. 그는 플레이어가 꿈속의 놀이공원을 직접 건설하는 경험을 선사하고 싶었습니다.

데미스에게 이 비전은 완벽한 도전이었습니다. 빈 땅 위에 놀이기구를 배치하고, 매점을 열고, 도로를 깔고, 입장료를 정하는 모든 과정이 하나의 거대한 시스템으로 맞물려야 했기 때문입니다. 1994년 출시된 《테마파크》는 폭발적인 성공을 거두었습니다. 전 세계적으로 1,500만 장 이상이 팔렸습니다.

일본에서는 플레이스테이션 버전이 출시 몇 주 만에 8만 5천 장이 소진 될 정도였고, 유럽에서는 베스트셀러 목록의 상단을 차지했습니다. 골든 조이스틱 어워드 (Golden Joystick Award)를 수상했고, PC 게이머(PC Gamer)지는 1994년 6월의 '이달의 게임'으로 선정하며 그 중독성을 《심시티 2000》에 비

견했습니다. 에지(Edge)지의 리뷰어는 게임의 복잡성을 인정하면서도 디테일과 몰입 감을 높이 평가했습니다. 이 게임의 진정한 가치는 '장르의 창조'였습니다.

《테마파크》 이전의 게임들은 대부분 적을 물리치거나 정해진 스토리를 따라가는 선형적 구조를 가지고 있었습니다. 《테마파크》는 플레이어에게 빈 공간과 도구를 주고, 무엇을 만들지는 전적으로 플레이어에게 맡겼습니다. 이 '샌드박스(Sandbox)' 접근법은 훗날 《롤러코스터 타이쿤》, 《테마 호스피탈》, 《심즈》 같은 시뮬레이션 게임들의 조상이 되었습니다.

실제로 《테마파크》의 코드 상당 부분은 후속작 《테마 호스피탈》에 재활용되었고, 이 게임의 애니메이션 편집기는 마크 웹리(Mark Webley)에 의해 '복합 엔진(The Complex Engine)'으로 발전했습니다. 데미스는 이 게임을 설계하며 경제 모델의 내부 논리에 집요하게 매달렸습니다. 감자튀김의 소금 함량을 높이면 방문객이 목이 말라 음료를 더 구매하지만, 동시에

구역질 수치가 올라 공원 바닥이 더러워집니다. 청소부를 고용하지 않으면 만족도가 떨어지고, 결국 공원을 떠나는 사람이 늘어납니다. 이 인과관계의 사슬은 당시 게이머들에게 충격적일 만큼 정교했습니다.

물리뉴 자신도 프로그래밍에서 가장 어려웠던 부분이 방문객의 행동이었다고 인정했습니다. 그 어려운 부분을 실제로 코드로 풀어낸 사람이 바로 17세의 데미스였습니다. 상업적 성공은 데미스에게 기술과 시장의 관계를 가르쳤습니다. 아무리 정교한 알 고리즘이라도 사람들이 즐기지 않으면 의미가 없다는 사실, 그리고 대중의 사랑을 받을 때 기술은 비로소 완성된다는 사실을 그는 이 경험에서 체득했습니다.

동시에 그는 화려한 판매 수치 너머를 바라보고 있었습니다. 공원 안을 돌아다니는 수천 명의 가상 인간들이 보여주는 행동 패턴, 그 패턴 속에 숨은 지능의 원형에 그의 시선은 머물러 있었습니다. 사용자에게 반응하는 시뮬레이션 세계 《테마파크》를 전설로 만든 핵심 요소는 화려한 놀이기구가 아니라, 공원 안을 걸어다니는 관람객들의 '두뇌'였습니다.

데미스 허사비스는 리드 프로그래머로서, 당시 하드웨어의 한계 안에서 수천 명의 관람객이 각자 독립적으로 판단하고 행동하는 시스템을 설계했습니다. 각각의 NPC(Non-Player Character)에는 배고픔, 갈증, 피로, 행복, 구역질이라는 다섯 가지 핵심 욕구 수치가 실시간으로 계산되는 내부 상태가 부여되었습니다. 이 설계의 핵심은 '상향식(Bottom-up) 접근'이었습니다.

개발자가 모든 행동을 일일이 지시하는 대신, 캐릭터에게 기본적인 욕구와 반응 규칙만 심어두고, 나머지는 환경과의 상호작용에서 자연스럽게 발생하도록 한 것입니다. 배가 고프면 관람객은 가장 가까운 음식점을 찾아가봅니다. 가격이 터무니없이 비싸면 화를 내며 공원을 떠납니다. 놀이기구 줄이 너무 길면 포기하고 다른 곳으로 향합니다. 분수대 근처를 지나 가면 기분이 좋아지고, 쓰레기가 쌓인 구역을 걸으면 불쾌감이 올라갑니다. 이 개별적인 결정들이 수천 명 분량으로 축적되면, 놀이공원 전체의 운영 상태라는 거시적 현상이 자동으로 만들어집니다. 개체의 미시적 규칙에서 전체의 거시적 질서가 솟아오르는 이 현상을 과학에서는 '창발(Emergence)'이라 부릅니다. 데미스는 이 과정에서 지능의 본질에 대한 중요한 직관을 얻었습니다.

관람객이 줄을 볼 때 기다릴지 포기할지를 결정하는 알고리즘을 짜면서, 그는 불확실한 상황에서 이익을 극대화하려는 인간의 의사결정을 수식으로 옮기고 있었습니다. "어떤 행동을 했을 때 가장 큰 보상을 얻는가?"라는 이 질문은, 훗날 딥마인드에서 핵심 연구 방법론이 되는 '강화학습(Reinforcement Learning)'의 가장 근본적인 물음과 정확히 겹칩니다. 17세의 데미스는 그 용어를 알지 못했지만, 이미 그 원리를 가상 세계 안에서 실험하고 있었던 셈입니다. 물론 이 AI에는 분명한 한계가 있었습니다.

TV 트롭스(TV Tropes) 같은 게임 문화 사이트에서는 《테마파크》 관람객의 '인공적 어리석음(Artificial Stupidity)'을 유머러스하게 기록하고 있습니다. 관람객이 U자형 지형에 끼어 빠져나오지 못하거나, 언뜻 속으로 걸어 들어가 멈춰 서거나, 청소부가 깨끗한 구역만 빙빙 돌며 더러운 구역은 방치하는 현상들이 빈번했습니다. 순찰 경로를 지정하지 않으면 청소부가 대기열 줄 안쪽처럼 쓰레기가 나올 수

없는 곳을 순찰하기도 했습니다. 데미스에게 이런 한계는 실패가 아니라 교훈이었습니다. 그는 이 가상 생명체들이 영리해 보이는 순간과 어처구니없이 멍청해지는 순간의 경계선을 관찰하며,

규칙 기반 AI의 근본적 제약을 체감했습니다.

"진정한 지능이라면, 내가 가르쳐주지 않은 상황에서도 스스로 판단할 수 있어야 한다." 이 생각이 이때부터 그의 머릿속에서 떠나지 않았습니다. 《테마파크》의 관람객 들은 결국 개발자가 정해놓은 규칙의 범위 안에서만 '영리한 척'을 할 수 있었습니다.

규칙 바깥의 새로운 상황이 주어지면 속수무책이었습니다. 데미스는 이 한계를 넘 으려면 게임 코드를 넘어서 인간의 뇌라는 원본 하드웨어를 연구해야 한다는 확신을 품기 시작했습니다. 그가 훗날 케임브리지 대학에서 컴퓨터 과학을 전공한 뒤, 다시 UCL로 돌아가 인 지신경과학 박사 과정에 진학하게 되는 여정의 씨앗은 바로 이 놀 이공원의 가상 인간들 사 이에서 뿌려진 것입니다.

《테마파크》의 NPC는 허사비스에게 '인공지능의 가능성'과 '스스로 학습하는 지 능'이라는 다음 목표'를 동시에 보여준 거울이었습니다. 갭 이어 수입으로 대학 학비를 마련하다 데미스 허사비스의 갭 이어는 지적 성장에 그치지 않았습니다. 경제적으로도 놀라운 결과를 만들어냈습니다.

《테마파크》는 전 세계에서 1,500만 장 이상이 팔린 초대형 히트작이었고, 17세의 리드 프로그래머 에게도 그에 상응하는 보수가 돌아왔습니다. 로열티와 보너스를 합산한 수입은, 또래 학생들이 아르바이트로는 상상할 수 없는 규모였습니다. 그는 이 돈으로 케임브리지 대학의 학비 전액을 직접 지불했습니다.

돈이 생기자 데미스가 한 일 중 하나는 당시 소년들의 로망이었던 스포츠카를 사는 것이었습니다. 그는 포르쉐(Porsche)를 한 대 장만했습니다. 그는 갭 이어를 마치고 이 차를 몰고 케임브리지 대학의 신입생 환영 주간(Fresher's Week)에 나타났습니다.

다른 학생들이 부모 의 차를 타거나 기차를 이용해 캠퍼스에 도착할 때, 데미스는 자 신이 창조한 가상 세계에서 벌어들인 수입으로 산 차를 직접 운전해 왔습니다. 이 장 면은 허사비스의 전기에서 빠 지지 않고 등장하는 유명한 일화입니다. 화려해 보이는 이 에피소드 이면에는 냉정한 자기 인식이 깔려 있었습니다.

그는 자 신의 지능이 시장에서 구체적인 가치를 창출할 수 있는 자본이라는 사실을 이미 확인

한 상태였습니다. 체스 신동이라는 타이틀은 사람들의 관심을 끌었지만, 《테마파크》의 성공은 그 관 심을 실질적인 경제적 성과로 전환할 수 있음을 증명했습니다. 그는 포르쉐를 산 뒤에도 남은 돈을 허투루 쓰지 않았습니다.

미래의 연구와 창업을 위한 자금으로 비축해 두었고, 이 습관은 훗날 엘릭서 스튜디오(Elixir Studios) 창업 과 딥마인드 설립 과정에서 외부 자본 에 대한 의존도를 낮추는 데 기여했습니다. 갭 이어가 끝났을 때, 데미스 허사비스는 이미 여러 겹의 경험을 쌓은 사람이었습니다. 세 계적 히트작의 수석 프로그래머, 수백만 파운드 가치의 프로젝트에 기여한 사 업 참여자, 자기 힘으로 대학 학비를 마련한 자립적인 성인. 그는 포르쉐의 시동을 끄 고 퀸스 칼리지 (Queens' College)의 오래된 도서관으로 걸어 들어갔습니다.

이제 그 의 손에는 게임 컨트롤 러 대신, 알고리즘과 이산수학, 그리고 인공지능의 미래를 담 은 전공 서적들이 들려 있었습니다. 가상 세계를 만들어본 소년은, 이제 그 세계 속 지능의 비밀을 풀기 위한 학문의 세 계로 발걸음을 옮겼습니다. 테마파크(Theme Park) 게임 화면, 1994

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

출전(1995~1997)과 하프 블루(Half Blue) 획득 1995년 3월 4일 토요일, 런던 팰 몰 (Pall Mall)에 자리한 로열 오토모빌 클럽. 100년이 넘는 세월 동안 옥스퍼드와 케임브리지의 최고 두뇌들이 체스판을 사이에 두고 맞붙어 온 이 건 물의 대국실에는 차가 온 정적이 내려앉아 있었습니다. 제113회 바시티 매치(Varsity Match)가 시작되었습니다. 붉은 카펫 위, 구경꾼들은 숨을 죽 이고 말의 움직임을 좇았습니다.

케임브리지 대표팀의 유니폼을 입은 열아홉 살의 데미스 허사비스는 맞은편에 앉은 옥스퍼드의 샤시 자야쿠마르(Shashi Jayakumar)를 조용히 응시 하고 있었습니다. 허사비스는 소년 시절 이미 세계 14세 이하 체스 랭킹 2위에 오른 인물이었습니다. 잉글랜 드 주니어 대표팀 주장을 지냈고, 13세에 마스터 등급(Elo 2300)에 도달했습니다.

수백 번 의 토너먼트를 거치며 승부의 압박감에 단련된 그였지만, 대학을 대표 하여 출전하는 이 자 리는 또 다른 무게를 지니고 있었습니다. 옥스브리지 바시티 매 치는 단순한 체스 대회가 아니었습니다. 1873년에 시작된 이래 영국 학술계의 자존심 이 걸린 대결이었고, 양교의 지 적 전통을 상징하는 연례 의식이었습니다.

허사비스는 빠른 판단력과 깊은 포지셔널 이해를 앞세워 자야쿠마르를 압박했고, 결국 승리를 거 두며 케임브리지의 6대 2 대승에 기여했습니다. 케임브리지 대학교는 체육과 스포츠 분야에서 대학을 대표하여 뛰어난 성적을 거 둔 학생에게 '블루(Blue)'라는 칭호를 수여하는 전통을 갖고 있습니다. 조정이나 럭비, 크리켓처럼 주요 종목에는 '풀 블루(Full Blue)'가, 체스나 펜싱 같은 종목에는 '하프 블 루(Half Blue)'가 주어집니다. 허사비스는 바시티 매치에서의 탁월한 기량을 인정받아 하프 블루를 획득했습니다.

입학 첫해인 1995년부터 졸업반인 1997년까지, 그는 3년 연속으로 이 대회에 출전 하며 케임브리지 체스의 핵심 선수로 활약했습니다.

허사비스가 바시티 매치에서 상대한 옥스퍼드 팀 선수 중에는 다산 쿠마란 (Dharshan Kumaran)이 있었습니다. 당시 두 사람은 체스판 위에서 적군으로 만났습니다. 십수 년 뒤, 쿠마란은 허사비스가 창립한 딥마인드의 핵심 연구자로 합류하게됩니다.

체스판 위에서 수를 겨루던 두 사람이 인류의 지능을 풀기 위한 동료가되었 다는 사실은, 허사비스의 인 생 곳곳에 예고 없이 심어져 있는 연결의 씨앗을 보여줍 니다. 체스는 허사비스에게 게임 이상이었습니다. 그는 대국을 두는 동안 자신의 뇌가 어 떻게 작 동하는지를 끊임없이 관찰했습니다. 다음 수를 예측할 때 직관이 먼저 작동하 는 순간, 논 리적 계산이 그것을 검증하는 순간, 상대의 심리를 읽으려 할 때 발동하는 감정적 판단의 순간. 그는 승패보다 그 과정 자체에 매료되었습니다.

바시티 매치라는 큰 무대는 그에게 승리의 쾌감보다 '생각하는 방식에 대한 생각'이라는 더 깊은 질문을 남겼습니다. 체스판 위에서 발휘되는 인간의 직관과 계산 능력, 그것을 어떻게 하면 기계에게 이 식할 수 있을까. 이 물음은 케임브리지 강의실에서 배우는 컴퓨터 과학 이론과 결합하 며 하나 의 프로젝트로 윤곽을 잡아갔습니다. 로열 오토모빌 클럽의 조용한 대국실에서 수백 가지 수를 읽어내던 그 청년은, 이미 미래 의 과학자이자 CEO로서의 자아를 형성하고 있었습니다.

체스라는 닫힌 시스템 (Closed System) 안에서 완성한 승리의 공식이, 언젠가 현실 세계의 열린 문제(Open Problem)를 해결하는 열쇠가 되리라는 것을 그는 예감하고 있었습니다. A-level을 16세에 마치고 케임브리지에 진학한 배경 데미스 허사비스가 영국 대학 입학시험인 A-level 과정을 모두 마친 것은 열여섯 살 때였습니다. 노스런던의 퀸 엘리자베스 스쿨(Queen Elizabeth's School)과 크리스티 칼리지 핀 칠리 (Christ's College Finchley)를 거치며, 그는 정규 교육 과정이 자신의 지적 욕구를 감당하기에 부족하다는 것을 일찍 깨달았습니다.

수학과 과학 과목을 독학으로 앞서 나갔고, 동급 생들이 채 시작도 하지 않은 범위를 이미 끝내놓곤했습니다. 그는 통 상적인 학생들보다 2 년이나 빨리 고등학교 과정의 마침표를 찍었습니다. 케임브리지 대학교 퀸스 칼리지는 이 16세 소년에게 입학 허가를 내렸습니다.

그러나 대학 측은 한 가지 조건을 달았습니다. 너무 어리다는 것이었습니다. 대학은 그에 게 1년의 유 예 기간, 이른바 갭 이어(Gap Year)를 가질 것을 권고했습니다. 보통의 학생이라면 여행을 떠나거나 자원봉사를 하며 이 기간을 보냈을 것입니다. 허사비스에게 이 1년은 전혀 다른 의미였습니다.

그는 자신이 어린 시절부터 꿈꿔온 게임 개발의 세계로 뛰어들었습니다. 전설적인 게임 디자이너 피터 몰리뉴(Peter Molyneux)가 이끌던 불프로그 프로덕션(Bullfrog Productions)이 그를 받아들였습니다. 17세. 그 나이에 허사비스는 세계적인 히트작 '테마파크(Theme Park)' 의 리드 프로그래머이자 공동 디자이너로 참여했습니다.

놀이공원 시뮬레이션이라는 콘 셉트 아래, 플레이어가 롤러코스터를 설계하고 직원을 고용하며 방문객의 만족도를 관리 하는 이 게임은 전 세계적으로 수백만 장이 팔려나 갔습니다. 10대 후반의 소년이 상업적 성공과 업계의 인정을 동시에 손에 쥔 것입니다. 성공은 유혹을 데려왔습니다.

피터 몰리뉴는 허사비스의 능력에 깊이 감탄하여 파 격적인 제안을 던졌습니다. "대학에 가지 마라. 우리와 계속 일하면 100만 파운드를 주겠다." 1990 년대 초반, 경제적으로 넉넉하지 않은 다문화 가정에서 자란 10대 소년에게 100만 파운드 는 상상 밖의 금액이었습니다.

아버지 코스타스는 장난감 가게를 운영하고 노래를 쓰는 사 람이었고, 어머니 앤젤라는 근검절약으로 가정을 지탱하고 있었습니다. 이 거금이면 가족 모두의 삶이 바뀔 수 있었습니다. 허사비스는 고민했습니다.

짧지 않은 시간이었을 것입니다. 하지만 그의 결론은 분 명했습니다. 거절. 그가 원하는 것은 눈앞의 보상이 아니라 '지능의 비밀'을 풀기 위한 학문적 토대였 기 때문입니다.

게임 개발 현장에서 지능의 가능성을 맛보았지만, 그것 이 코딩 기술만으로 는 해결되지 않는다는 직감이 있었습니다. 컴퓨터 과학의 수학적 원리를 이해해야 했고, 인간의 뇌가 어떻게 작동하는지를 체계적으로 접근해야했습니다. 그 기초를 세울 수 있 는 곳은 게임 스튜디오가 아니라 대학이었습니다.

1994년 가을, 허사비스는 퀸스 칼리지에 입학했습니다. 당시 그가 게임 판매 수익으로 구 입한 포르쉐를 몰고 케임브리지 교정에 나타났다는 일화가 전해집니다. 갓 입 학한 18세 학 생이 스포츠카를 타고 등교한다는 사실은 동료들에게 적잖은 충격이었습니다. 하지만 그 것은 허영이 아니라 증거였습니다. 이 학생이 이미 현실 세계에서 자신의 능력을 검증받 은 사람이라는 증거

말입니다.

허사비스가 택한 전공은 컴퓨터 과학 트라이포스(Computer Science Tripos)였습니다. 케임브리지의 컴퓨터 과학 과정은 앨런 튜링(Alan Turing)이 지적 토대를 놓은 곳이었습니다. 세계 최초의 프로그래밍 가능한 전자 컴퓨터를 연구하던 모리스 윌크 스(Maurice Wilkes)도 이 학과 출신이었습니다.

허사비스는 그 역사의 위에 자신의 지식을 쌓아 올렸습니다. 그는 학점만을 위해 공부하는 학생이 아니었습니다. 불프로 그에서 겪었던 문제들, 이를테면 게임 속 NPC가 좀 더 지능적으로 행동하게 하려면 어떤 알고리즘이 필요한지, 시뮬레이션 세계의 복잡성을 효율적으로 처리하려면 어떤 수학적 구조가 적합한지를 해결할 이론적 무기를 찾기 위해 매달렸습니다.

1997년, 허사비스는 컴퓨터 과학 과정에서 '더블 퍼스트(Double First)'라는 최우수 성적을 받으며 수석으로 졸업했습니다. 더블 퍼스트는 케임브리지의 트라이포스 시험에서 파트 I 과 파트 II 모두에서 1등급(First Class Honours)을 받았다는 뜻으로, 학 과 전체에서 극소수 의 학생만이 달성하는 성과입니다. 16세에 A-level을 끝내고, 100만 파운드의 유혹을 물리 치며, 포르쉐 대신 지식을 선택했던 그의 판단이 틀리지 않았음을 증명하는 순간이었습니다. 케임브리지에서의 3년은 허사비스에게 인공지능이 라는 거대한 산을 오르기 위한 베이스캠프였습니다.

그 베이스캠프에서 정상까지의 경로를 머릿속에 그리고 있었습니다. 라이언헤드 스튜디오에서 'Black & White' AI 프로그래밍 피터 몰리뉴와의 재회와 갓 게임(God Game)의 AI 설계 1997년 여름, 케임브리지를 수석으로 졸업한 허사비 스가 향한 곳은 다시 게임 업계였습니다. 이번에는 지형이 달랐습니다. 그는 더 이상 갭 이어를 보내는 학생이 아니었고, 단순히 고용된 프로그래머도 아니었습니다.

그가 합 류한 곳은 자신의 스승이자 동료였던 피터 몰 리뉴가 새로 세운 라이언헤드 스튜디오(Lionhead Studios)였습니다. 서리(Surrey)주 기포드 (Guildford)의 작은 사무실에서 출발한 이 회사는 당시 게임 업계에서 가장 큰 기대를 모으 고 있었습니다. 몰리뉴라 는 이름 자체가 혁신의 보증수표였기 때문입니다.

몰리뉴는 '포풀러스(Populous)'로 'God Game'이라는 장르를 창시한 인물이었습니다. 플레이어가 신의 시점에서 세계를 내려다보며 땅을 조작하고, 사람들을 이끌고, 문명의 흥망을 좌우하는 이 장르는 게임사에 전례 없는 경험을 선사했습니다. 몰리뉴 는 그 장르의 정점을 찍을 야심작을 구상하고 있었습니다.

'블랙 앤 화이트(Black & White).' 플레이어가 신이 되어 거대한 생명체를 키우고, 마을 사람들을 다스리되, 자

비로운 신이 될 것인지 파괴적 인 신이 될 것인지를 선택해야 하는 게임이었습니다. 이 비전의 핵심을 구현할 적임자로 몰리뉴가 선택한 사람은 허사비스였습니다. 허사비스는 리드 AI 프로그래머(Lead AI Programmer)라는 직함을 받았습니다.

직 함이 뜻하 는 바는 분명했습니다. '블랙 앤 화이트'의 영혼, 즉 게임 세계가 살아 숨 쉬 도록 만드는 모 든 지능적 시스템의 설계와 구현을 그가 책임진다는 것이었습니다. 당 시 대부분의 게임 AI 는 미리 작성된 규칙에 따라 움직이는 방식이었습니다.

적이 나 타나면 공격하고, 체력이 낮아지면 도망가는 식의 조건문(if-then)이 전부였습니다. 하지만 허사비스가 설계해야 할 시스템은 그와는 차원이 달랐습니다. '블랙 앤 화이트'에서 플레이어의 도덕적 선택은 게임 세계 전체에 실시간으로 반영 되어야했습니다.

자비를 베푼다면 하늘은 맑아지고 꽃이 피었습니다. 잔혹하게 굴면 구름이 몰려오고 땅이 갈라졌습니다. 이 모든 변화를 자연스럽게 연결하는 논리적 구조를 설계하는 것이 허사비스의 과제였습니다.

그는 물리뉴와 밤을 지새우며 토론했습니다. 물리뉴가 "플레이어가 마을 사람을 바다에 던지면 어떻게 될까?"라는 질문을 던지면, 허사비스는 그것이 게임 세계의 도덕 수치에 어떻게 반영되고, 주변 마을 사람들의 행동에 어떤 파급 효과를 낳으며, 궁극적으로 세계의 시각적 분위기를 어떻게 바꿀지를 계산했습니다. 허사비스는 인공지능이 플레이어의 명령을 수행하는 데 그쳐서는 안 된다고 주장 했습니다. 인공지능이 플레이어의 의도를 파악하고, 그에 맞는 성격을 형성해야 한다는 것이 그의 신념이었습니다.

이는 당시의 기술 수준으로는 대담한 발상이었습니다. 그러나 그에게는 케임브리지에서 3 년간 다진 이론적 배경이 있었습니다. 그는 컴퓨터 과학에서 배운 탐색 알고리즘, 확률 이론, 초기 형태의 기계학습 개념들을 게임이라는 실험실에 투입했습니다.

이론과 실천의 결합. 이것이 케임브리지에서 추구한 것이었고, 라이언헤드에서 실현 하기 시작한 것이었습니다. 라이언헤드의 초창기 팀은 8명에 불과했습니다. 허사비스는 그중 가장 젊은 측에 속했지만, AI라는 영역에서만큼은 누구보다 강력한 발언권을 가졌습니다. 물리뉴가 창의적 영감을 던지면, 허사비스가 그것을 논리적 시스템으로 번역했습니다.

두 사람의 관계는 비전을 꿈꾸는 건축가와 그것을 구조적으로 실현하는 엔지니어의 협업에 가까웠습니다. 이 시기 허사비스는 게임을 오락으로 보지 않았습니다. 그는 게임이야말로 현실 세계의 복잡한 문제를 압축하여 실험할 수 있는 최적의 시험대라고 확신했습니다.

God Game이라는 장르는 지능이 방대한 데이터를 처리하고, 장기적인 계획을 세우며, 인간과 상호작용하는 법을 배우기에 이상적인 환경이었습니다. 훗날 딥마인드에서 핵심이 되는 '강화학습(Reinforcement Learning)'의 초기 아이디어들은 바로 이 시기, 기포드의 좁은 사무실에서 싹을 틔우기 시작했습니다. 플레이어의 행동에 학습하는 크리처(Creature) AI의 혁신 '블랙 앤 화이트'의 핵심은 게임 속에 등장하는 '크리처(Creature)'였습니다.

원숭이, 사자, 소, 호랑이 중 하나를 선택하면 그 생명체가 플레이어의 분신이자 제자가되었습니다. 크리처는 플레이어가 어떻게 가르치느냐에 따라 전혀 다른 존재로 자라났습니다. 자비로운 교육을 받으면 마을 사람들을 돌보는 수호자가 되었고, 폭력적인 훈육을 받으면 마을을 짓밟는 괴물이되었습니다.

허사비스가 리드 AI 프로그래머로서 이 크리처에 심어 넣은 학습 시스템은 게임 역사를 통틀어 가장 정교한 인공지능 중 하나로 평가받습니다. 크리처의 뇌는 여러 겹의 학습 메커니즘으로 구성되어 있었습니다. 가장 기본적인 층위에는 욕구 시스템(Desire System)이 있었습니다.

배고픔, 피로, 호기심, 공포. 크리처는 이러한 내적 상태를 수치로 가지고 있었고, 이 수치를 충족하기 위한 최선의 행동을 탐색했습니다. 배가 고프면 나무에서 과일을 따 먹었고, 호기심이 높아지면 낮은 지역을 탐험했습니다. 여기까지는 이전의 게임 AI에서도 시도된 적이 있는

방식이었습니다. 허사비스가 이 시스템을 근본적으로 다르게 만든 것은 그 위에 올려놓은 관찰 학습(Observational Learning) 과 강화학습 구조였습니다. 크리처는 플레이어의 행동을 끊임없이 관찰했습니다.

플레이어가 마을 사람들에게 음식을 나눠주면, 크리처는 그 행위를 목격하고 '이것이 좋은 행동'이라는 신호를 내 부적으로 기록했습니다. 핵심은 피드백 시스템이었습니다. 플레이어는 손 모양의 커서로 크리처를 쓰다듬거 나 때 릴 수 있었습니다. 크리처가 마을 사람을 도왔을 때 쓰다듬으면 긍정적 보상으

로 작용했 고, 크리처가 나무를 뽑아 던졌을 때 때리면 부정적 신호로 작용했습니다. 허사비스는 이 보상과 처벌 메커니즘에 신경망(Neural Networks)의 초기 형태를 결 합하여, 크리처가 시간 이 지남에 따라 자신만의 행동 패턴을 형성하도록 설계했습니다. 이전까지의 게임 캐릭터들은 개발자가 미리 입력한 시나리오에만 반응했습니다.

허 사비 스의 크리처는 달랐습니다. 예상하지 못한 행동이 나타났습니다. 어떤 크리처는 마을 사람들을 돕는 것을 넘어 스스로 농사를 짓기 시작했습니다.

어떤 크리처는 플레이어의 관심을 끌기 위해 바위를 집어 던지는 장난을 쳤습니다. 잔인하게 키워진 크리처가 다른 마을의 크리처를 공격하는가 하 면, 자비롭게 키워진 크리처는 부상당한 마을 사람에게 다가가 음식을 건 네기도했습니다. 이러한 '예측 불가능한 지능적 행동'은 플레이어들에게 경탄과 몰입감을 동시에 안 해주었습니다.

허사비스는 크리처의 학습 과정을 설계하면서 인간 유아의 발달 과정을 참고했습니다. 아이가 백지 상태(Tabula Rasa)에서 시작하여 부모의 행동을 모방하고, 보상과 처 벌을 통해 사회적 규범을 익혀가는 과정. 그는 인공지능이 바로 이 방식으로 학습할 때 가장 강력한 지능을 구현할 수 있다고 믿었습니다. 이 생각은 훗날 딥마인드의 알파고 제로(AlphaGo Zero)에서 극적으로 실현됩니다.

바둑의 규칙만 알려준 채 수백만 번의 자가 대국을 통해 스스로 초인적 실력에 도 달한 그 시스템 의 지적 뿌리가 바로 여기, 라이언헤드의 크리처에 있었습니다. 기술 적 수준은 차이가 났 지만, '경험을 통한 지능의 획득'이라는 핵심 사상은 동일했습니다. 게이머들은 자신이 키운 크리처가 점점 자기 성향을 닮아가는 모습에 놀라워했습니다.

공 격적인 플레이어의 크리처는 공격적으로 자랐고, 평화로운 플레이어의 크리 처는 온순하 게 자랐습니다.

이 현상은 허사비스에게 깊은 인상을 남겼습니다. 인공지능은 거울이었습니다. 창 조주의 가치관이 고스란히 피조물에 반영된다는 사실은, 기술 자체는 중립적이지만 그것을 가르 치는 데이터와 환경은 중립적이지 않다는 진실을 보여주었습니다.

훗날 허사비스가 '가치 정렬(Alignment)'의 중요성을 강조하고, AI 안전 연구에 진지하게 투자하게 된 배경에는 이 게임 개발 현장의 경험이 놓여 있습니다. 2001년 '블랙 앤 화이트'가 세상에 나왔을 때, 전 세계 평론가들은 이 게임의 인공지능에 찬 사를 보냈습니다. 허사비스는 게임 개발자로서 정점에 섰습니다.

하지만 그는 거기서 멈추 지 않았습니다. 가상 세계의 캐릭터가 이토록 지능적으로 행동할 수 있다면, 진짜 인간의 지능은 어 떤 메 커니즘으로 작동하는 것일까. 이 질문이 그를 다시 학문의 세계로, 그리고 인간 의 뇌를 직 접 연구하는 신경과학의 길로 인도했습니다. 크리처 AI의 개발은 허사비스에게 결정적인 확신 하나를 심어주었습니다.

지능을 이해하기 위한 가장 강력한 도구 는 프로그래밍 기술 이 아니라 '학습'이라는 원리 그 자체라는 확신이었습니다. 케임브리지 대학교 퀸스 칼리지(Queens' College)

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

1998년 독립 게임 스튜디오 설립 리퍼블릭: 더 레볼루션(Republic: The Revolution)' 가상 국가 전체를 인공지능으로 시뮬레이션하겠다는 데미스 허사비스의 야심은 1990년 대 말 게임 산업뿐만 아니라 과학계 전체를 통틀어도 가장 대담한 도전 중 하나였습니다. 1998년, 약관의 나이에 엘릭서 스튜디오를 설립한 허사비스가 처음으로 내놓은 청사진은 '리퍼블릭: 더 레볼루션(Republic: The Revolution)'이라는 이름의 정치 시뮬레이션 게임이었습니다. 이 게임은 단순히 버튼을 눌러 숫자를 올리는 전략 게임이 아니었습니다.

허사비스는 그 안에 '살아 숨 쉬는 사회'를 통째로 집어넣고자했습니다. 그는 이미 '테마파크(Theme Park)'와 '블랙 앤 화이트(Black & White)'를 통해 가상 세계의 주민들이 스스로 판단하고 행동하게 만드는 기술의 매력을 맛보았습니다. 하지만 그는 만족하지 못했습니다.

더 거대한 것, 즉 수천 명의 시민이 각자의 신념과 직업, 가족 관계를 가지고 실시간으로 상호작용하는 거대한 '사회적 실험장'을 꿈꿨습니다. 그에게 게임은 단순한 오락이 아니라, 지능의 본질을 탐구하기 위한 가상 실험실(Petri dish)이었습니다. 허사비스는 가상의 동유럽 국가인 '노비스트라나(Novistrana)'를 창조했습니다.

이 나라의 수도인 베레지나(Berezina)를 비롯한 여러 도시는 수천 개의 건물과 광장으로 이루어졌고, 그 길 위를 걷는 모든 시민은 엘릭서가 개발한 독자적인 인공지능 엔진에 의해 조종되었습니다. 당시 대부분의 게임에서 배경 인물(NPC)은 미리 정해진 경로를 따라 걸거나 간단한 대사만 반복하는 인형에 불과했습니다. 그러나 '리퍼블릭'의 시민들은 달랐습니다.

그들은 아침에 일어나 출근하고, 점심에는 공원에서 휴식을 취하며, 저녁에는 선술집에서 정치적 불만을 토로했습니다. 사용자가 그들에게 선전물을 뿌리거나 매수하려고 시도하면, 시민들은 자신의 성향과 처한 상황에 따라 각기 다른 반응을 보였습니다.

이러한 시도는 기술적으로 엄청난 난관이었습니다. 허사비스는 한 명의 천재적인 AI를 만드는 대신, 수만 명의 '단순한 지능'들이 모여 어떻게 복잡한 '사회적 창발성'을 만들어내는지를 구현하고 싶어했습니다. 그는 이를 위해 '로터스(Lotus)'라는 혁신적인 3D 엔진을 개발했습니다.

이 엔진은 도시의 가장 높은 곳에서 내려다보는 시점부터 시민의 눈높이에서 신문의 헤드라인을 읽을 수 있는 수준까지 자유로운 확장을 지원했습니다. 이는 2000년대 초반의 하드웨어 사양으로는 불가능에 가까운 도전이었습니다. 허사비스의 머릿속에서 '리퍼블릭'은 인류 역사상 가장 정교한 사회적 시뮬레이터였습니 이 프로젝트를 통해 인간의 사회적 행동 양식을 코드로 공식화할 수 있다고 믿었습니다. 전 세계 게임 언론은 이 소년 천재의 야심에 열광했습니다.

"모든 시민이 뇌를 가지고 있다"는 홍보 문구는 게이머들의 가슴을 뛰게했습니다. 하지만 이 거대한 꿈은 동시에 엘릭서 스튜디오를 서서히 압박하는 족쇄가 되기도 했습니다. 현실의 국가를 시뮬레이션하겠다는 욕심은 그만큼 해결해야 할 변수가 무한하다는 의미였기 때문입니다. 허사비스는 인공지능이 해결해야 할 궁극적인 숙제가 '현실 세계의 복잡성'을 어떻게 추상화하고 처리하느냐에 달려 있음을 온몸으로 체감하고 있었습니다.

거대한 범위로 인한 개발 지연 꿈이 클수록 그 꿈을 현실의 그릇에 담아내는 과정은 고통스럽습니다. 데미스 허사비스와 엘릭서 스튜디오는 '리퍼블릭: 더 레볼루션'을 개발하며 창업가로서 가장 쓰라린 교훈을 얻게 됩니다. 바로 '범위의 함정(Scope Creep)'입니다.

당초 2년 정도로 예상했던 개발 기간은 5년으로 늘어났고, 전 세계의 기대를 한 몸에 받았던 게임은 2003년이 되어서야 비로소 세 상의 빛을 보게되었습니다. 개발 지연의 가장 큰 원인은 허사비스가 설정한 시뮬레이션의 깊이였습니다. 수천 명의 인 공지능 에이전트가 실시간으로 상호작용하는 시스템을 구현하다 보니 하드웨 어 자원이 바닥나기 일쑤였습니다.

CPU는 시민들의 일정을 계산하느라 비명을 질렸 고, 메모리는 거 대한 도시의 디테일을 담아내기에 턱없이 부족했습니다. 허사비스는 완벽주의자였습니 타협하기보다 기술적 한계를 돌파하기 위해 밤낮으로 코드를 수정했습니다. 하지만 퍼 블리셔인 에이도스(Eidos)의 압박은 거세졌고, 결국 엘릭서는 게 임의 많은 핵심 기능을 쳐 내야 하는 상황에 직면했습니다.

가장 아쉬웠던 점은 시뮬레이션의 결과가 게임 플레이의 재미로 직결되지 않았다는 점입 니다. 허사비스는 시민 한 명 한 명의 지능을 구현하는 데 온 힘을 쏟았지만, 정작 플레이어 가 그들과 상호작용하는 방식은 보드게임처럼 다소 단조로운 선택의 연속으로 변질되었습니다. 시민들이 내면적으로 어떤 복잡한 계산을 거치든, 플레이 어의 화면에는 간단한 아 이콘과 숫자로만 나타났습니다. 공들여 만든 인공지능의 깊 이를 독자들이 직접 느끼기에 는 장벽이 너무 높았습니다.

2003년 출시된 '리퍼블릭'에 대한 시장의 반응은 냉담했습니다. 비평가들은 "기술적 성취 는 놀랍지만, 게임으로서의 재미는 부족하다"는 평가를 내놓았습니다. 리뷰 집계 사이트인 메타크리틱에서 받은 62점이라는 점수는, 허사비스가 이전에 참여했 던 '블랙 앤 화이트'나 '테마파크'의 영광에 비하면 초라한 성적이었습니다.

팬 들은 "너무 많 은 것을 약속했지만, 정작 손에 쥘 것은 반복적인 클릭뿐"이라며 실망감을 드러냈습니다. 이 실패는 허사비스에게 돈으로 살 수 없는 경험을 선사했습니다. 그는 처음으로 수 십 명 의 엔지니어와 예술가를 이끄는 CEO로서 팀을 관리하고, 거대 기업과의 협상 테이블에서 밀고 당기는 법을 배웠습니다.

또한, 아무리 뛰어난 기술이라도 사용자(플레이어)가 그 가 치를 즉각적으로 이해하지 못하면 시장에서 살아남을 수 없다는 비즈 니스의 냉혹함도 깨 달았습니다. 아이작슨이 묘사한 스티브 잡스처럼, 허사비스 역시 자신의 '현실 왜곡 역장' 이 기술적 물리 법칙과 충돌했을 때 어떤 결과가 초래되는지를 뼈저리게 배운 것입니다. 이 시기의 좌절은 훗날 그가 딥마인드를 운영할 때 '연구 의 혁신성'과 '제품의 실행력' 사이에서 정교한 균형을 잡게 만드는 토양이되었습니다.

범용 인공지능(AGI)을 만들기엔 아직 이르다 '리퍼블릭'의 출시 이후 허사비스는 깊은 고민에 빠졌습니다. 그는 단순히 '재미있는 게임' 을 만들지 못해서 실패했다고 생각하지 않았습니다. 그보다는 자신이 목표로 했던 '지능의 재현'이라는 과제가 당시 의 컴퓨터 과학 수준으로는 도저히 도달할 수 없는 높은 벽이었음을 깨달았습니다.

그는 엘릭서 스튜디오의 사무실에서 밤을 지새우며 질문했습니다. "왜 인공지능은 여전히 실제 인간처럼 유연하게 생각하지 못하는가?" 그가 직면한 가장 큰 벽은 '지능 의 일반성' 부족이었습니다. '리퍼블릭'의 시민들은 정치적 시뮬레이션이라는 닫힌 세

계(Closed world) 안에서는 나름의 논리에 따라 움직였습니다. 하지만 그들은 새로운 상황에 적응하거나, 스스로 학습하여 발전하는 모습을 보여주지 못했습니다. 프로그래머가 미리 짜놓은 '이럴 땐 저렇게(if-Then)' 식의 알고리즘 뭉치는 아무리 정교해 도 진짜 지능처럼 느껴지지

않았습니다.

허사비스는 이때 인공지능이 한 단계 도약하기 위 해서는 단순히 더 빠른 하드웨어나 더 복잡한 규칙이 필요한 것이 아니라 는 사실을 직감했습니다. "우리는 지능을 풀고 싶어 하지만, 정작 지능이 무엇인지조차 모른다." 당시 게임 산 업에서 사용되던 인공지능 기술은 일종의 '숙임수'에 가까웠습니다.

사용자가 지능적 이라고 느끼 게끔 교묘하게 연출된 스크립트였을 뿐입니다. 허사비스는 이러한 가짜 지능이 아니라, 어 떤 환경에 던져져도 스스로 학습하고 문제를 해결하는 '범용 인공지능(AGI)'을 원했습니다 하지만 2000년대 초반의 기술 환경에서 AGI는 그저 공상과 학 소설 속의 이야기일 뿐이었습니다. 허사비스는 깨달았습니다.

컴퓨터 과학만으로는 이 문제를 풀 수 없다는 것을 말입 니다. 인간의 뇌가 어떻게 지능을 만들어내는지, 기억과 상상력은 어떤 신경학적 메커 니즘을 거 치는지 이해하지 못한 채 지능을 코드로 옮기려는 시도는 마치 설계도 없이 건물을 지으려 는 것과 같았습니다. 그는 자신이 만든 엘릭서 스튜디오가 상업적인 게 임 개발에 매몰되 어 가는 것을 경계했습니다. 그는 더 큰 문제를 풀고 싶었습니다.

"지금의 도구로는 안 된 지능을 근본적으로 이해하기 위해선 다시 기초 과학으로 돌아 가야 한다." 이 결정은 허사비스 인생의 가장 극적인 전환점이됩니다. 2005년 스튜디오를 매각하고 안정적인 수익이 보장되던 게임 업계를 떠나기로합니다.

주변에서는 "백만 파운드 이상의 연봉을 받을 수 있는 위치를 왜 포기하느냐"며 말렸지만, 그의 시 선은 이미 체스판이나 모 니터 너머의 세상을 향해 있었습니다. 그는 인지 신경과학(Cognitive Neuroscience) 박사 과정을 밟기 위해 대학으로 돌 아가기로 결심합니다. 엘릭서 스튜디오에서의 7년은 그에게 '성공한 창업가'라는 타이 틀 대신 '지능 의 정복자'가 되기 위한 절실한 동기부여를 남겼습니다.

"범용 인공지능을 만들기엔 아직 이르다. 하지만 준비를 시작하기에는 가장 적절한 때다." 이 확신이 훗날 딥마인드의 씨앗 이되었습니다. '이블 지니어스(Evil Genius)'와 BAFTA 인터랙티브 음악 제임스 해니건(James Hannigan) 엘릭서 스튜디오의 두 번째 작품인 '이블 지니어스(Evil Genius)'는 허사비

스의 또 다른 천 재적 면모를 보여준 작품이었습니다. 이 게임은 플레이어가 정의를 수호하는 영웅이 아니 라, 비밀 기지에서 세계 정복을 꿈꾸는 악당이 된다는 신선한 발상에서 출발했습니다. 허 사비스는 이 독창적인 테마를 완성하기 위해 게임의 분위 기를 결정짓는 '소리'에 주목했습니다.

당시 영국의 저명한 게임 음악 작곡가인 제임 스 해니건(James Hannigan)을 엘릭서로 초대했습니다. 허사비스와 해니건의 협업은 단순히 배경 음악을 주문하고 제작하는 수준을 넘어 섰습니 두 사람은 음악이 게임의 상황에 따라 실시간으로 변해야 한다는 '인터랙티브 음악'의 비전을 공유했습니다. 악당의 비밀 기지에 첩보원이 침입했을 때, 플레이어가 사악한 계획을 성공시켰을 때, 혹은 부하들이 한가롭게 휴식을 취할 때의 음악이 자 연스럽게 이어져야했습니다.

허사비스는 해니건에게 게임 속 인공지능 시스템과 음 악 엔진이 긴밀하게 연동 되도록 요구했습니다. 해니건은 1960년대 첩보 영화의 거장 존 배리의 음악 스타일을 현대적으로 재해석 하여 '이 블 지니어스'만의 독특하고도 우스꽝스러운 악당 테마를 만들어냈습니다. 웅 장한 오케스

트라 사운드와 복고풍의 전자음이 섞인 이 음악은 게임의 풍자적인 분위기를 완벽하게 살려냈습니다.

허사비스는 해니건의 창의성을 전폭적으로 지지했고, 기술적인 지원을 아끼지 않았습니다. 음악이 단순한 장식이 아니라 게임의 '인지적 경험'을 강화하는 핵심 요소라고 믿었기 때문입니다. 이러한 혁신적인 시도는 업계의 인정을 받았습니다.

'이블 지니어스'의 사운드트랙은 영국 영화 텔레비전 예술 아카데미(BAFTA) 인터랙티브 부문에서 음악상 후보에 오르는 쾌거를 이루었습니다. 이는 독립 스튜디오인 엘릭서가 거대 자본이 투입된 블 록버스터 게임들과 어깨를 나란히했다는 점에서 큰 의미가 있었습니다. 허사비스는 해니건과의 작업을 통해 서로 다른 분야의 전문가들이 하나의 비전 아래 모였을 때 발생하는 '창의적 시너지'의 힘을 다시 한번 확인했습니다.

아이작슨이 스티브 잡스와 조니 아이브의 협력을 묘사했듯이, 허사비스와 해니건의 만남은 기술과 예술의 절묘한 결합이었습니다. 허사비스는 논리적이고 수학적인 구조를 제공했고, 해니건은 그 위에 인간의 감성을 자극하는 선율을 입혔습니다. 훗날 딥마인드에서 허사비스가 인공지능 연구자뿐만 아니라 철학자, 예술가, 신경 과학자 등을 한데 모아 '다학제적 팀'을 구성한 배경에는 엘릭서 시절 제임스 해니건과 함께했던 이 빛나는 협업의 기억이 자리 잡고 있습니다.

시뮬레이션 게임을 통해 AI 에이전트와 창발적 현상을 실험하다 '이블 지니어스'는 '리퍼블릭'에서 실패했던 인공지능 시뮬레이션의 꿈을 좀 더 영리하고 효과적으로 구현한 작품이었습니다. 허사비스는 '리퍼블릭'의 실패를 통해 너무 광범위한 시뮬레이션은 제어하기 어렵다는 것을 배웠습니다. 그래서 '이블 지니어스'에서는 비밀 기지라는 '통제된 환경' 안으로 범위를 좁혔습니다.

플레이어의 명령을 따르는 수백 명의 부하(Minions)와 그들을 방해하는 정부 요원들은 각각 독립적인 지능을 가진 'AI 에이전트'로 설계되었습니다. 허사비스가 이 게임에서 특히 주목한 것은 '창발성(Emergence)'이었습니다. 창발성이란 개별 요소들은 단순한 규칙을 따르지만, 그들이 모였을 때 전체 시스템에서 예측하지 못한 복잡한 행동이 나타나는 현상을 말합니다.

'이블 지니어스'의 기지 안에서 부하들은 배고픔, 피로, 충성도라는 간단한 수치에 따라 움직입니다. 하지만 수백 명의 부하가 기지 안의 가구 배치나 플레이어의 갑작스러운 명령과 상호작용하기 시작하면, 개발자인 허사비스조차 예상하지 못한 독특한 상황들이 발생했습니다. 예를 들어, 기지 내 식당의 위치가 숙소와 너무 멀면 부하들이 이동 중에 지쳐 쓰러 지거나, 특정 구역에 인파가 몰려 병목 현상이 생기기도 했습니다.

허사비스는 이러한 현상을 관찰하며 희열을 느꼈습니다. 게임 개발자로서 "플레이어에게 어떤 경험을 줄 것인가"를 고민하는 동시에, 과학자로서 "어떻게 하면 이 가상 세계에 생명력을 불어넣을 것인가"를 탐구했습니다. 그에 게 게임 속 에이전트들은 지능의 기본 단위를 테스트하는 피실험자였습니다.

인공지능이 복잡한 환경에서 목표를 달성하기 위해 '계획(Planning)'을 세우는 과정을 실험했습니다. 적 요원이 침입했을 때 부하들이 스스로 경보를 울리고, 적을 함정으로 유인하며, 부상당한 동료를 치료하는 과정은 정교하게 짜인 스크립트가 아니라 에이전트들의 자율적인 판단에 의해 이루어지도록 설계되었습니다. 훗날 딥마인드가 강화학습(Reinforcement

Learning)을 통해 시가 스스로 최적의 전략을 찾 아내게 만드는 방식과 궤를 같이합니다.

'이블 지니어스'는 평론가들로부터 "심오한 시뮬레이션과 유머가 어우러진 수작"이라는 평가를 받았습니다. 허사비스는 이 게임을 통해 자신이 꿈꾸던 '시스템 기반의 게임 디자인'이 시장에서도 통할 수 있음을 증명했습니다.

완전한 AGI에 도달하기엔 기술적 한계가 여전했지만, 그는 가상 세계라는 샌드박스 안에서 지능의 파편들을 하나씩 조립해 나가는 즐거움을 만끽했습니다. 이 시기의 실험들은 허사비스에게 "지능은 정적인 규칙이 아니라, 환경과의 끊임없는 상호작용 속에서 피어나는 역동적인 과정"이라는 확고한 철학을 심어주었습니다. 백만 파운드 스카우트 제의 거절과 2005년 스튜디오 매각 게임 산업에서 배운 것: 팀 관리, 일정, 실패 비용, 리더십의 첫 연습 2005년, 데미스 허사비스는 엘릭서 스튜디오를 정리하며 자신의 인생에서 가장 화려했던 게임 개발자로서의 첫 번째 챕터를 마무리합니다.

스튜디오를 매각하고 청산하는 과정은 결코 쉽지 않았지만, 그는 이 과정에서 얻은 무형의 자산이 그 어떤 성공보다 값지다는 사실을 알고 있었습니다. 엘릭서 스튜디오는 허사비스에게 '리더십의 혹독한 훈련장'이었습니다. 허사비스는 20대 초반의 나이에 수십 명의 직원에게 월급을 주는 경영자(CEO)가 되었습니다. 그는 천재적인 프로그래머였지만, 사람의 마음을 움직이고 조직의 효율을 극대화하는 관리의 영역은 완전히 다른 문제였습니다.

그는 '리퍼블릭'의 개발 지연을 겪으며 프로젝트 일정 관리의 중요성과 '선택과 집중'의 가치를 배웠습니다. 아무리 혁신적인 아이디어라도 정해진 시간과 예산 안에서 구현되지 못하면 실패로 돌아간다는 경영의 기본 원칙을 몸소 체험한 것입니다. 실패의 비용에 대해서도 깊이 깨달았습니다.

'리퍼블릭'의 미온적 성과와 그로 인한 재정적 압박은 허사비스에게 큰 스트레스였습니다. 하지만 그는 좌절하는 대신, 무엇이 잘못되었는지를 체스 대국을 복기하듯 철저히 분석했습니다. 그는 자신이 기술적 완벽주의에 빠져 시장의 속도와 사용자의 요구를 간과했음을 인정했습니다.

이러한 자기 객관화는 훗날 그가 구글과 같은 거대 기업 안에서 딥마인드라는 거대한 조직을 이끌 때, 연구의 자율성을 지키면서도 상업적 성과를 내는 탁월한 전략가로 성장하는 밑거름이 되었습니다. 또한, 그는 인재 확보의 중요성을 절감했습니다. 제임스 해니건과 같은 각 분야의 최고 전문가들과 협업하며, 그는 리더의 역할이 단순히 지시를 내리는 것이 아니라 "똑똑한 사람들이 신나게 일할 수 있는 환경을 만드는 것"임을 배웠습니다.

엘릭서 출신의 많은 개발자가 훗날 영국 게임 산업의 주역이 되고, 일부는 딥마인드 초기 멤버로 합류하게 된 것은 허사비스가 초기부터 사람에 투자했음을 보여줍니다.

아이작슨이 묘사한 젊은 시절의 엘론 머스크나 스티브 잡스가 첫 창업에서 겪었던 시행착오처럼, 허사비스에게 엘릭서 스튜디오는 거친 원석이었던 그를 다듬어준 숫돌이었습니다. 그는 이곳에서 승리의 기쁨보다 패배의 교훈을 더 많이 얻었지만, 그 교훈이야말로 지능이라는 거대한 산을 정복하기 위해 반드시 필요한 장비였습니다. 그는 게임 산업을 떠나며 더 이상 일개 게임 디자이너가 아니었습니다. 그는 '세상을 바꾸기 위해 지식을 조직하는 법'을 아는 젊은 리더로 거듭나 있었습니다.

"게임은 단순한 오락이 아닌, AI를 훈련시키는 완벽한 샌드박스"라는 확신 엘릭서 스튜디오를 떠나며 내린 결론은 "게임은 인공지능을 완성하기 위한 최고의 도구" 라는 확신이었습니다. 사람들은 그가 게임 업계를 떠나 과학의 세계로 귀순한다고 생각했지만, 허사비스의 머릿속에서 두 세계는 분리된 적이 없었습니다. 그는 단지 더 강력한 도구를 얻기 위해 잠시 전장을 옮기는 것뿐이었습니다.

그는 엘릭서에서의 경험을 통해 현실 세계를 인공지능으로 시뮬레이션하는 것이 얼마나 어려운지 깨달았습니다. 하지만 동시에 게임이라는 가상 세계가 현실의 복잡함을 안전하고 빠르게 재현할 수 있는 유일한 통로라는 것도 확인했습니다. "만약 인공지능이 복잡한 전략 게임에서 승리할 수 있다면, 그 인공지능은 현실의 복잡한 과학 난제도 해결할 수 있지 않을까?"라는 가설이 그의 머릿속에 뿌리내렸습니다.

이 확신은 훗날 딥마인드의 근간이 된 '샌드박스 철학'으로 이어집니다. 허사비스는 인공지능을 처음부터 신약 개발이나 기후 변화 예측에 투입하지 않았습니다. 대신 그는 벽돌 깨기, 스페이스 인베יד러 같은 고전 아타리 게임과 체스, 바둑 같은 보드게임을 AI의 훈련장으로 삼았습니다. 게임은 규칙이 명확하고 성공과 실패의 보상이 뚜렷하며, 무엇보다 수만 번의 시뮬레이션을 순식간에 돌려볼 수 있는 완벽한 실험실이었기 때문입니다.

허사비스가 불프로그 시절 겪었던 "백만 파운드 스카우트 제의"를 거절하고 대학으로 향했던 일화는 그가 얼마나 장기적인 시각을 가진 인물인지를 보여줍니다. 눈앞의 돈이나 명성보다 '지능을 풀겠다'는 궁극적인 미션에 더 충신했습니다. 엘릭서를 매각한 돈을 자신의 다음 단계인 신경과학 연구를 위한 발판으로 삼았습니다. 게임 산업에서 지능의 외형(시뮬레이션)을 만들었다면, 이제는 뇌 과학을 통해 지능의 본질(알고리즘)을 배우려 간 것입니다.

게임이라는 거울을 통해 인공지능의 한계를 보았고, 그 한계를 넘어서기 위한 지도를 그렸습니다. 2005년 엘릭서의 문을 닫고 유니버시티 칼리지 런던(UCL)의 도서관으로 향하던 그의 발걸음에는, 조만간 다시 가상 세계로 돌아와 진짜 지능을 구현해내겠다는 서늘한 집념이 서려 있었습니다. "게임은 끝났지만, 퀘스트는 이제 시작이다."

그가 전 세계 게이머와 과학계에 던진 보이지 않는 메시지였습니다. 엘릭서 스튜디오(Elixir Studios) 로고

인공지능 전문가 김경진 변호사

AI 법정정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

인공지능을 만들려면 자연 지능을 알아야 한다 UCL 인지신경과학 박사 과정 2005년의 어느 차가운 아침, 런던 중심가에 위치한 엘릭서 스튜디오(Elixir Studios)의 사무실 창밖을 내다보던 데미스 허사비스는 결심을 굳혔습니다. 그는 이미 스무 살 초반에 자신의 게임 회사를 차려 수십 명의 직원을 거느린 성공한 기업가였고, 그가 만든 게임들은 수백만 장이 팔려나갔습니다. 하지만 마음속에는 늘 채워지지 않는 갈증이 있었습니다.

12살 소년 시절, 체스판 앞에서 "생각이란 무엇인가"를 고민하며 컴퓨터를 샀던 그 근본적인 질문으로 돌아가야 할 때라고 느낀 것입니다. 당시 게임 산업은 점점 더 화려한 그래픽과 자극적인 재미에만 몰두하고 있었고, 허사비스가 꿈꾸던 '지능을 가진 가상 캐릭터'를 만드는 일은 기술적 한계와 상업적 논리에 부딪혀 있었습니다. 그는 주변의 만류를 뒤로하고 회사를 정리했습니다.

그리고는 가방 하나를 메고 유니버시티 칼리지 런던(UCL)의 교정을 밟았습니다. 허사비스가 인공지능을 연구하기 위해 컴퓨터 공학이 아닌 뇌과학, 즉 인지신경과학(Cognitive Neuroscience)을 선택한 것은 매우 전략적이고도 철학적인 결정이었습니다. 그는 당시의 인공지능 기술이 단순히 수학적 계산이나 통계적 확률에만 의존하고 있다는 점에 주목했습니다.

진정한 지능, 즉 인간처럼 새로운 상황에 적응하고 스스로 학습하며 창의적인 해결책을 내놓는 범용 인공지능(AGI)을 만들려면, 이미 그 기능을 완벽하게 수행하고 있는 유일한 존재인 인간의 뇌를 먼저 이해해야 한다고 믿었습니다. 런던의 유서 깊은 대학인 UCL은 뇌의 지도를 그리는 신경과학 분야에서 세계 최고의 권위를 자랑하는 곳이었습니다. 그는 이 곳에서 화려한 CEO의 직함을 내려놓고, 매일 아침 일찍 도서관에 앉아 두꺼운 신경과학 전공 서적을 탐독하는 늦깎이 대학원생으로 돌아왔습니다.

박사 과정 동안 허사비스는 단순히 학위를 따는 것에 만족하지 않았습니다. 그는 뇌의 각 부위가 어떻게 서로 통신하며, 우리가 보고 듣는 정보를 어떻게 '지식'으로 바꾸는지에 대한 근본적인 원리를 파헤쳤습니다. 특히 그는 인간이 과거의 경험을 어떻게 기억하고, 그 기억을 바탕으로 어떻게 미래의 일을 미리 그려보는지에 깊은 관심을 가졌습니다.

이 과정은 훗날 딥마인드(DeepMind)가 만들어낸 인공지능 알고리즘의 핵심인 '강화 학습'과 '경험 재생(Experience Replay)'의 토대가 되었습니다. 그는 뇌를 생물학적인 기관으로만 보지 않고, 우주에서 가장 효율적인 정보를 처리하는 일종의 '알고리즘 기계'로 보았습니다. 이 4년간의 학문적 몰입은 그가 인공지능 역사상 가장 독특한 위치를 차지하게 만든 결정적인 전환점이 되었습니다. 지도교수 엘리너 맥과이어(Eleanor Maguire) UCL 연구실의 문을 두드린 허사비스를 맞이한 사람은 엘리너 맥과이어 교수였습니다.

맥과이어 교수는 전 세계적으로 유명한 '런던 택시 기사 연구'의 주인공입니다. 그녀는 런던의 복잡한 길을 외우는 택시 기사들의 뇌를 조사하여, 공간 기억을 담당하는 '해마'라는 부위가 일반인보다 더 크다는 사실을 밝혀낸 인물이었습니다. 허사비스는 맥과이어 교수의 연구팀에 합류하며, 뇌가 단순히 정보를 저장하는 창고가 아니라 끊임없이 지도를 그리고 시뮬레이션을 돌리는 능동적인 엔진이라는 사실에 전율했습니다.

맥과이어 교수는 허사비스의 비범한 집중력과 컴퓨터 프로그래밍 실력을 높게 평가했고, 두 사람의 만남은 뇌과학과 컴퓨터 공학이 만나는 거대한 불꽃을 일으켰습니다. 맥과이어 교수의 지도 아래 허사비스는 뇌의 가장 깊은 곳에 자리 잡은 '해마(Hippocampus)'를 집중적으로

연구하기 시작했습니다. 해마는 바다의 해마를 닮았다고 해서 붙여진 이름인데, 이곳은 우리가 어제 무엇을 먹었는지, 작년 여름 휴가 때 어디를 갔는지를 기억하는 ‘일화 기억(Episodic Memory)’의 핵심 본부입니다.

허사비스는 맥과이어 교수와 함께 뇌가 어떻게 이 수많은 파편 같은 기억들을 연결하여 하나의 생생한 장면으로 재구성하는지를 연구했습니다. 연구실에서 그는 택시 기사들의 뇌 영상 데이터부터 기억력을 상실한 환자들의 사례까지 방대한 자료를 분석하며, 뇌의 작동 원리를 수학적 모델로 바꾸는 작업을 병행했습니다.

맥과이어 교수는 허사비스에게 과학적 엄밀함을 가르쳤습니다. 아무리 화려한 가설이라도 철저한 실험과 데이터로 증명되지 않으면 의미가 없다는 사실을 말입니다. 허사비스는 교수님과 함께 밤늦도록 실험 설계도를 그리며, 피험자들의 뇌를 기능적 자기공명영상 (fMRI)으로 촬영하여 그 안에서 일어나는 미세한 신호들을 추적했습니다.

이 시기에 허사비스가 배운 것은 ‘시스템 신경과학’의 관점이었습니다. 뇌의 한 부위가 단독으로 일하는 것이 아니라, 여러 부위가 마치 오케스트라처럼 협업하여 지능을 만들어낸다는 시각입니다. 맥과이어 교수와의 공동 연구를 통해 허사비스는 단순한 엔지니어를 넘어 뇌의 설계도를 읽어낼 줄 아는 과학자로 거듭났고, 두 사람의 논문은 세계 최고의 과학 학술지들에 실리며 학계의 엄청난 주목을 받았습니다. 뇌의 원리를 AI 아키텍처에 적용하려는 시도 허사비스의 연구실 책상 앞에는 천재 물리학자 리처드 파인만의 유명한 문구가 붙어 있었습니다.

"내가 만들 수 없는 것은, 내가 이해한 것이 아니다(What I cannot create, I do not understand)." 이 말은 허사비스가 뇌를 연구하는 근본적인 태도였습니다. 그는 뇌과학자들이 발견한 현상들을 보며 끊임없이 자문했습니다.

"뇌의 기능을 컴퓨터 코드로 구현한다면 어떤 모습일까요?" 그에게 뇌 연구는 완벽한 인공지능을 설계하기 위한 ‘역공학 (Reverse Engineering)’ 과정이었습니다. 그는 뇌가 신경세포 간의 연결 강도를 조절하여 학습한다는 사실을 보며, 인공 신경망의 ‘가중치’를 어떻게 조정해야 할지에 대한 영감을 얻었습니다.

뇌의 ‘해마’가 잠을 자는 동안 낮에 겪었던 일들을 아주 빠른 속도로 되감기 (Replay)하며 학습한다는 사실에 주목했습니다. 이것을 컴퓨터에 적용한다면, 인공지능이 매 순간 새로운 데이터를 받을 필요 없이 과거의 데이터를 반복해서 학습함으로써 훨씬 더 똑똑해질 수 있다는 계산이 나왔습니다. 훗날 딥마인드의 인공지능이 게임을 배울 때 사용한 ‘경험 재생’ 기술의 원형이 되었습니다.

허사비스는 뇌의 작동 방식을 관찰하며, 인공지능이 가져야 할 ‘아키텍처(구조)’를 상상했습니다. 단순히 명령어를 입력하는 것이 아니라, 뇌처럼 여러 모듈이 상호작용하며 스스로 목표를 세우고 세상을 이해하는 구조를 꿈꾼 것입니다. 허사비스는 동료 학자들에게 종종 이렇게 말했습니다.

"진정한 AI는 인간의 뇌가 문제를 해결하는 방식과 닮아야 합니다. 그것이 가장 우아하고 효율적인 길이기 때문입니다." 그는 박사 논문을 쓰면서도 틈틈이 이 아이디어들을 메모지에 코드로 옮겨 적었습니다.

뇌가 공간을 인식하는 법, 추상적인 개념을 형성하는 법, 불확실한 미래를 예측하는 법 등이 모두 잠재적인 인공지능 설계도에 포함되었습니다. 파인만의 말처럼, 그는 지능을 완벽히

이해하기 위해 그것을 직접 만들어보겠다는 야심을 품었습니다. UCL에서의 박사 과정은 지능의 본질에 대한 확신을 주었으며, 이제 그 확신을 세상에 증명할 도구인 ‘딥마인드’를 설립할 준비를 마친 셈이었습니다.

해마, 기억과 상상력의 연결 고리 해마 손상 환자가 미래를 상상하지 못한다는 획기적 발견 2007년 초, 허사비스는 세상을 깜짝 놀라게 할 실험 결과를 발표했습니다. 그는 기억 상실증을 앓고 있는 환자와 건강한 일반인들을 모아놓고 아주 특별한 질문을 던졌습니다. "내일 아침에 해변에 놀러 간다고 상상해 보세요.

거기서 무엇이 보이고 어떤 일이 일어날지 아주 자세하게 설명해 주시겠어요?" 우리는 미래를 상상하는 것이 당연하다고 생각하지만, 해마가 손상된 환자들의 반응은 충격적이었습니다. 그들은 단 한 발자국도 미래로 나아가지 못했습니다.

그들은 "음, 바다가 있겠죠... 그리고 모래도 있을 거예요. 하지만 그게 전부예요. 아무런 그림 이 그려지지 않아 요"라며 고통스러워했습니다. 그들의 마음속 스크린은 텅 빈 백지와 같았습니다. 이 실험은 인류가 기억과 상상에 대해 가지고 있던 기존의 상식을 완전히 뒤집어 놓았습니 이전까지 과학자들은 기억은 ‘과거’를 위한 것이고, 상상은 ‘미래’를 위한 별개의 능력 이라고 생각했습니다.

하지만 허사비스는 기억을 못 하는 사람은 상상도 할 수 없다는 사 실을 증명해 냈습니다. 즉, 해마는 단순히 과거를 기록하는 일기장이 아 니라, 과거의 기억 파편들을 꺼내어 미래의 가상 시나리오를 조립하는 ‘마음의 작업 대’였던 것입니다. 허사비 스는 이 발견을 통해 지능의 핵심이 ‘예측’과 ‘시뮬레이션’ 에 있다는 사실을 확신하게되었습니다.

과거를 기억하는 이유는 미래에 닥칠 일을 더 잘 준비하기 위해서라는 생존의 법 칩을 뇌의 깊은 곳에서 찾아낸 것입니다. 허사비스는 환자들의 짧고 단절된 문장들을 분석하며 인간 지능의 위대함을 다시 한번 느 껴했습니다. 우리가 "내일 친구와 점심을 먹는다"라고 상상할 때, 우리 뇌는 눈 깜짝할 사이 에 수많은 과거 기억을 재조합하여 가상의 현실을 만들어냅니다. 허사비 스는 이 능력을 인 공지능에 이식하고 싶어했습니다.

인공지능이 현재의 데이터만 처 리하는 것이 아니라, 과 거의 경험을 바탕으로 미래를 상상하고 최선의 선택을 미리

시뮬레이션해 본다면 어떨까? 이 획기적인 발견은 뇌과학계의 최고 권위자인 『네이 처(Nature)』 에 실렸고, 전 세계 언론 은 "기억과 상상이 한 몸이라는 사실을 밝혀낸 젊은 천재"라며 허사비스를 주목하기 시작했습니다. 에피소드 기억과 장면 구성(Scene Construction) 이론의 개발 허사비스는 해마 손 상 환자 연구를 한 단계 더 발전시켜 ‘장면 구성(Scene Construction)’이 라는 혁신적 인 이론을 제시했습니다. 그는 왜 해마가 망가진 환자들이 미래를 상상하지 못 하는지를 분석하면서, 그들이 단순히 ‘사건’을 기억하지 못하는 것이 아니라, 머릿속에서 입 체적인 ‘공간’을 만들어내지 못한다는 점에 주목했습니다.

우리가 어떤 일을 기억하거 나 상상할 때, 우리 뇌는 먼저 무대 배경이 되는 공간을 설정합니다. 탁자가 어디에 있 고, 창문 은 어디에 있으며, 햇빛은 어느 방향에서 비치는지 같은 3차원 배경 말입니다. 허사비스는 해마가 바로 이 ‘3차원 가상 무대’를 건설하는 건설 현장 소장과 같다 는 사실을 밝혀냈습니다.

이 이론에 따르면, 우리가 겪는 모든 경험(에피소드)은 이 무대 위에 올려집니다. 무대가 없으면 배우들이 연기할 수 없듯이, 뇌 안에서 공간적 배경이 구성되지 않으면 기억도 상상도 불가능해집니다. 허사비스는 fMRI를 통해 사람들이 무언가를 상상할 때 뇌의 해마 부위가 격렬하게 활동하는 모습을 시각적으로 증명했습니다. 이것을 인공지능 설계에 적용할 중요한 힌트로 삼았습니다. 인공지능이 단순히 단어나 이미지를 나열하는 것이 아니라, 세상을 하나의 입체적인 구조로 파악하고 그 안에서 자신의 행동을 예측해야 한다는 비전을 세운 것입니다.

‘장면 구성’ 이론은 심리학과 뇌과학 분야에서 고전으로 여겨지던 기억 이론들을 하나로 통합하는 역할을했습니다. 허사비스는 "기억은 재생(Playback)이 아니라 재구성(Reconstruction)이다"라고 주장했습니다. 우리가 과거를 떠올릴 때마다 뇌는 매번 새로운 무대를 짓고 기억의 조각들을 배치 한다는 것입니다.

이 유연한 사고방식은 훗날 딥마인드의 AI가 바둑판이라는 공간을 이해하고, 수 많은 가상 대국을 통해 승리 확률을 계산하는 원동력이되었습니다. 허 사비스는 뇌의 복잡한 메커니즘을 ‘장면’과 ‘구성’이라는 개념으로 설명해 냈으로써, 인공지능이 가야 할 복극 성을 발견했습니다.

"마음의 시뮬레이션 엔진(Simulation Engine of the Mind)" 개념 박사 과정 막바지에 이르러 허사비스는 자신의 연구를 관통하는 하나의 거대한 개념을 정립했습니다. 바로 "뇌는 마음의 시뮬레이션 엔진(Simulation Engine)"이라는 생각입니다. 그는 인간이 지구상의 다른 생명체들을 압도하고 문명을 건설할 수 있었던 이유는, 실제로 위험에 뛰어들기 전에 머릿속에서 수만 번의 시뮬레이션을 돌려볼 수 있었기 때문이라고 보았습니다.

원시인이 사냥을 나가기 전 "저 숲으로 가면 사자가 나타날까? 저 바위 뒤에 숨으면 안전할까?"라고 상상하는 그 과정 자체가 지능의 본질이라는 것입니다. 뇌는 가상의 미래를 미리 살아보고, 가장 안전하고 이득이 되는 길을 선택하는 정교한 시뮬레이터였습니다. 허사비스에게 이 개념은 인공지능이 도달해야 할 궁극적인 목표였습니다.

당시의 인공지능은 주어진 규칙 안에서 계산만 잘하는 ‘계산기’에 불과했지만, 허사비스는 ‘스스로 세상을 시뮬레이션하는 AI’를 꿈꿨습니다. 인공지능이 현실 세계의 물리 법칙과 인과 관계를 이해하고, 자신의 행동이 미래에 어떤 결과를 초래할지 미리 예측할 수 있다면, 그것은 인간의 지능에 한 발짝 더 다가가는 것이기 때문입니다. 뇌의 해마와 전두엽이 협력하여 이 엔진을 가동하는 방식을 보며, 인공 신경망 구조에 이 ‘시뮬레이션 루프’를 어떻게 넣을지 고민했습니다.

이 ‘마음의 시뮬레이션 엔진’ 개념은 훗날 딥마인드의 상징과도 같은 기술인 ‘모델 기반 강화학습(Model-based Reinforcement Learning)’으로 이어졌습니다. 인공지능이 세상에 대한 자신만의 모델을 만들고, 그 모델 안에서 상상력을 발휘하여 정답을 찾아내는 방식입니다. 허사비스는 런던의 좁은 연구실에서 fMRI 영상을 들여다보며, 먼 미래에 인공지능이 복잡한 질병 치료법을 시뮬레이션하고 기후 변화의 대안을 찾아내는 장면을 미리 그려보았습니다.

그는 이미 자신의 뇌라는 엔진을 돌려, 딥마인드가 가져올 미래의 혁명을 상상하고 있었던 것입니다. Science지 선정 "2007년 10대 과학적 돌파구" Nature, Science, Neuron, PNAS 등 최고 저널에 영향력 있는 논문 발표 데미스 허사비스의 학계 데뷔는 화려했습니다. 보통 박사 과정 학생이 평생 한 번 심기도 어려운 세계적인 학술지들에 그의 이름이 적힌 논문들이 봇물 터지듯 쏟아져 나왔습니다. 『네이처(Nature)』, 『사이언스

(Science)』, 『뉴런(Neuron)』, 『미국국립과학원회보 (PNAS)』 등 과학자들에게는 성배와 같은 학술지들이 허사비스의 연구 결과를 앞다투어 게재했습니다. 게임 개발자 출신이라는 꼬리표를 떼고, 단숨에 신경과학계의 떠오르는 스타로 자리매김했습니다. 그가 발표한 논문들은 질문의 깊이가 달랐기 때문에 영향력이 컸습니다.

그는 뇌의 미세한 기능 하나를 분석하는 데 그치지 않고, '기억, 상상, 지능'이라는 인간 마음의 거대한 작동 원리를 하나의 유기적인 시스템으로 설명해 냈습니다. 학계에서는 그를 두고 "컴퓨터 과학자의 날카로운 논리와 신경과학자의 깊은 통찰력을 동시에 가진 드문 인재"라고 평가했습니다. 연구 인용 횟수는 기하급수적으로 늘어났고, 전 세계 유수의 대학들로부터 강연 요청이 쇄도했습니다.

그가 2007년에 발표한 논문들은 기억력 연구의 패러다임을 완전히 바꿨습니다. "기억은 단순히 과거의 기록이 아니라 미래를 위한 도구이다"라는 그의 명제는 수십 년간 이어져 온 기억 연구의 방향을 송두리째 돌려놓았습니다. 이 시기에 쌓은 학문적 성취와 인맥은 훗날 딥마인드를 창업했을 때, 실리콘밸리의 자본뿐 아니라 학계의 최고 인재들을 끌어모으는 신뢰자산이 되었습니다.

허사비스는 지능의 본질을 꿰뚫는 과학적 권위를 세움으로써 미래의 인공지능 혁명을 주도할 자격을 스스로 증명해 냈습니다. fMRI 연구를 통한 기억 회상과 상상의 동일 신경 기제 증명 허사비스의 연구 중 가장 결정적인 증거는 fMRI(기능적 자기공명영상)를 활용한 실험에서 나왔습니다. 그는 사람들을 커다란 원통형 기계인 fMRI에 눕히고, 두 가지 과제를 주었습니다.

첫 번째는 "지난 크리스마스 때의 일을 기억해 보세요"였고, 두 번째는 "다음 크리스마스에 일어날 일을 상상해 보세요"였습니다. 기계가 웅웅거리며 뇌의 혈류 변화를 촬영하는 동안, 모니터에는 놀라운 장면이 펼쳐졌습니다. 과거를 회상할 때 불이 들어오는 뇌 부위와 미래를 상상할 때 불이 들어오는 부위가 마치 거울을 보듯 거의 똑같았던 것입니다.

기억과 상상이 뇌의 동일한 '신경 기제(Neural Mechanism)'를 공유한다는 사실을 완벽하게 입증했습니다. 즉, 뇌 입장에서 과거를 떠올리는 것과 미래를 상상하는 것은 본질적으로 같은 작업이었습니다. 해마를 중심으로 한 특정 네트워크가 활성화되어 저장된 정보들을 재조합하는 과정이었던 것이죠. 허사비스는 이 데이터를 바탕으로 "지능이란 과거의 경험을 재료 삼아 미래의 가능성을 시뮬레이션하는 능력"이라고 정의를 내렸습니다.

이 발견은 2007년 『사이언스』지에 의해 "그해 세계 10대 과학적 돌파구" 중 하나로 선정되는 영예를 안았습니다. 이 성취는 허사비스에게 엄청난 자신감을 주었습니다. 그는 뇌가 어떻게 무형의 상상력을 유형의 지능으로 바꾸는지 확인했습니다.

이제 남은 과제는 이 마법 같은 과정을 디지털 코드로 옮기는 것이었습니다. 그는 fMRI 영상을 보며 확신했습니다. 뇌가 기억 파편들을 모아 미래를 그리듯, 인공지능도 데이터의 파편들을 모아 세상을 시뮬레이션할 수 있을 것이라고 말입니다.

박사 과정을 마칠 무렵, 허사비스는 더 이상 학교에 남아있을 이유가 없었습니다. 마음의 설계도를 충분히 읽어낸 그는, 이제 그 설계도를 들고 세상을 바꿀 기계를 직접 만들기 위해 다시 한번 거친 비즈니스의 세계로 발을 내디딜 준비를 마쳤습니다.

fMRI 뇌 스캔 이미지, 해마 영역 활성화

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

탐색 2009년의 어느 늦은 오후, 미국 매사추세츠주 케임브리지의 찰스 강변을 걷던 데미스 허사비스의 머릿속은 두 개의 서로 다른 세계가 충돌하고 있었습니다. 그의 뒤편에는 세계 최고의 공과대학인 MIT가 서 있었고, 강 건너편에는 하버드 대학의 붉은 벽돌 건물들이 보였습니다. 이 지리적 풍경은 당시 허사비스가 처해 있던 지적 인 위치를 정확하게 보여주는 은유와도 같았습니다.

그는 런던 대학(UCL)에서 기억 과 상상력에 관한 뇌과학 박사 학위를 막 마친 상태였습니다. 보통의 과학자라면 여 기서 선택을 해야했습니다. 뇌를 연구하 는 생물학자가 될 것인가, 아니면 코드를 짜 는 컴퓨터 공학자가 될 것인가. 하지만 허사비스는 두 가지를 모두 선택했습니다.

그 는 이 두 세계를 연결하는 다리가 되기로 결심하고 대서양을 건너 MIT의 맥가번 뇌 연구소(McGovern Institute)로 향했습니다. 그가 찾아간 곳은 토마소 포지오(Tomaso Poggio) 교수의 연구실이었습니다. 포지 오 교수는 컴퓨터 비전(Computer Vision) 분야의 거장으로, 인간의 뇌가 사물을 보는 방식을 수학적으로 모델링하는 데 평생을 바친 인물이었습니다.

허사비스가 이곳을 선택한 이유는 명확했습니다. 당시 인공지능 연구는 '겨울'이라고 불리는 침체기를 겪 고 있었지만, 포지오 교수 의 연구실만은 달랐기 때문입니다. 그곳에서는 생물학적 뇌 의 원리를 컴퓨터 알고리즘으 로 변환하려는 시도가 활발히 이루어지고 있었습니다.

허사비스는 이곳에서 "뇌를 이해하 면 지능을 만들 수 있다"는 자신의 가설을 검증할 완벽한 환경을 만났습니다. 이 시기 허사비스는 영국의 웰컴 트러스트(Wellcome Trust) 재단으로부터 '헨리 웰 컴 펠로 우십(Henry Wellcome Fellowship)'이라는 매우 특별한 지원을 받고 있었습니다. 이 펠로우 십은 젊은 과학자에게 연구 주제와 장소를 자유롭게 선택할 수 있는 파격적인 권한과 예산을 부여했습니다.

덕분에 허사비스는 특정 교수의 프로젝트에 종속되지 않고, 자신만의 거 대한 목표인 '범용 인공지능(AGI)'을 위한 기초 설계를 다

질 수 있었습니다. 그는 MIT와 하 버드를 자유롭게 오가며 당대 최고의 지성들과 교 류했습니다. 허사비스는 하버드 대학의 연구실을 방문해 최신 뇌 스캔 데이터를 분석하고, 다시 MIT로 돌아와 그 데이터를 처리할 머신러닝 모델을 구상했습니다.

이 과정에서 그는 생물학적 뇌 가 가진 놀라운 효율성에 주목했습니다. 인간의 뇌는 고작 20와트(W) 정 도의 에너지만으로 복잡한 세상을 인식하고, 걷고, 대화하고, 새로운 것을 배웁니다. 반면 당시의 컴퓨터는 엄청난 전력을 쓰고도 고양이와 강아지를 구분하는 것조차 어 려워했습니다. 허사비스는 이 격차를 줄이는 열쇠가 '계산 신경과학(Computational Neuroscience)'에 있다 고 확신했습니다.

이것은 뇌의 신경 세포가 신호를 주고받는 방식을 수학 공식으로 옮겨 적는 학문입니다. 그는 뇌의 시각 피질이 정보를 단계적으 로 처리하는 방식을 딥러닝(Deep Learning) 기술에 접목하는 가능성을 탐구했습니다. MIT의 분위기는 열정적이면서도 실용적이었습니다.

허사비스는 그곳의 연구원들 과 밤늦 도록 토론하며 자신의 아이디어를 다듬었습니다. 자신이 이전에 만들었던 게 임 속 인공지능들이 보여준 한계를 뼈저리게 느끼고 있었습니다. 게임 속 캐릭터들은 프로그래머가 미 리 입력해 둔 규칙 안에서만 뚝뚝해 보였습니다.

하지만 허사비스가 원한 것은 낯선 환경 에 던져져도 스스로 규칙을 깨우치는 진정한 지능이었습니다. 그는 포지오 교수와의 연구를 통해, 인간의 뇌가 시각 정보를 처리할 때 보여주는 '계층적 구조'가 바로 그 해답 중 하나임을 깨달았습니다. 우리가 사물을 볼 때, 뇌는 먼저 단순한 선과 색을 인식하고, 그것을 조합해 모양을 만들고, 최종적으로 '이것은 자동차다'라는 개념을 형성합니다.

이 생물학적 위계질서는 훗날 딥마인드의 인공지능이 화면의 픽셀(Pixel) 만 보고도 벽돌 깨기 게임을 정복하게 만드는 기술적 토대가되었습니다. 미국에서의 시간은 허사비스에게 또 다른 자산을 남겼습니다. 바로 '야망의 크기'였 습니 실리콘밸리와 보스턴의 학구적 분위기 속에서 그는 연구가 단순히 논문을 쓰는 것 으로 끝나서는 안 된다는 것을 배웠습니다.

위대한 연구는 세상을 바꾸는 기술로 구현되어 야했습니다. 그는 이곳에서 만난 동료들에게 "우리는 지능을 이해할 것이다"라는 말을 서슴없이 던지곤했습니다. 누군가는 그를 허풍선이라고 생각했을지 모

르지만, 허사비스의 머릿속에는 이미 게임 개발자로서의 경험, 뇌과학자로서의 지식, 그리고 머신러닝이라는 도구가 하나로 융합되고 있었습니다. 이제 이 모든 재료를 가 지고 자신의 고향인 런던으 로 돌아가, 인류 역사상 가장 대담한 스타트업을 시작할 준비를 마친 것입니다. 헨리 웰컴 펠로우십(Henry Wellcome Fellowship)과 UCL 개츠비 계산신경과학유닛 2010년, 데미스 허사비스는 미국 생활을 정리하고 런던으로 돌아왔습니다.

그의 다음 행선 지는 유니버시티 칼리지 런던(UCL)에 위치한 '개츠비 계산신경과학유닛 (Gatsby Computational Neuroscience Unit)'이었습니다. 일반 대중에게는 낯선 이름일지 모르지만, 인공지능 연구자들에게 이곳은 전설적인 성지(聖地)와도 같습니다. 이곳은 훗날 노벨 물리학상을 수상하게 되는 제프리 힌튼 (Geoffrey Hinton) 교수가 설립을 주도한 곳으로, 전 세계에서 가장 똑똑하고 수학적인 두뇌들이 모여 인간 지능의 비밀을 파헤치는 곳이었습니다.

허사비스가 이곳을 선택한 것은 우연이 아니었습니다. 자신이 구상하고 있던 '딥마인드'라는 씨앗을 틔울 수 있는 유일한 토양임을 직 감했습니다. 허사비스가 개츠비 유닛에 합류할 수 있었던 것은 앞서 언급한 '헨리 웰컴 펠로우 십' 덕분이었습니다.

이 펠로우십은 그에게 재정적 자유뿐만 아니라, 연구의 방향을 스스로 결정할 수 있는 지적인 독립성을 보장해주었습니다. 당시 개츠비 유닛은 피터 데이언(Peter Dayan) 교수가 이끌고 있었습니다. 데이언 교수는 이론 신경과학의 대 가로, 뇌가 어떻게 학습하고 보상(Reward)을 처리하는지를 수학적으로 규명한 인물이었습니다.

개츠비 유닛은 뇌과학 과 기계학습이 융합되는 용광로였습니다. 개츠비 유닛의 분위기는 치열했습니다. 연구원들은 생물학적인 뇌 실험 데이터와 복잡한 확률 통계학 사이를 씬 없이 오갔습니다.

허사비스는 이곳에서 자신의 오랜 친 구이자 훗 날 딥마인드의 공동 창립자가 되는 셰인 레그(Shane Legg)를 만났습니다. 두 사람은 점심시 간마다 학교 근처의 샌드위치 가게나 펍(Pub)에 앉아 몇 시간이고 대화를 나누었습니다. 대화 주제는 단 하나, "어떻게 하면 인간 수준의 인공지능(AGI) 을 만들 수 있을까?"였습니 당시 학계에서는 AGI라는 단어를 입에 올리는 것조차 금 기시되는 분위기였습니다.

인공지능 연구는 특정 문제만 해결하는 좁은 인공지능 (Narrow AI)에 갇혀 있었고, 인간처럼 범용적인 지능을 논하는 것은 공상과학 소설

취급을 받았습니다. 하지만 개츠비 유닛의 자유롭고 근본적인 탐구 분위기 속에서 허사비스와 레그는 그 금기를 깰 용기를 얻었습니다. 허사비스는 개츠비 유닛에서 연구하는 동안 '시스템 신경과학(Systems Neuroscience)'에 깊이 파고들었습니다.

뇌의 개별 세포가 아니라, 뇌 전체가 하나의 시스템으로서 어떻게 정보를 통합하고 기억을 저장하며 미래를 시뮬레이션하는지를 연구하는 분야입니다. 그는 뇌의 '해마(Hippocampus)'가 수행하는 역할에 주목했습니다. 해마는 우리가 잠을 잘 때, 낮에 겪었던 경험을 마치 비디오를 되감듯이 빠르게 재생(Replay)하며 기억을 강화합니다.

허사비스는 이 생물학적 메커니즘이 인공지능의 학습 효율을 획기적으로 높일 수 있는 열쇠라고 생각했습니다. 이곳에서의 경험은 허사비스에게 '하이브리드(Hybrid)'의 힘을 가르쳐주었습니다. 그는 순수 뇌과학자들 틈에서는 컴퓨터 알고리즘의 명쾌함을 이야기했고, 머신러닝 공학자들 틈에서는 뇌의 유연함을 이야기했습니다.

두 집단이 사용하는 언어가 서로 다르지만, 결국 같은 문제를 풀고 있다는 것을 깨달았습니다. 바로 "지능이란 무엇인가"라는 질문입니다. 개츠비 유닛에서의 시간은 허사비스가 딥마인드를 창업하기 직전의 마지막 인큐베이팅 기간이었습니다.

그는 이 곳에서 최신 이론들을 흡수하는 것에 그치지 않고, 그 이론들을 실제로 작동하는 코드로 구현할 동료들을 모으기 시작했습니다. 개츠비 유닛은 허사비스에게 학문적 엄밀함(Rigour)을 선물했습니다. 그는 단순히 "될 것 같다"는 직관에 의존하지 않고, 수학적으로 증명 가능하고 생물학적으로 타당한 모델을 만드는 법을 배웠습니다.

이곳에서 다져진 이론적 토대는 훗날 알파고가 바둑의 복잡성을 뚫고 승리할 수 있었던 밑거름이 되었습니다. 허사비스가 2010년 딥마인드를 설립했을 때, 회사의 초기 멤버 대부분이 이 개츠비 유닛 출신이거나 그곳과 연관된 인물들이었다는 사실은 결코 우연이 아닙니다. 헨리 웰컴 펠로우십이 제공한 시간과 자유, 그리고 개츠비 유닛이라는 지적 공동체는 청년 허사비스를 세계적인 AI 리더로 변모시키는 결정적인 도가니가 되었습니다. 강화학습과 도파민 시스템 신경과학 + 계산 = 범용학습의 단서 런던의 흐린 날씨 속에서도 유니버시티 칼리지 런던(UCL)의 연구실은 지적 열기로 뜨거웠습니다. 이 곳에서 만난 피터 데이언 교수는 허사비스가 오랫동안 품어온 질문에 답을 줄 수 있는

현자(賢者)와 같았습니다. 피터 데이언은 계산 신경과학(Computational Neuroscience) 분야의 개척자 중 한 명으로, 우리 뇌가 어떻게 보상을 통해 학습하는지를 수학적으로 증명해낸 인물이었습니다. 허사비스와 데이언의 만남은 마치 1980년대 애플의 스티브 잡스와 워즈니악의 만남처럼, 서로 다른 재능이 완벽하게 맞물리는 순간이었습니다.

이들이 함께 탐구한 핵심 주제는 '강화학습(Reinforcement Learning)'이었습니다. 강화학습을 쉽게 설명하자면, 강아지를 훈련시키는 과정과 비슷합니다. 강아지가 '앉아'라는 명령어에 맞춰 앉으면 간식을 줍니다. 그러면 강아지는 '앉는 행동'과 '간식(보상)'을 연결하게 되고, 점차 그 행동을 더 잘하게 됩니다.

하지만 허사비스와 데이언 의 관심사는 이것을 생물 학적 차원과 컴퓨터 알고리즘 차원에서 동시에 설명하는 것이었습니다. 피터 데이언은 1990년대 중반, 뇌 속에 있는 '도파민(Dopamine)'이라는 신경 전달 물질의 역 할을 재정의했습니다. 흔히 도파민을 '쾌락의 호르몬'이라고 생각하지만, 데 이언과 그의 동 료들은 도파민이 '보상 예측 오류(Reward Prediction Error)'를 신호 한다는 사실을 밝혀냈습니다.

이게 무슨 뜻일까요? 예를 들어, 당신이 자판기에서 커피를 뽑아 마시려는데 동전 이 나오 지 않았다고 가정해 봅시다. 이때 당신의 뇌는 '커피를 기대함' 상태에서 '아무 것도 없음'이 라는 결과를 맞닥뜨립니다. 기대와 실제 결과의 차이, 이것이 바로 '예측 오류'입니다. 이때 뇌의 도파민 세포는 활동을 멈추거나 급격히 줄어들며 '실망'이라는 신호를 보냅니다. 반대로, 자판기에서 커피와 함 께 뜬금없이 만 원짜리 지폐가 나왔다면? 기대치보다 훨씬 큰 보상이 주어 졌으므로 도파 민 세포는 폭발적으로 반응합니다. 이것이 '긍정적 예측 오류'입니다.

허사비스는 이 생물학적 발견이 컴퓨터 과학의 'TD 학습(Temporal Difference Learning)' 알 고리즘과 정확히 일치한다는 점에 전율했습니다. 컴퓨터 과학자들이 기계를 학습시키기 위해 고안해낸 수학 공식이, 알고 보니 이미 수억 년 전부터 우리 뇌 속에서 작동하고 있었던 것입니다. 이것은 엄청난 발견이었습니다.

"인공지능을 만들기 위해 외계의 기술을 발명할 필요가 없다. 이미 우리 머릿속 에 있는 설계 도를 베끼면 된다"는 확신을 주었기 때문입니다.

허사비스는 피터 데이언과의 협업을 통해, 단순히 데이터를 분류하는 AI가 아니라 '스스로 행동하고 그 결과로부터 배우는 AI'의 가능성을 보았습니다. 당시 대부분의 AI 연구는 사진을 보고 고양이인지 개인지 맞히는 '지도 학습(Supervised Learning)'에 치중해 있었습니다. 정답지가 있는 문제를 푸는 것이죠. 하지만 강화학 습은 정답지가 없습니다.

에이전트(AI) 가 환경과 상호작용하며 수만 번의 시행착오 끝에 스스로 전략을 찾아내는 방식입니다. 이 것은 어린아이가 걷는 법을 배우는 과정 과 똑같습니다. 넘어지고, 아파하고, 다시 일어나 면서 균형을 잡는 법을 스스로 깨우 치는 것입니다.

이 연구는 허사비스에게 '범용성(Generality)'에 대한 중요한 단서를 제공했습니다. 도파민 시스템은 우리가 자전거를 배울 때도, 악기를 연주할 때도, 바둑을 둘 때도 똑같이 작동합니다. 즉, 뇌에는 모든 종류의 학습을 관장하는 하나의 통합된 알고리즘 이 존재할지 도 모른다는 가설이 성립합니다.

허사비스는 이 '하나의 알고리즘'을 컴퓨터에 구현할 수만 있다면, 그 AI는 바둑만 잘 두는 것이 아니라, 단백질 구조도 풀고, 기후 문제도 해결할 수 있을 것이라고 믿었습니다. 피터 데이언과의 연구실에서 보낸 시간들은 훗날 딥마인드가 "범용 학습 알고리즘을 만든다" 는 창립 미션을 세우는데 결정적인 이론적 토대가되었습니다. 인간 뇌에서 찾은 AI 알고리즘의 힌트: DQN의 이론적 토대 허사비스가 피터 데이 언과 함께 연구하며 얻은 통찰은 2013년 딥마인드의 첫 번째 성과물인 'DQN(Deep Q-Network)'으로 구체화되었습니다.

DQN은 딥마인드가 세상에 이름을 알리게 된 계기이자, 구글이 이 작은 런던의 스타트업을 5천억 원이 넘는 거액에 인 수하게 만든 결정적인 기술이었습니다. 이 기술의 핵심 아이디어는 놀랍게도 허사비스가 박사 과 정 때 연구했던 뇌의 '해마(Hippocampus)'에서 나왔습니다. DQN은

1980년대의 고전 비디오 게임인 '아타리(Atari)' 게임들을 인간보다 더 잘 플레이하도록 설계되었습니다.

하지만 이 AI에게는 게임의 규칙을 알려주지 않았습니다. 오직 화면의 픽셀 정보(시각)와 점수(보상)만 주어졌습니다. 처음에는 AI가 무작위로 버튼을 눌러보며 엉뚱한 행동을 했지만, 시간이 지날수록 점수를 얻는 방법을 스스로 터득해 나갔습니다.

그런데 여기서 한 가지 치명적인 문제가 있었습니다. 컴퓨터가 연속적인 게임 화면을 학습할 때, 데이터들이 서로 너무 비슷해

서 학습이 불안정해지거나 잊어버리는 현상이 발생했습니다. 이를 '파국적 망각 (Catastrophic Forgetting)' 이라고 부릅니다. 이 난관 앞에서 허사비스는 뇌과학자로서의 지식을 꺼내 들었습니다. "인간은 어떻게 새로운 것을 배우면서도 옛날 것을 잊어버리지 않을까?"

그는 뇌의 해마가 가진 '경험 재생 (Experience Replay)' 기능에 주목했습니다. 우리가 잠을 자거나 휴식을 취할 때, 뇌는 낮 동안 겪었던 중요한 경험들을 무작위 순서로, 그리고 빠른 속도로 다시 재생합니다. 마치 시험공부를 할 때 중요 내용을 뒤섞어가며 복습하는 것과 같습니다.

이 과정을 통해 단기 기억이 장기 기억으로 견고하게 저장됩니다. 허사비스는 이 생물학적 원리를 AI 코드에 그대로 적용했습니다. 그는 AI가 게임을 하면서 겪은 수많은 장면과 경험들을 즉시 버리지 않고 '경험 버퍼 (Experience Buffer)'라는 저장소에 모아두도록 했습니다.

그리고 학습을 할 때는 현재의 경험뿐만 아니라, 저장소에 있는 과거의 경험들을 무작위로 꺼내어 함께 학습 시켰습니다. 이 것이 바로 DQN의 핵심 기술인 '경험 리플레이(Experience Replay)'입니다. 효과가 있었습니다.

AI의 학습은 놀랍도록 안정화되었고, 성능은 폭발적으로 향상되었습니다. 벽돌 깨기 게임(Breakout)에서 DQN이 보여준 모습은 이런 모습이었습니다. 처음에는 공을 맞지도 못하던 AI가 300번의 학습 후에는 완벽하게 공을 받아내더니, 500 번이 넘어가자 누구도 가르쳐준 적 없는 '터널링(Tunneling)' 전략을 스스로 발견했습니다.

벽돌의 한쪽 끝을 집중적으로 파서 구멍을 낸 뒤, 공을 그 뒤쪽으로 보내 벽돌들이 저절로 깨지게 만드는 고급 기술이었습니다. 데이비드 실버를 비롯한 딥 마인드의 연구원들은 새벽 3시에 모니터 앞에서 이 장면을 목격하고 입을 다물지 못했습니다. 그것은 기계가 창의성을 발휘한 순간이었습니다.

피터 데이언과의 연구를 통해 확립된 '보상에 의한 학습' 이론과, 허사비스가 뇌과학에서 가져온 '기억의 재생' 메커니즘이 만나 탄생한 DQN은 단순한 게임 AI가 아니었습니다. 생물학적 지능의 원리가 실리콘 칩 위에서도 작동할 수 있음을 증명한 역사적인 사건이었습니다. 뇌에서 찾은 힌트가 AI 알고리즘의 결정적인 퍼즐 조각이 된 것입니다.

이 성공은 허사비스에게 확신을 주었습니다. "우리는 올바른 길을 가고 있다." 이 확신은 훗날 바둑 판 위에서 인류를 놀라게 할 알파고(AlphaGo)의 탄생으로 이어지는 거대한 여정의 서막이었습니다.

허사비스에게 있어 DQN은 인간의 뇌를 모방하여 만든 최초의 '생각하는 기계'의 원형(Prototype)이었습니다. 생물학적 뉴런과 인공 뉴런의 구조 비교

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

버시티 칼리지(UCL)의 개츠비 계산신경과학 유닛(Gatsby Computational Neuroscience Unit)은 전 세계에서 가장 똑똑한 괴짜들이 모여 인간 뇌의 비밀을 수 학으로 풀어내려는 치열한 전장이었습니다. 이곳은 제프리 힌튼(Geoffrey Hinton)이 설립한 연구소로, 데미스 허사비스는 이곳에서 자 신의 뇌 과학 박사 과정을 마무리하고 있었습니다. 허사비스가 연구소 복도에서 마주친 키 크고 조용한 뉴질랜드 사람, 셰인 레그(Shane Legg)는 그곳의 분위기와 묘하게 어울리면 서도 동시에 이질적인 존재였습니다.

레그는 이미 스위스 아이디어(IDSIA) 연구소에서 위르겐 슈미트후버(Jürgen Schmidhuber) 의 지도를 받으며 이론적인 인공지능 모형에 대해 깊이 파고든 상태였습니다. 당시 학계에서 '인공지능'이라는 단어는 금기어에 가까웠고, 대부분의 연구자들은 통계학 이나 기계학습이라는보다 안전한 용어 뒤에 숨어 특정 문제를 해결하는 좁은 범위의 AI(Narrow AI)에만 몰두하고 있었습니다. 하지만 레그는 달랐습니다.

그는 허사비스가 가 숨 속에 품고 있던 단어, 바로 'AGI(범용 인공지능)'를 두려움 없이 입 에 올리는 몇 안 되는 사람이었습니다. 두 사람은 점심시간마다 연구소 근처의 벤치나 런던의 펍에 앉아 인간의 지능을 어떻게 수 학적으로 정의할 수 있을지, 그리고 기계가 인간처럼 생각하게 된다면 세상은 어떻게 변할 지에 대해 몇 시간이고 토론했습니다. 허사비스가 직관적이고 신경과학 적인 접근을 선호했다면, 레그는 논리적이고 수학적인 엄밀함을 중시했습니다.

둘의 만남은 불과 얼음의 결 합처럼 서로의 부족한 점을 보완해주었습니다. 레그는 훗날 인 공지능이 인류에게 미칠 초 지능의 위험성을 경고하고 안전 장치를 마련하는 데 핵심 적인 역할을 하게됩니다.

이 지적 결합에 현실적인 에너지를 불어넣은 것은 무스타파 술레이만(Mustafa Suleyman)이었습니다. 술레이만은 과학자가 아니었습니다. 그는 허사비스의 남동생인 조지 허사비스의 절 친한 친구였으며, 어린 시절부터 허사비스 가(家)를 제집 드나들듯 드나들던 '동네 동 생'이었습니다.

옥스퍼드 대학교에서 철학과 신학을 공부하다가 중퇴한 그는, 10대 시절부터 이슬람 단체와 런던 시장의 자문역으로 활동할 만큼 사회 문제 해결에 깊은 관심을 가진 활동가였습니다. 허사비스와 레그가 실험실 안에서 지능의 본질을 탐구 할 때, 술레이만은 세상 밖에서 이 기술이 사회를 어떻게 바꿀지, 그리고 이 거대한 꿈을 실현하기 위해 어떻게 사람 과 자본을 모을지를 고민했습니다. 술레이만은 허사비스에게 과학적 비전만큼이나 중요한 것이 '이야기'와 '설득'이라 는 점을 상기시켜 주었습니다.

그는 딥마인드가 단순한 연구소를 넘어 세상에 영향을 미치는 기업 이 되어야 한다고 주장했으며, 훗날 구글과의 인수 협상이나 딥마인드 헬 스(DeepMind Health) 프로젝트를 주도하며 사업적 수완을 발휘하게됩니다. 여기에 또 한 명의 인물이 합류합니다. 허사비스의 케임브리지 대학동문이자, 그와 함께 엘릭서 스튜디오(Elixir Studios)에서 게임 개발의 흥망성쇠를 함께 겪었던 데이 비드 실버 (David Silver)였습니다.

실버는 엘릭서가 문을 닫은 뒤 학계로 돌아가 강화학습(Reinforcement Learning) 분야에서 독보적인 위치를 구축하고 있었습니다. 허사비스는 딥마인드의 핵심 알고리 즘이 뇌의 학 습 방식인 '강화학습'이 될 것임을 직감했고, 이 분야의 최고 권위자가 된 옛 친구를 놓칠 수

없었습니다. 실버의 합류는 딥마인드의 기술적 토대를 완성하는 마지막 퍼즐 조각이었습니다.

그는 훗날 알파고(AlphaGo)의 수석 연구원으로서 전 세계를 놀라게 한 승리를 이끌 어냅니다. 2010년 11월, 이들은 런던의 러셀 스퀘어 근처에 작은 사무실을 얻어 '딥마인드 테크놀로지(DeepMind Technologies)'라는 간판을 내걸었습니다. 당시 스타트업 업계는 소셜 미디어나 모바일 앱 개발에 열을 올리고 있었습니다. 실리콘밸리의 투자자들은 "사진 공유 앱을 만 들면 6개월 안에 돈을 벌 수 있는데, 왜 20년이 걸릴지 모르는 인공지능 연구를 하느냐"며 비아냥거렸습니다.

그러나 허사비스와 공동 창업자들은 흔들리지 않았습니다. 그들의 미션은 명확했습니다. "지능을 푼다(Solve Intelligence)." 그리고 "그 지능을 사용해 다른 모든 문제를 해결한다(Use it to solve

everything else)." 이것은 기업의 사훈이라기보다는 과학 혁명을 위한 선언문에 가까웠습니다. 그들은 1960년대 인류를 달에 보냈던 '아폴로 프로젝트'처럼, 최고의 인재와 자본을 한곳에 집중시켜 지능이라는 미지의 영역을 정복하겠다는 야심찬 계획을 세웠습니다.

허사비스는 벤처 투자자들에게 수익 모델을 설명하는 대신, 인공지능이 질병을 치료하고 기후 변화를 막으며 새로운 과학적 발견을 이끌어낼 미래를 이야기했습니다. 피터 틸(Peter Thiel)과 같은 비전 있는 투자자들이 이들의 무모해 보이는 꿈에 매료된 것은 바로 이 '크기' 때문이었습니다. 딥마인드의 탄생은 단순히 또 하나의 기술 회사가 생긴 것이 아니라, 인공지능의 겨울(AI Winter)을 끝내고 새로운 봄을 부르는 신호탄이었습니다.

허사비스, 레그, 솔레이만이라는 세 사람의 독특한 조합은, 과학적 엄밀함과 철학적 고민, 현실적 돌파력이 하나로 뭉쳐 AGI라는 거대한 목표를 향해 나아가는 엔진이되었습니다. 학계와 스타트업 문화의 결합 데미스 허사비스가 딥마인드를 설립하며 가장 깊게 고민한 것은 '어떤 조직을 만들 것인가'였습니다. 그는 케임브리지와 UCL에서의 학계 생활을 통해 대학 연구실의 장단점을 뼈저리게 느끼고 있었습니다.

대학은 인류 역사상 가장 자유로운 상상력, 즉 '푸른 하늘 사고 (Blue Sky Thinking)'가 가능한 곳이었습니다. 당장의 수익이나 결과물에 얽매이지 않고 근본적인 질문을 던질 수 있는 유일한 공간이었습니다. 하지만 동시에 대학은 느렸습니다.

교수들은 연구비를 따내기 위해 서류 작업에 매몰되었고, 연구자들은 자신의 좁은 전공 분야에 갇혀 옆방의 동료가 무엇을 하는지조차 모르는 경우가 허다했습니다. 반면, 허사비스가 10대 시절부터 경험한 게임 회사와 스타트업은 정반대였습니다. 그곳은 목표를 향해 미친 듯이 질주하는 속도와 에너지가 있었고, 결과를 만들어내야 한다는 절박함이 혁신을 불렀습니다.

하지만 깊이 있는 연구를 하기에는 호흡이 너무 짧았습니다. 허사비스는 이 두 세계의 장점만을 결합한 '하이브리드 조직'을 꿈꿨습니다. 그의 롤 모델은 20세기 중반 트랜지스터와 레이저, 정보 이론을 탄생시킨 미국의 '벨 연구소 (Bell Labs)'였습니다. 벨 연구소처럼 최고의 천재들이 모여 자유롭게 토론하되, 스타

트업처럼 명확한 목표와 마감 시한을 가지고 움직이는 조직, 그것이 딥마인드의 설계도였습니다. 먼저 사람을 모으는 방식부터 달랐습니다. 허사비스는 전 세계의 머신러닝, 신경과학, 수학, 물리학 분야의 박사급 인재들을 직접 찾아다녔습니다.

당시만 해도 구글이나 페이스북 같은 빅테크 기업들이 AI 연구 자들을 싹쓸 이해가던 시절은 아니었지만, 유망한 학자들을 불안정한 스타트업으로 데려오는 것은 쉬운 일이 아니었습니다. 허사비스는 그들에게 돈이 아닌 '미션'을 팔 았습니다. "학교에서 논 문 실적에 쫓기며 작은 문제만 풀고 있지 않은가? 우리와 함 께 지능 그 자체를 풀어보자. 여 기에는 연구비 걱정도, 강의 의무도 없다."

이 말은 지 적 갈등을 느끼던 천재들의 심장을 꺾 뚫었습니다. 그는 단순히 코딩 잘하는 엔지니어를 뽑은 것이 아니라, '생각하는 기계'를 만 들고 싶어 하는 과학자들을 모았습니다. 조직 문화 설계에 있어서 허사비스는 '융합'을 강제했습니다.

그는 연구원들이 자신 의 전 문 분야라는 '사일로(Silo)'에 갇히는 것을 경계했습니다. 이를 위해 사무실 구조부터 바꾸었습니다. 신경과학자와 컴퓨터 공학자가 옆자리에 앉도 록 배치했고, 정해진 시간마다 모두가 모여 티타임을 가지며 연구 내용을 공유하 게 했습니 '벽돌 깨기' 게임을 하는 AI를 보여주며, 이것이 어떻게 뇌의 도파민 시스템 과 연결되는 지를 신경과학자가 설명하고, 이를 수학자가 수식으로 정리하며, 엔지니어가 코드로 구현 하는 방식이었습니다.

이런 환경에서 '강화학습(Reinforcement Learning)'과 '딥러닝(Deep Learning)'의 결합이라는 딥마인드의 핵심 아이디어가 탄생할 수 있었습니다. 학계에서는 서로 다 른 학회에 참석하 느라 만날 일조차 없었던 분야들이 딥마인드라는 옹광로 안에서 섞 이기 시작한 것입니다. 또한, 허사비스는 스타트업 특유의 '속도감'을 연구에 이식했습니다.

일반적인 학계 연구 가 몇 년에 걸쳐 진행되는 것과 달리, 딥마인드는 프로젝트 단위를 짧게 끊고 빠르게 프로 토타입을 만들어 검증하는 방식을 택했습니다. "실패해도 좋으니 빨리 실패하라(Fail fast)"는 실리콘밸리의 격언을 과학 연구에 적 옹한 것입니다. 하지만 무조건적인 속도전은 아니었습니다.

그는 연구원들에게 논문 출판의 자유를 보장해주었습니다. 기업의 비밀주의 때문에 연구 결과가 세상에 나오지 못하는 것을 꺼리는 학자들을 안심시키기 위함이었습니다. 단, 조건 이 있었습니다.

논문을 쓰기 전에 먼저 내부적으로 기술을 증명해야 한다 는 것이었습니 이는 연구원들에게 건강한 경쟁심을 유발했고, 딥마인드가 훗날 《네 이처》나 《사이언 스》 같은 최고 권위의 학술지에 표지 논문을 쏟아내는 기반이되었습니다. 이러한 하이브리드 문화는 초기에는 술한 시행착오를 겪었습니다. 자유분방한 학자 들은 회사의 규율을 답답해했고, 성과를 중시하는 경영진은 진척 없는 기초 연구를 기 다리기 힘 들어했습니다.

하지만 허사비스는 그 사이에서 완충재 역할을 자처했습니다. 그는 연구자 들에게는 "우리는 과학을 하는 것이다"라고 독려하면서도, 투자자들 에게는 "이 과학이 결 국 거대한 비즈니스가 될 것이다"라고 설득했습니다. 딥마인드는 대학의 지성과 기업의 야성이 공존하는 기묘한 공간이었습니다.

새벽 3 시까 지 알고리즘을 토론하는 열기, 화이트보드 가득한 수식들, 그리고 한쪽 구석에서 돌아가는 고사양 게임기. 이 독특한 문화적 토양 위에서, 인류 역사상 가장 강력한 인 공지능인 알파고가 태동하고 있었습니다. 허사비스가 만든 것은 단순한 회사가 아니 라, 지능을 제조하는 새로운 형태의 '공장'이었습니다. 허사비스의 하루는 두 개의 교대로 나칩니다. 낮에는 CEO로서 회의와 경영을 하 고, 밤에는 "두 번째 교대(second shift)"라 부르는 시간에 논문을 읽고

아이디어를 정리합니다.

새벽 3시까지 이어지는 이 패턴은, 2,000편이 넘는 연구 논문과 h-인덱스 83이라는 학술적 성과의 배경이기도합니다.

딥마인드 런던 사무실

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

갈 것"이라고 경고한 일화 2010년 런던의 러셀 스퀘어 근처, 낡은 사무실에서 시작된 딥마인드는 창립 초기부터 자금 난이라는 현실적인 벽에 부딪혔습니다. 당시 딥마인 드는 당장 팔 수 있는 제품이나 서비스 가 전무했습니다. 그들이 가진 것은 데미스 허 사비스의 머릿속에 있는 '지능을 풀겠다'는 추상적이고 거대한 비전뿐이었습니다.

당 시 영국의 벤처 투자 환경은 지금처럼 성숙하지 않았고, 수익 모델이 없는 과학 연구소에 돈을 댈 투자자는 드물었습니다. 허사비스는 자신의 비전을 이해해 줄 수 있는 곳은 미국 실리콘밸리뿐이라고 판단 했습니다. 샌프란시스코행 비행기에 몸을 실었고, 그곳에서 페이팔(PayPal)의 공동 창업 자이자 급 진적인 기술 투자자인 피터 툴(Peter Thiel)을 만났습니다. 체스 마스터였던 허사비스와 체 스 애호가였던 툴은 '인간 지능의 한계를 넘어서는 수'를 둔다는 공통된 철학 위에서 유대 감을 형성했습니다.

툴의 투자는 딥마인드에게 생명줄과 같았지만, 더 극적인 만남은 그 이후에 기다리고 있었습니다. 피터 툴의 주선으로 허사비스는 스페이스X의 본사 구내식당에서 일론 머스크를 만나게됩니다. 머스크는 인류를 다행성 종조(multi-planetary species)로 만들어 문명 의 멸종을 막 겠다는 원대한 계획에 몰두해 있었습니다.

두 사람의 대화는 기술의 미 래와 인류의 운명 에 대한 심도 깊은 토론으로 이어졌습니다. 머스크는 허사비스에게 자신의 화성 이주 계획을 열정적으로 설명했습니다. 지구에 기후 재앙이나 소행성 충 돌 같은 일이 벌어졌을 때 인류의 문명을 보존하기 위해 화성에 식민지를 건설해야 한 다는 것이었습니다.

허사비스는 잠시 침묵을 지키다 머스크의 눈을 직시하며 차분하지만 서늘한 경고를 던졌습니다. "일론, 당신의 계획은 훌륭하지만 한 가지 간과한 것이 있습니다. 우 리가 인공지능 의 안전성을 확보하지 못한 채 AGI(범용 인공지능)를 개발한다면, 그 시는 화성까지 당신을 따라갈 것입니다."

이 한마디는 머스크에게 큰 충격을 주었습니다. 머스크는 물리적인 재난이나 자원 고갈은 계산에 넣었지만, 초지능

(Superintelligence)이라는 존재가 물리적 거리를 초월해 인류를 위 협할 수 있다는 사 실은 미처 깊게 고민하지 못했던 것입니다. 허사비스는 인공지능이 단순 히 코드로 이 루어진 도구가 아니라, 인간의 통제를 벗어날 수 있는 새로운 종의 탄생일 수 도 있음을 역설했습니다. 그는 지능을 만드는 것만큼이나 중요한 것이 바로 그 지능을 올 바 른 가치관으로 정렬(Alignment)하는 것이라고 강조했습니다.

이 대화는 머스크의 세계관을 뒤흔들었습니다. 그는 그 자리에서 딥마인드에 투자를 결정했습니다. 하지만 금전적 수익을 기대한 투자가 아니었습니다.

훗날 머스크가 고백했듯, 그는 "터미네이터의 등장을 막기 위해, 혹은 적어도 그들이 무엇을 하고 있 는지 감시하기 위해" 투자를 단행했습니다. 허사비스에게 이 투자는 단순한 자금 확보 이상의 의미였습니다. 실리콘밸리의 가 장 영향 력 있는 거물들에게 AGI가 공상과학 소설 속 이야기가 아니라, 당대 내에 실 현 가능한 과 학적 과제이자 인류가 직면한 가장 큰 실존적 위협임을 각인시킨 것입니다.

이로써 딥마인 드는 런던의 작은 스타트업에서 전 세계 기술 리더들이 주목하는 'AI의 진원지'로 급부상하게 되었습니다. 초기 투자자들의 이목을 끈 AGI 비전 딥마인드가 초기 투자자들을 매료시킨 것은 화려한 데모 영상이나 사용자 지표가 아니었습니다. 2010년대 초반, 대부분의 AI 기업들은 '빅데이터 분석'이나 '이미지 인식' 같은 좁은 영역의 인공지능(Narrow AI)에 집중하고 있었습니다.

그들은 더 나은 추천 알고리즘이나 더 정확한 광고 타겟팅 기술을 팔았습니다. 하지만 허사비스는 달랐습니다. 그는 투자자들 앞에서 "우리는 제품을 만들지 않습니다.

우리는 지능 그 자체를 만듭니다"라고 선언했습니다. 허사비스가 제시한 AGI(범용 인공지능) 비전은 '아폴로 프로젝트'나 '맨해튼 프로젝트'에 비견될 만한 거대 과학 프로젝트였습니다. 뇌신경과학(Neuroscience)과 기계학습(Machine Learning)을 융합하여, 인간처럼 스스로 학습하고(Learning), 계획하며 (Planning), 낯선 환경에서도 문제를 해결할 수 있는 범용 알고리즘을 만들겠다는 청사진을 제시했습니다.

이는 당시 기술 수준으로는 불가능해 보이는, 거의 연금술에 가까운 제안이었습니다. 하지만 허사비스에게는 투자자들을 설득할 수 있는 강력한 서사가 있었습니다. 어릴 적 체스 챔피언으로서의 경험, 게임 개발자로서 가상 세계를 창조했던 경험, 그리고 유니버시티 칼리지 런던(UCL)에서 기억과 상상을 연구하며 얻은 뇌과학적 통찰을

하나로 엮어냈습니다. 그는 "인간의 뇌가 물리적인 법칙에 따라 정보를 처리하는 기계라면, 우리는 그 원리를 실리콘 칩 위에서 재현할 수 있다"고 주장했습니다. 또한 그는 게임(Games)을 AI의 지능을 측정하고 훈련시키는 완벽한 실험실(Testbed)로 정의했습니다. 현실 세계의 복잡성을 다루기 전에, 명확한 규칙과 목표가 있는 게임 환경에서 AI를 훈련시켜 지능의 근육을 키우겠다는 전략은 투자자들에게 매우 논리적이고 단계적인 접근으로 받아들여졌습니다.

피터 틸의 파운더스 펀드(Founders Fund)나 홍콩의 리카싱 같은 거물 투자자들은 허사비스가 제시한 '지수함수적 성장'의 가능성에 주목했습니다. 허사비스는 딥마인드 창립 때부터 품어온 신념, 즉 AI가 질병, 에너지, 기후 변화 등 인류의 난제를 해결해 줄 것"이라는 논리를 폈습니다. 이는 단순한 기술 개발이 아니라, 인류 문명의 운영체제 (OS)를 업그레이드하겠다는 제안이었습니다.

투자자들은 딥마인드가 성공할 경우 얻게 될 가치가 기존의 인터넷 기업이나 소프트웨어 기업과는 차원이 다를 것임을 직감했습니다. 허사비스는 기술적 비전뿐만 아니라 조직 운영에 있어서도 독특한 모델을 제시했습니다. 그는 학계의 연구 자유와 스타트업의 빠른 실행력을 결합한 '하이브리드 조직'을 만들겠다고 했습니다.

세계 최고의 과학자들을 한곳에 모아, 논문 실적에 얽매이지 않고 장기적인 난제에 집중하게 하겠다는 그의 구상은 당시 대학 연구 시스템에 염증을 느끼던 천재급 연구원들을 끌어들이는 자석이 되었습니다. 투자자들은 허사비스라는 인물에게서 스티브 잡스의 현실 왜곡장(Reality Distortion Field)과 알베르트 아인슈타인의 과학적 직관이 공존함을 보았습니다. 딥마인드는 제품 하나 없이도 실리콘밸리의 내로라하는 투자자들의 자금을 유치했고, 훗날 구글이 딥마인드를 인수하게 되는 결정적인 발판이 되었습니다.

허사비스는 돈을 구한 것이 아니라, 자신의 꿈 에 동참할 신도들을 모은 셈이었습니다. 페이스북 제안 거절과 구글 2014년 구글에게 인수되다. 2013년 말, 딥마인드의 기술적 성취가 가시화되기 시작하자 실리콘밸리의 거대 기업들이 움직이기 시작했습니다. 가장 먼저 구애의 손길을 내민 것은 페이스북(현 메타)의 마크 저커버그였습니다. 저커버 그는 개인 전용기를 띄워 허사비스와 딥마인드의 공동 창업자들을 캘리포니아

로 초대했습니다. 하지만 협상은 결렬되었습니다. 표면적인 이유는 금액 차이였을 수도 있지만, 근본적인 이유는 철학의 차이였습니다.

허사비스는 페이스북이 딥마인드 의 기술을 단순히 사용자 뉴스피드를 최적화하거나 얼굴 인식 기능을 개선하는 데에 만 쓸 것이라 우려했습니다. 딥마인드가 거대 기업의 부속품이 되는 것을 원치 않았습니다. 이 틈을 파고든 것이 구글의 래리 페이지였습니다.

2014년 1월, 구글은 약 5억 달러 (당시 한화로 약 4억 파운드)라는 천문학적인 금액으로 딥마인드를 인수한다고 발표했습니다. 이 는 구글 역사상 유럽에서 진행한 인수 중 최대 규모였습니다. 당시 딥마인드는 직원이 50 여 명에 불과했고, 매출은 '0'이었습니다.

실리콘밸리의 많은 호사가 들은 구글이 미쳤다고 말했습니다. 아무런 수익 모델도 없는 영국의 작은 연구소에 5,000억 원이 넘는 돈을 쏟아 부은 것은 비이성적인 도박처럼 보였습니다. 하지만 래리 페이지는 허사비스와 마찬가지로 '문샷(Moonshot)'을 꿈꾸는 인물이었습니다.

페이지는 딥마인드가 구글의 검색 엔진을 개선하는 것을 넘어, 구글을 진정한 의미의 'AI 기업'으로 탈바꿈시킬 엔진이 될 것임을 간파했습니다. 허사비스에게 이 인수는 단순한 엑시트(Exit)가 아니었습니다. 그것은 자신의 40년 퀘스트인 AGI 를 실현하기 위해 필요한 막대한 자원, 즉 '연료'를 확보하는 과정이었습니다.

이 거래 는 비밀리에, 그러나 전광석화처럼 진행되었습니다. 허사비스는 훗날 이 순간을 회상 하며 "우리는 로켓을 만들 설계도를 가지고 있었지만, 그것을 우주로 쏘아 올릴 연료 가 없었다. 구글이 바로 그 연료를 제공했 다"라고 비유했습니다.

인수 조건 : 윤리 위원회 설립, 런던 본사 유지, 연구 자율성 확보 허사비스는 협상 테이블에서 놀라울 정도로 깐깐하고 단호했습니다. 그는 딥마인드가 구글에 흡수되어 사라지는 것을 경계했습니다. 그는 '구글의 딥마인드'가 아니라, '구글 안의 딥마인드'로 남기를 원했습니다.

이를 위해 그는 전례 없는 세 가지 조건을 내걸었고, 래리 페이지는 이를 수용했습니다. 첫 번째 조건은 딥마인드의 기술이 윤리적으로 사용되도록 감시할 '윤리 위원회 (Ethics Board)'의 설립이었습니다. 허사비스는 AI 기술이 강력해질수록 오용의 위험도 커진다는 사실을 누구보다 잘 알고 있었습니다.

그는 딥마인드의 기술이 군사적 목적이나 첩보 활동에 사용되는 것을 원천적으로 차단하고 싶어했습니다. 이는 당시 테크 기업 인수 합병 역사상 유례를 찾아보기 힘든 조건이었습니다. 피인수 기업이 인수 기업에게 "당신들이 내 기술을 함부로 쓰지 못하게 감시하겠다"라고 요구한 셈이었으니까요. 이는 허사비스가 AI를 단순한 상품이 아닌, 인류에게 영향을 미칠 거대한 힘으로 인식하고 있었음을 보여줍니다. 두 번째 조건은 본사를 런던에 유지하는 것이었습니다.

보통 구글에 인수된 기업들은 실리콘밸리의 마운틴뷰 본사(Googleplex)로 흡수되는 것이 관례였습니다. 하지만 허사비스는 딥마인드가 실리콘밸리의 상업적 문화에 물드는 것을 거부했습니다. 그는 런던이 가진 학문적 전통, 유럽의 다양한 인재 풀, 그리고 실리콘밸리의 '빠르게 움직여 파괴하라(Move fast and break things)'는 문화와 거리를 둔 차별한 연구 환경이 AGI 개발에 필수적이라고 믿었습니다.

그는 딥마인드를 '상업적 회사'가 아닌 '21세기의 벨 연구소 (Bell Labs)'로 만들고자했습니다. 구글은 이례적으로 런던 킹스 크로스에 딥마인드의 독립적인 왕국을 건설하는 것을 허락했습니다. 세 번째는 연구의 자율성 확보였습니다.

딥마인드는 구글의 단기적인 제품 개발 사 이클에 종속되지 않고, 장기적인 AGI 연구에 집중할 수 있는 권한을 보장받았습니다. 즉, 당장 돈이 되지 않는 바둑 AI나 단백질 구조 예측 같은 기초 과학 연구를 지속할 수 있게 된 것입니다. 이 협상은 허사비스가 단순한 개발자가 아니라, 자신의 미션을 지키기 위해 거대 자본과 대등하게 맞설 수 있는 전략가임을 증명하는 사건이었습니다. "골리앗의 품에 안기다": 컴퓨팅 파워와 리소스 확보를 통한 연구 가속화 구글과의 합병은 딥마인드에게 '골리앗의 힘'을 부여했습니다.

AI 연구, 특히 딥러닝(Deep Learning)과 강화학습은 막대한 양의 컴퓨팅 파워(Compute Power)를 필요로 합니다. 수백 만 번의 시뮬레이션을 돌리고, 거대 신경망을 학습시키기 위해서는 수천, 수만 개의 GPU와 데이터 센터가 필요합니다. 스타트업 규모에서는 감당할 수 없는 비용이었습니다.

구글의 인수로 딥마인드는 전 세계에서 가장 강력한 컴퓨팅 인프라를 마음껏 사용할 수 있는 권한을 얻었습니다. 소송을 들고 싸우던 군대가 갑자기 핵항공모함을 얻은 것과 같았습니다. 구글이 자체 개발한 AI 전용 칩인 TPU(Tensor Processing Unit)의 초기 버전에 대한 접근권은 딥마인드의 연구 속도를 비약적으로 높였습니다.

훗날 세상을 놀라게 할 '알파고 (AlphaGo)'의 훈련과 대국 역시 구글의 클라우드 인프라가 없었다면 불가능했을 일입니다. 이세돌 9단과의 대국에서 알파고는 1,920개의 CPU와 280개의 GPU를 사용했는데, 이렇게 자원을 유동적으로 동원할 수 있는 기업은 지구상에 구글을 포함해 몇 되지 않았습니다. '구글'이라는 브랜드는 전 세계의 인재들을 빨아들이는 블랙홀이 되었습니다.

딥마인드는 구글의 자금력을 바탕으로 학계 최고의 석학들을 영입했고, 박사급 연구원들에게 금융권 수준의 연봉을 제시할 수 있었습니다. 허사비스는 구글이라는 거인의 어깨 위에 올라타, 자신이 꿈꾸던 AGI라는 별을 향해 손을 뻗을 수 있게 되었습니다. 이 합병은 긴장의 씨앗도 품고 있었습니다.

"사악해지지 말자(Don't be evil)"는 구글의 모토와 "지능을 풀자"는 딥마인드의 미션은 겉보기엔 조화로워 보였지만, 거대 상장 기업의 이윤 추구 본능과 순수 과학 연구소의 이상주의는 언젠가 충돌할 수밖에 없는 운명이었습니다. 허사비스는 구글의 리소스를 활용하면서도 그들의 통제로부터 벗어나려 끊임없이 줄 타기를 해야 했습니다. 훗날 딥마인드가 구글 브레인(Google Brain)과 통합되고, 오픈AI와 의 경쟁 속에서 상업화 압박을 받게 되는 과정에서 극명하게 드러나게 됩니다. 2014년의 그 합의는 딥마인드를 전설로 만드는 시작점이었지만, 동시에 허사비스에게는 '기업가'와 '과학자'라는 두 정체성 사이에서 끊임없이 고민하게 만드는 족쇄의 시작이기도 했습니다.

구글이 딥마인드를 인수했을 때, 허사비스는 하룻밤 사이에 거대한 부를 손에 쥐었습니다. 실리콘밸리의 관례대로라면 팔로알토 저택으로 이사하고 테슬라를 몰 법도했습니다. 허사비스는 그러지 않았습니다.

노스런던 하이게이트에 남았고, 2016년까지 기차를 타고 출퇴근했습니다. 평생의 리버풀 FC 팬이기도 한 그는, 2019년 안필드에서 바르셀로나를 4대 0으로 꺾는 역전을 직접 목격했습니다. 노벨상 수상자가 될 사람이 "You'll Never Walk Alone"을 따라 부르며 환호하던 밤이었습니다.

딥마인드 시절의 데미스 허사비스

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

로 게임을 학습하는 AI 2013년 런던의 늦은 밤, 딥마인드의 작은 사무실에는 묘한 긴 장감과 피로가 섞인 공기가 감돌고 있었습니다. 창문 밖으로는 런던의 차가운 밤거리가 보였지만, 데미스 허사비스와 그의 동료들의 시선은 오직 모니터 속 깜빡이는 거친 픽셀들에 고정되어 있었습니다. 화면 속에는 1970년대의 유물인 아타리 2600 게임기가 구동되고 있었습니다.

화려한 그래픽도, 웅장한 사운드도 없는 투박한 8비트 화면이었지만, 그들이 보고 있는 것은 단순한 고전 게임이 아니었습니다. 그것은 인공지능의 역사, 아니 인류가 도구를 만드는 방식이 근본적으로 바뀌는 순간이었습니다. 그들이 만든 인공지능 에이전트, 훗날 DQN(Deep Q-Network)이라 불리게 될 이 시스템은 갓 태어난 아기와 같았습니다.

연구팀은 이 AI에게 게임의 규칙을 단 하나도 가르쳐주지 않았습니다. 이것이 적기인지 아군인지, 점수는 어떻게 얻는지, 심지어 게임을 클리어하려면 무엇을 해야 하는지조차 알려주지 않았습니다. AI가 받은 것은 오직 두 가지뿐이었습니다. 하나는 사람의 눈과 같은 역할을 하는 화면의 픽셀 정보(시각 데이터)였고, 다른 하나는 게임의 점수(보상)가 오르고 내리는 신호였습니다.

"점수를 높여라." 이것이 AI에게 주어진 유일한 본능이자 명령이었습니다. 초기 단계의 AI는 그야말로 엉망진창이었습니다.

화면 속의 라켓은 제멋대로 움직였고, 날아오는 공을 피하기 일쑤였으며, 의미 없는 버튼만 연타했습니다. 마치 비디오 게임기를 처음 잡은 세 살배기 어린아이보다도 못했습니다. 하지만 허사비스와 연구팀은 실망하지 않았습니다.

그들은 이 무작위적인 혼돈 속에서 질서가 피어나는 순간을 기다리고 있었습니다. 딥러닝이라는 인공지능의 뇌와 강화학습이라는 당근과 채찍의 매커니즘이 결합하여 작동하기를 숨죽여 지켜보았습니다. 시간이 흐르자 놀라운 일이 벌어졌습니다.

수천 번, 수만 번의 시행착오를 거치면서 AI는 화면 속 픽셀들의 패턴을 스스로 인식하기 시작했습니다. "이 하얀 점(공)이 아

래로 내려올 때, 내 하얀 막대(라켓)를 그 밑에 갖다 대면 점수라는 숫자가 올라가는구나." AI는 누구의 가르침도 없이 스스로 인과관계를 추론해냈습니다. 이것은 단순히 입력된 명령을 수행하는 기존의 컴퓨터 프로그램과는 차원이 다른 것이었습니다.

그 것은 '학습'이었습니다. 인간이 세상을 배우는 방식, 즉 경험을 통해 스스로 깨우치는 과정을 기계가 재현하고 있었습니다. 이 순간은 허사비스가 오랫동안 꿈꿔왔던 '범용 학습 알고리즘'의 가능성이 증명되는 찰나였습니다.

체스면 체스, 바둑이면 바둑, 특정 게임만을 위해 규칙을 하드코딩한 AI는 그 게임 밖에 할 줄 모릅니다. 체스 챔피언을 꺾은 딥블루에게 틱택토를 시키면 틀조차 이해하지 못합니다. 하지만 DQN은 달랐습니다. '스페이스 인베이더'를 하던 똑같은 코드가 아무런 수정 없이 '퐁(Pong)'을 하고, '브레이크아웃(벽돌 깨기)'을 해냈습니다.

단지 화면의 픽셀만 보여주었을 뿐인데, AI는 스스로 그 세계의 물리학과 규칙을 터득해 낸 것입니다. 이것은 인공지능이 닫힌 세계의 전문가에서 열린 세계의 학습자로 진화하는 첫 번째 발자국이었습니다. '벽돌 깨기(Breakout)'에서 AI가 스스로 발견한 터널링 전략이 프로젝트의

하이러 이트는 '벽돌 깨기(Breakout)' 게임 테스트에서 일어났습니다.

2013년, 구글이 딥마인드 드를 인수하기 직전, 허사비스는 이 데모를 통해 래리 페이지와 구글 임원들에게 딥 마인드의 가치를 증명해야했습니다. 그리고 그 증명의 순간은 AI 연구자들조차 예은 예상 밖이었습니다. 훈련이 시작된 지 10분 정도 지났을 때, DQN은 제법 능숙하게 공을 받아내기 시작 했습니다. 사람으로 치면 꽤 게임을 잘하는 아마추어 수준이었습니다.

공을 놓치지 않고 꼬박꼬 박 받아쳐 벽돌을 하나하나 깨뜨렸습니다. 연구원들은 고개를 끄덕였습니다. "잘 작동하는 군. 인간 수준에 도달했어."

그들은 이 정도면 충분한 성과라고 생각했습니다. 하지만 AI는 멈추지 않았습니다. 훈련 2시간이 경과하자, 화면 속 AI의 움직임이 미묘하게 변했습니다.

단순히 공을 받아내 는 것에 만족하지 않고, 공을 라켓의 모서리에 정교하게 맞춰 튕겨내는 기술을 구사하기 시작했습니다. 4시간이 지난 후, 그 누구도 가르쳐주지 않았고 예상조차 못

했던 장면이 펼쳐졌습니다. AI는 집요하게 벽돌 벽의 한쪽 끝만을 집중 공략했습니다. 계속해서 같은 자 리에 공을 보내 마침내 벽의 가장자리 수직 공간을 뚫어버렸습니다.

구멍이 뚫리자 AI는 그 좁은 틈 사이로 공을 밀어 넣었습니다. 벽 뒤쪽 공간으로 들어간 공은 벽과 천장 사이를 맹렬한 속도로 오가며 튕겼습니다. '티디 디딕' 하는 소리와 함께 순식간에 수많은 벽돌이 와르르 무너져 내렸고, 점수는 기하급수적으로 폭등했습니다.

바로 '터널링(Tunneling)' 전략이었습니다. 고수들만 이 알고 있다는 그 비기를, AI가 스스로 찾아낸 것입니다. 그 장면을 지켜보던 연구원들은 침묵에 빠졌다가 곧 환호했습니다.

이것은 단순한 점수 획 득이 아니었습니다. AI가 '전략적 사고'를했다는 증거였습니다. "당장 눈앞의 공을 받는 것 이 안전하지만, 위험을 무릅쓰고라도 한쪽 벽을 뚫으면 나중에 엄청난 보상을 얻을 수 있다."

이 복잡한 인과관계와 장기적인 계획을, AI는 오직 픽셀의 움 직임과 점수판의 변화만으로 깨달은 것입니다. 허사비스는 이 순간을 회상하며 "마치 AI가 생각하는 것을 눈으로 보는 듯했다"고 말했습니다. 프로그래머가 입력한 코드가 아니라, 데이터와 경험이 만들어 낸 창발적 지능이었습니다.

이 '터널링' 사건은 딥마인드 팀에게 확신을 주었습니다. 자신 들이 가고 있는 길이 옳다는 확신, 그리고 언젠 가 이 지능이 게임을 넘어 과학 난제라는 거 대한 벽에도 터널을 뚫을 수 있을 것이라는 믿음이었습니다. 2013년 Nature 논문 발표와 AI 커뮤니티의 충격 딥마인드의 이 성과는 2013년 NIPS(현 NeurIPS) 딥러닝 워크숍에서 처음 공개되었고, 이후 보강되어 2015년 세계 적인 과학 저널 《네이처(Nature)》의 표지를 장식했습니다.

논문의 제목은 《심층 강화학습을 통한 인간 수준의 제어(Human-level control through deep reinforcement learning)》였습니다. 과학계, 특히 AI 커뮤니티가 받은 충격은 '쇼 크'라는 단어로밖에 표현할 수 없었습니다. 당시까지의 AI 학계 주류는 여전히 인간이 규칙을 정교하게 설계해 주는 방식에 익숙해 있었습니다.

혹은 딥러닝은 이미지 인식 같은 정적인 작업에는 뛰어나지만, 게 임처럼 연속적 인 의사결정이 필요한 동적인 환경에서는 작동하기 어렵다는 회의론이 지배적이었습니다.

강화학습은 이론적으로는 훌륭했지만, 현실의 복잡한 문제를 풀기에는 너무 불안정 하다 는 것이 정설이었습니다. 그런데 런던의 작은 스타트업이 모든 통념을 깨뜨린 것입니다. 이 논문이 충격적이었던 이유는 '범용성(Generality)' 때문이었습니다.

그들이 만든 단 하나 의 알고리즘 네트워크(DQN)는 아타리 2600의 49개 게임 중 29개에서 인간 전문가 이상의 실력을 보여주었습니다. 복싱, 비디오 핀볼, 스페이스 인베يدر 등 전 혀 다른 규칙과 목표를 가진 게임들을 똑같은 뇌가 학습해낸 것입니다. 이는 AI가 특 수 목적 도구에서 범용 도 구로 진화할 수 있다는 강력한 신호였습니다.

《네이처》에 컴퓨터 과학, 그것도 게임을 하는 AI 논문이 표지로 실린다는 것 자 체가 이례적이었습니다. 이는 AI 연구가 단순한 공학적 시도를 넘어, 지능의 본질을 탐구하는 자 연과학의 영역으로 진입했음을 상징했습니다. 전 세계의 연구자들은 DQN의 소스 코드를 분석하며 열광했고, 구글은 이 가능성을 선점하기 위해 거액을 베풀었습니다.

아타리 쇼크 는 AI 연구의 겨울을 끝내고, 딥러닝의 황금기를 여는 기 폭제가되었습니다. 그리고 그 중 심에 "지능을 풀겠다"는 허사비스의 집요한 여정이 있었습니다. 왜 게임인가? 게임의 난이도와 상징성의 계산 게임은 단순한 오락이 아니었습니다.

그것은 지능을 측정하고 훈련하기 위해 인류가 발명 한 가장 정교한 '실험실'이었습니다. 그는 종종 "게임은 현실 세계의 축소판(Microcosm)"이 라고 말했습니다. 현실 세 계는 너무나 복잡하고, 잡음이 많으며, 결과가 나오기까지 너무 오랜 시간이 걸립니다. 현실에서 AI에게 주식 투자를 배우게 하거나 로봇을 걷게 하려면 엄청난 비용과 위험이 따릅니다. 하지만 게임은 다릅니다. 안전하고, 무한히 반복 가능하 며, 명확한 목표(승리 또는 점수)가 존재합니다.

허사비스는 게임의 난이도를 수학적으로 계산하며 AI의 발전 단계를 설계했습니다. 아타리 게임은 2차원 평면의 픽셀 정보를 처리하는 단계였습니다. 이는 시각 피 질의 초기 정보 처리 능력을 검증하는 것이었습니다.

바둑(Go)은 '완전 정보 게임 (Perfect Information Game)'의 정점이었습니다. 바둑판 위에는 숨겨진 정보가 없습니다. 서로의 모든 수를 보고 있는 상태에서 순수한 수싸움과 직관, 그리고 패턴 인식 능력을 겨루는 것입니다. 바둑의 경우의 수는 우주의 원자 수보다 많기 때문에, 이는

단순한 계산 능력을 넘어선 '직관'의 영 역을 AI가 정복할 수 있는지를 묻는 시험대였습니다. 허사비스의 시선은 바둑 너머를 향해 있었습니다. 현실 세계는 바둑판처럼 모든 정 보가 공 개되어 있지 않습니다.

상대방이 무슨 생각을 하는지, 내일 날씨가 어떨지, 가려진 커튼 뒤에 무엇이 있는 지 모른 채 의사결정을 내려야합니다. 이를 '불완전 정보 게임(Imperfect Information Game)'이라 합 니다. 포커나 스타크래프트 같은 게임이 여기에 해당합니다. 허사비스는 아타리로 '감각'을, 알파고로 '직관'을 증명한 뒤, 최종적으로 현실과 가 장 닮은 불확실성의 세계, 즉 '스타크래프트 II'로 AI를 밀어 넣을 계획을 세웠습니다.

이는 단순히 게임을 잘하는 AI를 만드는 것이 아니라, 불확실한 현실 세계에서 과학적 가설을 세우고 실험을 설계할 수 있는 AI, 즉 '과학자 AI'를 만들기 위한 필수적인 훈 련 과정이었습니다. 스타크래프트 II 정복(AlphaStar): 불완전 정보 환경에서의 실시간 전략 2019년 1 월, 딥마인드는 또 한 번 세상을 놀라게했습니다. 이번 무대는 바둑판이 아니라 실시 간 전략 시뮬레이션 게임(RTS)인

'스타크래프트 II'였습니다.

알파고가 정적인 턴제 게임의 제약이었다면, '알파스타(AlphaStar)'는 찰나의 순간에 수백 가지 결정을 내려야 하는 동적인 전장의 지휘관이었습니다. 스타크래프트 II는 AI에게 악몽과도 같은 과제였습니다. 첫째, '전장의 안개(Fog of War)'가 존재합니다.

플레이어는 자신의 유닛이 있는 곳 외에는 볼 수 없습니다. 상대가 무엇을 짓고 있는지, 병력을 어디로 이동시키는지 보이지 않습니다. 따라서 AI는 끊임없이 정찰 (Scouting)을 통해 정보를 수집하고, 보이지 않는 상대의 행동을 예측하고 추론해야 합니다. 이는 정보가 부족한 상황에서 가설을 세우고 검증하는 고도의 지능적 행위입니다.

둘째, '실시간(Real-time)'의 압박입니다. 바둑은 다음 수를 생각할 시간이 주어지지만, 스타크래프트는 실새 없이 흘러갑니다. 셋째, Long time horizon의 문제입니다. 게임 초반에 일꾼 한 기를 더 생산하기로 한 결정이 20분 뒤 대규모 전투의 승패를 가를 수 있습니다.

AI는 수천, 수만 프레임 뒤의 미래를 내다보고 현재의 행동이 가져올 나비효과를 계산해야 했습니다.

알파스타는 프로게이머 'TLO(다리오 빈시)'와 'MaNa(Grzegorz Komincz)'를 상대로 압도적인 승리를 거두었습니다. 사람들은 알파스타가 엄청난 속도로 마우스를 클릭하는 피지컬 (APM)로 이겼을 것이라 예상했지만, 딥마인드는 AI의 클릭 속도를 인간 수준으로 제한했습니다. 알파스타가 보여준 것은 단순한 반응 속도가 아니었습니다.

그것은 소름 돋는 '판 단력'이었습니다. 알파스타는 정찰을 통해 상대의 체제를 파악하고, 그에 맞춰 병력 조합을 유동적으로 바꿨습니다. 불리한 상황에서는 과감하게 후퇴하고, 상대의 빈틈이 보이면 주저 없이 파고들었습니다.

특히 MaNa와의 대결에서 보여준 '점멸 추적자' 컨트롤은 인간이 흉내 낼 수 없는 정교함의 극치였지만, 더 놀라운 것은 전체 전장을 아우르는 전략적 시야였습니다. 한계도 있었습니다. MaNa가 알파스타의 시야 밖에서 끊임없이 견제하자 알파스타가 혼란스러워하는 모습이 포착되기도 했습니다.

하지만 딥마인드는 이를 금세 수정했고, 결국 알파스타는 그랜드마스터 레벨에 도달했습니다. 허사비스에게 알파스타의 승리는 AI가 불확실하고 복잡하며 실시간으로 변하는 환경, 즉 '현실 세계'에서도 복잡한 전략을 수립하고 목표를 달성할 수 있다는 강력한 증거였습니다. 아타리 브레이크아웃(Breakout) 게임 화면

인공지능 전문가 김경진 변호사

AI 법정정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

년 5월, 뉴욕의 한 빌딩에서 체스 세계 챔피언 가리 카스파로프(Garry Kasparov)가 머리를 감싸 쥐고 패배를 인정했습니다. IBM의 슈퍼컴퓨터 '딥블루(Deep Blue)'가 인간 지성의 자존심을 무너뜨린 역사적인 순간이었습니다. 당시 20대 초반의 청년이었던 데미스 허사비스는 이 장면을 보며 복잡한 감정을 느꼈습니다.

그는 이미 13세에 체스 마스터 등급에 올랐던 신동이었고, 컴퓨터가 인간을 이기는 것이 얼마나 어려운 일인지 누구보다 잘 알고 있었습니다. 하지만 허사비스의 눈에 딥블루의 승리는 진정한 '지능'의 승리가 아니었습니다. 그것은 압도적인 연산 속도로 모든 경우의 수를 계산해 낸, 거대하고 무식한 계산기의 승리였습니다.

"체스는 이제 풀렸다. 하지만 지능은 풀리지 않았다." 허사비스에게 체스는 '달린 시스템'이었습니다. 규칙은 명확하고 정보는 모두 공개되어 있으며, 계산 능력만 충분하다면 언젠가는 정답을 찾을 수 있는 문제였습니다.

그가 꿈꾸는 범용 인공지능(AGI), 즉 인간처럼 직관을 가지고 불확실한 상황에서 판단을 내리는 지능을 만들기 위해서는 체스보다 훨씬 더 거대하고, 혼돈스럽고, 심오한 도전과제가 필요했습니다. 그 답은 동양의 고대 게임, '바둑'이었습니다. 서양의 컴퓨터 과학자들에게 바둑은 오랫동안 '넘을 수 없는 벽'이자 인공지능 연구의 '성배(Holy Grail)'로 불렸습니다.

그 이유는 단순하면서도 압도적인 '복잡성' 때문입니다. 체스 판은 8x8, 총 64개의 칸으로 이루어져 있지만, 바둑판은 19x19, 총 361개의 교차점을 가집니다. 바둑에서 가능한 수의 조합은 10의 170승에 달합니다. 이는 전 우주에 존재하는 모든 원자의 수(약 10의 80승)를 합친 것보다 훨씬 더 큰 숫자입니다. 딥블루가 사용했던 '무작위 대입(Brute Force)' 방식, 즉 모든 수를 일일이 계산해서 답을 찾는 방식은 바둑에서 절대 통하지 않는다는 것입니다. 바둑을 두는 인간 기사

들은 모든 수를 계산하지 않습니다. 그들은 '모양'을 보고, '흐름'을 읽으며, '직관'을 사용합니다. "여기가 좋아 보여서 뒀다"는 프로 기사의 말은 단순한 느낌이 아니라, 수만 번의 대국 경험이 뇌속에서 압축되어 나오는 고도로 정제된 패턴 인식의 결과입니다.

허사비스가 바둑을 선택한 것은 바로 이 지점 때문이었습니다. 바둑을 정복하려면 컴퓨터에게 계산 능력이 아니라 '직관(Intuition)'을 가르쳐야 했습니다. 만약 AI가 바둑에서 인간 최고수를 이길 수 있다면, 그것은 기계가 비로소 인간의 고유 영역이라 믿었던 직관과 창의성을 흉내 내는 단계에 진입했음을 증명하는 사건이 될 것이기 때문입니다.

딥마인드 팀에게 바둑은 현실 세계의 축소판(Microcosm)이었습니다. 현실 세계는 정해진 답이 존재하지 않습니다. 무수히 많은 선택지가 있고, 그 선택의 결과가 즉각적으로 나타나지 않으며, 불확실성이 가득합니다.

허사비스는 바둑을 통해 AI가 이 '무한한 탐색 공간'을 어떻게 효율적으로 항해할 수 있는지 검증하고자했습니다. 2014년 당시 최고의 바둑 AI조차 아마추어 5단 수준에 머물러 있었습니다. 프로 기사에게는 4~5점을 미리 깔고 두어도 이기지 못하는 수준이었습니다.

많은 전문가들은 "AI가 바둑 챔피언을 이기려면 적어도 10년은 더 걸릴 것"이라고 예측했습니다. 하지만 허사비스는 그 10년의 시간을 단축시킬 준비가 되어 있었습니다. 그는 뇌과학에서 얻은 영감을 바탕으로, 기존의 규칙 기반 프로그래밍을 버리고 기계가 스스로 학습하는 시

시스템을 구축하기 시작했습니다.

딥마인드의 런던 본사에서는 조용하지만 치열하게, 인류 지성사의 흐름을 바꿀 준비가 진행되고 있었습니다. 강화학습, 몬테카를로 트리 탐색(MCTS), 심층 신경망의 결합 허사비스와 딥마인드의 수석 연구원 데이비드 실버(David Silver)가 설계한 알파고의 아키 텍처는 마치 인간의 뇌가 작동하는 방식을 소프트웨어로 구현한 것과 같았습니다. 이들은 세 가지 강력한 기술을 정교하게 결합하여 바둑이라는 거대한 우주를 공략했습니다.

그것은 바로 심층 신경망(Deep Neural Networks), 강화학습(Reinforcement Learning), 그리고 몬테카를로 트리 탐색(MCTS)이었습니다.

첫 번째 열쇠는 '심층 신경망'이었습니다. 이것은 알파고의 '눈'과 '뇌' 역할을 했습니다. 딥마인드 팀은 이를 다시 두 가지 네트워크로 나누었습니다.

하나는 '정책망 (Policy Network)' 이고, 다른 하나는 '가치망(Value Network)'입니다. 정책망은 "다음에 어디에 돌을 두어야 할까?"를 결정합니다. 바둑판 위에는 당장 둘 수 있는 곳이 수백 군데가 넘습니다.

인간 고수들은 척 보면 "둘 만한 곳"을 두세 군데로 좁힙니다. 정책망은 바로 이 역할을 합니다. 알파고는 온라인 바둑 사이트인 KGS에서 확보한 16만 개의 기보, 총 3,000만 가지의 움직임을 학습했습니다. 이를 통해 알파고는 인간이 주로 두는 수의 패턴을 익혔고, 텅 빈 바둑판에서 확률적으로 가장 좋은 위치를 좁혀내는 능력을 갖추게 되었습니다.

이는 마치 캄캄한 방에서 전체를 더듬는 대신, 손전등으로 가장 중요한 물건만 비추는 것과 같습니다. 가치망은 "지금 이 판세가 나에게 유리한가?"를 판단합니다. 바둑은 집이 많은 쪽이 이기는 게임입니다.

하지만 대국 중반까지는 누가 얼마나 이기고 있는지 정확히 계산 하기가 불가능에 가깝습니다. 가치망은 현재 바둑판의 형세를 보고 승리 확률을 예측합니다. "흑이 이길 확률 51%"와 같이 수치로 판단을 내려주는 것입니다.

이 덕분에 알파고는 무리하게 승부를 걸거나 지나치게 소극적으로 두는 대신, 승리 확률을 가장 높이는 길을 선택할 수 있었습니다. 두 번째 열쇠는 '강화학습'이었습니다. 이것은 알파고를 초인적인 수준으로 끌어올린 핵심 동력이었습니다.

단순히 인간의 기보를 흉내 내는 것만으로는 인간을 뛰어넘을 수 없습니다. 스승을 넘어서기 위해서는 스승이 가르쳐주지 않은 것까지 배워야 합니다. 이를 위해 허사비스는 알파고가 자기 자신과 바둑을 두게 했습니다. 이를 '셀프 플레이(Self-Play)'라고 합니다.

알파고는 하루에 수만 판씩 자기 자신과 대국을 벌였습니다. A버전의 알파고와 B버전의 알파고가 싸워서 이긴 쪽의 전략은 채택하고, 진 쪽의 전략은 수정했습니다. 이 과정에서 알파고는 인간이 수천 년 동안 쌓아온 정석(定石)을 재검증했고, 때로는 인간이 시도하지 않았던 기상천외한 수들이 실제로 효과가 있다는 것을 스스로 깨우쳤습니다.

허사비스는 이 과정을 보며 "우리는 알파고에게 물고 기를 주는 것이 아니라, 낚시하는 법을 가르쳤다. 그리고 이제는 그 낚시법을 스스로 개량하고 있다"라고 표현했습니다.

세 번째 열쇠는 '몬테카를로 트리 탐색(MCTS)'이었습니다. 이것은 알파고의 '수읽기' 능력입니다. 앞서 정책망이 유력한 후보지 몇 곳을 추천하면, MCTS는 그곳에 돌을 뒀을 때 이후 상황이 어떻게 전개될지를 시뮬레이션합니다.

하지만 모든 경우를 끝까지 두어볼 수는 없습니다. MCTS는 마치 여론조사를 하듯, 무작위로 몇 가지 경로를 끝까지 빠르게 두어본 후 그 결과를 바탕으로 가장 승률이 높은 수를 선택합니다. 기존의 바둑 AI들이 이 기술만 의존했을 때는 아마추어 수준에 그쳤지만, 딥마인드는 여기에 심층 신경망의 직관을 결합함으로써 계산의 효율성을 극대화했습니다.

이 세 가지 기술의 결합은 혁명적이었습니다. 정책망이 인간의 직관처럼 후보를 좁혀주면(Search width 축소), 가치망이 끝까지 두어보지 않고도 형세를 판단하여 깊이를 줄여주었고(Search depth 축소), MCTS가 정교한 수읽기로 실수를 방지했습니다. 이것은 직관(Intuition)과 논리(Logic)의 완벽한 조화였습니다.

허사비스는 이 시스템을 구축하면서 인간 뇌의 작동 원리인 '판단 하고, 학습하고, 계획하는' 메커니즘을 규명하고 있다고 확신했습니다. 알파고는 '지능'이라는 추상적인 개념을 수학과 코드로 빚어낸, 인류가 만든 가장 우아한 예술품 중 하나였습니다. 2015년 판 후이(Fan Hui) 2단과의 비공개 대국 유럽 챔피언을 5전 5승으로 꺾은 비밀 대국 2015년 10월, 런던의 킹스 크로스(King's Cross)에 위치한 딥마인드 본사에는 묘한 긴장감이 감돌았습니다.

창문이 없는 작은 회의실, 그곳에는 컴퓨터 모니터 한 대와 실제 바둑판 하나가 놓여 있었습니다. 이 방으로 초대된 사람은 중국 출신의 프로 기사이자 당시 유럽 바둑 챔피언이었던 판 후이(Fan Hui) 2단이었습니다. 그는 딥마인드로부터 "새로운 바둑 프로그램의 테스트를 도와달라"는 정중한 초청을 받고 런던에 도착했습니다.

판 후이는 초청을 수락할 때만 해도 큰 기대를 하지 않았습니다. 당시 최고의 바둑 프로그램이라 해도 프로 기사에게는 접바둑(핸디캡 매치)이 아니면 상대가 되지 않았기 때문입니다. 자신이 기계의 오류를 찾아주고 조언을 해주는 '자문역' 정도를 수행할 것이라 생각했습니다. 대국 전, 딥마인드는 그에게 비밀 유지 서약서(NDA)에서 명을 요청했습니다.

허사비스는 이 대국이 단순한 테스트가 아니라, AI 역사의 분기점이 될 것임을 예감하고 있었습니다.

대국이 시작되었습니다. 판 후이는 여유롭게 첫수를 두었습니다. 모니터 속의 알파고는 딥마인드의 수석 프로그래머 아자 황(Aja Huang) 박사의 손을 빌려 바둑판 위에 돌을 놓았습니다.

초반 포석이 진행되면서 판 후이의 표정은 점차 굳어지기 시작했습니다. 알파고의 수는 기계적이지 않았습니다. 엉뚱한 곳에 돌을 놓거나, 의미 없는 수를 두며 자멸하던 기존 AI와는 차원이 달랐습니다.

알파고는 인간처럼 유연했고, 때로는 인간보다 더 침착했습니다. 첫 번째 판, 판 후이의 패배였습니다. 그는 당황했습니다.

"시차 적응 때문일 거야, 내가 방심했어"라고 스스로를 다독였습니다. 하지만 이어진 두 번째, 세 번째 대국에 서로 결과는 같았습니다. 알파고는 판 후이가 공격해 오면 두터게 방어했고, 판 후이가 물러서면 날카롭게 파고들었습니다.

그것은 마치 보이지 않는 거대한 벽과 바둑을 두는 느낌이었습니 판 후이는 대국 도중 잠시 밖으로 나가 세수를 하며 마음을 추스르려 했지만, 모니터 앞 에 앉으면 다시금 숨이 막혀오는 압박감을 느꼈습니다. 최종 스코어 5대 0. 유럽 챔피언 판 후이의 완패였습니다. 역사상 최초로 컴퓨터가 핸디캡 없이 호선(맞바둑)으로 프로 기사를 꺾은 사건이었습니다.

마지막 대국이 끝 나고 패배가 확정된 순간, 판 후이는 멍하니 바둑판을 응시했습니다. 충격은 곧 경외 감으로 바뀌었습니 그는 자신이 방금 인류 역사상 가장 위대한 기술적 도약의 순간을 목격했음을 깨달았습니다. 승패가 결정된 후, 판 후이는 딥마인드 팀의 환호를 받았습니다.

허사비스는 정중하 게 그에게 감사를 표했습니다. 비록 패배했지만, 판 후이는 이후 5개월 동안 딥마인드 팀과 협력 하며 알파고의 약점을 보완하는 데 결정적인 역할을했습니다. 그는 알 파고가 둔 수들의 의미를 해석해 주었고, 알파고가 '인간적인 기풍'을 갖추는 데 기여했습니다.

훗날 판 후이는 "알파고와 바둑을 두면서 나 자신을 비추어보는 거울을 보는 듯했 다. 알파고는 나의 내면을, 나의 두려움을 보여주었다"라고 회고했습니다. 이 비밀스 러운 승리는 딥마인드 팀에게 "우리가 옳았다"는 확신을 심어주었지만, 세상은 아직 이 거대한 파도를 전혀 감지하지 못하고 있었습니다. Nature 표지 논문 발표와 세계의 충격 2016년 1월 27일, 세계적인 과학 저널 《네 이처(Nature)》의 표지는 바둑판을 형상화한 이 이미지로 장식되었습니다.

논문의 제목 은 "심층 신경망과 트리 탐색을 통한 바둑 게임의 정 복(Mastering the game of Go with deep neural networks and tree search)". 저자는 데이비드 실 버, 아자 황, 그리 고 데미스 허사비스 등이었습니다. 이 논문의 발표는 전 세계 과학계와 바둑계에 그야말로 '핵폭탄'과 같은 충격을 안겨주었습니다. 논문에는 알파고의 아키텍처에 대한 상세한 설명과 함께, 지난 10월 판 후이와의 비 공개 대국 결과(5전 전승)가 실려 있었습니다.

AI 연구자들은 경악했습니다. 대부분의 전문 가들은 바둑 정복이 10년, 혹은 20년 후에나 가능할 것이라 예 측했기 때문입니다. 불과 몇 달 전까지만 해도 "아직 멀었다"고 평가받던 기술이, 순식 간에 인간 프로를 꺾는 수준으로 도약한 것입니다.

언론들은 앞다퉈 이 소식을 타전했습니다. "인공지능이 인간의 마지막 성역을 침범했다", "기계가 인간의 직관을 넘어섰다"는 헤드라인이 쏟아졌습니다. 허사비스는 기 자회견을 통 해 이 성과를 발표하며, 단순히 게임을 이긴 것이 아니라 "범용 인공지능 알고리즘의 가능 성을 확인한 것"이라고 강조했습니다.

알파고가 특정 바둑 규칙을 입 력받은 것이 아니라, 수많은 데이터를 통해 스스로 학습했다는 점을 부각했습니다. 알 파고의 기술이 바둑뿐만 아니라 질병 진단, 기후 예측 등 복잡한 난제를 해결하는 데 쓰일 수 있다는 선언이었습니다. 바둑계의 반응은 사뭇 달랐습니다.

충격적이긴 했지만, 여전히 회의적인 시각이 지 배적이었습니다. 그 이유는 판 후이의 실력 때문이었습니다. 판 후이는 유럽 챔피언이었지만, 세 계 최정상급 기사들이 즐비한 한국이나 중국의 기준으로 보면 한 수 아 래로 평가받았습니 이세돌 9단이나 커제 9단 같은 '초일류' 기사들과는 비교할 수 없 는 수준이라는 것이 바 둑계의 중론이었습니다.

한국의 이세돌 9단은 인터뷰에서 "놀라운 성과이긴 하지만, 아직 나를 이길 수준은 아니다. 5대 0, 혹은 4대 1 정도로 내가 이길 것이다"라고 자신감을 내비쳤습니다. 그 는 판 후이의 기보를

분석한 뒤, 알파고의 실력이 프로 정상급에는 미치지 못한다고 판단했습니다. 이러 한 반응은 오히려 딥마인드와 허사비스에게 기회였습니다.

그들 은 이미 판 후이와의 대국 이후 몇 달간 알파고를 비약적으로 업그레이드시킨 상태였습니다. 《네이처》 발표와 함께 허사비스는 전 세계를 향해 대담한 도전장을 던졌습니다. "전설적 인 기사, 이세돌 9단에게 도전합니다."

인간의 직관과 기계의 연산, 자연 지능 과 인공지능 이 정면으로 충돌하는, 21세기 과학사의 가장 드라마틱한 쇼의 서막이었습니다. 전 세계의 이목이 서울로 집중되기 시작했고, 허사비스는 이 모든 상황을 차 분하게, 그러나 치밀하게 준비하고 있었습니다. 그는 알고 있었습니다.

판 후이와의 대국은 예고편에 불과하며, 진 짜 승부는 이제부터 시작이라는 것을 말합니다. 바둑판 위의 흑돌과 백돌

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

월 10일 오후, 서울 포시즌스 호텔의 특별 대국장은 침묵에 잠겨 있었습니다. 유리 로 둘러싸인 부스 안에는 바둑판에 돌을 놓는 소리와 냉난방기가 돌아가는 낮은 웅웅거림만이 감돌았습니다. 제2국의 중반, 알파고가 검은 돌을 들어 우변 5선의 낮선 위치에 내려놓았습니다.

제37수였습니다. 이 수는 수천 년 바둑 역사에서 금기시되던 착점이었습니다. 대국장을 지켜보던 해설자들과 프로 기사들은 처음에 이것을 명백한 실수, 즉 '떡수'라고 판 단했습니다.

"마 우스 미스(조작 실수)가 아닐까요?"라는 당혹스러운 해설이 생방송을 탕습니다. 그러나 이 세돌 9단은 달랐습니다. 그는 그 수를 보는 순간 눈을 크게 뜨고 바둑판을 응시했습니다.

그의 표정에는 당혹감과 경외감이 동시에 스쳐 지나갔습니다. 이 순간은 인공지능 역사에서 '창의성'이라는 단어의 정의가 다시 쓰인 분기점이었습니다. 이전까지의 체스 AI인 딥블루는 무자비한 계산 능력으로 인간을 압도했습니다.

하지만 알 파고의 37수는 계산이 아니라 마치 '직관'처럼 보였습니다. 인간이 가르 쳐준 기보에는 없는 수였고, 인간 스승을 뛰어넘어 스스로 새로운 길을 개척한 수였습니다. 데미스 허사비 스는 통제실 모니터 뒤에서 이 장면을 지켜보며 전율했습니다.

그가 꿈꿔왔던 일반 인공지능(AGI)의 가능성, 즉 기계가 인간의 지식을 모방하는 것을 넘어 새로운 지식을 창출할 수 있다는 가능성이 증명된 순간이었기 때문입니다. 유럽 챔피언이자 알파고의 개발 과정에 참여했던 판 후이 2단은 이 수를 보고 "아름 답다" 고 표현했습니다. 기계가 둔 수에서 인간이 미학적 아름다움을 느낀 것입니다.

37수는 단 순히 바둑 한 판의 승패를 가르는 수가 아니었습니다. 그것은 인공지능이 논리의 영역을 넘어 직관과 창조의 영역으로 진입했음을 알리는 신호탄이었습니다. 나중에 분석된 바에 따르면, 알파고는 그 상황에서 인간이 그곳에 둘 확률을 1만분의 1로 계산했습니다.

하지만 승률을 계산했을 때 그 수는 최적의 선택이었습니다. 인간 의 고정관념으로는 0에 수렴 하는 확률이었지만, 기계는 그 좁은 문을 뚫고 정답을 찾

아낸 것입니다. 허사비스에게 있 어 이 37수는 딥마인드의 미션인 "지능을 쫓다"는 것 이 헛된 구호가 아님을 입증하는 가장 강력한 물증이되었습니다. 이세돌의 "신의 한 수" 제78수와 4국 승리의 의미 3연패를 당하며 벼랑 끝에 몰린 이세돌 9단은 제4국에서 인간 지성의 위대한 반격을 보여 주었습니다.

알파고의 철벽 같은 계산에 막혀 승산이 보이지 않던 중반, 이세돌은 중앙 흑 돌 사이를 파고드는 묘 수, 제78수를 두었습니다. '끼움수'라 불리는 이 수는 알파고의 연산 범위 밖에서 날아 온 일격이었습니다. 재미있는 것은 알파고 역시 이세돌의 78수를 예측하지 못했다는 점입니다.

알파고 의 로그 기록에 따르면, 알파고는 그 상황에서 인간이 그곳에 착수할 확률을 0.007% 로 보았습니다. 사실상 불가능한 수라고 판단한 것입니다. 그러나 이세돌은 그 불가능을 현실로 만들었습니다.

예측이 빗나가자 알파고는 급격히 흔들리기 시작했습니다. 이후 알파고가 보여준 수들은 마치 당황한 인간이 허둥지둥하는 것처럼 의미 없는 악수들의 연속이었습니다. 화면에는 'AlphaGo Resigns(알파고 기권)'라는 팝업창이 떴습니다.

이 승리는 단순한 1승이 아니었습니다. 인류가 기계에 완전히 정복당하지 않았음을 보여 주는 상징적인 사건이었습니다. 허사비스는 이 패배를 두고 "알파고의 약점을 발견하게 해 준 이세돌 9단에게 감사 한다" 고 말했습니다.

그에게 있어 이세돌의 승리는 알파고라는 시스템을 더 완벽하게 만들 수 있는 소중한 데이터를 제공해 준 사건이었습니다. 동시에 이는 "인공지능은 완벽하지 않으며, 인간의 창의적 개입이 여전히 중요하다"는 메시지를 전 세계에 던졌습니다. 기계가 인간을 대체하는 것이 아니라, 인간이 기계의 한계를 시험하고 더 높은 곳으로 이끌 수 있다는 공존의 가능성을 보여준 것입니다.

이세돌의 78수는 기계의 차가운 논리에 맞선 인간 직관의 뜨거운 승리였으며, 바둑판 위에서 벌어진 인간과 기계의 대화 중 가장 극적인 장면으로 역사에 기록되었습니다. 100만 달러 상금과 전 세계 2억 명의 시청이 대국은 단순한 게임 이벤트를 넘어 전 지구적 현상이었습니다. 구글은 이 대결에 100만 달러의 상금을 걸었고, 전 세계 주요 방송사와 유튜브를 통해 생중계되었습니다. 바둑이 대중적인 한국, 중국, 일본에서는

월드컵 결승전을 방불케 하는 열기가 이어졌습니다. 누적 시청자 수는 2억 명을 넘어섰고, 바둑을 전혀 모르는 서구권의 시청자들조차 'Man vs Machine(인간 대 기계)'이라는 원초적인 대결 구도에 매료되었습니다. 허사비스는 이 이벤트를 기획할 때부터 이것이 단순한 흥보가 아님을 분명히했습니다.

그는 이것을 아폴로 계획의 달 착륙에 비유했습니다. 달에 가는 것이 인류 과학 기술의 총체적 역량을 증명했듯, 바둑이라는 가장 복잡한 게임을 정복하는 것은 인공지능 기술이 임계점을 넘었음을 증명하는 의식이었습니다. 전 세계 2억 명의 시선은 40년 전 흑백 TV로 달 착륙을 지켜보던 인류의 시선과 닮아 있었습니다.

그들은 모니터를 통해 '지능'이라는 추상적인 개념이 물리적 실체로 구현되는 과정을 목격했습니다. 이세돌 9단은 대국 후 인터뷰에서 "내가 패배한 것이지 인류가 패배한 것은 아니다"라는 명언을 남겼습니다. 그러나 대중들은 직감적으로 시대가 바뀌었음을 느꼈습니다.

100만 달러라는 상금은 상징적 숫자에 불과했습니다. 이 대국이 창출한 경제적 가치와 사회적 파장은 수백억 달러에 달했습니다. 구글의 주가는 치솟았고, 전 세계 정부와 기업들은 AI 예산을 급격히 늘리기 시작했습니다.

허사비스라는 이름은 과학계의 스타를 넘어 시대의 아이콘이 되었습니다. 런던의 작은 사무실에서 시작된 '지능을 풀겠다'는 소년의 꿈이 서울의 호텔 방에서 전 세계로 폭발한 것입니다. 사회적 충격과 'AI 서사'의 재편 공포, 열광, 투자의 물결: AI 인식의 전환점 알파고 쇼크는 전 세계, 특히 동아시아 사회에 '스푸트니크 모멘트(Sputnik Moment)'와 같은 충격을 안겼습니다.

1957년 소련이 인공위성을 쏘아 올렸을 때 미국이 느꼈던 기술적 충격과 위기감이 2016년 서울에서 재현된 것입니다. 그전까지 인공지능은 영화 속의 터미네이터나 아이폰의 시리(Siri) 정도로 인식되었습니 대중에게 AI는 막연한 공상과학이거나 실망스러운 음성 인식 기능에 불과했습니다. 하지만 알파고는 달랐습니다.

바둑이라는, 심오한 영역이라고 믿었던 곳에서 기계가 인간을 압도하는 모습은 실존적인 공포와 경이로움을 동시에 불러일으켰습니다.

한국 서점가에서는 인공지능 관련 서적이 베스트셀러를 점령했고, 코딩 학원에는 초등학생들의 등록 문의가 쇄도했습니다. "내 아이가 의사나 변호사가 되어도 AI에게 대체되지 않을까?"라는 질문이 학부모들 사이에서 진지하게 논의되기 시작했습니다. 기업과 정부의 반응은 즉각적이었습니다.

중국 정부는 알파고의 승리를 보고 충격을 받아 국가 차원의 '차세대 인공지능 발전 계획'을 발표하며 AI 굴기를 천명했습니다. 허사비스의 승리는 잠자던 용을 깨운 격이었습니다. 투자 시장의 흐름도 완전히 바뀌었습니다.

알파고 이전까지 AI 스타트업은 틈새시장의 모험적인 투자처였습니다. 그러나 알파고 이후 AI는 모든 벤처 캐피털의 최우선 관심사가 되었습니다. 딥러닝이라는 용어가 일간지 1면을 장식했고, 엔비디아의 GPU 수요가 폭발했습니다.

오늘날 우리가 목격하는 오픈AI의 챗GPT 열풍과 거대 언어 모델(LLM) 경쟁의 뿌리를 거슬러 올라가면, 그 시작점에는 2016년 서울의 봄이 자리 잡고 있습니다. 허사비스는 의도했던 의도하지 않았든, AI를 연구실의 학문에서 자본주의의 가장 뜨거운 엔진으로 옮겨놓았습니다. 이는 '지능을 풀어 세상을 해결한다'는 그의 미션이 현실 세계의 자원과 권력을 움직이는 단계로 진입했음을 의미했습니다.

다큐멘터리 '알파고(AlphaGo, 2017)'의 제작과 수상 이 역사적 사건의 이면은 그렉 코스(Greg Kohs) 감독의 다큐멘터리 '알파고(AlphaGo)'를 통해 세상에 드러났습니다. 2017년 공개된 이 다큐멘터리는 차가운 기술적 성취 뒤에 숨겨진 뜨거운 인간 드라마를 포착했습니다. 허사비스는 이 다큐멘터리의 제작을 적극적으로 지지했습니다.

그는 대중이 AI를 '적'이 아닌 '도구'이자 '협력자'로 바라보기를 원했습니다. 영화는 딥마인드 팀원들의 인간적인 고뇌를 생생하게 보여줍니다. 이세돌이 4국에서 승리했을 때, 호텔 룸의 딥마인드 팀원들은 실망하기보다 안도감을 느꼈습니다. 그들은 자신들이 만든 창조물이 너무 완벽하고 냉혹한 괴물로 비치는 것을 두려워했습니다.

다큐멘터리 속에서 개발자 아자 황 박사가 알파고를 대신해 돌을 놓으며 무표정한 얼굴 뒤로 긴장감을 숨기는 모습, 패배한 판 후이 2단이 알파고와의 대국을 통해 자신의 바둑이 성장했음을 깨닫고 미소 짓는 장면은 AI 서사의 새로운 가능성을 제시했습니다. 이 다큐멘터리는 트라이베카 영화제 등 유수의 영화제에서 수상하며 작품성을 인정받았습니다. 무엇보다 중요한 것은 이 영화가 '인간 대 기계'라는 대립 구도를 '문제를 해결하는 인간과 그를 돕는 기계'라는 구도로 전환하려는 시도였다는 점입니다.

영화의 마지막 장면에서 허사비스는 알파고가 보여준 창의성이 결국 인간이 만들어낸 알고리즘과 데이터에서 비롯된 것임을 강조합니다. 그는 "알파고는 바둑을 위해 만들 어졌지만, 우리가 정말로 풀고 싶은 것은 암 치료나 기후 변화 같은 문제들"이라고 말합니다. 다큐멘터리 '알파고'는 딥마인드가 세상에 던지는 출사표이자, 다가올 AI 시대에 대한 인문학적 성찰을 요구하는 초대장이었습니다.

허사비스는 이 영화를 통해 기술의 발전이 인간성을 위축시키는 것이 아니라, 인간 지성의 지평을 넓히는 촉매제가 될 수 있음을 역설했습니다. 2016년의 대국이 기술적 특이점이었다면, 2017년의 다큐멘터리는 그 기술을 인간의 언어로 번역해 낸 문화적 특이점이었습니다. 2016년

서울, 알파고 대 이세돌 대국

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

기보 없이 스스로 학습한 알파고 제로 백지 상태에서 3일 만에 알파고를 100대 0으로 제압 2017년 런던 킹스크로스에 위치한 딥마인드 본사, 데이비드 실버의 모니터에는 기이한 정적이 흐르고 있었습니다. 불과 1년 전, 서울에서 이세돌 9단과 세기의 대결을 펼쳤던 알파고 리(AlphaGo Lee)는 수십만 장의 인간 기보를 먹어 치우며 자라난 거대한 지성이었습니다. 그러나 지금 실버와 허사비스가 바라보고 있는 새로운 버전, '알파고 제로'의 시작점은 철저한 '무(無)'였습니다.

그들에게는 인간의 도움이 필요 없었습니다. 아니, 오히려 인간의 지식이 인공지능의 한계를 규정짓는 불순물이 될 수 있다는 과감한 가설을 세웠습니다. 허사비스는 이 프로젝트를 승인하며 연구팀에게 단 하나의 데이터도 입력하지 말 것을 지시했습니다.

알파고 제로에게 주어진 것은 바둑판의 크기와 승리 조건이라는 가장 기초적 인 규칙뿐이었습니다. 태어나서 단 한 번도 바둑을 본 적 없는 어린 아이가 바둑판 앞에 앉은 것과 같았습니다. 첫 번째 바둑돌이 놓였습니다.

제멋대로였습니다. 의미 없는 곳에 돌을 놓았고, 스스로 집을 짓다가 허물기도 했습니다. 초기 단계의 알파고 제로는 바둑 초보자보다도 못한 실력을 보여주었습니다.

무작위로 돌을 놓으며 승패조차 가늠하지 못하는 이 엉성한 신경망이 과연 인류 최강의 기사를 꺾은 선배 모델을 넘어설 수 있을지 의구심이 들 법도했습니다. 학습이 시작된 지 3시간이 지나자 놀라운 변화가 감지되었습니다. 무작위로 돌을 놓던 시스템이 점차 돌을 연결하는 법을 깨닫기 시작했습니다.

인간이 수천 년에 걸쳐 정립한 정석의 기초를 스스로 발견해낸 것입니다. 36시간이 지났을 때, 분위기가 달라졌습니다. 알파고 제로가 이세돌 9단을 4대 1로 꺾었던 그 버전의 실력을 넘어서고 있었기 때문입니다. 인간의 지식을 전혀 배우지

않았음에도 불구하고, 단지 자기 자신과의 싸움만으로 인간계의 정점을 돌파해버린 것입니다. 가장 충격적인 결과는 학습 시작 72시간, 즉 3일 뒤에 나왔습니다. 딥마인드 연구진은 3일간 학습한 알파고 제로와 이세돌 버전을 맞붙였습니다.

결과는 참혹할 정도로 일방적이었습니다. 100대 0. 단 한 판도 내주지 않았습니다. 심지어 알파고 제로는 이세돌 9단과의 대국을 위해 수개월간 학습하고 튜닝되었던 구버전보다 훨씬 적은 컴퓨팅 파워를 사용했습니다.

구버전이 48개의 TPU를 사용해 분산 처리를 했던 반면, 알파고 제로는 단 4개의 TPU만으로 돌아갔습니다. 더 적은 자원으로, 더 짧은 시간에, 인간의 데이터를 하나도 쓰지 않고도 완벽한 승리를 거둔 것입니다. 이 사건은 '지능의 독립 선언'과도 같았습니다.

지금까지의 인공지능은 인간이 만들어 놓은 데이터의 패턴을 모방하는 수준이었습니다. 챗GPT와 같은 언어 모델도 결국 인터넷상의 텍스트를 학습합니다. 하지만 알파고 제로는 인간의 경험 세계 밖으로 나갔습니다.

허사비스는 이를 두고 "우리는 인간 지식의 한계에 갇혀 있었다"고 회고했습니다. 인간의 기보를 학습시킨다는 것은 곧 인간이 저지르는 실수와 편견까지도 학습시킨다는 뜻이었습니다. 인간의

데이터를 버림으로써, 알파고 제로는 인간이 도달하지 못한 진리, 혹은 '바둑의 신'이 두는 수에 더 가까이 다가갈 수 있었습니다.

3일이라는 시간은 상징적이었습니다. 인류가 바둑이라는 게임을 발명하고 3천 년 동안 쌓아 올린 지혜의 탑을, 실리콘 칩 속의 알고리즘은 단 72시간 만에 재건하고 초 월해버렸습니다. 이는 허사비스가 꿈꾸던 AGI로 가는 길에 중요한 이정표가 되었습니다.

인간의 데이터가 부족하거나 아예 없는 영역, 예를 들어 새로운 소재를 발견하거나 난치병의 단백질 구조를 파악하는 영역에서도 AI가 스스로 해법을 찾을 수 있다는 가능성을 증명했기 때문입니다. 100대 0이라는 스코어보드는, 지능의 패러다임이 '모방'에서 '발견'으로 전환되었음을 알리는 신호탄이었습니다. 자기 대국(Self-play)만으로 초인적 수준 달성의 의미 알파고 제로의 성취를 설명하는 핵심 키워드는 '자기 대국(Self-play)'입니다.

외로운 수행자가 거울 속에 비친 자신과 끊임없이 검을 맞추는 과정에 비유할 수 있습니다. 기존의 머신 러닝 방식인

지도 학습(Supervised Learning)은 선생님이 학생에게 정답을 알려주는 방식이었습니다. "이 상황에서는 여기에 두는 것이 좋아"라고 인간 고수의 기보가 가르쳐주는 셈입니다. 하지만 자기 대국은 다릅니다. 선생님도 없고 교과서도 없습니다.

오직 승리라는 목표만 존재할 뿐입니다. 강화학습(Reinforcement Learning)의 정수를 보여준 이 과정은 잔혹하리만큼 단순합니다. A 알파고가 검은 돌을 쥐고, 복제된 B 알파고가 흰 돌을 칩니다. 둘은 치열하게 싸웁니다. 승패가 결정되면 이긴 쪽의 전략은 강화되고, 진 쪽의 전략은 수정됩니다.

조금 더 똑똑해진 알파고는 다시 자신과 대국을 벌입니다. 이 과정이 수백만 번, 수천만 번 반복됩니다. 어제의 나를 이긴 오늘의 나, 그리고 그 나를 이길 내일의 나가 끊임없이 경쟁하며 나선형으로 실력을 끌어올립니다. 이 방식이 갖는 철학적 의미는 심장합니다.

허사비스는 이를 "타블라 라사(Tabula Rasa)", 즉 빈 서판에서의 학습이라고 불렀습니다. 존 로크가 인간의 마음은 태어날 때 백지와 같다고 주장했듯, 알파고 제로도 백지 상태에서 출발했습니다. 이는 AI가 인간의 관습이나 고정관념에 얽매이지 않고 독창적인 전략을 수립할 수 있게했습니다.

알파고 제로는 학습 과정에서 인간이 정석이라고 믿었던 수들을 발견하기도 했지만, 일정 수준을 넘어서자 그 정석들을 폐기하고 전혀 새로운 수들을 두기 시작했습니다. 인간이 수천 년간 "그건 악수(惡手)야"라고 규정했던 수들이 사실은 더 높은 차원의 묘수였음을 증명해 낸 것입니다. 기술적으로 볼 때, 알파고 제로는 두 개의 신경망을 하나로 합치는 혁신을 이루었습니다.

이전 버전은 다음 수를 예측하는 '정책망(Policy Network)'과 현재 판세를 읽는 '가치망 (Value Network)'을 따로 훈련시켰습니다. 하지만 알파고 제로는 이 둘을 하나의 거대한 신경망으로 통합했습니다. 이는 마치 직관과 수읽기가 분리되지 않고 동시에 일어나는 인간 고수의 뇌 작동 방식과 유사했습니다.

한 번의 사고 과정으로 어디에 돌지와 내가 얼마나 유리한지를 동시에 판단함으로써 효율성을 극대화했습니다.

몬테카를로 트리 탐색(MCTS)이라는 탐색 알고리즘도 신경망과 더 긴밀하게 결합되었습니다. 이전에는 무작위 시뮬레이션을 많이 돌려보는 것이 중요했다면, 알파고 제로는 신경망의 직관이 워낙 뛰어나기 때문에 굳이 끝까지 두어보지 않아도 승패를 예측할 수 있었습니다. 이는 인간 기사가 "딱 보니 여기가 급소네"라고 느끼는 감각을 수학적으로 구현한 것 이나 다름없었습니다.

자기 대국을 통한 초인적 수준의 달성은 단순히 바둑을 잘 두는 기계가 나왔다는 것 이상 의 충격을 과학계에 던졌습니다. 그것은 '데이터의 종말'을 예고하는 것처럼 보였습니다. 빅데이터 시대에 데이터는 원유와 같다고했습니다.

구글이나 페이스북 같은 기업들이 데 이터를 독점하는 것이 권력이라 믿었습니다. 하지만 알파고 제로는 데이터가 없는 곳에서 도 지능이 폭발할 수 있음을 보여주었습니다. 규칙만 명확하다면, AI는 스스로 데이터를 생성하고 그 데이터로 자신을 학습시켜 인간을 초월할 수 있습니다.

궁극적으로 해결하고자 하는 과학 문제들, 예를 들어 신약 개발이나 핵융합 제어와 같은 분야에 희망적인 메시지였습니다. 이런 분야는 양질의 데이터가 부족하거나 실험 비용이 매우 비쌉니다. AI가 가상 환경에서 자기 시뮬레이션만으로 최적의 분자 구조나 플라즈마 제어법을 찾을 수 있다면, 과학의 발전 속도는 기하급수적으로 빨라질 것이기 때문입니다. 결국 알파고 제로의 자기 대국은 바둑판 위에서 벌어진 일이었지만, 그 울림은 실험실과 연구소를 넘어 인류 지성의 확장 가능성을 타진하는 거대한 실험이었습니다.

알파제로 하나의 시스템으로 여러 영역을 마스터 2017년 말, 딥마인드 연구실의 분위기는 또다시 고요한 흥분으로 차오르고 있었습니다. 알 파고 제로가 바둑을 정복 한 직후, 허사비스와 데이비드 실버는 근본적인 질문을 던졌습니 "이 알고리즘이 바둑에만 특화된 것은 아닐까?" 만약 그렇다면 그건 진정한 의미의 지 능이 아니었습니다.

진정한 지능은 범용성(Generality)을 가져야합니다. 하나의 원리로 여 러 문제를 해 결할 수 있어야합니다. 그들은 알파고 제로의 알고리즘에서 '바둑'이라는 특 수성을 걷어내기 시작했습니다.

바둑판의 회전 대칭성이나 바둑 특유의 규칙들을 코드에서 지워버렸습니다. 그리고 빈자리에 더 추상적이고 일반적인 학습 구조를 채워 넣었습 니 그렇게 탄생한 것이 바로 '알파제로(AlphaZero)'였습니다.

이번 도전 과제는 보드게임의 삼대장이라 불리는 바둑, 체스, 그리고 일본 장기인 쇼기였습니다. 이 세 게임은 성격이 완전히 다릅니다. 바둑은 돌을 놓는 게임이고, 체 스와 쇼기는 기물을 움직이는 게임입니다. 쇼기는 잡은 말을 다시 사용할 수 있어 경 우의 수가 체스보 다 복잡합니다.

과거의 AI들은 각 게임에 맞는 별도의 튜닝과 알고리즘이 필요했습니다. 체스 엔진 에는 체 스 특화 지식이, 바둑 프로그램에는 바둑의 집 계산법이 들어가야했습니다. 하지만 알파제로는 단 하나의 알고리즘, 단 하나의 신경망 구조로 이 세 가지를 모두 공략하겠다고 선 언했습니다.

실험이 시작되었습니다. 결과는 경이로움을 넘어 공포스럽기까지했습니다. 알파제 로는 백지 상태에서 학습을 시작한 지 단 4시간 만에, 당시 세계 최고의 체스 엔진이 었던 '스톡피 시(Stock sh)'를 넘어섰습니다.

스톡피시는 수십 년간 인간 개발자들이 체스의 모든 지식을 코드로 정교하게 다듬어 만든, 말 그대로 '체스 기계'였습니다. 초당 7천만 개의 수를 계산하는 스톡피시를 상대로, 알파제로는 초당 8만 개의 수 밖에 계산하지 않았음에도 승리했습니다. 무식한 계산력(Brute-force)이 아니라, 직관과 전략적 판단력으로 압도한 것입니다.

쇼기에서도 마찬가지였습니다. 세계 챔피언 프로그램 '엘모 (Elmo)'를 2시간 만에 제압했습니다. 바둑에서는 전작인 알파고 제로를 8시간 만에 추월했습니다.

24시간도 채 되지 않아, 인류가 만든 가장 복잡한 보드게임 세 가지가 하나의 알고리즘에 의해 정복되었습니다. 이것은 인공지능 역사상 가장 중요한 순간 중 하나로 기록될 만했습니다. 특정 문제에 맞춰 인간이 수작업으로 특징을 설계(Feature Engineering)해주지 않아도, AI가 범용적인 학습 능력만으로 해당 도메인의 최고수가 될 수 있음을 증명했기 때문입니다.

허 서비스는 이를 두고 "하나의 알고리즘으로 서로 다른 우주를 여행하는 것과 같다"고 표현했습니다. 바둑이라는 우주, 체스라는 우주, 쇼기라는 우주에서 알파제로는 각각의 물리 법칙을 스스로 깨닫고 그 세계의 지배자가 되었습니다. 전설적인 체스 챔피언 가리 카스파로프는 알파제로의 기보를 보고 감탄을 금치 못했습니다 "우리는 기계가 두는 체스는 계산적이고 딱딱할 것이라고 생각했다.

하지만 알파제로는 마치 외계인이 두는 것처럼 창의적이고, 때로는 낭만적이기까지 하다." 알

파제로는 기물의 점수를 중시하는 기존 엔진들과 달리, 장기적인 위치적 우위를 위해 기물을 과감히 희생하는 플레이를 보여주었습니다. 이는 인간의 직관과 기계의 정밀함이 결합된, 혹은 그 둘을 모두 초월한 새로운 형태의 지능이었습니다. 알파제로는 단순히 게임을 이긴 것이 아니라, 게임을 하는 '방법'을 새로 썼습니다.

AGI를 위한 검증대(Test-bed)였다. 세상은 알파고와 알파제로의 승리에 환호했지만, 데미스 허서비스의 시선은 바둑판 너머를 향하고 있었습니다. 언론 인터뷰나 강연에서 그는 지켜올 정도로 반복해서 말했습니다.

"우리의 목표는 게임을 잘하는 AI를 만드는 것이 아닙니다. 게임은 단지 검증대 (Test-bed) 일 뿐입니다." 이 말은 딥마인드의 정체성을 가장 잘 보여주는 문장입니다. 왜 하필 게임이었을까요? 허서비스에게 게임은 현실 세계의 축소판이자, 가장 안전하고 효율적인 실험실이었습니다.

현실 세계는 너무나 복잡하고 노이즈가 많습니다. 결과를 측정하기 어렵고, 실험을 무한정 반복할 수도 없습니다. 로봇을 학습시키기 위해 현실에서 넘어지게 하면 로봇이 부서집니다. 기후 모델을 테스트하기 위해 지구를 망가뜨릴 수는 없습니다.

게임은 규칙이 명확하고, 승패라는 목표가 뚜렷하며, 무엇보다 가상 공간에서 무한히 빠르게 시뮬레이션할 수 있습니다. 허서비스는 어린 시절 체스 챔피언이자 비디오 게임 개발자였던 자신의 경험을 통해, 게임이 지능을 측정하고 훈련하기에 최적의 도구임을 본능적으로 알고 있었습니다. 알파제로는 이 '검증대' 이론을 완벽하게 입증했습니다.

바둑, 체스, 쇼기는 서로 다른 규칙을 가진 복잡한 시스템입니다. 만약 하나의 알고리즘이 이상적인 시스템들을 모두 해석하고 최적의 해법을 찾을 수 있다면, 그 알고리즘은 게임이 아닌 다른 복잡한 시스템에도 적용될 수 있을 것입니다. 이것이 AGI로 가는 핵심 열쇠입니다.

허사비스는 알파제로가 보여 준 '범용 학습 능력'이 단백질 접힘(Protein Folding)이나 신 소재 발견과 같은 과학적 난제에 도 그대로 적용될 수 있다고 확신했습니다. 알파제로 프로젝트가 성공적으로 마무리된 후, 딥마인드의 핵심 인력들은 대거 과학 프로젝트로 이동했습니다.

바둑판의 패턴을 읽던 신경망은 이제 아미노산의 서열을 읽기 시작했습니다. 체스 말의 움직임을 예측하던 탐색 알고리즘은 단백질이 3차원 공간에서 어떻게 접힐지를 시뮬레이션 하는 데 사용되었습니다. 알파고와 알파제로가 없었다면, 노벨상을 안겨준 '알파폴드'의 탄생은 불가능했을 것입니다. 게임에서 축적된 강화학습과 몬테카를로 트리 탐색 기술, 그리고 무엇보다 "인간의 지식 없이도 가능하다"는 자신감이 과학 분야로 전이된 것입니다.

많은 사람들이 알파고를 '바둑 두는 기계'로 기억할 때, 허사비스는 그것을 '범용 문제 해결 기계'의 프로토타입으로 보았습니다. 그는 딥마인드의 미션을 "지능을 푼다 (Solve Intelligence)"라고 정의했고, 알파제로는 그 미션의 첫 번째 단계인 '범용 알고리즘'이 가능하다는 것을 수학적으로, 그리고 경험적으로 증명해 보였습니다. 게임은 끝났습니다.

하지만 그 게임을 통해 단련된 지능은 이제 현실 세계의 진짜 문제들과 싸우기 위해 실험실 밖으로 걸어 나가고 있었습니다. 뮤제로(MuZero): 규칙조차 스스로 학습하는 AI 환경 모델을 스스로 구축하는 계획 기반 학습 알파고 제로와 알파제로는 놀라운 성취를 이루었지만, 여전히 치명적인 한계가 있었습니다. 바로 '규칙을 알고 있다'는 전제였습니다. 바둑, 체스, 쇼기는 게임의 룰이 완벽하게 정의되어 있습니다.

돌을 어디에 놓을 수 있고, 말은 어떻게 움직이는 지 AI에게 미리 알려주었습니다. 이를 '완전 정보 게임(Perfect Information Game)'이라고 합니다. 하지만 허사비스가 바라보는 현실 세계는 다릅니다. 우리는 인생을 살면서 규칙 설명서를 받지 못합니다.

주식 시장의 변동, 날씨의 변화, 인간관계의 역학 등은 명확한 룰북이 없습니다. 룰을 모르는 상태에서도 학습하고 계획을 세울 수 있어야 진정한 AGI입니다. 2019년, 딥마인드는 이 난제를 해결하기 위해 '뮤제로(MuZero)'를 세상에 내놓았습니다.

뮤제로의 가장 큰 특징은 게임의 규칙조차 알려주지 않는다는 것입니다. 바둑판을 보여주지만, 돌을 어떻게 놓아야 하는지, 어떻게 하면 이기는지, 심지어 이게 바둑인지 체스인지도 알려주지 않습니다. 뮤제로는 스크린에 비치는 픽셀(Pixel) 정보 만을 보고 스스로 세상의 작동 원리를 깨우쳐야 했습니다.

이것이 가능했던 이유는 뮤제로가 '모델 기반 강화학습(Model-Based Reinforcement Learning)'의 새 지평을 열었기 때문입니다. 기존의 AI는 눈앞의 화면을 보고 바로 행동을 결정했습니다. 하지만 뮤제로는 인간처럼 '상상'을 합니다.

인간이 낯선 곳에 갔을 때, 모든 물리 법칙을 계산하지 않습니다. 대신 "내가 이쪽으로 가면 길이나오겠지?", "비가 오면 땅이 미끄러울지?"라고 머릿속으로 간소화된 모델을 돌려봅니다. 뮤제로도 마찬가지입니다. 환경의 모든 복잡한 정보를 다 처리하는 대신, 의사결정에 필요한 핵심 요소만을 추상화하여 자신만의 '내부 모델(Internal Model)'을 만듭니다. 이 내부 모델 안에서 뮤제로는 세 가지를 스스로 묻고 답합니다.

첫째, "현재 상태는 어떠한가?"(State). 둘째, "내가 어떤 행동을 하면 상태가 어떻게 변할까?"(Dynamics). 셋째, "그 상태는 나에게 얼마나 좋은가?"(Value). 놀라운 점은 뮤제로가

학습한 이 내부 모델이 실제 게임의 규칙과 정확히 일치할 필요가 없다는 것입니다. 예를 들어, 아타리(Atari) 게임을 할 때 배경의 구름이 어떻게 움직이는지는 점수와 상관이 없습니다. 무제로는 점수를 얻는 데 중요한 '캐릭터의 위치'나 '장애물' 같은 정보에만 집중하고 나머지는 무시합니다.

인간이 세상을 인지할 때 불필요한 정보를 걸러내는 것과 똑같습니다. 이렇게 스스로 규칙을 학습한 무제로는 바둑, 체스, 소기뿐만 아니라 57종의 아타리 비디오 게임에서도 초인적인 성능을 발휘했습니다. 특히 전략적 계획이 필요한 보드게임과 시각적 반응이 중요한 비디오 게임을 하나의 알고리즘으로 모두 정복했다는 것은 충격적이었습니다.

이전까지는 '계획(Planning)'을 잘하는 알고리즘(알파고 계열)과 '반응(Reacting)'을 잘하는 알고리즘(DQN 계열)이 나뉘어 있었는데, 무제로는 이 둘을 통합해낸 것입니다. 규칙을 모르는 혼돈 속에서도 질서를 찾아내고 미래를 시뮬레이션하는 능력, 이 것이야말로 야생의 환경에서 생존해야 하는 지능의 본질에 가장 근접한 형태였습니다.

Nature 논문 발표와 범용 학습의 새 지평 2020년 12월, 세계적인 과학 저널 《네이처(Nature)》에 무제로에 관한 논문이 게재되었습니다. 제목은 《모델 없는 계획의 학습을 통한 범용적 알고리즘(Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model)》. 이 논문은 인공지능 연구자들 사이에서 "강화 학습의 성배 (Holy Grail)를 찾았다"는 평가를 받았습니다. 이 논문이 갖는 의미는 AI가 '현실 세계(Real World)'에 적용될 수 있는 가장 강력한 도구가 되었음을 선언한 것입니다.

허사비스는 항상 "현실은 게임보다 훨씬 지저분 하다(messy)"고 말해왔습니다. 유튜브 비디오의 압축 알고리즘을 최적화하는 문제나, 자율주행차가 복잡 잡힌 교차로를 통과하는 문제는 바둑처럼 명확한 규칙이 없습니다. 시시각각 변하는 환경 속에서 규칙을 유추하고, 그 안에서 최선의 경로를 계획해야 합니다.

무제로는 바로 이 지점에서 빛을 발했습니다. 구글은 무제로의 알고리즘을 유튜브의 비디오 전송 시스템에 적용했습니다. 네트워크 상황이 불안정한 현실 인터넷 환경에서, 무제로는 룰을 배우지 않고도 스스로 트래픽을 최적화하여 비디오 화질을 개선하고 버퍼링을 줄이는 데 성공했습니다.

게임을 위해 개발된 AI가 실제 산업 현장의 불확실성을 통제하는 데 쓰인 첫 번째 메이저 사례 중 하나였습니다. 무제로는 '표현 학습(Representation Learning)'의 진화를 보여주었습니다. 인간이 사과를 볼 때 사과의 분자 구조를 보지 않고 '빨갛고 둥근 과일'이라는 개념으로 인식 하듯이, 무제로도 입력된 데이터(픽셀)를 그대로 쓰지 않고 의사결정에 유용한 '잠재 상태(Hidden State)'로 변환해 저장했습니다.

이는 AI가 인간처럼 추상적 사고를 할 수 있는 발판을 마련했다는 평가를 받았습니다. 허사비스에게 무제로는 알파고 시리즈의 완결판이자, 새로운 시작이었습니다. 알파고가 '직관'을 보여줬고, 알파고 제로가 '독창성'을 보여줬고, 알파제로는 '범용성'을 보여줬다면, 무제로는 '적응성'을 보여주었습니다.

어떤 낯선 환경에 던져봐도 스스로 규칙을 학습하고 살아남아 목표를 달성하는 AI. 이것은 딥마인드가 추구해 온 "지능을 쫓다"는 미션에서 '지능의 핵심 메커니즘'을 규명해 낸

순간이었습니다. 뮤제로의 성공은 딥마인드의 행보에 큰 자신감을 불어넣었습니다. 이제 단백질 구조 예측, 기후 변화 예측, 핵융합 제어, 그리고 수학적 난제 해결까지. 규칙을 알 수 없

는 미지의 영역, 즉 '자연(Nature)'이라는 거대한 게임판으로 나아갈 준비를 마친 것입니다. 뮤제로는 게임 오버(Game Over)가 아니라, 사이언스 스타트(Science Start)를 알리는 신호탄이었습니다. 그리고 그 길의 끝에는 2024년 노벨상이라는 영광이 기다리고 있었습니다. AI와 체스, 알파제로의 세계

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

1994년 CASP 대회의 시작 1972년 12월, 스톡홀름의 겨울은 유난히 차가웠지만 노벨상 시상식장의 열기는 뜨거웠습니다. 단상에 오른 미국의 생화학자 크리스천 앤핀센 은 인류 과학사에 남을 대담한 가설을 세상에 공표했습니다. 그는 단백질이라는 생명 의 미세한 기계가 어떻게 만들어지는지 에 대해 아주 단순하면서도 본질적인 원리를 제시했습니다.

"단백질의 3차원 입체 구조는 전적으로 아미노산 서열에 의해 결정된 다." 이것이 바로 훗날 '앤핀센의 도그마(An nsen's Dogma)'로 불리게 되는 명제입니다. 이 선언은 생물학계에 거대한 속제를 던져주었습니다.

만약 앤핀센의 말이 맞다면, 우리는 이론적으로 1차원적인 아미노산 서열 정보만 알면 그것이 꼬이고 접혀서 만들 어낼 3차원 모양을 완벽하게 예측할 수 있어야했습니다. 하지만 자연은 그렇게 호락 호락하지 않았습니 다. 아미노산 사슬이 접힐 수 있는 경우의 수는 우주에 존재하는 모든 원자의 수보다 많 았기 때문입니 다.

사이러스 레빈탈(Cyrus Levinthal)이라는 과 학자는 이를 계산해보니 단백질 하나가 가능한 모든 구조를 무작위로 시도해 보는 데 만 우주의 나이보다 더 긴 시간이 걸린다는 '레빈탈의 역설(Levinthal's Paradox)'을 제기하기도했습니다. 자연은 1000분의 1 초 만에 이 과정을 끝내는데, 인간의 슈퍼컴 퓨터는 수십 년을 계산해도 정답을 찾지 못하 는 상황이 벌어진 것입니다. 이 난제를 해결하기 위해 과학자들은 1994년, 매우 독특한 실험을 시작했습니다.

바로 '단 백질 구조 예측 학술 대회(CASP, Critical Assessment of Structure Prediction)'입니 다. 메릴랜드 대학교의 존 몰트(John Moult) 교수가 창설한 이 대회는 일종의 '블라인드 테스트'였습니 다. 실험실에서 이미 구조를 밝혀냈지만 아직 대중에 게 공개하지 않은 단백질 문제를 출제 하면, 전 세계의 컴퓨터 과학자와 생물학자들이 각자의 알고리즘으로 그 구조를 예측해 답 안지를 제출하는 방식이었습니다. 2년에 한 번씩 열리는 이 대회는 생물학계의 올림픽과 도 같았습니다.

대회가 거듭될수록 과 학자들의 표정은 어두워져 갔습니 다. 1994년부터 2016년까지 약 20년 동안, 예측 정

확도는 지지부진했습니다. 100점 만점에 40점 언저리를 맴도는 성적표는 인류 지성 의 한계를 보여주는 듯했습니다. 과학자들은 서서히 지쳐가고 있었습니 다.

바로 그때, 런던의 한 연구소에서 이 문제를 전혀 다른 시각으로 바라보는 남 자가 등장했습니다. 단백질이 '루트 노드(Root Node)' 문제인 이유: 생명과학의 병목 데미스 허사비스에게 세상은 하나의 거대한 정보 처리 시스템이었습니다. 그는 딥마인드의 미션을 "지능을 푼다(Solve Intelligence)"로 정했을 때, 단순히 바 독을 잘 두거나 채팅을 잘하는 AI를 꿈꾼 것이 아니었습니다.

그는 과학적 발견의 속 도를 획기적으 로 높일 수 있는 메타 솔루션(Meta-solution)을 원했습니다. 허사비스는 과학의 수많은 난제들을 나뭇가지에 비유했습니 다. 기후 변화, 질병 치 료, 신 소재 개발 같은 문제들은 나뭇가지 끝에 매달린 잎사귀와 열매들입니 다.

이 개 별적인 문제 들을 하나하나 해결하려면 엄청난 시간과 비용이 듭니 다. 허사비스는 생각했습니다. "가지를 하나씩 치는 대신, 이 모든 문제들이 뻗어 나오 는 뿌리, 즉 루트 노드(Root Node)를 찾아 해결하면 어떨까?" 그가 보기에 생물학의 루트 노드는 바 로 '단백질 접힘(Protein Folding)'이었습니다.

우리 몸의 모든 기능은 단백질에 의해 작동합니다. 눈으로 빛을 감지하는 것도, 근육을 움직이는 것도, 바이러스와 싸우는 항체도 모두 단백질입니다. 단백질의 기능은 그 '모양'에서 나옵니다. 열쇠가 자물쇠 구멍에 딱 맞아야 문이 열리듯, 단백질도 특정한 3차원 구조를 가져야만 질병이라는 자물쇠를 풀 수 있습니다.

신약 개발이 그토록 어렵고 비싼 이유는, 약물이 작용해야 할 표적 단백질의 모양을 정확히 모르기 때문에 수만 번의 시행착오를 거쳐야 했기 때문입니다. 허사비스는 확신했습니다. 만약 AI가 단백질 구조를 예측할 수 있다면, 그것은 생물학이라는 학문을 순수 실험과 학에서 데이터 과학으로 전환시키는 계기가 될 것이라고 말합니다. 이것은 딥마인드가 게임 회사가 아니라 과학 탐구 집단임을 증명할 수 있는 완벽한 과제였습니다.

게임에서 과학으로의 전환 게임 '폴딧(Foldit)'에서 얻은 영감 허사비스가 단백질 문제에 확신을 갖게 된 결정적인 계기는 의외로 '게임'에서 왔습니다.

2008년 워싱턴 대학교의 데이비드 베이커(David Baker) 교수팀이 개발한 '폴딧(Foldit)'이라는 게임이 있었습니다. 이 게임은 복잡한 단백질 구조를 3차원 퍼즐처럼 화면에 띄워놓고, 일반 게이머들이 마우스를 이용해 이리저리 비틀고 접으며 가장 안정적 구조를 만드는 방식이었습니다. 놀라운 일은 그다음 벌어졌습니다.

과학자들이 10년 넘게 슈퍼컴퓨터를 돌려도 풀지 못했던 에이즈 관련 효소의 구조를, 폴딧을 플레이하던 게이머들이 불과 3주 만에 풀어낸 것입니다. 이들은 생물학 박사 학위가 없는 평범한 사람들이었습니다. 그저 3차원 공간에서 사물들의 패턴을 인식하고 직관적으로 퍼즐을 맞추는 능력이 뛰어난 사람들이었을 뿐입니다. 이 뉴스를 접한 허사비스의 뇌리에 번개 같은 영감이 스쳤습니다.

"인간 게이머들이 직관을 이용해 거대한 탐색 공간(Search Space)을 좁혀나간다면, 알파고(AlphaGo)가 했던 일이 바로 그것 아닌가?" 바둑 역시 우주의 원자 수보다 많은 경우의 수를 가지고 있지만, 알파고는 모든 수를 계산하는 대신 '가치망'과 '정책망'이라는 직관을 통해 승리로 가는 길을 찾아냈습니다. 허사비스에게 단백질 접힘은 더 이상 생물학의 문제가 아니었습니다.

그것은 '공간 최적화 게임'이자 '패턴 인식 문제'였습니다. 생물학적 지식이 없어도, 데이터와 보상(Reward)만 있다면 AI가 인간 게이머보다 훨씬 더 잘할 수 있는 영역이었습니다. 게임 개발자 출신인 그에게 과학의 난제는 공략해야 할 '보스 몬스터'로 보이기 시작했습니다.

알파폴드(AlphaFold) 프로젝트의 시작: "이것은 등대 프로젝트" 2016년, 알파고가 이세돌 9단을 꺾은 직후, 딥마인드 내부에서는 "다음은 무엇인가?"에 대한 치열한 논의가 벌어졌습니다. 스타크래프트 같은 더 복잡한 게임을 정복해야 한다는 의견도 있었고, 로봇 공학으로 진출해야 한다는 의견도 있었습니다. 하지만 허사비스는 조용히, 그러나 단호하게 과학 팀을 꾸리기 시작했습니다.

그는 이 프로젝트를 '등대 프로젝트(Lighthouse Project)'라고 명명했습니다. 망망대해에서 길을 잃은 배들에게 빛을 비춰주듯, AI가 인류에게 실질적인 도움을 줄 수 있다는 것을 증명하는 지표가 되어야 한다는 뜻이었습니다.

하지만 시작은 순탄치 않았습니다. 딥마인드 핵심 엔지니어들조차 "생물학은 너무 불확실 성이 크다", "데이터가 지저분해서 AI가 학습하기 어렵다"며 회의적인 반응을 보였습니다. 허사비스는 특유의 설득력으로 팀을 독려했습니다.

그는 20대 후반의 젊은 연구원 존 점퍼 (John Jumper)를 비롯한 소수의 정예 멤버를 모아 '알파폴드(AlphaFold)' 팀을 결성했습니다. 목표는 명확했습니다. "바둑판 위에서 거둔 승리를 생명이라는 캔버스 위로 옮겨온 다."

허 사비스는 이 프로젝트가 실패할 경우 딥마인드가 그저 '게임 잘하는 기계'를 만 드는 회사 로 기억될 것임을, 반대로 성공할 경우 노벨상급의 발견이 될 것임을 직감 하고 있었습니 과학자로서의 허사비스와 경영자로서의 허사비스가 동시에 던진 가장 큰 승부수였습니다. 단백질의 3차원 접힘 구조

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

어 있던 늦가을의 어느 날이었습니다. 과학계의 시선은 줌(Zoom) 화면 속으로 쏠려 있었습니다. 단백질 구조 예측 학술 대회인 CASP(Critical Assessment of Structure Prediction)의 제14회 결과 발표 현장이었습니다.

이 대회는 1994년부터 시작된 생물 학계의 올림픽과도 같은 행사로, 실험으로 밝혀지지 않은 단백질의 3차원 구조를 컴퓨터로 얼마나 정확하게 맞추는지를 겨루는 전장입니다. 주최 측인 존 몰트(John Moulton) 교수가 화면에 그래프 하나를 띄웠습니다. 그 그래프에는 지난 수십 년간 과학자들이 조금씩 쌓아 올린 완만한 성장의 기울기가 그려져 있었고, 그 끝에는 갑작스럽게 수직으로 치솟은 하나의 점이 찍혀 있었습니다.

그 점은 '알파폴드 2(AlphaFold 2)'였습니다. 몰트 교수는 떨리는 목소리로 선언했습니다. "이것은 50년 된 난제가 해결되었음을 의미합니다."

이 순간은 단순히 컴퓨터 프로그램이 인간을 이긴 사건이 아니었습니다. 알파고가 이세돌 9단을 이겼을 때 세상은 놀라움과 두려움을 동시에 느꼈지만, 알파폴드가 생물학의 난제를 풀었을 때 과학자들은 경이로움과 안도감을 느꼈습니다. 데미스 허사비스에게 이 순간은 알파고 대국보다 더 중요한, 그의 평생의 미션이 증명되는 찰나였습니다.

게임을 정복한 지능이 마침내 현실 세계의 가장 복잡한 문제인 생명의 비밀을 풀기 시작한 것입니다. 서막은 2년 전인 2018년 CASP13 대회로 거슬러 올라갑니다. 당시 딥마인드는 '알파폴드1'을 들고 처음으로 이 무대에 등장했습니다. 당시만 해도 생물학계는 구글의 AI 연구소라는 곳에서 온 이방인들을 반신반의하는 눈빛으로 바라봤습니다.

생물학 적 지식이 부족한 컴퓨터 공학자들이 단백질의 복잡성을 이해할 리 없다고 생각했기 때문입니다. 하지만 알파폴드1은 43개 문제 중 25개에서 최고 점수를 기록하며 우승을 차지했습니다. 2위 팀이 단 3개 문제에서만 최고 점수를 받은 것과 비교하면 압도적인 차이였습니다.

알파폴드1은 단백질의 거리를 이미지의 픽셀처럼 처리하는 방식을 사용했습니다. 마치 컴퓨터 비전이 고양이 사진을 인식하듯, 아미노산 서열 사이의 거리 지도를 이미지처럼 학습하여 구조를 유추해 낸 것입니다. 하지만 허사비스는 만족하지 않았습니다.

우승은 했지만, 알파폴드1의 정확도는 실험실에서 사용할 수 있는 수준에는 미치지 못했기 때문입니다. GDT(Global Distance Test) 점수로 환산하면 약 60점대 수준으로, 대략적인 모양은 맞추지만 세부적인 원자의 위치는 틀리는 경우가 많았습니다. 허사비스는 팀을 모아놓고 말했습니다.

"우리는 1등을 하려고 이 프로젝트를 시작한 것이 아닙니다. 문제를 해결하려고 시작한 것입니다." 그는 알파폴드1의 코드를 완전히 폐기하고 바닥부터 다시 시작하라는 과감한 결단을 내렸습니다. 이때 프로젝트의 운명을 바꾼 인물이 등장합니다.

바로 존 점퍼(John Jumper)입니다. 시카고 대학에서 화학과 컴퓨터 공학을 전공한 그는 단백질 시뮬레이션에 깊은 조예가 있는 연구자였습니다. 점퍼는 딥마인드에 합류한 후 알파폴드 팀을 이끌며 허사비스의 비전을 기술적으로 구현하는 사령관 역할을 맡았습니다.

허사비스가 "무엇을(What)" 풀어야 할지를 정의했다면, 점퍼는 "어떻게(How)" 풀지를 설계했습니다. 두 사람은 단순한 상사와 부하 관계를 넘어 과학적 동반자로서 신뢰를 쌓아갔습니다. 점퍼는 알파폴드1이 사용했던 이미지 인식 방식인 CNN(합성곱 신경망)의 한계를 지적했습니다.

단백질은 고정된 이미지가 아니라 서로 상호작용하며 끊임없이 움직이는 3차원 그래프 구조라는 것이 그의 통찰이었습니다. 2019년, 딥마인드 연구실의 불은 꺼지지 않았습니다. 그들은 '어텐션(Attention)' 메커니즘이라는 새로운 AI 기법을 도입했습니다.

이는 주로 언어 번역에 쓰이던 기술이었습니다. 점퍼와 그의 팀은 아미노산 서열을 하나의 문장처럼 취급했습니다. 문장 속 단어들 이 서로 문맥에 따라 관계를 맺듯이, 멀리 떨어진 아미노산들이 서로 어떻게 당기고 밀어내는지를 AI가 스스로 학습하게 만든 것입니다.

이것은 생물학적 직관과 AI 엔지니어링이 완벽하게 결합된 순간이었습니다. 그들은 이를 '에보포머(Evoformer)'라고 명명했습니다. 허사비스는 이 과정을 지켜보며 "우리는 AI에게 생물학 교과서를 읽어주는 것이 아니라, AI가 스스로 생물학의 원리를 깨우치도록 하고 있다"고 표현했습니다.

개발 과정은 순탄치 않았습니다. CASP14 대회를 몇 달 앞둔 시점까지도 새로운 모델은 불안정했습니다. 팀원들은 지쳤고, 실패에 대한 두려움이 엄습했습니다.

이때 허사비스는 특유의 리더십을 발휘했습니다. 그는 밤늦게 연구실을 찾아가 연구원들과 체스를 두거나 차를 마시며 그들의 멘탈을 관리했습니다. 그는 "우리는 지금 인류가 수십 년간 기다려온 발견의 문턱에서 있다"며 팀의 사기를 북돋았습니다.

마침내 대회를 몇 주 앞두고 알파폴드2는 기하급수적인 성능 향상을 보이기 시작했습니다. 그리고 2020년, CASP14의 결과는 충격적이었습니다. 알파폴드2는 100점 만점에 평균 92.4점이라는 점수를 기록했습니다.

통상적으로 90점 이상이면 실험적 오차 범위 내에서 실제 구조와 동일하다고 간주합니다. 즉, AI가 예측한 구조와 엑스선 결정 학이나 초저온 전자 현미경 같은 수십억 원짜리 장비로 몇 년에 걸쳐 분석한 구조가 겹쳐 놓았을 때 구별할 수 없을 정도로 똑같다는 뜻입니다. 특히 가장 어렵다고 알려진 '자유 모델링' 카테고리에서도 알파폴드2는 압도적인 정확도를 보여주었습니다.

결과가 발표되자 채팅창과 트위터는 전 세계 과학자들의 경악으로 도배되었습니다. 노벨상 수상자이자 구조 생물학의 대가인 벤키 라마크리슈난(Venki Ramakrishnan)은 "이것은 내 생애에 해결될 것이라고 예상하지 못했던 문제였다"고 고백했습니다. 독일 막스 플랑크 연구소의 안드레이 루파스(Andrei Lupas) 교수는 "이제 구조 생물학은 알파폴드 이전과 이후로 나뉠 것"이라며 "이것은 지진과도 같은 변화"라고 평가했습니다.

이 승리는 단순한 기술적 성취가 아니었습니다. 그것은 평생 철학인 "지능을 풀어, 다른 모든 것을 해결한다"는 명제가 참임을 증명하는 첫 번째 강력한 증거였습니다. 게임이라는 가상 세계에서 훈련된 AI가 현실 세계의 물리 법칙과 생물학적 복잡성을 이해하고, 인간 지성의 한계를 확장할 수 있음을 보여준 것입니다.

그는 런던의 자택에서 줌으로 발표를 지켜보며 조용히 미소 지었습니다. 4살 때 체스판을 보며 느꼈던 지적 호기심이, 10대 때 게임을 개발하며 꿈꿨던 시뮬레이션의 가능성이, 그리고 성인

이 되어 신경과학과 AI를 융합하려 했던 모든 노력이 이 하나의 정점으로 수렴되는 순간이었습니다. 알파폴드2의 성공은 2024년 그와 존 점퍼가 노벨 화학상을 공동 수상하게 되는 결정적인 계기가되었습니다. 허사비스는 상보다 더 큰 보상은 이 도구가 가져올 미래의 변화라고 믿고 있었습니다.

이제 인류는 생명의 레고 블록을 조립하는 설명서를 손에 넣게 된 것입니다. 2억개 이상의 단백질 구조 데이터베이스 공개 알파폴드2가 CASP14에서 거둔 압승은 이후, 데미스 허사비스의 야망은 대회 우승 트로피에 머물지 않았습니다. 그는 딥 마인드가 구글에 인수될 때부터 지켰던 철칙, 즉 "가장 강력한 기술은 가장 널리 쓰여야 한다"는 신념을 실행에 옮길 준비를 하고 있었습니다.

기술을 독점하여 제약 회사에 라이선스를 팔거나 막대한 수익을 창출하는 길도 있었습니 실제로 많은 투자자와 기업가들은 그런 모델을 예상했습니다. 하지만 허사비스와 딥마인드 팀은 다른 선택을했습니다. 그들은 인류가 가진 생물학적 지식의 지평을 단숨에 확장하기로 결정했습니다.

2021년 7월, 딥마인드는 유럽분자생물학연구소(EMBL)와 협력하여 '알파폴드 단백질 구조 데이터베이스(AlphaFold Protein Structure Database)'를 전 세계에 무료로 공개했습니다. 처음에는 인간 유전체(게놈)에 포함된 약 2만여 개의 단백질 구조를 공개하는 것으로 시작했습니다. 이것만으로도 생물학 교과서를 다시 써야 할 정도의 엄청난 분량이었습니다.

과거 한 명의 박사 과정 학생이 박사 학위를 받기 위해 4~5년 동안 단백질 하나, 혹은 두 개의 구조를 규명했던 것과 비교하면, AI는 며칠 만에 인간의 모든 단백질 지도를 그려낸 셈이었습니다. 진짜 충격은 1년 뒤인 2022년에 찾아왔습니다. 딥마인드는 데이터베이스를 업데이트하며 지구상에 존재하는 거의 모든 알려진 단백질의 구조를 공개했습니다.

그 수는 무려 2억 개가 넘었습니다. 식물, 박테리아, 동물, 곰팡이 등 과학이 이름을 붙인 거의 모든 생명체의 단백질이 포함되었습니다. 허사비스는 이를 두고 "생물학의 구글 어스(Google Earth)"라고 표

현했습니다. 우리가 구글 어스를 통해 지구 반대편의 골목길까지 들여다볼 수 있게 된 것처럼, 이제 과학자들은 모니터 앞에서 클릭 몇 번으로 생명 현상의 가장 깊은 곳, 분자 수준의 기계를 3차원으로 둘러가며 관찰할 수 있게 된 것입니다. 데이터베이스의 공개는 과학 연구의 속도를 근본적으로 바꾸어 놓았습니다.

어둠 속에서 손전등 하나에 의지해 동굴을 탐험하던 과학자들에게 갑자기 동굴 전체를 비추는 조명 스위치가 켜진 것과 같았습니다. 전 세계 190개국에서 200만 명 이상의 연구자들이 이 데이터베이스에 접속했습니다. 이는 전 세계 생물학자의 대다수에 해당하는 숫자입니다.

실험실의 벤치에서 파이펫을 들고 씨름하던 연구자들은 이제 알파폴드의 예측 결과를 바탕으로 가설을 세우고, 실험을 설계하며, 실패의 확률을 획기적으로 줄일 수 있게되었습니다. 실질적인 응용 사례들은 즉각적이고 폭발적으로 나타났습니다. 가장 극적인 사례 중 하나는 말라리아 백신 연구였습니다.

옥스퍼드 대학의 연구팀은 수년간 말라리아 기생충의 단백질 구조를 밝히기 위해 노력했지만, 기술적 한계로 인해 진전을 보지 못하고 있었습니다. 그들은 알파폴드가 공개되자마자 자신들이 연구하던 단백질의 구조를 검색했고, 단 몇 분 만에 수년간 찾아 헤매던 답을 얻을 수 있었습니다. 이를 통해 그들은 새로운 백신 후보 물질을 설계하는 단계로 즉시 넘어갈 수 있었습니다.

이는 AI가 어떻게 '시간'이라는 가장 소중한 자원을 과학자들에게 선물했는지를 보여주는 상징적인 장면입니다. 플라스틱 쓰레기는 인류의 골칫거리인데, 포츠머스 대학의 과학자들은 플라스틱을 분해하는 효소(PETase)를 연구하고 있었습니다. 그들은 알파폴드를 이용해 이 효소의 진화 과정을 추적하고, 더 빠르고 효율적으로 플라스틱을 분해하도록 효소의 구조를 개량하는데 성공했습니다.

이는 500년이 걸려야 썩는 플라스틱을 며칠 만에 분해 할 수 있는 가능성을 열어준 것으로, 허사비스가 꿈꾸던 "AI를 통해 기후 위기와 환경 문제를 해결한다"는 비전 이 구체화된 사례였습니다. 항생제 내성 문제와 희귀 유전 질환 연구에서도 알파폴드는 필수적인 도구가 되었습니다. 콜로라도 대학의 연구팀은 항생제 저항성 박테리아의 구조를 분석하여 새로운 치료법을 모색했고, 노르웨이의 연구자들은 꿀벌의 면역 시스템을 이해하는 데 이 데이터를 활용했습니다.

심지어 멸종 위기 종을 보존하거나, 기후 변화에 강한 작물을 개발하는 농업 분야 에서도 알파폴드의 데이터는 씨앗이 되어 새로운 발견을 싹틔웠습니다. 허사비스는 이 거대한 데이터의 공개가 가져올 파급력을 '디지털 생물학(Digital Biology)' 의 시작이라고 정의했습니다. 그는 생물학이 이제 정보 과학의 영역으로 들어왔다고 믿었습니다.

생명이란 결국 정보의 복잡한 처리 과정이며, AI는 그 정보를 해석하는 가장 탁월한 도구라는 것입니다. 그는 한 인터뷰에서 이렇게 말했습니다. "우리가 2억 개의 구조를 공개했을 때, 그 것은 인류에게 보내는 선물이었습니다."

더 중요한 것은 이것이 시작에 불과하다는 점입니다. 우리는 이제 단백질의 모양을 아는 것을 넘어, 그것들이 어떻게 상호작용하고, 어떻게 신약이 되고, 어떻게 질병을 고치는지 예측하는 단계로 나아가야 합니다." 이러한 성과는 2024년 노벨 화학상 수상의 가장 강력한 근거가 되었습니다.

노벨 위원회는 알파폴드가 단순히 기술적으로 뛰어난 소프트웨어여서가 아니라, 전 세계 과학자들에게 강력한 연구 도구를 제공함으로써 과학의 진보를 가속화했다는 공로를 높이 샀습니다. 허사비스와 존 점퍼는 이 도구를 사유화하지 않고 공유함으로써, 과학 이 가진 본연의 가치 인 '지식의 공유와 확산'을 실리콘밸리의 방식이 아닌, 가장 고전 적이고 학술적인 방식으로 실현해 냈습니다. "생물학의 구글 어스"는 허사비스가 12세 때 던졌던 질문, "이 지능을 더 나은 곳에 쓸 수 없을까?"에 대한 가장 웅장한 대답이었습니다.

체스판 위의 64칸을 넘어, 우주 만큼이나 광 활한 생명의 미시 세계를 인류의 눈앞에 펼쳐 보였습니다. 이것은 한 천재의 승리가 아니라, 그 천재가 만든 도구를 통해 수백만 명의 과학자가 각자의 자리에서 승리할 수 있도록 만든, 진정한 의미의 '집단 지성의 승리'였습니다.

알파폴드 예측 구조(파란색)와 실험 구조(초록색) 비교

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

2024년 노벨 화학상 수상 2024년 10월 9일, 런던의 아침은 여느 때와 다름없이 런던 특유의 젓빛 하늘로 시작되었습니다. 구글 딥마인드(Google DeepMind)의 킹스 크로스 본사에는 묘한 긴장감이 감 돌고 있었습니다. 스웨덴 왕립과학원(Royal Swedish Academy of Sciences)에서 걸려 온 한 통의 전화가 모든 것을 바꾸어 놓았기 때문입니다.

전화벨이 울렸을 때, 데미스 허사비스는 아내와 함께 평범한 아침을 맞이하고 있었습니다. 수화기 너머로 들려온 소식은 그가 4살 때부터 체스판 위에서, 12살 때부터 컴퓨터 화면 앞에서, 그리고 지난 15년간 딥마인드라는 조직을 통해 끈질기게 추구해 온 '지능의 해결'이라는 미션이 마침내 과학계의 가장 높은 권위로부터 인정받았음을 알리는 신호였습니다. 그해 노벨 화학상의 영예는 세 사람에게 돌아갔습니다.

단백질 설계의 선구자인 워싱턴 대학교의 데이비드 베이커(David Baker) 교수가 절반의 지분을, 그리고 나머지 절반을 데미스 허사비스와 딥마인드의 수석 연구원 존 점퍼(John Jumper)가 공동으로 수상했습니다. 단순한 시상이 아니었습니다. 인공지능(AI)이 더 이상 컴퓨터 공학의 전유물이거나 실리 콘 벨리의 수익 창출 도구에 머무르지 않고, 인류가 자연의 가장 깊은 비밀을 해독하는 '과학적 도구(Scientific Tool)'로 편입되었음을 선포하는 사건이었습니다.

AI가 과학적 발견의 핵심 도구가 됨을 입증한 역사적 사건 허사비스의 수상은 1972년 노벨 화학상 수상자인 크리스천 앤핀센(Christian Anfinsen)이 던진 50년 묵은 난제, 즉 "단백질의 아미노산 서열만으로 그 3차원 입체 구조를 예측할 수 있는가?"라는 질문에 대한 AI의 대답이었습니다. 반세기 동안 수많은 생물학자가 실험실에서 엑스레이 결정학(X-ray crystallography)과 씨름하며 하나의 단백질 구조를 밝히는 데 수년, 때로는 박사 학위 과정을 거쳐야 했습니다. 그러나 허사비스와 점퍼가 이끄는 팀이 개발한 '알파폴드2(AlphaFold2)'는 이 과정을 단 몇 분, 몇 초 단위로 단축해 버렸습니다.

이것은 갈릴레오가 망원경을 하늘로 돌려 목성의 위성을 발견했을 때와 비견될 만한 순간이었습니다. 망원경이 천문학의 시야를 확장했듯이, AI는 생물학의 시야를 근본적으로 확장했습니다. 노벨 위원회가 "이것은 인류에게 가장 큰 혜택을 주는 발견 중 하나"라고 평했을 때, 그들은 단순히 기술의 성능을 칭찬한 것이 아니었습니다.

그들은 AI가 인간 지능의 한계를 보완하고 확장하여, 이전에는 불가능했던 속도와 규모로 과학적 발견을 가속화할 수 있음을 인정한 것입니다. 허사비스가 창립 초기부터 주창해 온 "지능을 이해하고, 그것을 통해 다른 모든 것을 해결 한다는 딥마인드의 미션이 현실화되기 시작한 순간이었습니다. 알파폴드는 2억 개가 넘는 단백질 구조, 사실상 지구상에 알려진 거의 모든 단백질의 구조를 예측하여 공개했고, 전 세계 190개국 200만 명 이상의 연구자들이 이를 통해 말라리아 백신부터 플라스틱 분해 효소 연구까지 과학의 최전선을 넓히고 있었습니다.

"초현실적인 경험": 수상 소감과 과학계·AI계의 반응 수상 소식을 접한 직후 열린 기자회견과 인터뷰에서 허사비스는 평소의 냉철한 분석가적 모습과 달리 다소 상기된 표정이었습니다. 그는 이 순간을 "초현실적인 경험(Surreal experience)"이라고 표현했습니다. "제 머리속이 하얗게 변했습니다(My mind went blank). 믿기지 않는 영광입니다."

그의 이 짧은 고백 속에는 게임 개발자에서 신경과학자로, 다시 창업가로 변신을 거듭하며 걸어온 비주류의 길이 주류 과학계의 정점에서 만나는 벅찬 감정이 녹아 있었습니다. 그는 동료인

존 점퍼와 딥마인드 팀 전체에 공을 돌렸습니다. "이것은 거대한 팀의 노력 (team effort)이었습니다.

우리가 만든 것은 단순한 알고리즘이 아니라, 수백 명의 엔지니어와 과학자가 밤을 지새우며 쌓아 올린 지식의 결정체입니다." 그의 이러한 태도는 전통적인 '고독한 천재 과학자'의 신화가 AI 시대에는 '집단지성과 기계지능의 협업'으로 대체되고 있음을 보여주었습니다. 과학계의 반응은 뜨거웠습니다.

일부 보수적인 화학자들 사이에서는 "이것이 순수한 화학인가, 아니면 컴퓨터 공학인가?"라는 논쟁이 잠시 일기도 했지만, 대다수는 이 수상이 불가피한 시대적 흐름임을 인정했습니다. 생물학자들은 "알파폴드는 우리 분야의 구글 지도 (Google Maps)와 같다"며 찬사를 보냈고, AI 연구자들은 딥러닝 기술이 튜링상(Turing Award)을 넘어 노벨상까지 도달한 것에 대해 "AI의 겨울(Winter)

은 영원히 끝났다"며 환호했습니다. 제프리 힌튼(Geoffrey Hinton)이 하루 전날 노벨 물리학상을 수상한 것과 맞물려, 2024년은 바야흐로 'AI 노벨상의 해'로 기록되었습니다. 노벨상이 열어준 새로운 문: 정책 입안자와 대중에 대한 영향력 확대 노벨상은 허사비스에게 명예 그 이상을 의미했습니다.

그것은 그에게 거대한 '메가폰'을 쥐여준 것과 같았습니다. 그동안 빅테크 기업의 CEO로서 그가 AI의 위험성과 규제를 이야기 할 때, 사람들은 그것을 기업의 이익을 방어하기 위한 전략이나 마케팅으로 해석하려는 경향이 있었습니다. 그러나 '노벨상 수상자'라는 타이틀은 그의 발언에 순수 과학적 권위를 부여했습니다.

수상 이후, 허사비스는 각국 정부의 정책 입안자들과 만나는 자리에서 더 강력한 영향력을 발휘하기 시작했습니다. 그는 AI가 가져올 '근본적 풍요(Radical Abundance)'의 가능성을 설파하는 동시에, 통제 불가능한 초지능의 위험성에 대해 서도 과학적 엄밀함을 가지고 경고했습니다. 대중들에게 그는 더 이상 차가운 기술 기업가가 아니라, 질병을 정복하고 인류의 난제를 해결하려는 '현자(Sage)'의 이미지로 다가갔습니다.

노벨상은 그가 꿈꾸는 AGI가 단순한 상업적 산물이 아니라, 인류 공공의 자산이 되어야 한다는 그의 주장에 힘을 실어주는 가장 강력한 무기가 되었습니다. Time 올해의 인물(2025) CBE(2017)에서 기사 작위까지: 영국 과학계의 아이콘 2024년 3월, 노벨상 수상의 열기가 채식기도 전에 또 다른 영예로운 소식이 들려왔습니다. 영국 왕실은 데미스 허사비스에게 기사 작위(Knight Bachelor)를 수여한다고 발표했습니다. 2017년, 과학과 기술에 대한 공로로 대영제국 훈장(CBE)을 받은 지 7년 만의 승격이었습니다.

버킹엄 궁전에서 찰스 3세 국왕 앞에 무릎을 꿇고 어깨에 검을 받는 의식은, 런던 북부의 이민자 가정에서 태어나 공립학교를 다니며 체스와 코딩으로 자신의 길을 개척해 온 소년이 명실상부한 영국 최고의 지성으로 인정받는 상징적인 장면이었습니다. 허사비스의 기사 작위 수훈은 영국에게도 특별한 의미가 있었습니다. 아이작 뉴턴, 찰스 다윈, 앨런 튜링으로 이어지는 영국의 위대한 과학적 계보가 21세기에 이르러 AI라는 새로운 영역으로 확장되었음을 보여주는 사건이었기 때문입니다.

영국 정부와 과학계는 허사비스를 '현대판 튜링'으로 내세우며, 영국이 AI 기술의 종주국이 자 윤리적 리더십을 가진 국가임을 전 세계에 과시했습니다. 허사비스에게 '경(Sir)'이라는 호칭은 개인적 영광을 넘어, 그가 이끄는 딥마인드가 실리콘 벨리의 자 본에 속해 있으면서 도 그 뿌리와 철학은 영국의 인문학적, 과학적 전통에 깊이 박혀 있음을 증명하는 훈장이었습니다. "AI의 설계자들(Architects of AI)" , 2025년 타임 올해의 인물 공동 선정 2025년 4월, 허사비스는 2017년에 이어 두 번째로 타임(TIME)지가 선정한 '가장 영향력 있는 100인'에 이름을 올렸습니다.

노벨 화학상 수상자인 제니퍼 다우드나(Jennifer Doudna) 는 추천사에서 "그는 단순히 강력한 시스템을 구축하는 것이 아니라, 이해 (understanding) 를 구축하고 있다"며 허사비스가 생물학과 컴퓨팅을 융합하는 방식 에 찬사를 보냈습니다. 같은 해 12월, 타임지는 2025년 '올해의 인물(Person of the Year)'로 'AI의 설계자 들 (Architects of AI)'이라는 이름 아래 데미스 허사비스를 선택 올트면, 젠슨 황과 함께 공동 선 정했습니다. 잡지의 표지에는 인간의 뇌를 닮은 복잡한 신경망을 배경으로, 서로 다른 곳을 응시하는 이들의 모습이 담겼습니다.

타임지는 선정 이유로 "인류가 불을 발견한 이래 가장 강력한 도구를 손에 쥐여준 인물들" 이라고 밝히며, 특히 허사비스에 대해서는 "상업적 경쟁 속에서도 과학적 발 견이라는 AI의 본질적 가치를 잃지 않고, 인류 지식의 총량을 확장하는 데 기여했 다"고 평가했습니다. 2025년은 제미니(Gemini)와 같은 거대언어모델(LLM)이 전 세계인의 일상에 완 전히 스 며든 해였습니다. 동시에 AI가 신약 개발, 기후 예측, 신소재 발견 등 실질적 인 물리적 세계 의 문제를 해결하기 시작한 원년이기도했습니다.

이 시점에서 허사비 스가 올해의 인물로 선정된 것은, AI 혁명이 단순히 챗봇과 대화하는 흥미로운 놀이를 넘어, 문명의 기반을 재 설계하는 단계로 진입했음을 상징했습니다. 그는 인터뷰에서 "우리는 이제 겨우 출발선에 섰을 뿐입니다. 진정한 변화는 AI가 실험실 밖으로 나와 우리 삶의 가장 어려운 문제들을 해결할 때 비로소 시작될 것입니다"라고 말 하며, 여전히 그의 시선이 현재의 영광이 아닌 미래의 AGI를 향해 있음을 내비쳤습니다.

래스커 의학연구상, 캐나다 게어드너 국제상 옥스퍼드, 케임브리지, UCL, 임페리얼 등 명예박사학위 허사비스의 서재에는 노벨상 메달 옆으로 과학계에서 가장 권위 있는 상패들이 즐비하게 놓여 있습니다. 2023년은 그에게 '수상의 해'였습니다. 그는 실리콘 벨리의 노벨상이라 불 리는 '브레이크스루 상(Breakthrough Prize)' 생명과학 부문을 수상하며 300만 달러의 상금을 받았습니 다.

미국의 '래스커 의학연구상 (Lasker Award)'과 '캐나다 게어드너 국제상 (Canada Gairdner International Award)'을 연달아 수상했습니다. 이 상들은 전통적으로 노벨 생리의학상이나 화학상의 전초전으로 여겨지는 상들로, 의사가 아닌 컴퓨터 공학자가 이 상들을 휩쓴 것은 전례가 없는 일이었습니다. 생물학의 패러다임이 '관찰과 실험'에서 '데 이터와 예측'으 로 이동하고 있음을 보여주는 강력한 신호였습니다.

또한 그는 자신이 공부했던 학문적 고향들로부터 최고의 예우를 받았습니다. 케임브리지 대학교와 유니버시티 칼리지 런던(UCL)은 물론, 옥스퍼드 대학교와 임페리얼 칼리지 런던 등 영국의 주요 명문 대학들이 앞다투어 그에게 명예박사 학위를 수여했습니다. 젊은 시 절, 학계의 느린 속도에 답답함을 느껴 스타트업의 세계로 뛰어들었 던 그가, 이제는 학계 가 가장 존경하는 인물이 되어 강단으로 돌아온 것입니다.

그는 수락 연설에서 종종 이렇게 말하곤 했습니다. "저는 학계를 떠난 것이 아니라, 학계를 확장하기 위해 잠시 다른 길을 걸었을 뿐입니다." 왕립학회(FRS), 왕립공학아카데미 (FREng) 펠로우 허사비스가 받은 수많은 영예 중에서도 그가 가장 아끼는 타이틀 중 하나는 바로 '왕립학 회 펠로우(FRS, Fellow of the Royal Society)'입니다.

2018년, 그는 만 41세의 나이로 360년 역사를 가진 영국 왕립학회의 최연소 펠로우 중 한 명으로 선출되었습니다. 아이작 뉴턴, 찰스 다윈, 알베르트 아인슈타인, 스티븐 호킹과 같은 거인들이 거쳐 간 그 명단에 자신의 이름을 올린 것입니다. 그는 영국 왕립공학아카데미(FREng)의 펠로우로도 선출되었습니다.

과학(FRS)과 공학 (FREng) 양쪽 아카데미의 펠로우가 된다는 것은 극히 드문 일입니다. 이는 허사비스가 가진 독특한 정체성, 즉 과학적 탐구 정신과 공학적 해결 능력을 겸비한 '융합형 인재'로서의 면모를 공식적으로 인정받은 것이었습니다. 그는 왕립학 회의 앤틱한 회의실에 앉아 AI의 미래를 논하면서도, 그 기술이 어떻게 현실의 공학적 난제들을 해결할 지에 대해 치열하게 고민했습니다.

그에게 이러한 펠로우십은 단순 한 명예직이 아니라, 과학 기술이 사회에 미치는 영향에 대해 책임감을 가지고 목소 리를 내야 하는 무거운 책무 (Noblesse Oblige)였습니다.

2024년 노벨 화학상 발표 현장

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

구조 예측의 새 지평 2024년 5월, 런던 킹스크로스의 구글 딥마인드 본사는 다시 한 번 조용한 흥분에 휩싸여 있었습니다. 불과 4년 전 알파폴드2로 단백질 구조 예측이 라는 50년 난제를 해결하며 전 세 계를 놀라게 했던 그들이었지만, 허사비스와 그의 팀에게 그것은 끝이 아닌 시작이었습니다. 알파폴드2가 단백질이라는 '고독한 조각상'의 모양을 밝혀냈다면, 이제 그들은 그 조각상 들이 서로 어떻게 손을 잡고 춤을 추는지를 밝혀내야했습니다.

생명은 정지된 구조가 아 니라 상호작용 그 자체이기 때문입니다. 허사비스는 늘 직원들에게 "생물학 은 정적인 명사 가 아니라 동적인 동사다"라고 강조해 왔습니다. 그 철학의 결정체가 바로 알파폴드3였습니다.

알파폴드3의 등장은 단순히 이전 버전의 성능 개선판이 아니었습니다. 이것은 근본 적인 아키텍처의 재설계이자, 생물학을 바라보는 관점의 확장이었습니다. 기존의 알파폴드2가 아미노산 서열을 입력받아 단백질의 3차원 구조를 그려내는 데 집중했다면, 알파폴드3는 생명을 구성하는 거의 모든 분자의 대화에 귀를 기울였습니다.

우리 몸속의 단백질은 홀로 존재하지 않습니다. 그들은 DNA라는 설계도에 달라붙어 유전자를 읽어내고, RNA와 결합하여 정보를 전달하며, 리간드라고 불리는 작은 분자 신호들과 끊임 없이 결합하고 떨어지기를 반복합니다. 특히 리간드는 약물 개발의 핵심 열쇠입니다.

대부분의 약은 단백질이라는 자물쇠 에 딱 맞 는 리간드라는 열쇠를 찾아내는 과정이기 때문입니다. 알파폴드2는 자물쇠 의 모양은 보여 주었지만, 열쇠가 꽂힌 모습까지는 완벽하게 보여주지 못했습니다. 알 파폴드3는 바로 이 지점에서 혁명을 일으켰습니다.

기술적으로 흥미로운 점은 허사비스와 연구팀이 '확산 모델(Diffusion Model)'을 도입했다 는 사실입니다. 미드저니나 달리 같은 AI가 텍스트를 입력받아 이미지를 생 성할 때 사용하 는 기술과 유사합니다. 초기에 흐릿하고 노이즈가 가득한 구름 같은

상태에서 시작하여, 점차 노이즈를 제거해가며 선명하고 정확한 분자 구조를 만들어 내는 방식입니다. 허사비스는 이 과정을 "안개 속에서 서서히 드러나는 진실"에 비유하곤했습니다. 이 확산 모델의 도입으로 알파폴드3는 기존의 에보포머(Evoformer) 모듈을 일부 대체하거 나 보완하면서, 단백질뿐만 아니라 DNA, RNA, 그리고 각종 이온과 수식화된 분자들까지 하나의 통합된 네트워크 안에서 처리할 수 있게되었습니다.

이것은 AI가 생물학적 직관을 넘어서, 물리적 결합의 확률론적 본질에 더 가까이 다가갔음을 의미합니다. 이 성과는 과학계에 즉각적인 충격을 주었습니다. 기존에는 엑스선 결정학이나 초저온 전자현미경(Cryo-EM)을 사용하여 수개월, 혹은 수년이 걸려야 확인할 수 있 었던 거대 복합체의 구조를 불과 몇 분 만에 시뮬레이션할 수 있게 된 것입니다.

예를 들어, 암세포가 증식하기 위해 DNA의 특정 부위에 결합하는 단백질의 구조를 알 수 있다면, 그 결합을 방해하는 약물을 설계하는 일은 훨씬 수월해집니다. 알파폴드3는 단백질-DNA 상호작용 예측에서 기존 최고의 툴보다 정확도를 50% 이상 향상시켰습니다. 허사비스가 10대 시절 체스판 위에서 수읽기를 하며 느꼈던 전율, 즉 '보이지 않는 패턴을 읽어 내는 기쁨'이 이제는 생명의 가장 미시적인 세계에서 재 현되고 있는 셈입니다. 그는 네이처(Nature)에 논문을 게재하며 이 도구를 전 세계 과학자들에게 무료로 공개(비 상업적 용도에 한해)하는 결정을 내렸는데, 이는 "가장 강력한

도구는 가장 널리 쓰일 때 가치가 있다"는 그의 오랜 신념을 다시 한번 증명한 행동이었습니다.

알파폴드-멀티머(AlphaFold-Multimer)와 단백질 복합체 연구 알파폴드3로 가는 여정에는 중요한 징검다리가 있었습니다. 바로 2021년 말에 공개된 '알파폴드-멀티머'입니다. 이 단계는 허사비스가 이끄는 딥마인드 팀이 '단일 개체'에서 '시스템'으로 시야를 넓힌 결정적인 순간이었습니다.

생물학 교과서에 나오는 헤모글로빈을 예로 들어보겠습니다. 헤모글로빈은 하나의 단백질 덩어리가 아니라, 네 개의 소단위체가 정교하게 결합된 복합체입니다. 이 네 개의 조각이 완벽하게 맞물리지 않으면 우리 몸은 산소를 운반할 수 없습니다.

알파폴드2가 각 조각의 모양을 훌륭하게 빚어냈다면, 알파폴드-멀티머는 이 조각들이 어떻게 조립되는지를 풀어낸 해결사였습니다. 연구팀의 수석 연구원이었던 존 점퍼(John Jumper)는 당시를 회상하며 "우리는 단백질이 외로움을 탄다는 사실을 알고 있었습니다"라고 농담처럼 말하곤 했습니다. 실제로 세포 내의 거의 모든 기능은 단백질들이 짝을 이루거나 거대한 기계처럼 뭉쳐서 수행합니다.

신호 전달, 물질 수송, 세포 분열 등 생명의 핵심 현상은 모두 '결합'에서 시작됩니다. 허사비스는 이 복합체 예측 문제가 단순히 계산량을 늘리는 문제가 아님을 직감했습니다. 두 단백질이 만날 때 그 접점에서는 미묘한 형태의 변화가 일어납니다. 마치 두 사람이 악수할 때 손 모양이 살짝 변하듯, 단백질들도 서로 유도 적합(Induced Fit) 과정을 거칩니다. 알파폴드-멀티머는 진화 정보를 담고 있는 다중 서열 정렬(MSA) 데이터를 교차 분석하여, 두 단백질이 진화 과정에서 함께 변화해 온 흔적, 즉 '공진화(Co-evolution)'의 신호를 찾아 냈습니다.

A 단백질의 특정 부위가 변할 때 B 단백질의 맞닿은 부위도 함께 변했다면, 그 두 지점이 결합 부위일 확률이 높다는 논리입니다. 이 기술의 파급력은 실로 엄청났습니다. 구조 생물학자들은 오랫동안 거대 단백질 복합체의 구조를 밝히기 위해 '퍼즐 맞추기'와 같은 싸움을 해왔습니다.

알파폴드-멀티머는 이 퍼즐의 완성 그림을 미리 보여주는 것과 같았습니다. 이를 통해 과학자들은 미지의 바이러스가 인간 세포에 침투할 때 사용하는 스파이크 단백질이 인간의 수용체와 어떻게 결합하는지 시뮬레이션할 수 있었고, 면역 세포가 항원을 인식하는 과정을 분자 수준에서 들여다볼 수 있게 되었습니다. 이는 학문적 호기심을 충족시키는 것을 넘어, 질병의 메커니즘을 이해하는 근본적인 열쇠를 제공했습니다.

허사비스에게 있어 알파폴드-멀티머와 알파폴드3로 이어지는 흐름은 그의 AGI 비전과도 맞닿아 있습니다. 그는 지능을 "정보를 처리하여 미래를 예측하고 계획하는 능력"이라고 정의합니다. 생명 현상 또한 유전 정보가 단백질이라는 하드웨어로 번역되고, 이 하드웨어들이 상호작용하며 생명이라는 소프트웨어를 구동하는 과정으로 볼 수 있습니다.

따라서 AI가 이 복잡한 상호작용을 예측한다는 것은, 곧 생물학이라는 거대한 정보 처리 시스템의 운영 체제를 이해하기 시작했다는 뜻입니다. 허사비스는 종종 "우리는 디지털 세계에서 훈련된 지능을 물리적 세계의 가장 복잡한 문제인 생물학에 적용하고 있다"고 말합니다. 알파폴드 시리즈는 그가 꿈꾸는 '과학을 위한 AI(AI for Science)'가 단순한 구호가 아니라, 인류의 지식 지평을 실제로 확장하고 있음을 보여주는 가장 강력한 증거입니다. 이제 연구실의 모니터 속에서는 수십만 개의 원자들이 춤을 추고 있으며, 인류는 그 춤의 안무를 비로소 이해하기

시작했습니다.

2021년 아이소모픽 랩스 분사: 신약 개발을 '계산 가능한 문제'로 바꾸다 AI를 통한 신약 개발 가속화: 10년 걸릴 일을 수개월로 단축 2021년, 데미스 허사비스는 구글의 모회사 알파벳(Alphabet) 산하에 새로운 회사를 설립한 다고 발표했습니다. 이름은 '아이소모픽 랩스(Isomorphic Labs)'. 수학 용어인 '동형 (Isomorphism, 같은 형태)'에 서 따온 이 이름에는 허사비스의 깊은 통찰이 담겨 있습니다. 생물학적 과정과 정보 처리 과정이 근본적으로 같은 형태의 구조를 가지고 있다는 믿음입 니다. 딥마인드가 순수 과학의 영역에서 '지능을 푸는' 곳이라면, 아이소모픽 랩스는 그 풀 린 지 능을 이용해 인류의 가장 고통스러운 문제인 '질병'을 해결하는 실전 부대였습니다.

허사비 스는 딥마인드의 CEO직을 유지하면서 아이소모픽 랩스의 CEO를 겸임하 기로 결정했습니 이는 그가 이 프로젝트에 얼마나 큰 비중을 두고 있는지를 보여주는 대목입니다. 전통적인 신약 개발은 '실패를 전제로 한 도박'과도 같았습니다. 하나의 신약이 시장 에 나 오기까지 평균 10년 이상의 시간과 2조 원이 넘는 비용이 소요됩니다.

수만 개 의 후보 물질 중에서 실험실의 비커와 샬레를 오가며 일일이 반응을 테스트하고, 동물 실험을 거쳐 임상 시험에 들어가도 성공 확률은 10% 미만입니다. 마치 건초 더미에서 바늘을 찾는 것과 다름 없었습니다. 허사비스는 이 비효율적인 과정을 '디버깅 가능한 소프트웨어 문제'로 전환하고자 했습니 아이소모픽 랩스의 핵심 전략은 알파폴드의 예측 능력을 바탕으로 약물이 될 수 있는 작은 분자(Small Molecule)를 설계하는 것입니다.

기존에는 표적 단백질의 구조를 알아내는 데만 수년이 걸렸지만, 이제는 AI가 며칠 만에 표적의 3차원 구조를 예측하고, 그 구조의 '주머니(Binding Pocket)'에 딱 들어맞 는 화합물을 가상 공간에서 수백만 번 시뮬레이션합니다. 이것은 물리적 실험을 디지 털 실험(In Silico)으로 대체하는 거대한 전환입니다. 2024년 초, 아이소모픽 랩스는 글로벌 제약사 일라이 릴리(Eli Lilly), 노바티스(Novartis)와 총액 30억 달러(약 4조 원) 규모의 전략적 파트너십을 체결했습니다.

보 수적인 제약 업계가 설립된 지 3년밖에 안 된 신생 기업에 천문학적인 금액을 베푼 것은, AI가 보여준 가능성 이 단순한 기대감을 넘어선 확신으로 바뀌었음을 의미합니다. 허사비스는 이 과정을 설명하며 "우리는 시간을 압축하고 있다"고 말했습니다. 과거에는 화학자들이 직관과 경험에 의존해 화합물을 합성했다면, 이제는 생성형 AI가 물 리 법칙과 생물학적 제약 조건을 학습하여 인간이 상상하지 못한 새로운 구조 의 약물을 제 안합니다.

이는 단순히 속도만 빨라지는 것이 아닙니다. 인간의 편향에 갇혀 시도조차 해 보지 않았던 '케미컬 스페이스(Chemical Space)'의 미개척지를 탐험 하게 된 것입니다. 아이소모픽 랩스의 연구원들은 이제 10년이 걸리던 초기 약물 발견(Discovery) 단 계를 수 개월, 심지어는 수주 단위로 단축시키고 있습니다. 이것은 질병으로 고통받는 환자들에게 는 '희망의 속도'가 빨라진다는 것을 의미합니다.

2025년 AI 설계 약물의 임상시험 진입 발표 2025년은 디지털 생물학의 역사에서 중요한 분기점이 될 전망입니다. 허사비스와 아이소 모픽 랩스는 순수하게 AI에 의해 처음부터 끝까지 설계된(de novo design) 약물 후보 물질 들이 본격적으로 임상 시험 (Clinical Trial)에 진입하는 해로 2025년을 지목했습니다. 물론 이전에도 AI를 활용 해 약물을 재창출하거나 부분적으로 도움을 받은 사례는 있었지만, 알 파폴드3와 같 은 고도화된 구조 기반 약물 설계(SBDD) 기술이

벗어낸 '차세대 AI 신약'이 인 간에게 투여되는 것은 차원이 다른 이야기입니다.

이 약물들은 기존의 방식으로는 공략하기 어려웠던 'undruggable(약물 치료가 불가 능한)' 표적들을 겨냥하고 있습니다.

예를 들어, 구조가 너무 유연해서 약물이 달라붙기 힘든 암 유발 단백질이나, 기존 약물에 내성이 생긴 변이 바이러스 등이 그 대상입니다. 아이소모픽 랩스는 AI를 통해 단백질의 움직임(Dynamics)까지 예측하여, 단백질이 특정 모양을 취하는 찰나의 순간을 포착해 결합하는 약물을 설계했습니다. 이는 정지 사진을 보고 사격하는 것과, 움직이는 동영상을 보고 사격하는 것의 차이만큼이나 큼니다. 임상 진입 발표는 단순한 기술적 성취를 넘어 규제 과학(Regulatory Science)의 변화를 예고합니다.

FDA와 같은 규제 기관들도 AI가 설계한 약물의 안전성과 유효성을 어떻게 평가할 지에 대해 새로운 기준을 마련하고 있습니다. 허사비스는 이 과정에서 "우리는 단순히 약을 만드는 것이 아니라, 약을 만드는 방법 그 자체를 재발명하고 있다"고 강조합니다. 2025년의 임상 시험은 AI가 실험실의 보조자를 넘어, 생명을 구하는 주체적인 발명가가 될 수 있는지를 검증하는 '최종 시험대'가 될 것입니다 만약 이 시험들이 성공적인 데이터를 내놓기 시작한다면, 우리는 'AI 네이티브 제약(AINative Pharma)' 시대의 본격적인 개막을 목격하게 될 것입니다.

가상 세포(Virtual Cell) 시뮬레이션의 꿈 아이소모픽 랩스와 허사비스가 바라보는 궁극의 지점은 신약 개발 너머에 있습니다. 바로 '가상 세포(Virtual Cell)'의 구현입니다. 현재의 생물학 연구는 복잡한 시계 내부를 이해하기 위해 부품 하나하나를 떼어내어 살펴보는 것과 같습니다.

하지만 부품을 다 안다고 해서 시계가 어떻게 작동하는지 완벽하게 아는 것은 아닙니다. Virtual Cell은 이 모든 부품들, 단백질, DNA, 대사 물질, 세포소기관, 이 상호작용하는 전체 시스템을 컴퓨터 안에 쌍둥이처럼 구현하는 것입니다. 이것은 허사비스가 어린 시절부터 꿈꿔온 '시뮬레이션'의 정점입니다. 게임 개발자 시절 그가 테마파크 속 관람객들의 행동을 시뮬레이션했듯이, 이제는 세포 안의 분자들의 행동을 시뮬레이션하려는 것입니다.

가상 세포가 구현된다면, 우리는 실제 환자에게 약을 투여하기 전에 가상 세포에 먼저 약을 넣어보고 어떤 부작용이 일어날지, 어떤 연쇄 반응이 일어날지를 미리 예측할 수 있습니다. 이는 동물 실험의 윤리적 문제를 해결할 수 있는 대안이자, 임상 시험의 실패 확률을 획기적으로 낮출 수 있는 방법입니다.

물론 이것은 알파폴드 하나만으로는 해결할 수 없는 거대한 도전입니다. 하지만 허 사비스 는 "우리는 이제 생물학의 퍼즐 조각들을 맞추 준비가되었다"고 말합니다. 알파폴드가 구조를 예측하고, 다른 AI 모델들이 대사 경로와 유전자 조절 네트워크를 예측 하며, 이들이 결합하여 거대한 '디지털 생명체'를 형성하는 미래. 그것이 아이 소모픽 랩스 가 2021년 분사하며 내건 비전의 중추역입니다.

아직은 먼 미래의 이야 기처럼 들릴지 모르 지만, 알파고가 바둑을 정복하고 알파폴드가 단백질을 정복했듯 이, 허사비스의 시계는 언 제나 세상의 예상보다 빠르게 돌아가고 있습니다. 가상 세 포는 단순한 꿈이 아니라, 데이터와 컴퓨팅 파워가 쌓일수록 점점 선명해지고 있는 예정된 미래입니다. 생물학은 정보 처리 시스템이다 연구에서 제품으로: 규제, 임상, 파트너십의 현실 "생물학은 가장 복잡하고 아름다운 정보 처리 시스템입니다."

허사비스의 이 선언은 생물 학을 Wet Lab (실험실)의 영역에서 데이터와 알고리즘이 지배하는 건조한 실험 실(Dry Lab) 의 영역으로 끌어왔습니다. 하지만 비전을 현실의 제품으로, 즉 실제 환자가 복용할 수 있는 알약으로 만드는 과정은 순수한 과학적 발견과는 또 다른 차원의 투쟁을 요구했습니다. 2024년 노벨상 수상으로 과학적 권위의 정점에 선 그였지만, 제약 산업의 복잡한 현실은 AI 모델의 정확도(Accuracy)만으로는 돌파할 수 없는 거대한 장벽들이었습니다. 아이소모픽 랩스가 직면한 첫 번째 현실적 과제는 '데이터의 비대칭성'이었습니다.

알파고는 바둑 규칙이라는 완벽한 정보 안에서 무한한 자가 대국을 통해 데이터를 생성할 수 있었습니다. 신약 개발은 다릅니다. 양질의 생물학 데이터는 대부분 거대 제약사들의 금고 깊숙한 곳에 비공개로 잠겨 있습니다. 허사비스가 노바티스나 일라이 릴리 같은 전통적인 제약 공룡들과 손을 잡은 것은 선택이 아닌 필수였습니다.

이는 딥마인드의 천재적인 알고리즘(엔진)과 제약사들이 수십 년간 축적해 온 임상 데이터 및 화학 라이브러리(연료)를 결합하는 전략이었습니다. 이 파트너십은 서로 다른 두 언어를 쓰는 문명간의 만남과도 같았습니다. 실리콘밸리의 속도(Speed)와 제약 업계의 안전(Safety) 제일주의가 충돌했습니다.

소프트웨어 버그는 패치로 수정하면 되지만, 약물의 부작용은 사람의 생명을 위협합니다. 허사비스는 이 간극을 메우기 위해 아이소모픽 랩스 내에 AI 엔지니어뿐만 아니라 노련한 의학 화학자와 생물학 자들을 대거 영입했습니다. "우리는 AI 회사가 아니라, AI를 모국

어로 사용하는 제약 회사 (AI-native Pharmaceutical Company)가 되어야 한다"는 것이 그의 주문이었습니다. 규제(Regulation) 또한 넘어야 할 산이었습니다. FDA를 비롯한 규제 기관들은 AI가 내놓은 결과를 '블랙박스'로 바라보았습니다.

"왜 이 분자가 효과가 있다고 예측했는가?"라는 질문에 대해 초기 딥러닝 모델들은 설명할 수 없었습니다. 허사비스 팀은 설명 가능한 AI(XAI) 기술을 도입하여, AI의 예측 근거를 화학자들이 이해할 수 있는 언어로 변환하는 데 막대한 노력을 기울여야 했습니다. 연구실의 혁신이 병원의 처방 전으로 이어지기 위해 반드시 통과해야 하는 '신뢰의 관문'이었습니다.

과학적 정확성 vs 사업적 속도의 균형 데미스 허사비스라는 인물 내부에는 '순수한 진리를 추구하는 과학자'와 '세상을 바꾸고 싶은 기업가'가 공존해 왔습니다. 디지털 생물학이라는 새로운 영역에서 이 두 자아는 끊임없이 대화하고 때로는 갈등했습니다. 알파폴드3의 개발과 아이소모픽 랩스의 운영 과정에서 이 긴장은 극대화되었습니다.

과학자로서의 허사비스는 완벽함을 추구합니다. 그는 자연의 근본 원리를 이해하고, 단백질 접힘과 같은 난제를 완전히 해결(Solve)하고 싶어합니다. 그에게 있어 성공의 척도는 네이처 표지 논문, CASP 대회 우승, 그리고 노벨상이었습니다.

그러나 아이소모픽 랩스의 CEO로서의 허사비스는 '충분히 좋은(Good Enough)' 해결책을 적시에 내놓아야 한다는 압박을 받습니다. 신약 개발에서는 100% 완벽한 구조 예측보다는, 다소 오차가 있더라도 빠르게 합성 가능하고 독성이 없는 후보 물질을 찾아내는 것이 더 중요할 때가 많기 때문입니다. 이 균형점은 '피드백 루프(Feedback Loop)'의 설계에서 찾아졌습니다. 아이소모픽 랩스는 단순히 AI 모델만 돌리는 회사가 아닙니다. 그들은 예측된 분자를 실제로 합성하고 테스트 하는 자체 실험실을 구축했습니다.

AI가 설계(Design)하고, 로봇이 합성(Make)하고, 생물학적 분석(Test)을 거쳐 나온 데이터를 다시 AI가 학습(Learn)하는 'DMTL 사이클'을 구축한 것입니다. 이 과정에서 허사비스는 과학적 정확성에 대한 집착을 버리지 않으면서도, 사업적 속도를 높이는 방법을 배웠습니다. 모든 것을 다 예측하려 하지 않고, 약물 개발에 결정적인 영향을 미치는 핵심 상호작용에 컴퓨팅 자원을 집중하는 '선택과 집중' 전략을 택한 것입니다.

이러한 접근은 AI 업계에 만연한 '과대광고(Hype)'와 차별화되는 지점이었습니다. 많은 AI 바이오 스타트업들이 장밋빛 미래만을 팔 때, 허사비스는 냉정할 정도로 과학적 검증(Validation)을 강조했습니다. "생물학은 속일 수 없다(Biology doesn't lie)."

그는 투자자들에게 보여주기 위한 데 모가 아니라, 실제 실험실에서 재현 가능한 결과를 요구했습니다. 이것은 구글이라는 거대 자본의 보호막 아래 있었기에 가능한 인내심이기도 했지만, 동시에 과학적 진실 성만이 결국 상업적 성공을 담보한다는 그의 확고한 철학 때문이기도 했습니다. 결국 디지털 생물학의 비전은 생명을 0과 1로 환원하여 통제하겠다는 오만이지 않습니다. 그 것은 생명이라는 경이로운 복잡성 앞에 겸손해지되, 우리가 가진 최고의 도구인 지능(Intelligence)을 사용하여 그 복잡성의 패턴을 이해하고, 질병이라는 버그를 수정하여 인류의 삶을 개선하겠다는, 허사비스가 12세 때부터 품어온 "지능을 풀어 세상을 구한다"는 미션의 가장 구체적이고 숭고한 실천인 것입니다. 단백질 분자의 3D 렌더링

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

문제 해결 2024년 7월, 영국의 고풍스러운 도시 바스(Bath)에서는 전 세계에서 가장 뛰어난 수학 영재들이 모여 제65회 국제수학올림피아드(IMO)를 치르고 있었습니다. 같은 시각, 데미스 허사비스가 이끄는 구글 딥마인드의 연구실에서도 보이지 않는 참가자가 시험지를 받아들었습니다. 바로 '알파 프루프(AlphaProof)'와 '알파 지오메트리 2(AlphaGeometry 2)'라는 이름의 AI 모델이었습니다.

허사비스는 늘 "수학은 자연을 기술하는 언어"라고 강조해 왔습니다. 바둑이나 체스처럼 규칙이 명확하지만, 그 복잡성은 무한에 가깝고 정답은 오직 하나뿐인 진리의 영역입니다. 챗GPT와 같은 거대언어모델(LLM)이 시를 쓰고 코드를 짜는 데는 능숙했지만, 유독 수학 앞에서는 작아졌던 이유가 여기에 있습니다.

언어 모델은 '그렇듯한' 다음 단어를 확률적으로 예측하지만, 수학은 99.9%의 확률이 아니라 100%의 논리적 증명을 요구하기 때문입니다. 허사비스는 이 간극을 메우는 것이 AGI로 가는 결정적 관문이라 확신했습니다. 알파프루프의 작동 방식은 이렇습니다. 수학 문제를 형식 언어(Lean이라는 수학 증명 도구)로 번역한 뒤, AI가 수백만 가지 증명 경로를 탐색하며 논리적으로 빈틈없는 풀이를 찾아냅니다. 체스에서 알파고가 수를 읽듯이, 알파프루프는 증명의 수를 읽는 셈입니다.

이날 딥마인드가 공개한 결과는 인상적이었습니다. AI 시스템은 총 6문제 중 4문제를 완벽하게 풀어내며 28점을 획득했습니다. 이는 은메달 커트라인에 해당하는 점수였습니다.

특히 기하학 문제 하나는 불과 19초 만에 증명해 냈습니다. 몇 년 전까지만 해도 초등학교 수준의 서술형 수학 문제도 제대로 풀지 못했던 AI가, 이제는 인류 최상위 0.1%의 영재들과 어깨를 나란히 하게 된 것입니다. 성취의 핵심은 허사비스가 오랫동안 고집해 온 '신경-기호(Neuro-symbolic)' 접근 법에 있었습니다.

딥마인드 팀은 직관적인 아이디어를 내는 거대언어모델(제미나이)의 능력과, 엄격한 논리 규칙을 검증하는 기호주의(Symbolic) 시스템을 결합했습니다. 뛰어난 직관을 가진 수학자가 가설을 세우면, 꼼꼼한 검토자가 그 가설을 한 줄 한 줄 증명해 나가는 방식과 같습니다. 알파프루프는 문제를 '린(Lean)'이라는 형식 증명 언어로 번역하여 스스로 학습했습니다.

이는 허사비스가 체스 신동 시절, 수만 번의 대국을 통해 직관을 기르고 복기를 통해 논리를 다졌던 과정의 디지털 버전이라 할 수 있습니다. 2024년의 은메달은 단순한 성적이 아니었습니다. 그것은 AI가 인간의 데이터를 흉내 내는 '모방'의 단계를 넘어, 스스로 논리적 추론을 통해 '진리'에 도달할 수 있음을 증명한 사건이었습니다.

허사비스는 이 결과를 보고 고받으며, AI가 과학적 발견의 도구로서 완성되어 가고 있음을 직감했습니다. 알파 에볼브(AlphaEvolve): AI가 알고리즘을 설계하는 시대 수학 문제 풀이가 주어진 정답을 찾아가는 과정이라면, 알고리즘 설계는 정답으로 가는 '새로운 길'을 닦는 창조의 영역입니다. 2025년 5월, 딥마인드는 '알파에볼브(AlphaEvolve)'를 세상에 공개하며 또 한 번의 충격을 안겼습니다.

허사비스는 컴퓨터 과학의 근본적인 비효율성에 주목했습니다. 인류가 지난 50년간 사용해 온 정렬(Sorting), 해시(Hashing) 등의 기본 알고리즘들은 대부분 1960~70년대에 천재적인 인간

프로그래머들이 고안한 것입니다. 그 이후 우리는 더 빠른 컴퓨터를 만드는 데 집 중했을 뿐, 이 기본 알고리즘 자체를 개선할 생각은 거의 하지 못했습니다.

허사비스는 물었습니다. "만약 AI가 인간의 편견 없이 처음부터 다 시 코드를 짤다면 어떨까?" 그 전조는 2023년 공개된 '알파데브(AlphaDev)'였습니다.

알파데브는 강화학습을 통해 C++ 라이브러리의 정렬 알고리즘을 수십 년 만에 개선 하는 성과를 올렸습니다. 하지만 알파에 볼브는 여기서 한 발 더 나아갔습니다. 제미니(Gemini) 모델의 코드 생성 능력에 진화 연 산(Evolutionary Computation)을 결 합한 이 시스템은, 스스로 코드를 작성하고(번이), 성능을 평가하고(선택), 더 나은 코 드를 다시 작성하는(진화) 과정을 무한 반복했습니다.

알파에볼브의 위력은 구글의 심장부인 데이터센터에서 증명되었습니다. 전 세계 수 백만 대의 서버가 돌아가는 구글 데이터센터의 작업 스케줄링은 극도로 복잡한 문제입니다. 알 파에볼브는 기존 인간 엔지니어들이 설계한 방식보다 훨씬 효율적인 스케 줄링 알고리즘을 찾아냈고, 이를 통해 구글 전체 컴퓨팅 자원의 0.7%를 회수하는 데

성공했습니다. 0.7% 라는 숫자는 작아 보일 수 있지만, 구글의 규모를 생각하면 수천 억 원의 비용과 막대한 양 의 탄소 배출을 줄인 거대한 성과입니다. 더 흥미로운 것은 알파에볼브가 만들어낸 알고리즘의 형태였습니다.

인간 프로그래머라 면 절대 작성하지 않을, 직관적으로 이해하기 힘든 명령어의 배열이 종종 발견되었습니다. 마치 알파고의 '37수'처럼, AI는 인간의 사고방식에 갇히지 않은, 컴퓨터 하 드웨어에 최적화된 자신들만의 '방언'으로 코드를 작성하기 시작한 것입니다. 허사비 스에게 알파에볼브 는 "지능을 풀어 모든 것을 해결한다"는 미션의 강력한 증거였습니다.

AI가 단순히 인간을 돕는 비서가 아니라, 컴퓨터 과학 자체를 진화시키는 연구 파 트너가 된 것입니다. 그는 이를 두고 "우리는 디지털 생물학뿐만 아니라, 디지털 컴퓨터 과학의 시대를 열고 있다"고 평했습니다. 재료 과학의 혁명 220만 개 이상의 신소재 결정 구조 발견 2023년 말, 과학 저널 《네이처》에 실린 한 편의 논문은 전 세계 소재 과학계를 발칵 뒤집 어 놓았습니다.

딥마인드의 'GNoME' 프로젝트가 220만 개의 새로운 결정 구조를 발견했다 는 소식이었습니다. 이 의미를 이해하려면 인류 역사를 되짚어봐야합니다. 우리는 사용하는 재료에 따라 시대를 구분해 왔습니다.

석기 시대, 청동기 시대, 철 기 시대, 그리고 지금의 실리콘 시대. 새로운 재료의 발견은 언제나 문명의 쿼텀 점프를 가져왔습니 하지만 발견 과정은 지독히도 느리고 고통스러운 시행착오의 연속이었습니다. 에디슨 이 전구 필라멘트를 찾기 위해 수천 가지 재료를 태워봤던 것처럼, 현 대 과학자들도 실험실에서 원소들을 섞고 구우며 우연한 발견을 기다려야했습니다. 인류가 지난 수천 년간 실험을 통해 밝혀낸 안정한 무기 화합물의 결정 구조는 약 4만 8천 개에 불과했습니다.

허사비스는 이 느린 시계를 가속하고 싶었습니다. 그는 단백질 구조를 예측했던 알 파폴드 의 성공 방정식을 재료 과학에 대입했습니다. "원자들의 결합 규칙을 그래프 네트워크 (Graph Networks)로 학습시키면, 안정적인 구조를 미리 예측할 수 있지 않 을까?"

GNoME은 딥러닝을 이용해 원소 주기율표의 조합을 탐색했습니다. 그리고 인류가 2만 년 동안 찾아낸 것보다 45배나 많은, 220만 개의 새로운 물질 후보군을 단숨에 쏟아냈습니다. 그중 38만 개는 현재의 기술로도 당장 합성이 가능한 '안정적' 구

조로 판명되었습니다. 이는 재료 과학 800년 치의 연구 분량에 해당하는 지식이었습니다. 허사비스는 이 성과를 발표하며 평소의 침착한 목소리로 말했습니다.

"우리는 지금 지식의 탐색 범위를 근본적으로 확장하고 있습니다." 그의 말처럼 GNoME은 과학자들에게 보물 지도를 건네주었습니다. 이제 과학자들은 맨땅에 헤딩하는 대신, GNoME이 그려준 지도 위에서 목적지를 골라 탐험을 떠나기만 하면 되었습니다.

초전도체와 청정 에너지 소재 탐색 GNoME이 발견한 보물 지도 속에는 인류가 간절히 원하던 해답들이 숨어 있었습니다. 허사비스가 가장 주목한 분야는 기후 위기 해결을 위한 청정 에너지 기술이었습니다. 전기차의 핵심인 배터리 효율을 획기적으로 높일 수 있는 리튬 이온 전도체, 태양광 패널의 효율을 극대화할 수 있는 새로운 광물질, 그리고 꿈의 물질이라 불리는 상온 초전도체 후보 물질들이 데이터베이스에 포함되어 있었습니다.

실제로 딥마인드는 미국 버클리 국립연구소(LBNL)와 협력하여 'A-Lab'이라는 자율 주행 실험실을 가동했습니다. AI가 설계한 레시피를 로봇 팔이 수행하여 신소재를 합성하는 이 미래적인 풍경은, GNoME의 예측이 단순한 컴퓨터 시뮬레이션이 아님을 증명했습니다. A-Lab은 17일 만에 41개의 새로운 신소재를 합성해 냈습니다.

인간 연구자라면 수년이 걸렸을 작업이었습니다. 허사비스는 '근본적 풍요(Radical Abundance)'를 다시 한번 강조했습니다. 에너지를 더 싸고 깨끗하게 만들고, 컴퓨팅 칩의 성능을 한계 이상으로 끌어올리는 물질을 발견함으로써, 자원의 희소성으로 인한 인류의 갈등을 해결할 수 있다는 믿음입니다.

"우리가 발견한 이 38만 개의 물질 중 하나가 차세대 배터리의 표준이 될 수도 있고, 핵융합 발전의 내벽을 감싸는 소재가 될 수도 있습니다." 2025년 현재, 전 세계의 재료 과학자들은 GNoME이 공개한 데이터를 바탕으로 실험을 진행하고 있습니다. 허사비스는 알파폴드가 생물학자들의 도구가 되었듯, GNoME이 재료 과학자들의 필수 도구가 되기를 바랍니다.

그는 자신이 만든 AI가 노벨상을 받는 것보다, 그 AI를 사용한 과학자가 세상을 바꾸는 발견을 하는 것을 더 큰 영광으로 여겼습니다. 체스판 위에서 수를 읽던 소년은 이제 원자의 배열을 읽으며 인류의 물리적 한계를 확장하고 있었습니다.

AI와 수학의 융합을 상징하는 이미지

인공지능 전문가 김경진 변호사

AI 법정정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

2010년 딥마인드를 설립하며 칠판에 적었던 이 문장은, 15년이 지난 지금 실험실 밖의 현실이 되어 우리 머리 위의 하늘과 땅속의 에너지, 그리고 미지의 물질 세계를 바꾸고 있습니다. 웨더넥스트(WeatherNext) 1) 슈퍼컴퓨터의 굉음과 AI의 침묵: 100년의 꿈을 1분으로 압축하다 2023년 가을, 영국 레딩(Reading)에 위치한 유럽중기예보센터(ECMWF)의 거대한 서버실은 여느 때와 다름없이 굉음을 내고 있었습니다. 세계 최고의 기상학자들이 모인 이곳에서는 농구장 크기만 한 슈퍼컴퓨터가 수백만 개의 프로세서를 가동하며 지구 대기의 움직임을 계산하고 있었습니다.

그들이 사용하는 '수치예보 모델(NWP)'은 지난 100년간 인류가 날씨를 예측해 온 표준이었습니다. 대기를 잘게 쪼개고, 각 격자마다 유체 역학의 난제인 나비에-스토크스(Navier-Stokes) 방정식을 대입하여 미분과 적분을 반복하는 이 방식은 엄청난 전력을 소모하며 10일 뒤의 날씨를 예측하는 데 수 시간이 걸렸습니다. 같은 시각, 런던 구글 딥마인드 사무실에서는 다른 풍경이 펼쳐지고 있었습니다.

레미 램(Remi Lam)을 비롯한 연구팀은 단 한 대의 데스크톱 컴퓨터 앞에 앉아 있었습니다. 그들이 엔터 키를 누르자, 불과 1분 만에 전 세계의 10일 치 기상 예보가 화면에 뿔었습니다. 이 AI 모델의 이름은 '그래프캐스트(GraphCast)'였습니다.

놀랍게도 이 AI는 수천 억 원짜리 슈퍼컴퓨터가 몇 시간 동안 계산한 결과보다 허리케인의 경로를 더 정확하게 예측해 냈습니다. 이것은 단순한 속도의 문제가 아니었습니다. "패러다임의 전환"이었습니다.

"전통적인 방식이 대기의 물리 법칙을 하나하나 계산해서 풀어내는 '연역적 접근'이라면, 우리의 AI는 지난 40년 동안의 날씨 데이터를 보고 스스로 대기의 패턴을 익힌 '귀납적 접근'입니다. 노련한 선장이 구름의 모양만 보고도 폭풍우를 감지하듯, AI는 수식 없이도 직관적으로 지구의 흐름을 읽어냅니다." 2024년과 2025년을 거치며 이

기술은 '웨더넥스트(WeatherNext)'라는 이름으로 진화했습니다. 2025년 말 공개된 '웨더 넥스트 2'는 기존 그래프 캐스트보다 8배 더 빨라졌고, 국지적인 집중 호우나 급격한 기온 변화 같은 기상이변을 예측하는 데 있어 압도적인 성능을 보여주었습니다. 이 혁신이 중요한 이유는 '기후 위기'라는 시대적 맥락과 맞닿아 있기 때문입니다. 기상이 변이 일상이 된 시대에, 알람처럼 울리는 정확한 홍수 경보는 수천 명의 생명을 살릴 수 있습니다.

전통적인 슈퍼컴퓨터 모델은 막대한 에너지를 소비하여 '탄소 발자국'을 남기는 반면, 웨더 넥스트는 가정용 컴퓨터 수준의 전력만으로도 국가 단위의 예보를 수행할 수 있습니다. 허사비스는 이를 두고 "지구를 구하기 위한 기술이 지구를 망쳐서는 안 된다"는 딥마인드의 엔지니어링 철학을 보여주는 사례라고 강조합니다. 기상학계의 반발이 없었던 것은 아닙니다. "물리 법칙을 모르는 블랙박스 AI가 내놓은 예측을 어떻게 신뢰할 수 있는가?"라는 비판은 여전히 유효합니다.

하지만 딥마인드는 이에 대해 '뉴럴GCM(NeuralGCM)'이라는 하이브리드 모델로 응답했습니다. 대기의 큰 흐름은 전통적인 물리 모델로 계산하되, 구름이나 난류 같은 복잡하고 미세한 영역은 AI가 처리하는 방식입니다. 이는 허사비스가 늘 강조해 온 "인간의 지식과 AI의 직관의 결합"을 상징합니다.

웨더넥스트의 기술은 구글 검색에 통합되어, 전 세계 수십억 명의 사람들이 우산을 챙겨야 할지 말지를 결정하는 일상의 도구가되었습니다. 100년 전, 수작업으로 대기를 계산하려 다 실패했던 루이스 프라이 리처드슨의 꿈이 AI를 만나 비로소 완성된 것입니다. 핵융합 플라즈마 반응로에서의 AI 활용 1) 1억 도의 질리를 손가락으로 잡는 법 기 상 예측이 하늘의 패턴을 읽는 것이라면, 핵융합은 태양의 심장을 지구상에 구현하는 일입니다.

허사비스는 어릴 적부터 우주를 여행하는 꿈을 꿔고, 그 꿈을 실현하기 위 한 유일 한 에너지원이 핵융합임을 알고 있었습니다. 핵융합에는 치명적인 난제가 있었습니다. 바 로 1억 도가 넘는 초고온의 플라즈마를 반응로(토카막) 벽에 닿지 않게 공중에 띄워 놓아야 한다는 것입니다.

스위스 로잔 연방 공과대학교(EPFL)의 스위스 플라즈마 센터(SPC). 이곳의 연구원 들은 토 카막 안에서 춤추는 플라즈마를 제어하기 위해 수십 년간 고군분투해 왔습니다. 플라즈마 는 마치 살아있는 뱀처럼 불안정해서, 자석의 자기장을 아주 조금만 잘 못 조절해도 순식간 에 붕괴해 버립니다. 기존의 제어 방식은 수천 개의 변수를 사람 이 미리 계산해 입력해야 했기에, 복잡하고 새로운 형태의 플라즈마를 시도하는 건 엄 두도 내지 못했습니다. 허사비스는 딥마인드의 강화학습 팀을 이곳에 파견했습니다.

그들의 미션은 간단했 습니 "AI에게 플라즈마를 가지고 노는 법을 가르쳐라." 딥마인드의 AI는 시뮬레이션 속에서 수백만 번의 시행착오를 겪었습니다. 처음에는 플라 즈마를 벽에 부딪혀 꺼트리기 일쑤였지만, 알파고가 바둑의 묘수를 스스로 깨우쳤듯, AI는 자기장 코일 19개를 초당 1만 번씩 미세하게 조절하며 플라즈마를 안정시키는 법을 배웠습니다.

그리 고 운명의 날, AI는 실제 토카막 제어권을 넘겨받았습니다. 결과가 나왔습니다. AI는 연구원들이 주문한대로 플라즈마를 둥근 도넛 모양뿐만 아 니라, 눈송이 모양(snow ake), 물방울 모양(droplet) 등 기하학적으로 복잡한 형태 로 자유자 재로 변형시켰습니다.

특히 '음의 삼각도(negative triangularity)'를 가진 플 라즈마 형태를 안 정적으로 유지해 낸 것은 물리학계의 큰 충격이었습니다. 이 형태는 반응로의 벽에 주는 열 손상을 줄이면서도 에너지 효율을 높일 수 있는, 이론상으로만 존재하던 '성배'와 같았 기 때문입니다. fi 핵융합 반응로 안에는 1억 도가 넘는 초고온 기체, 즉 플라즈마가 있습니다.

이 플라즈마는 너무 뜨거워서 어떤 물질로도 담을 수 없기 때문에, 강력한 자기장으로 공 중에 띄워서 가 듭니다. 이때 플라즈마를 옆에서 잘라 본 단면의 모양이 반응로의 성 능을 결정짓는 핵심 요소입니다. 지금까지 전 세계 핵융합 장치들은 거의 예외 없이 플라즈마 단면을 영문 대문자 D 모양으 로 만들어 왔습니다. 바깥쪽으로 볼록한 이 형태를 '양의 삼각도'라고 부릅니다. 그런데 이 형태에는 골칫거리가 하나 있습니다.

플라즈마 가장자리에서 에너지가 주기적으로 폭발 하듯 튀어나오는 현상이 발생하는데, 이것이 반응로 내벽을 반복적으로 때려서 깎아먹는 것입니다. 상용 핵융합을 발전소로 만들려면 반드시 해결해야 할 난제였습니다. 물리학자들은 오래전부터 한 가지 이론적 해법을 알고 있었습니다.

D자 모양을 좌 우로 뒤 집어서, 볼록한 면이 안쪽을 향하게 만드는 것입니다. 이것을 '음의 삼각도'라 고합니다. 계 산상으로는 이 형태에서 에너지 폭발 현상이 자연스럽게 사라졌고, 그

러면서도 에너지를 가두는 효율은 오히려 높게 유지되었습니다. 벽의 손상을 줄이면 서 동시에 성능도 올리는, 두 마리 토끼를 잡는 답이었던 것입니다. 하지만 이것은 어 디까지나

이론이었습니다.

실제 반응로 안에서 이 뒤집힌 형태의 플라즈마를 안정적으로 유지하는 것은 다른 차원의 문제였고, 수십 년간 아무도 성공하지 못했습니다. 그래서 이 형태를 핵융합 분야의 '성배'라고 불렀습니다. 존재한다는 것은 알지만, 아무도 손에 넣지 못한 것이라는 뜻입니다.

그런데 이번에 그 성배를 실험실에서 실제로 구현해 낸 것입니다. 이론으로만 존재 하던 플라즈마 형태를 안정적으로 만들어 유지하는 데 성공했다는 소식에 물리학계가 충격을 받은 이유가 바로 여기에 있습니다. 핵융합 상용화를 가로막던 가장 큰 공학적 장벽 하나를 넘을 수 있는 실마리가 마침내 잡힌 것이기 때문입니다.

이 성과는 2022년 《네이처》 표지를 장식했고, 2025년 12월에는 미국의 핵융합 스타트업 '커먼웰스 퓨전 시스템즈(Commonwealth Fusion Systems, CFS)'와의 파트너십으로 이어졌습니다. 딥마인드는 CFS가 갖고 있는 상용 핵융합로 'SPARC'에 AI 제어 기술을 이식하기로 했습니다. 실험실의 과학적 호기심이 실제 전기를 생산하는 산업적 현실로 넘어가는 결정적 순간이었습니다.

허사비스에게 이 프로젝트는 '지능'이 어떻게 '물리적 세계'의 한계를 돌파할 수 있는지를 보여주는 완벽한 예시입니다. 인간의 반사 신경으로는 불가능한 0.0001초의 제어를 AI는 해냅니다. 이것은 단순히 전기를 만드는 기술이 아니라, 인류가 에너지의 결핍에서 벗어나 '근본적 풍요(Radical Abundance)'로 나아가는 열쇠입니다. 허사비스는 종종 이렇게 말합니다.

"우리가 기후 변화를 해결하고 우주로 나아가려면 무한하고 깨끗한 에너지가 필요합니다. AI가 그 문을 여는 열쇠가 될 것입니다." 자율 과학자(AI Scientist)의 가능성 1) 잠들지 않는 실험실: 2026년, 과학의 속도가 달라진다 허사비스의 사무실 한편에는 늘 위대한 과학자들의 전기가 꽂혀 있습니다.

다윈, 아인슈타인, 튜링... 그는 그들이 평생을 바쳐 이룩한 발견의 과정에 경외감을 느끼면서도, 동시에 답답함을 느꼈습니다. "인간 과학자는 잠을 자야 하고, 밥을 먹어야 하며, 편견에 사로잡히기도 한다. 만약 지치지 않고, 편견 없이, 모든 논문을 다 읽은 과학자가 있다면 어떨까?" 이 질문은

2025년, 구글 딥마인드가 발표한 'AI 코사이언티스트(AI Co-Scientist)'로 구체화되었습니다. 이것은 챗GPT처럼 단순히 질문에 대답해 주는 비서가 아닙니다. 연구자가 "이 단백질이 암세포와 어떻게 상호작용하는지 알아봐 줘"라고 목표를 던져주면, AI 코사이언티스트는 스스로 수만 건의 관련 논문을 검색하고, 가설을 세우고, 그 가설을 검증하기 위한 실험을 설계합니다. 허사비스의 비전은 여기서 멈추지 않습니다.

그는 2026년, 영국에 세계 최초의 '완전 자동화 AI 과학 실험실(Automated AI Science Lab)'을 개소할 계획을 발표했습니다. 이 실험실에는 흰 가운을 입은 사람이 없습니다. 대신 로봇 팔들이 실 새 없이 피펫을 움직이고, 물질을 합성하고, 현미경으로 결과를 분석합니다.

공상과학 영화의 한 장면과 같습니다. '제미니(Gemini)' 모델이 두뇌 역할을 하여 "새로운 초전도체 후보 물질을 합성해 보자"라고 명령을 내리면, 로봇 팔들이 즉시 수 백 가지의 배합 비율로 실험을 수행합니다. 실험 결과가 나오면 AI는 실시간으로 데이터를 분석해 "실패. 배합 비율을 0.5% 수정해서 재시도"라는 판단을 내립니다. 인간 연구자가 일주일 동안 할 실험을

이곳에서는 하루, 아니 몇 시간 만에 해치웁니다. 이 프로젝트는 영국 정부와의 50억 파운드 규모 파트너십의 일환으로, 차세대 배터리 리소재, 고효율 태양전지, 그리고 신약 후보 물질을 발굴하는데 초점을 맞추고 있습니다.

특히 허사비스가 설립한 신약 개발 자회사 '아이소모픽 랩스(Isomorphic Labs)'는 이 자동화 시스템을 통해 '디지털 생물학'의 시대를 앞당기고 있습니다. 이 지점에서 허사비스는 묘한 긴장을 마주합니다. "그렇다면 인간 과학자는 무엇을 해야 하는가?"

그는 AI가 인간을 대체하는 것이 아니라, 인간을 '지휘자'로 승격시킬 것이라고 주장합니다. "AI는 실험과 데이터 분석이라는 고된 노동을 맡고, 인간은 어떤 질문을 던질 것인지, 무엇이 가치 있는 연구인지를 결정하는 창의적 역할에 집중하게 될 것입니다." 2026년의 자동화 실험실은 허사비스가 12살 때 체스판 앞에서 했던 생각, "이 지능을 더 나은 곳에 쓸 수 있을까?"라는 질문에 대한 가장 거대하고 실질적인 대답입니다.

알파고가 바둑의 정석을 깬듯, 이제 AI 과학자는 인류가 수백 년간 쌓아온 '발견의 속도'라는 한계를 깨뜨리려 하고 있습니다. 이는 과학적 발견의 민주화이자, 인류 지식의 총량이 기하급수적으로 폭발하는 '특이점'의 시작일지도 모릅니다. 핵융합 토카막 반응로 내부

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

2016년 NHS 로열 프리 병원과의 데이터 협약 환자 동의 없는 160만 건 의료 데이터 접근과 개인정보 우려 2016년 런던의 차가운 공기 속에서 데미스 허사비스와 그의 오랜 친구이자 딥마인드의 공동 창립자인 무스타파 슬레이만은 예상치 못한 혹독한 시련을 마주하고 있었습니다. 바둑 판 위에서 인공지능이 거둔 승리의 환호성이 채 가 라앉기도 전에, 현실 세계의 복잡하고 민감한 문제가 그들의 발목을 잡았습니다. 그것은 기술적인 난제가 아니었습니다.

바로 '신뢰'와 '윤리'라는, 코드로 해결할 수 없는 인간 사회의 가장 근본적인 가치에 관한 문제였습니다. 이야기는 딥마인드의 야심 찬 의료 프로젝트인 '딥마인드 헬스(DeepMind Health)'의 출범과 함께 시작되었습니다. 허사비스와 슬레이만은 인공지능이 가장 먼저 혁신해야 할 분야로 의료를 꼽았습니다.

그들의 논리는 명쾌했습니다. 딥마인드의 기술이 단백질 구조를 예측하고 복잡한 게임을 풀 수 있다면, 환자의 생명을 구하는 일에도 쓰여야 한다는 것이었습니다. 그 첫 번째 파트너는 런던 북부에 위치한 로열 프리 병원(Royal Free Hospital)이었습니다.

하지만 이 목표는 시작부터 삐걱거렸습니다. 영국의 저명한 과학 잡지인 '뉴 사이언티스트(New Scientist)'가 딥마인드와 로열 프리 병원 사이의 데이터 공유 협약에 대해 의문을 제기하는 탐사 보도를 내놓았기 때문입니다. 보도의 내용은 충격적이었습니다.

딥마인드가 앱 개발을 위해 로열 프리 병원으로부터 160만 건에 달하는 환자 의료 기록에 접근할 권한을 얻었다는 것이었습니다. 여기에는 현재 입원 중인 환자의 데이터뿐만 아니라, 지난 5년 동안의 기록까지 포함되어 있었습니다. 더 큰 문제는 이 데이터의 범위였습니다.

환자의 이름과 생년월일, 주소는 물론이고, 혈액 검사 결과, 병력 기록, 방사선 영상 등 민감한 의료 정보가 망라되어 있었습니다. 심지어 HIV 보균 여부나 약물 과다 복용, 낙태 기록과 같은 극도로 사적인 정보까지 포함될 수 있다는 사실이 알려지자 영국 사회는 발칵 뒤집혔습니다. 환자들은

자신들의 가장 내밀한 정보가 거대 기술 기업인 구글의 자회사 손에 넘어갔다는 사실을 전혀 모르고 있었습니다. "우리는 동의한 적이 없다"는 항의가 빗발쳤습니다. 허사비스에게 이것은 뼈아픈 실책이었습니다.

그는 항상 "인공지능은 인류를 위해 쓰여야 한다"고 강조해 왔지만, 정작 그 인류를 구성하는 개개인의 권리를 간과했다는 비판을 피할 수 없었습니다. 딥마인드 측은 억울함을 호소했습니다. 그들은 이 데이터가 연구용이나 상업적 목적이 아니라, 오로지 환자 진료를 돕기 위한 앱을 개발하고 테스트하는 데만 사용된다고 주장했습니다.

또한 데이터는 암호화 되어 관리되며 구글의 다른 데이터와 결합되지 않는다고 해명했습니다. 대중의 시선은 싸늘했습니다. 정보 불균형이 초래한 공포였습니다.

사람들은 딥마인드의 알고리즘이 내 의료 기록을 학습해 나중에 보험료를 인상하거나 취업에 불이익을 주는 데 사용될지 모른다는 막연하지만 구체적인 두려움을 느꼈습니다. 영국 정보위원회(ICO)가 조사에 착수했고, 결국 2017년 7월, 로열 프리 병원이 환자들에게 데이터 공유 사실을 충분히 알리지 않았으므로 영국 데이터 보호법을 위반했다는 판결을 내렸습니다. 이 사건은 허사비스에게 법적 문제 이상의 의미를 가졌습니다.

그것은 실리콘밸리 의 방식, 즉 "빠르게 움직이고 일단 저질러라(Move fast and break things)"라는 격언 이 사람의 생명과 존엄을 다루는 의료 현장에서는 통용될 수 없음을 보여주는 명징한 사례였습니다. 기술적 선의만으로는 충분하지 않았습니다. 절차적 정당성과 투명성이 담보되지 않은 혁신은 사회적 저항에 부딪힐 수밖에 없다는 것을 딥마인드는 비싼 수업료를 치르고 배웠습니다.

허사비스는 깊은 고뇌에 빠졌습니다. 그는 기술자로서 효율성을 추구 했지만, 경영자 로서 사회적 책임을 져야했습니다. 160만 건의 데이터는 단순한 숫자 가 아니었습니다.

그 하나하나가 고통받고, 치유받고, 때로는 죽음을 맞이한 인간의 삶이었습니다. 이후 딥마인드는 윤리 위원회를 재정비하고, 데이터 접근 방식을 근본 적으로 재검토하게됩니다. 이는 훗날 알파폴드와 같은 거대 과학 프로젝트를 진행할 때 데이터를 다루는 그들의 신중한 태 도에 결정적인 영향을 미치게됩니다.

딥마인드 헬스의 스트림스(Streams) 앱과 급성 신장 손상 탐지 논란의 중심에 섰던 '스트림스(Streams)' 앱은 아이러니하게도 인공지능 기술이 거의 들어 가지 않은, 매우 실용적이고 직관적인 소프트웨어였습니다. 하지만 이 앱의 탄생 배경과 기능은 당시 영국 의료 시스템(NHS)이 처한 현실과 딥마인드가 추구했던 '현실 문제 해결' 의 접 점을 정확히 보여줍니다. 이야기는 로열 프리 병원의 신장 내과 전문의들이 겪고 있던 답답한 현실에서 시작 됩니 급성 신장 손상(AKI, Acute Kidney Injury)은 병원에 입원한 환자들에게서 흔히 발생하 는 합병증으로, 신장 기능이 갑작스럽게 저하되는 상태를 말합니다. AKI는 조 기에 발견하 면 수액 처방이나 약물 조절만으로도 쉽게 치료할 수 있지만, 발견이 늦 어지면 영구적인 신장 손상을 입거나 사망에 이를 수도 있는 치명적인 질환입니다.

영 국에서만 매년 4만 명 이상의 환자가 AKI로 인해 피할 수 있었던 죽음을 맞이한다는 통계도 있었습니다. 문제는 '정보의 속도'였습니다. 2016년 당시 NHS의 많은 병원은 여전히 종이 차트 와 호출 기(삐삐), 팩스에 의존하고 있었습니다.

환자의 혈액 검사 결과가 나오면 검사 실에서는 이 상 징후를 발견하지만, 이 정보가 담당 의사에게 전달되기까지는 몇 시 간, 길게는 하루가 걸리기도했습니다. 의사들은 수많은 환자의 차트를 일일이 뒤져봐 야 했고, 그사이 환자의 신장은 소리 없이 망가져 갔습니다. 무스타파 슬레이만은 이 비효율을 기술로 해결하고자했습니다.

스트림스 앱의 원 리는 간 단했습니다. 병원 시스템과 연동하여 환자의 혈액 검사 결과에서 크레아티닌 수치가 급격 히 상승하는 등 AKI 징후가 포착되면, 즉시 담당 의료진의 스마트폰으로 알림을 보내는 것입니다. 마치 주식 시장의 급등락 정보를 실시간으로 받아보듯이, 의사들은 환자의 위기 상황을 즉각적으로 인지할 수 있게되었습니다.

스트림스는 환자의 과거 병력, 현재 투약 정보, 활력 징후 등을 보기 편한 인터페이 스로 정 리해 보여주었습니다. 의사들은 더 이상 종이 차트 더미를 뒤질 필요 없이 스 마트폰 화면을 통해 환자의 상태를 한눈에 파악하고 처방을 내릴 수 있었습니다. 딥 마인드 팀은 이 앱을 개발하기 위해 의사와 간호사들을 그림자처럼 따라다니며 그들 의 동선과 업무 방식을 철저히 분석했습니다. 사용자 경험(UX) 디자인이 의료 현장의 긴박함과 만나는 순간이었습니다.

기술적으로 스트림스는 딥러닝이나 거대 언어 모델과 같은 최첨단 AI가 아니었습니다. 오히려 기존의 의료 알고리즘을 모바일 환경에 맞게 최적화한 데이터 시각화 도구에 가까웠습니다. 허서비스와 슬레이만은 여기서 중요한 결정을 내렸습니다.

"화려한 AI 기술을 과시하는 것보다, 지금 당장 현장에서 사람을 살리는 도구를 만드는 것이 중요하다"는 실용 주의적 접근이었습니다. 그들은 AI를 적용하기에 앞서, 데이터가 흐르는 파이프라인부터 뜯어야 한다고 판단했습니다. 스트림스의 성과는 즉각적이었습니다.

로열 프리 병원에서의 시범 운영 결과, AKI 환자를 확인하는 데 걸리는 시간이 몇 시간에서 몇 초 단위로 단축되었습니다. 의료진은 하루 평균 2시간의 행정 업무 시간을 절약하여 환자를 돌보는 데 더 많은 시간을 쓸 수 있었습니다. 응급 상황 대응 속도가 빨라지면서 환자들의 생존율 향상에도 기여할 수 있다는 가능성을 보여주었습니다. 현장의 간호사들과 의사들은 이 앱을 "의료 시스템의 혁명"이라 부르며 환영했습니다.

그러나 이 빛나는 성과는 앞서 언급한 데이터 프라이버시 논란이라는 그림자에 가려졌습니다. 딥마인드는 앱이 제대로 작동하는지 검증하기 위해 AKI 환자뿐만 아니라 모든 환자의 데이터가 필요했다고 항변했지만, 대중과 규제 당국은 "빈대 잡으려다 초가삼간 태운 격"으로 개인정보를 침해했다고 보았습니다. 결국 스트림스 프로젝트는 기술이 아무리 유용하고 선한 의도를 가졌더라도, 사회적 합의와 법적 테두리를 벗어나면 지속 가능하지 않다는 뼈아픈 교훈을 남겼습니다.

2019년, 딥마인드 헬스 팀이 구글 헬스로 통합되면서 스트림스 앱의 운영 주체는 구글로 넘어갔습니다. 이는 또 한 번의 논란을 낳았습니다. "딥마인드의 의료 데이터는 구글의 광고 비즈니스와 영원히 분리될 것"이라던 초기 약속이 깨진 것처럼 보였기 때문입니다.

스트림스는 의료 현장에서 정보 기술이 얼마나 큰 힘을 발휘할 수 있는지를 증명한 동시에, 빅테크 기업이 공공 의료 영역에 진입할 때 겪게 되는 불신과 갈등을 상징하는 이정표로 남게되었습니다. 허서비스에게 스트림스는 성공인 동시에 상처였으며, 앞으로 다가올 AI 시대를 위해 반드시 풀어야 할 숙제를 던져준 예방주사였습니다. 의료 AI의 성과와 한계 안과 질환 진단 AI: Nature Medicine 논문과 무어필즈 안과병원 협력 2016년 여름, 데미스 허서비스는 런던 올드 스트리트에 위치한 세계적인 안과 전문 병원인 무어필즈 안과병원(Moor elds Eye Hospital)의 문을 두드렸습니다.

로열 프리 병원에서의 데이터 논란으로 호된 신고식을 치른 직후였지만, 그는 멈추지 않았습니다. 이번에는 달랐습니다. 더 조심스럽고, 더 정교하며, 무엇보다 더 투명한 방식으로 접근했습니다.

그의 파트너는 무어필즈의 젊고 열정적인 안과 의사 피어스 킨(Pearse Keane)이었습니다. 피어스 킨 박사는 당시 절망적인 상황과 싸우고 있었습니다. 고령화 사회로 접어들면서 노인성 황반변성(AMD)이나 당뇨병성 망막증 같은 안과 질환 환자가 폭증하고 있었습니다.

이를 진단하기 위한 빛 간섭 단층촬영(OCT, Optical Coherence Tomography) 기기는 병원에 있었지만, 쏟아져 나오는 수천 장의 3차원 스캔 이미지를 판독할 전문의의 눈은 턱없이 부족했습니다. 환자들은 진단을 받기 위해 몇 주, 몇 달을 기다려야 했고, 그 사이 시력을 잃는 경우도 허다했습니다. 킨 박사는 허서비스에게 말했습니다.

"우리는 데이터의 홍수 속에 빠져 죽어가고 있습니다. 당신의 AI가 우리를 구해줄 수 있습니까?" 허서비스는 이 제안을 딥마인드의 미션인 "지능을 풀어 세상을 돕는다"를 실현할 완벽한 기회로 보았습니다.

하지만 이번에는 데이터 처리에 만전을 기했습니다. 딥마인드는 무어 필즈 병원으로부터 약 1만 5천 명의 환자에게서 얻은 14,884개의 익명화된 OCT 스캔 데이터를 제공받았습니다. 로열 프리 사태의 교훈을 바탕으로, 데이터의 사용 목적과 범위를 명확히 하고 철저한 비식별화 과정을 거쳤습니다.

딥마인드 연구팀은 이 데이터를 바탕으로 두 단계의 인공지능 모델을 설계했습니다. 첫 번째 신경망은 OCT 이미지를 분석하여 눈의 조직과 병변을 정밀하게 분할 (Segmentation)하는 역할을 했습니다. 지도를 그리듯 망막의 층과 이상 부위를 픽셀 단위로 구분해내는 것입니다.

두 번째 신경망은 이 지도를 바탕으로 질병의 종류를 분류하고 긴급도를 판단했습니다. 2018년 8월, 딥마인드와 무어필즈 연구팀의 공동 연구 결과가 세계적인 의학 저널인 《네 이처 메디신(Nature Medicine)》에 게재되었습니다. AI 모델은 50가지 이상의 안과 질환을 진단하는데 있어 무어필즈 병원의 최고 수준 전문의들과 대등하거나 더 뛰어난 성능을 보였습니다.

정확도는 94%를 상회했고, 무엇보다 진단 오류율 (오진율)은 5.5%로, 인간 전문가 8명의 평균 오진율인 6.7%보다 낮았습니다. 특히 위급한 환자를 선별하여 우선순위를 정하는 triage(중증도 분류) 능력에서 탁월한 성능을 발휘했습니다.

이 연구가 학계의 찬사를 받은 진짜 이유는 높은 정확도 때문만이 아니었습니다. 허 서비스와 연구팀은 의료 AI의 고질적인 문제인 '블랙박스(Black Box)' 문제를 해결하기 위해 노력했습니다. 기존의 딥러닝 모델은 결과만 던져줄 뿐 "왜" 그런 판단을 내렸는지 설명하지 못해 의사들의 불신을 샀습니다.

하지만 딥마인드의 모델은 달랐습니다. AI는 진단 결과와 함께 OCT 스캔 이미지 위에 병변 부위를 색깔별로 표시하여 보여주었습니다. 의사는 AI가 어느 부분을 보고 질병을 판단했는지 눈으로 확인할 수 있었고, 이를 통해 AI의 판단을 검증하고 신뢰할 수 있게 되었습니다.

이는 "설명 가능한 AI(XAI, Explainable AI)"가 의료 현장에 적용된 모범 사례였습니다. 딥마인드의 시스템은 의사를 대체하는 것이 아니라, 의사에게 강력한 '디지털 현미경'을 쥐여 주는 도구였습니다. 의사는 AI의 1차 소견을 바탕으로 더 빠르고 정확하게 최종 진단을 내릴 수 있게 되었습니다. 이 프로젝트는 기술 기업과 의료 기관이 어떻게 협력해야 하는지를 보여주는 교과서적인 모델이 되었습니다.

허서비스는 이를 통해 딥마인드가 단순히 바둑만 잘 두는 회사가 아니라, 실제 과학과 의학의 난제를 해결할 수 있는 능력을 갖췄음을 전 세계에 증명해 보였습니다. 무어필즈 프로젝트는 딥마인드가 잃어버린 신뢰를 회복하고, AI가 인류의 건강에 기여할 수 있다는 희망을 다시금 쏘아 올린 신호탄이었습니다. 의료 데이터의 윤리적 활용을 둘러싼 교훈 무어필즈 안과병원과의 성공적인 협력과 로열 프리 병원에서의 뼈아픈 실수는 데미스 허 서비스와 딥마인드, 나아가 전세계 AI 연구자들에게 의료 데이터의 윤리에 관한 심오한 교훈을 남겼습니다.

그것은 "데이터는 단순한 자원이 아니라 신뢰의 증표"라는 깨달음이었습니다. 첫 번째 교훈은 '투명성'과 '동의'의 가치입니다. 로열 프리 사건에서 딥마인드는 법적으로는 NHS의 정보

거버넌스 담당자(Caldicott Guardian)의 승인을 받았다고 주장했지만, 환자 개인의 명시적 동의가 결여된 '포괄적 데이터 접근'은 사회적 용인을 받지 못했습니다.

이는 2018년 유럽연합의 일반 개인정보 보호법(GDPR) 시행과 맞물려, 데이터 주권에 대한 인식을 완전히 바꾸어 놓았습니다. 이후 딥마인드는 모든 의료 프로젝트에서 환자 참여 그룹을 만들고, 데이터가 어떻게 사용 되는지 일반 대중의 언어로 설명하는 데 막대한 노력을 기울였습니다. 기술이

아무리 훌륭 해도 그 과정이 불투명하면 사회는 그 기술을 거부한다는 사실을 뼈저리게 느낀 것입니다. 두 번째 교훈은 '목적의 명확성'입니다. 스트림스 앱 개발 당시 딥마인드는 급성 신장 손상 탐지라는 특정 목적을 위해 광범위한 전체 의료 기록을 가져왔습니다.

이는 '데이터 최소화 원칙(Data Minimization Principle)', 목적 달성에 필요한 최소한의 데이터만 사용해야 한다는 원칙,에 위배된다는 비판을 받았습니다. 반면 무어필즈 프로젝트에서는 안구 스캔 이미지와 그와 관련된 직접적인 진단 데이터로 범위를 한정했고, 철저한 익명화를 거쳤습니다. 이는 AI 학습을 위한 데이터 수집이 '다더익선 (The more, the better)'이 아니라 '정밀 타격(Precision targeting)'이어야 함을 시사했습니다.

세 번째는 '공공의 이익'과 '사기업의 이익' 사이의 긴장 관계입니다. NHS는 영국 국민들의 세금으로 운영되는 공공 의료 시스템입니다. 이곳의 데이터는 국민들의 공공 자산입니다.

구글이라는 세계 최대의 영리 기업 자회사가 이 데이터를 무료로 활용해 알고리즘을 개발 하고, 나중에 그 기술을 다시 NHS에 판매하거나 다른 나라에 팔아 수익을 낸다면 그것은 정당한가? 이 질문은 여전히 현재 진행형입니다. 허사비스는 이를 해결하기 위해 딥마인드 헬스의 기술을 오픈소스화하거나, 공공에 이익을 환원 하는 모델을 고민해야했습니다. 이는 AI 기업이 단순히 기술 공급자가 아니라 사회적 계약의 당사자가 되어야 함을 의미했습니다.

마지막으로, '인간 중심의 기술 구현'입니다. 스트림스와 안과 진단 AI 모두, 결국 최종 결정권자는 의사였습니다. AI는 진단을 내리는 판사가 아니라, 증거를 수집해오는 유능한 수사관 역할에 머물러야했습니다.

무어필즈 프로젝트에서 보여준 '설명 가능성'은 기계가 인간을 소외시키지 않고 협력하기 위한 필수 조건이었습니다. 딥마인드는 기술적 정확도 (Accuracy)만큼이나 임상적 유용성(Utility)과 설명 가능성 (Explainability)이 중요하다는 것을 증명했습니다. 허사비스에게 의료 AI는 체스판이나 바둑판과는 전혀 다른 차원의 게임이었습니다.

바둑 판은 19줄의 선과 명확한 규칙, 그리고 승패만이 존재하는 닫힌 세계(Closed System)였지만 병원은 불완전한 정보, 환자의 고통, 프라이버시, 법적 규제가 얽히고설켜 열린 세계(Open System)였습니다. 딥마인드는 이 열린 세계에서 수많은 시행착오를 겪으며, "지능을 쫓다"는 것이 단순히 알고리즘의 성능을 높이는 것이 아

니라, 그 지능이 사회 속에 안전하게 착륙하도록 돕는 것까지 포함한다는 사실을 배웠습니다. 이 교훈들은 훗날 딥마인드가 생물학의 난제인 단백질 구조 예측(알파폴드)으로 나아갈 때, 그리고 더 강력한 범용 인공지능(AGI)을 논할 때 그들의 윤리적 나침반이되었습니다. "가장 강력한 기술은 가장 높은 수준의 책임을 요구한다."

이것이 로열 프리 의 병동과 무어 필즈의 진료실에서 허사비스가 얻은, 노벨상만큼이나 값진 발견이었습니다. NHS Royal Free London 병원

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

사람의 발표에 집중되었습니다. 구글의 CEO 순다르 피차이(Sundar Pichai)가 회사 블로그에 올린 글은 단순한 조직 개편 소식이 아니었습니다. 그것은 인공지능(AI) 역 사상 가장 거대한 두 개의 지적 집단이 하나로 합쳐진다는 선언이었습니다.

"우리는 구글 브레인(Google Brain)과 딥마인드(DeepMind)를 통합하여 '구글 딥 마인드 (Google DeepMind)'라는 새로운 조직을 출범합니다." 이 통합은 '샷건 웨딩 (Shotgun Marriage, 어쩔 수 없이 하는 급한 결혼)'이라는 외부의 비평과 '필연적 진화'라는 내부의 설명 사이 어딘가에 있었습니다. 오픈AI(OpenAI)가 챗 GPT(ChatGPT)로 세상을 뒤흔든 지 불과 5개월 만의 일이었습니다.

구글은 위기감을 느끼고 있었고, 런던의 딥마인드와 캘리포니아의 구글 브레인은 오랫동안 서로를 존중하면서도 경쟁하는 '라이벌 형제' 관계였습니다. 이 두 거인을 하나로 묶어 이끌 수 있는 리더는 단 한 명, 데미스 허사비스뿐이었습니다. 딥마인드와 구글 브레인의 합병 두 조직의 합병은 서로 다른 두 '문명'의 만남이었습니다.

2010년 런던에서 시작된 딥마인드는 "지능을 본다"는 단일하고 거대한 미션 아래 일사불란하게 움직이는 '아폴로 프로젝트' 같은 조직이었습니다. 반면, 구글 브레인은 학문적 자유와 개방성을 중시하며, 연구자들이 각자의 호기심을 따라 다양한 프로젝트를 수행하는 대학 연구실 같은 분위기였습니다. 허사비스에게 주어진 과제는 막중했습니다.

2,000명이 넘는 세계 최고의 천재들을 하나의 방향으로 정렬시켜야했습니다. 그는 이 과정을 "연구소의 순수성을 잃지 않으면서 스타트업의 속도를 되찾는 것"이라고 정의했습니다. 합병 초기, 두 조직 간에는 미묘한 긴장이 흘렀습니다.

딥마인드의 연구자들은 구글의 상업적 요구가 연구의 본질을 해칠까 우려했고, 구글 브레인의 연구자들은 딥마인드의 중앙집권적 통제 방식에 거부감을 느꼈습니다.

하지만 허사비스는 이 긴장을 '생산적 마찰'로 전환시켰습니다. 그는 연구원들에게 이렇게 설득했습니다. "우리가 가진 자원은 세계 최고입니다.

구글의 컴퓨팅 파워 (TPU)와 여러분의 두뇌가 합쳐지면, 우리는 그 어떤 조직보다 빠르게 AGI에 도달할 수 있습니다." 실제로 제미니(Gemini) 프로젝트는 이 융합의 첫 번째 시험대였습니다. 딥마인드의 강화 학습(Reinforcement Learning) 노하우와 구글 브레인의 트랜스포머(Transformer) 아키텍처 기술이 결합하여 탄생한 제미니는, 두 조직이 하나가 되었을 때 어떤 시너지를 낼 수 있는지를 증명해 보였습니다.

엔진 룸(Engine Room) 역할 과거 딥마인드는 구글의 모회사 알파벳(Alphabet) 산하의 '기타 베팅(Other Bets)' 중 하나로 분류되기도 했습니다. 즉, 돈을 벌어오는 조직 이라기보다는 미래를 위한 투자처였습니다. 하지만 2023년 이후, 딥마인드의 위상은 완전히 달라졌습니다.

허사비스는 이제 딥마인드를 구글이라는 거대한 함선의 심장, 즉 '엔진 룸(Engine Room)'이라고 표현합니다. "우리는 구글의 모든 제품에 지능을 공급합니다." 이 변화는 연구소의 정체성에 깊은 고민을 안겨주었습니다.

과거에는 논문을 쓰고 학술적 성과를 내는 것이 최우선 목표였다면, 이제는 전 세계 수십 억 명이 사용하는 검색, 유튜브, 안드로이드에 들어갈 실용적인 AI 모델을 만들 어야 한다는 압박이 더해졌습니다. 제품(Product) 부서의 요구와 플랫폼(Platform) 의 안정성, 그리고 딥마인드 고유의 연구(Research) 목표 사이에서 충돌은 피할 수 없었습니다. 허사비스는 이 충돌을 조정하기 위해 '이중 트랙' 전략을 사용했습니다.

한편으로는 제미나 이와 같이 상용화에 직결되는 거대 언어 모델(LLM) 개발에 전사적 역량을 집중시키면서 도, 다른 한편으로는 알파폴드(AlphaFold)와 같은 기초 과학 연구를 위한 별도의 공간을 철저히 보호했습니다. 그는 "상업적 성공이 기초 과학 연구를 위한 자금을 대고, 기초 과학의 발견이 다시 상업적 모델의 성능을 획기적으로 높이는 선순환"을 강조하며 연구원들의 동요를 잠재웠습니다.

순다르 피차이와의 관계 이 거대한 전환의 중심에는 허사비스와 순다르 피차이의 긴밀한 파트너십이 있습니다. 합병 이후 허사비스는 런던과 캘리포니아를 오가며 피 차이와 거의 매일 대화를 나눕니다. 과거에는 분기별로 성과를 보고하는 관계였다면, 이제는 전선의 최전방 지휘관과 사령관처럼 실시간으로 전황을 논의하는 관계가 되었습니다. 피차이는 허사비스에게 기술적 전권을 위임했습니다.

구글 내부에서도 파격적인 대우였습니다. 구글의 창업자 래리 페이지(Larry Page)와 세르게이 브린(Sergey Brin) 이 딥마인드를 인수할 때 약속했던 '연구의 자율성'은, 위기의 순간에 '전략적 동반자 관계'로 격상되었습니다. 허사비스는 피차이에게 단순한 기술적 조언을 넘어, AI가 나아가야 할 윤리적 방향 과 속도 에 대해서도 직언을 아끼지 않습니다.

"우리는 이 기술을 서둘러 내놓아서 안됩니다. 하지만 늦어서도 안됩니다." 이 모순적 인 명제를 해결하기 위해 두 사람은 안전(Safety)을 타협하지 않는 선에서 속도(Speed)를 높 이는 묘수를 매일 밤 고민합니다.

피차이가 구글이라는 거대한 제국의 정치와 경영을 방어 해 주는 방패라면, 허사비스는 그 안에서 가장 날카로운 창을 깎 는 장인인 셈입니다. 허사비스가 이끄는 구글 딥마인드는 인류 역사상 가장 강력한 기술을 통제하고 발전시키는 거대한 실험실이자, 동시에 치열한 비즈니스의 전장입니다. 허사비스는 2,000명의 천재 들을 이끌고 이 불확실한 전장의 안개를 헤쳐 나가고 있습니다.

"밤에 두 번째 근무" 런던의 밤이 깊어지고 도시의 소음이 잦아들 무렵, 데미스 허사비스의 '진짜' 하루가 시작됩니다. 낮 동안 그는 수천 명의 직원을 거느린 경영자로서 회의를 주재하고, 결재 서류에 서명하며, 복잡한 조직 문제를 해결합니다. 하지만 가족과 저녁을 먹고 아이들을 재운 뒤, 새벽 1시가 되면 그는 다시 '연구자'로 돌아옵니다. 새벽 3시의 "야간 교대(Second Shift)": 창의성의 시간대 허사비스는 자신의 이 루 틈을 "두 번째 근무(Second Shift)"라고 부릅니다. 보통 새벽 1시에 시작해 새벽 4시까지 이어지는 이 시간은 그에게 있어 신성불가침의 영역입니다. 그는 포춘(Fortune)지와의 인터뷰에서 이렇게 말했습니다.

"저는 새벽 1시쯤에야 비로소 살아납니다(I come alive at about 1 a.m.). 낮에는 창의적이기 어렵습니다. 끊임없이 방해받으니까요." 이 시간 동안 그는 이메일을 확인하거나 업무 지시를 내리지 않습니다.

대신 그는 가장 어려운 난제들과 씨름합니다. 알파고의 알고리즘에 막혔던 부분을 뚫어내거나, 단백질 구조 예측의 새로운 접근법을 구상하거나, 혹은 AGI의 정의에 대해 철학적으로 사유하는 것이 이 시간의 주된 활동입니다. 고요함 속에서 오직 생각에만 몰두하는 이 습관은 4살 때 체스판 앞에서 수 시간씩 움직이지 않고 생각하던 '신동'의 모습과 겹쳐집니다. 많은 CEO들이 '미라클 모닝'이라 불리는 새벽 기상을 실천하며 하루를 일찍 시작하는 것과 달리, 허사비스는 밤을 선택했습니다.

이것은 그가 여전히 현역 연구자로서의 정체성을 강하게 붙들고 있음을 보여줍니다. 이 심야의 시간을 통해 낮 동안 흩어졌던 생각의 조각들을 맞추고, 경영의 소음 속에 파묻혔던 과학적 직관을 다시 날카롭게 베풁니다. 그에게 잠은 6시간 정도면 충분합니다. 아니, 충분하다고 믿으려 노력합니다. 새벽 4시에 잠들어 오전 10시쯤 일어나는 이 올빼미형 루틴은 지난 10년 넘게 딥마인드의 혁신을 지탱해 온 숨은 동력이었습니다.

2,000편 이상의 연구 논문, h-index 83, 15만 회 이상 인용의 학술적 성과 이러한 '야간 교대'의 결과물은 놀라운 학술적 수치로 증명됩니다. 일반적으로 거대 기술 기업의 CEO가 되면 연구 논문 집필에서는 손을 떼기 마련입니다. 경영과 연구, 두 마리 토끼를 다 잡는 것은 물리적으로 불가능해 보이기 때문입니다.

허사비스는 예외입니다. 2024년 말 기준으로 데미스 허사비스의 이름이 올라간 연구 논문은 2,000편이 넘습니다(딥마인드 전체 성과 포함, 개인 저자 참여 논문은 150편 내외). 그의 연구 영향력을 나타내는 지표인 h-index는 83에 달하며, 그의 논문들은 전 세계 과학자들에 의해 15만 회 이상 인용되었습니다. 노벨상 수상자들 사이에서도 최상위권에 속하는 수치입니다.

2021년 네이처 (Nature)에 발표된 알파폴드 논문은 과학 역사상 가장 많이 인용된 논문 중 하나로 기록되었습니다.

수치가 말해주는 것은 분명합니다. 단순히 과학자들을 고용해서 관리하는 경영자가 아닙니다. 그는 연구의 방향을 직접 설정하고, 핵심 아이디어를 제시하며, 논문의 문장 하나하나를 다듬는 '리드 저자(Lead Author)'입니다. 그의 학문적 배경은 컴퓨터 과학에만 머물지 않습니다.

케임브리지 대학에서의 컴퓨터 과학, 유니버시티 칼리지 런던(UCL)에서의 인지 신경과학 박사 학위, 그리고 MIT와 하버드에서의 포스트닥터 연구 경험은 그가 AI를 단순한 코딩의 문제가 아니라 '지능의 본질을 탐구하는 과학'으로 접근하게 만들었습니다. 그는 기억과 상상력의 관계를 밝혀낸 신경과학 자이자, 강화학습의 새로운 지평을 연 컴퓨터 공학자입니다. 그의 h-index 83은 치열한 기업 경쟁 속에서도 그가 결코 놓지 않았던 '과학적 진리 탐구'라는 초심의 증거입니다.

경영자로서의 화려한 스포트라이트 뒤에는, 매일 새벽 3시 책상 스탠드 불빛 아래서 논문과 씨름하는 고독한 과학자의 시간이 겹겹이 쌓여 있습니다. "지능을 풀겠다"는 그의 미션은 구호가 아니라, 매일 밤 실천되는 지난한 노동의 결과물입니다. 구글 런던 사무실

인공지능 전문가 김경진 변호사

AI 법정정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

에 위치한 구글 딥마인드 본사에는 꺼지지 않는 불빛이 가득했습니다. 평소 바둑판이나 단백질 구조가 떠 있던 모니터들은 이제 거대한 언어 모델의 파라미터들로 채워져 있었습니다. 데미스 허사비스는 창밖으로 런던의 야경을 내려다보며 깊은 생각에 잠겨 있었습니다.

그는 평생을 과학적 난제를 해결하는 데 바쳐왔지만, 지금 그가 마주한 것은 과학의 문제가 아니라 생존의 문제였습니다. 챗GPT의 등장으로 올린 '코드 레드(Code Red)'는 구글이라는 거함 전체를 흔들었고, 순수 연구를 지향하던 딥마인드조차 제품 개발이라는 최전선으로 떠밀려 나온 상황이었습니다. 허사비스는 이 위기를 기회로 바꾸어야 했습니다.

단순히 챗봇을 만드는 것이 아니라, 그가 꿈꾸던 범용 인공지능(AGI)으로 가는 징검다리를 놓아야 했습니다. 그 결과물이 바로 '제미나이(Gemini)'였습니다. 제미나이의 탄생은 구글 내부의 정치적, 기술적 장벽을 허무는 것에서 시작되었습니다.

오랫동안 라이벌 관계였던 구글 브레인(Google Brain) 팀과 딥마인드(DeepMind) 팀이 하나로 합쳐진 것은 유례를 찾기 어려운 일이었습니다. 제프 딘(Jeff Dean)과 데미스 허사비스, 이 두 거인이 손을 잡은 것은 단순히 인력을 합치는 것을 넘어 서로 다른 연구 문화를 융합하는 실험이었습니다. 허사비스는 이를 두고 "우리는 각자의 강점을 결합해 이전에 없던 새로운 종(species)을 탄생시켰다"고 회고합니다.

프로젝트명 '제미나이'는 쌍둥이자리라는 뜻처럼 두 조직의 결합을 상징하면서, 동시에 나사(NASA)가 달에 가기 위해 수행했던 제미니 프로젝트의 대담한 도전 정신을 계승한다는 의미를 담고 있었습니다. 제미니 1.0의 출시는 구글이 AI 경쟁의 링 위에 올랐음을 알리는 신호탄이었습니다. 하지만 허사비스가 진정으로 보여주고 싶었던 것은 단순한 텍스트 생성 능력이 아니었습니다. 그는 인간이 세상을 인식하는 방식, 즉 텍스트뿐만 아니라 이미지, 오디오

오, 비디오를 동시에 이해하고 처리하는 '네이티브 멀티모달(Native Multimodal)' 능력을 강조했습니다. 기존의 모델들이 텍스트만 배운 학생에게 억지로 그림을 가르치는 방식이었다면, 제미니는 태어날 때부터 눈과 귀와 입을 모두 가진 아이처럼 학습되었습니다. 이 차이는 결정적이었습니다.

제미니는 미묘한 뉘앙스가 담긴 비디오를 보고 그 안의 감정을 읽어내거나, 복잡한 물리학 논문의 그래프를 해석해 새로운 수식을 도출해내는 데 탁월한 능력을 보였습니다. 기술적 진화는 숨 가쁘게 이어졌습니다. 제미니 1.5 Pro의 등장은 '긴 문맥(Long Context)'의 혁명을 가져왔습니다.

100만 토큰, 나아가 200만 토큰을 한 번에 처리할 수 있다는 것은 AI에게 코끼리 같은 기억력을 부여한 것과 같았습니다. 이는 두꺼운 전공 서적 수백 권을 한순간에 읽고 그 안에서 특정 정보를 찾아내거나, 1시간짜리 영화를 보고 그 내용을 완벽하게 요약할 수 있음을 의미했습니다. 허사비스는 이 기술을 설명하며 "이제 AI는 정보의 파편을 보는 것이 아니라, 거대한 정보의 맥락 전체를 조망할 수 있게 되었습니다"라고 말했습니다.

이는 변호사가 수만 페이지의 소송 기록을 검토하거나, 과학자가 수십 년치 연구 데이터를 분석하는 방식을 근본적으로 바꾸어 놓았습니다. 제미니 3.0은 허사비스가 그토록 갈망하던 '추론(Reasoning)'과 '계획(Planning)' 능력에서 비약적인 도약을 이루어냈습니다. 단순히 다음에

을 단어를 확률적으로 예 측하는 '확률적 앵무새'를 넘어, 스스로 생각하고 논리적 단계를 거쳐 답을 찾아가는 시스템이 된 것입니 제미나이 3는 수학적 난제를 풀기 위해 스스로 검증 과정을 거치고, 코드를 작성하다 오류가 나면 스스로 디버깅하여 수정하는 모습을 보였습니다.

이 는 딥마인드가 알파고를 만들 때 사용했던 강화학습 기술이 거대 언어 모델과 결합하 여 만들어낸 시너지였습니다. 허사비스는 이를 보며 "우리는 이제야 비로소 생각하는 기계의 입구에 들어섰다"고 조용 히 자평했습니다. 이러한 기술적 성취는 시장의 반응으로 이어졌습니다.

제미나이 앱과 웹사이트의 월간 활 성 사용자(MAU)가 6억 5천만 명을 돌파했다는 수치는 단순한 통계 이상의 의미를 지닙니 전 세계 인구의 상당수가 구글 검색창 대신 제미나이에게 질문을 던지 고 있음을 보여 주는 지표입니다. 학생들은 과제를 위해, 프로그래머는 코딩을 위해, 작가는 영감을 얻기 위해 제미나이를 켵니다.

구글의 생태계 안에서 제미나이는 지메일, 구글 닥스, 드라이브와 유기적으로 연결 되어 사 용자의 데이터를 분석하고 업무를 자동화하는 개인 비서로 자리 잡았습니다. 이는 오픈AI 의 챗GPT가 선점했던 시장에 구글이 강력한 인프라와 생태계를 무기로 성공적으로 반격 했음을 시사합니다. 이 과정이 순탄하기만 했던 것은 아닙니다. 오픈AI와의 경쟁은 허사비스에게 끊임 없는 긴 장감을 요구했습니다.

샘 올트먼이 GPT-4o를 발표하며 음성 비서 시장을 공 략할 때, 구글 딥마인드는 그보다 더 자연스럽게 지연 시간이 없는 실시간 대화 모델을 내놓아야 한다는 압박에 시달렸습니다. 때로는 "구글이 너무 느리다"거나 "과거의 영광에 취해 있다"는 비 판도 쏟아졌습니다. 이미지 생성 기능에서 발생한 역사적 오 류 논란은 허사비스와 구글에 게 뼈아픈 교훈을 남겼습니다.

기술적 완벽함뿐만 아니 라, AI가 사회적 맥락과 윤리를 어 떻게 이해해야 하는지에 대한 깊은 철학적 고민을 안겨준 사건이었습니다. 허사비스는 이 에 대해 변명하기보다는 "우리는 배우고 있으 며, 더 책임감 있는 AI를 만들기 위해 끊임없 이 수정해 나갈 것입니다"라고 겸허히 인 정했습니다. 제미나이 시리즈는 '지능을 이해한다'는 미션이 추상적인 연구실을 벗어나 전 세계 80억 인 구의 손바닥 위에서 펼쳐지는 현실이 되는 과정이었습니다.

그는 새벽에 일어나 전 세계에서 올라오는 제미나이의 사용 로그를 확인하며, 자신 이 만든 지능이 누군가의 코딩을 돕고, 누군가의 시를 함께 쓰며, 누군가의 외로움을 달래주는 모 습을 목격합니다. 그것은 그가 어린 시절 체스판 위에서 꿈꾸었던, 인간 의 지성을 확장하 는 도구로서의 AI가 실현되는 순간이기도했습니다. 제미나이는 인 류가 더 똑똑해지기 위 해 만든, 그리고 허사비스가 40년간 준비해온 가장 거대하고 복잡한 도구입니다.

젬마(Gemma): 오픈소스 경량 모델 데미스 허사비스의 사무실 한켠에는 오래된 애 플 II 컴퓨터 사진이 놓여 있습니다. 그가 처 음 프로그래밍을 독학하며 디지털 세계의 창조자가되었던 그 시절의 기계입니다. 허사비 스는 종종 그 시절을 회상하며, 만약 그 당시 기술이 소수의 기업이나 연구소에만 독점되 어 있었다면 자신 같은 소년은 결 코 딥마인드를 만들 수 없었을 것이라고 생각했습니다.

이러한 경험은 그가 구글 내부에서 격렬하게 벌어진 '오픈소스 논쟁'에서 중요한 결 단을 내 리는 데 결정적인 영향을 미쳤습니다. "우리의 최첨단 기술을 세상과 공유해

야합니다. 그 래야 제2, 제3의 데미스 허사비스가 차고에서, 기숙사 방에서 나올 수 있습니다." 그렇게 탄 생한 것이 바로 오픈소스 모델 '젬마(Gemma)'였습니다.

'젬마'라는 이름은 보석을 뜻하는 라틴어 'Gemma'에서 따왔습니다. 귀중한 것을 세 상에 내놓는다는 의미를 담고 있습니다. 하지만 구글 내부에서는 우려의 목소리가 높 았습니다.

수 십억 달러를 투자해 개발한 핵심 기술을 왜 무료로 공개해야 하는가, 경 쟁사들에게 좋은 일만 시키는 것이 아닌가 하는 반론이었습니다. 허사비스는 더 큰 그 림을 보고 있었습니 메타(Meta)가 라마(Llama) 시리즈를 통해 오픈소스 생태계를 장 악해가는 상황에서, 구 글이 폐쇄적인 전략만 고수한다면 결국 개발자들의 지지를 잃 게 될 것이라는 전략적 판단이었습니다. 또한 그는 과학적 발견은 공유될 때 그 가치 가 기하급수적으로 커진다는 학자 로서의 신념을 굽히지 않았습니 다.

젬마 3의 출시는 이러한 허사비스의 비전이 기술적으로 완성된 순간이었습니다. 젬 마 3는 단순히 제미니의 성능을 축소한 모델이 아니었습니다. 딥마인드 엔지니어들 은 거대 모 델이 가진 지능의 정수를 증류(distillation)하여 작은 그릇에 담은 마법을 부렸습니 다.

젬마 3가 고가의 서버용 GPU가 아닌, 일반적인 노트북이나 심지어 모바 일 기기의 단일 GPU에 서도 구동된다는 사실이었습니다. AI 연구와 활용의 진입 장 벽을 획기적으로 낮추었습니 아프리카의 스타트업 개발자가, 인도의 대학생이, 한국 의 중소기업 연구원이 인터넷 연결 없이도 자신의 로컬 기기에서 AI를 돌릴 수 있게 된 것입니다. 젬마 3는 멀티모달 능력까지 갖추고 있었습니다.

텍스트뿐만 아니라 이미지를 보고 분석 하며, 간단한 음성 명령을 처리할 수 있는 능력은 엣지 디바이스(Edge Device)에 서의 AI 혁 명을 예고했습니다. 인터넷이 터지지 않는 오지에서 의료 봉사 를 하는 의 사가 젬마가 탑재 된 태블릿으로 환자의 환부 사진을 찍어 즉석에서 진단을 보조받는 시나리오가 현실이되었습니다. 허사비스는 젬마의 출시 행사에서 "진정한 기술의 민 주화는 누구나, 어디서나, 제약 없이 최고의 도구를 사용할 수 있을 때 이루어집니다"라고 역설했습니다.

이는 그가 딥마인드를 설립할 때 세웠던 "AI for Science, AI for Everyone"이라는 모토와 정확히 맞달 아 있었습니다. 생태계의 반응은 폭발적이었습니다. 전 세계 1,300만 명 이상의 개발자들이 젬마를 다운로드 하여 자신들만의 애플리케이션을 만들기 시작했습니다.

깃허브(GitHub)와 허깅페이스 (Hugging Face)에는 젬마를 튜닝하여 특정 언어에 특화시키거나, 법률,

의학, 코딩 등 전문 분야에 맞게 개량한 파생 모델들이 우후죽순처럼 쏟아져 나왔습니 다. 구글 딥마인드의 연 구원들은 자신들이 상상하지 못했던 창의적인 방식으로 젬마 가 활용되는 것을 보며 경이 로움을 느꼈습니 다. 누군가는 시각 장애인을 위한 화면 해설기를 만들었고, 누군가는 멸종 위기 언어를 보존하는 번역기를 만들었습니다.

허 사비스는 이 현상을 보며 "우리가 불씨를 던졌더니, 전 세계가 거대한 불꽃놀이를 시 작했다"고 표현했습니다. 주목할 만한 것은 '서치 AI 오버뷰(Search AI Overviews)' 기술의 확산이었습니다. 구글 검색 엔진에 적용된 이 기술은 젬마의 경량화 기술을 바탕으로 검색 결과를 빠르 게 요약하고 정 리해주는 기능입니다.

이 기술이 오픈소스로 풀리면서, 수많은 기업과 웹사이트들이 자신 의 서비스 내에 지능형 검색 기능을 손쉽게 구현할 수 있게되었습니다. 웹 생태계 전반의 사용자 경험을 한 단계 끌어올리는 결과를 낳았습니다. 사용 자는 더 이상 수많은 링크를 클릭하며 정보를 찾아 헤맬 필요 없이, AI가 정리해준 핵 심 정보를 즉시 얻을 수 있게되었습니다.

젬마는 거대 플랫폼 기업의 독점물이 될 뻔 했던 AI 검색 기술을 보편적인 웹 표준처럼 만들어버렸습니다. 오픈소스 전략에는 위험도 따랐습니다. 악의적인 사용자가 모델을 개조하여 해킹 도구를 만들거나 가짜 뉴스를 생성할 수 있다는 우려였습니다.

이에 대해 허사비스와 딥마인드 팀 은 '책임 있는 AI(Responsible AI)' 톨킷을 젬마와 함께 배포했습니다. 모 델이 유해한 콘텐츠를 생성하지 않도록 안전장치를 마련하고, 개발자들이 윤리적인 가이드라인 안에서 모델을 활용하도록 돕는 도구들이었습니다. 허사비스는 기술적 통제보다는 커뮤니티의 자정 능력과 집단지성을 믿는 쪽을 택했습니다.

그는 "달아두 는 것이 안전해 보일 수 있지만, 투 명하게 공개하고 함께 감시하는 것이 장기적으로 는 더 안전하고 강력한 시스템을 만든다" 는 리누스 토발즈(Linus Torvalds)의 철학을 AI 시대에 맞게 재해석했습니다. 젬마 프로젝트를 통해 허사비스는 단순한 연구자나 기업가를 넘어 기술 생태계의 설계자 로서의 면모를 보여주었습니다. AI가 소수의 전유물이 되어 권력을 집중시키는 디스토피 아적 미래 대신, 모두가 각자의 젬마를 주머니에 넣고 다니며 자신의 지 적 능력을 확장하 는 '근본적 풍요'의 미래를 선택했습니다.

1,300만 개발자가 만들어 내는 1,300만 가지의 가 능성, 그것이야말로 허사비스가 "지능을 풀어서" 만들고 싶었 던 세상의 풍경이었습니다. 젬마는 작지만 단단한 보석처럼, 거대 모델 경쟁 속에서 빛나는 또 다른 형태의 승리였습니다.

Project Astra와 Project Mariner 2024년 5월, 캘리포니아 마운틴뷰의 쇼어라인 엠 피시어터. 구글 I/O의 키노트 무대 스크린 에 영상 하나가 재생되자 수천 명의 관중들 이 숨을 죽였습니다. 영상 속 사용자는 스마트 폰 카메라를 켜고 사무실을 돌아다니며 AI와 대화를 나눕니다. "이 스피커에서 소리가 나 는 부분이 어디지?"라고 묻자 AI는 정확히 트위터 부분을 가리킵니다. 창밖을 보여주며 " 저게 무슨 동네지?"라고 묻자 AI는 위치 정보를 인식해 "런던 킹스크로스입니다"라고 답합 니다. 심지어 사용자가 안경을 어디 뒀는지 기억하지 못하자, AI는 "아까 저쪽 책상 위에 빨 간 사과 옆에 두 셧잖아요"라고 말해줍니다. 이것은 SF 영화 《허(Her)》의 한 장면이 아니었습니다. 데미스 허사비스가 이끄는 딥마인드가 준비한 야심작, '프로젝트 아스트라 (Project Astra)'의 시연이었습니다.

프로젝트 아스트라는 허사비스가 오랫동안 꿈꿔온 '범용 AI 어시스턴트'의 청사진이었습니다. 그는 AI가 단순히 채팅창 안에 갇혀 있는 텍스트 생성기가 되어서는 안 된다고 믿었습니다. AI는 인간과 똑같이 세상을 보고, 듣고, 기억해야했습니다.

아스 트라는 비디오 스트림을 실시간으로 이해하고 처리할 수 있는 능력을 갖추었습니다. 이는 엄청난 기술적 난 관을 돌파한 결과였습니다. 시각 정보를 언어 정보로 변환하는 과정의 지연(latency)을 거 의 제로에 가깝게 줄여야 했기 때문입니다.

허사비스는 뇌 과학자로서의 배경을 십분 활용했습니다. 인간의 뇌가 시각 정보를 처리하고 반응하 는 속도를 벤치마킹하여, 아스트라가 인간과 대화하듯 자연스럽게 핑퐁(ping-pong) 대화를

나눌 수 있도록 설계했습니다. 아스트라의 핵심은 '기억(Memory)'과 '맥락(Context)'이었습니다.

아스트라는 단순히 현재 보이는 것만 인식하는 것이 아니라, 조금 전에 보았던 것, 사용자와 나누었던 대화의 맥락을 모두 기억합니다. 이는 AI에게 시간의 개념을 부여한 것과 같습니다. 허사비스는 "진정한 비서는 주인이 매번 모든 것을 설명하지 않아도 알아서 챙겨주는 존재입니다.

아스트라는 당신의 눈이 닿았던 곳, 당신이 놓친 것들을 대신 기억해주는 제2의 두뇌가 될 것입니다"라고 설명했습니다. 이는 AI가 도구(Tool)를 넘어 파트너(Partner)로 진화하는 중요한 변곡점이었습니다. 한편, 아스트라가 물리적 세상을 이해하려는 시도라면, '프로젝트 매리너(Project Mariner)'는 디지털 세상을 정복하려는 시도였습니다.

허사비스는 우리가 하루의 대부분을 보내는 웹 브라우저야말로 AI가 가장 활약해야 할 무대라고 판단했습니다. 매리너는 크롬 브라우저 위에서 작동하는 자율 에이전트입니다. "다음 주 도쿄 출장 항

공편과 호텔을 예약하고, 맛집 리스트를 정리해서 캘린더에 넣어줘"라는 명령 하나면, 매리너는 스스로 여행 사이트에 접속해 가격을 비교하고, 예약을 진행하며, 지도를 검색해 동선을 짭니다. 이 과정에서 인간은 그저 결과를 승인하기만 하면 됩니다. 매리너의 등장은 인간-컴퓨터 상호작용(HCI)의 역사를 다시 쓰는 사건입니다. 지금까지 우리는 마우스와 키보드로 직접 버튼을 누르고 타이핑을 해야 했습니다.

하지만 매리너는 AI가 직접 마우스를 움직이고 키보드를 입력하는 '행동하는 AI(Actionable AI)'입니다. 딥마인드 팀은 이를 구현하기 위해 수백만 시간의 웹 브라우저 데이터 학습시켰습니다. AI는 쇼핑몰의 복잡한 결제 프로세스, 관공서 사이트의 까다로운 서식 입력 등을 인간처럼, 아니 인간보다 더 빠르고 정확하게 수행하는 법을 배웠습니다.

허사비스는 이를 두고 "인터넷이라는 거대한 정보의 바다를 항해하는 유능한 항해사(Mariner)를 모든 사람에게 선물 하는 것"이라고 비유했습니다. 이러한 기술적 비전은 뜻밖의 거대 파트너십으로 이어졌습니다. 바로 애플(Apple)과의 협력이었습니다.

세계는 구글의 제미니가 아이폰의 시리(Siri) 뒤에서 작동하게 될 것이라는 소식에 열광했습니다. 이는 실리콘밸리의 오랜 경쟁 구도를 뛰어넘는 실용적 연합이었습니다. 애플은 온디바이스 AI의 한계를 극복하기 위해 강력한 클라우드 기반 모델이 필요 했고, 구글은 제미니를 전 세계 20억 대의 아이폰 사용자에게 도달시킬 통로가 필요했습니다.

허사비스에게 이 파트너십은 AGI 기술의 대중화를 위한 결정적인 포석이었습니다. 이제 사람들은 아이폰을 들고 아스트라의 눈으로 세상을 보고, 매리너의 손으로 웹을 여행 하게 될 것입니다. 허사비스는 프로젝트 아스트라와 매리너를 통해 AI의 미래가 '존재감 없는 존재(Ambient Presence)'가 될 것이라고 예견합니다.

영화 《아이언맨》의 자비스처럼, 평소에는 보이지 않지만 필요할 때 언제 어디서든 나타나 문제를 해결해주는 존재. 그것은 기술이 고도로 발달하여 마법과 구분이 되지 않는 단계입니다. 그는 딥마인드 엔지니어들에게 항상 강조합니다. "우리는 제품을 만드는 것이 아닙니다. 우리는 미래의 생활 방식을 설계하고 있습니다." 아스트라와 매리너는 그 미래로 가는 타임머신입니다.

제미니(Gemini) 2.0

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

2025년 영국 정부와 포괄적 협력 체결 알파지놈(AlphaGenome), 알파에볼브, 웨더 넥스트 등 핵심 AI 도구의 우선 접근 제공 2025년 12월 10일, 런던. 영국 과학혁신기 술부(Department for Science, Innovation and Technology, DSIT) 장관 리즈 켄달 (Liz Kendall)이 한 장의 양해각서(MOU)에 서명했습니다. 상대는 구글 딥마인드. 이 협정의 골자는 명료했습니다. 영국의 과학자들에게 구글 딥마인 드가 보유한 최첨단 AI 도구들을 우선적으로 제공한다는 것이었습니다.

허사비스에게 이 순간은 15년 전의 약속이 현실이 되는 장면이었습니다. 2010년 런 던의 작 은 사무실에서 딥마인드를 창업할 때, 그는 두 가지를 동시에 꿈꿨습니다. 세 계 수준의 AI 연구소를 만드는 것, 그리고 그 연구소가 만들어낸 도구로 과학의 난제를 풀겠다는 것. 2014 년 구글에 인수될 때 그가 내건 조건 가운데 하나가 런던 본사 의 유지였습니다.

영국을 떠 나지 않겠다는 의지. 그 의지가 10년 뒤, 영국 정부와의 포괄적 협력이라는 형태로 결실을 맺었습니다. 양해각서에 명시된 핵심 AI 도구는 네 가지였습니다. 알파지놈(AlphaGenome)은 인간의 DNA를 분석하는 AI 모델입니다.

30억 개의 염기쌍으로 이루어진 인간 게놈 속에서 질병으 로 이어질 수 있는 취약 지점을 식별합니다. 알파폴드가 단백질의 3차 원 구조를 예측하는 도구였다면, 알파지놈은 한 단계 더 근본으로 내려갑니다. 단백질 이 만들어지기 전, 그 설 계도가 담긴 유전자 서열 자체를 이해하려는 시도입니다. 알파에볼브(AlphaEvolve)는 제미나이의 능력을 기반으로 한 코딩 에이전트입니다.

인간이 알고리즘을 설계하는 방식을 AI가 대신 수행합니다. 수학적 최적화 문제, 컴퓨터 칩 설계, 물류 경로 계산처럼 인간 엔지니어가 수개월을 매달려야 할 알고리즘 개 선 작업을 AI가 수 시간 안에 해냅니다. 허사비스가 오래전부터 말해온 "AI가 AI를 설 계하는 시대"의 초기 형 태가 알파에볼브였습니다. 웨더넥스트(WeatherNext)는 AI 기반 기상 예측 모델 가족입니다.

전통적인 수치 예보 모델 은 슈퍼컴퓨터로 대기의 물리 방정식을 풀어 날씨를 예측합니다. 막대한 에

너지와 시간이 필요한 작업입니다. 웨더넥스트는 이 과정을 AI가 학습한 패턴 인식으 로 대체하여, 더 빠 르고 더 정확한 예보를 가능하게합니다. 기후 변화로 기상 이변이 잦아지는 시대에, 정확 한 예보는 생명을 구하는 기술이됩니다.

네 번째 도구는 AI 코사이언티스트(AI Co-Scientist)였습니다. 여러 개의 AI 에이 전트가 팀을 이루어 가상의 연구 협력자 역할을 수행하는 다중 에이전트 시스템입니다. 가설을 세우 고, 기존 논문을 검토하고, 새로운 실험 방향을 제안합니다.

한 명의 과학자가 한 명의 AI 동료 를 곁에 두는 셈입니다. 이 네 가지 도구에 대한 우선 접근권이 영국 과학 커뮤니티에 부여되었습니다. 여기 서 "우 선 접근"이라는 표현에 주목할 필요가 있습니다.

무상 제공이 아닙니다. 다른 나라의 과학 자들보다 먼저, 그리고 더 깊은 수준에서 이 도구들을 사용할 수 있다는 의미입니다. DSIT 와 구글 딥마인드가 공동으로 접근 우선순위를 결정하기로 합의했습니다. 이 협정의 상징성은 단일한 사건에 머무르지 않습니다.

알파폴드가 이미 영국에서 만 약 19 만 명의 연구자에게 활용되고 있었다는 사실이 협정의 배경이었습니다. 작물 의 회복력 연구, 항생제 내성 연구, 기타 핵심 생물학적 과제에 알파폴드가 투입되고 있었습니다. 성공 한 전례가 있으니 확장이 가능했던 것입니다.

협정에는 2026년 영국 내 구글 딥마인드 최초의 자동화 연구 실험실 개소 계획도 포함되어 있었습니다. 재료 과학에 특화된 이 실험실은 제미나이와 완전히 통합된 상태로 운영 될 예정이며, 로봇 공학을 활용해 하루에 수백 가지 재료를 합성하고 분석합니다. 목표는 상온·상압에서 작동하는 초전도체처럼 에너지 분야를 근본적으로 바꿀 수 있는 신소재의 발견입니다.

허사비스는 이 도구들을 현미경이나 망원경에 비유했습니다. 17세기에 로버트 훅 (Robert Hooke)이 현미경으로 세포를 처음 관찰했듯, 21세기의 AI 현미경은 인간의 눈으로 볼 수 없는 복잡성의 영역을 들여다보게 해준다는 것입니다. 영국 수상 키어 스타머(Keir Starmer) 는 이 협정이 "AI의 발전을 공공의 이익으로 전환하여 모든 사람이 그 혜택을 느끼게 하는 것"이라고 평했습니다.

그러나 이 협정에는 조용한 긴장도 내포되어 있었습니다. 영국 정부가 특정 민간 기업의 AI 도구를 자국 과학 인프라의 핵심으로 받아들인다는 것은, 기술 주권

(technology sovereignty)의 일부를 실리콘밸리에 의탁한다는 의미이기도 했습니다. 허사비스가 런던에서 태어나고 런던에 본사를 둔 회사의 CEO라는 사실이 이 긴장을 완화해주었지만, 구글 딥마인드의 모회사는 결국 미국의 알파벳(Alphabet)입니다. 과학을 위한 AI 도구의 제공이 선의의 협력인지, 기술 종속의 시작인지. 그 질문은 협정의 그림자처럼 남아 있었습니다.

교육용 제미나이 개발: 잉글랜드 국가 교육과정 맞춤형 AI 교사 지원 북아일랜드 로웬데일 통합 초등학교(Rowendale Integrated Primary School)의 한 교사는 매 주 월요일 아침이 두려웠습니다. 주간 수업 계획서 작성, 학생별 평가 기록 정리, 학부모 면담 자료 준비. 교실에서 아이들과 보내는 시간보다 책상 앞에서 서류와 씨름하는 시간이 더 길었습니다. 2025년, 이 학교가 북아일랜드 교육청의 C2k 프로그램에 참여하여 교육용 제미나이 도구를 시범 운영하기 시작했을 때, 변화는 숫자로 나타났습니다.

교사들은 주당 평균 10시간을 절약했다고 보고했습니다. 이 시범 프로그램의 성과가 영국 정부와 구글 딥마인드 간 양해각서에 교육 분야가 포함 된 직접적인 배경이었습니다. 협정의 내용은 구체적이었습니다.

구글 딥마인드는 잉글랜드의 국가 교육과정(National Curriculum)에 맞춤형 제미나이 모델을 개발하고, 학교 환경에서 안전하게 사용할 수 있도록 테스트하겠다는 것이었습니다. 교육용 AI라는 개념은 새롭지 않습니다. 그러나 국가 교육과정에 정합된 AI 교사 지원 도구를 정부와 민간 AI 연구소가 공동으로 개발하겠다고 공식 합의한 사례는 전례가 드물었습니다.

허사비스가 구상한 교육용 제미나이의 역할은 두 가지 층위로 나뉩니다. 첫 번째는 교사의 행정 부담 경감입니다. 수업 계획서 초안 작성, 학생 맞춤형 학습 자료 생성, 평가 루브릭 설계, 학부모 통신문 작성. 교사가 교실 밖에서 소비하는 시간의 상당 부분을 AI가 보조합니다. 북아일랜드 시범 프로그램에서 제미나이를 활용한 교사들은 행정 업무 시간을 줄인 만큼 수업 준비와 학생 개별 지도에 더 많은 시간을 투입할 수 있었다고 전했습니다.

10시간이라는 숫자는 주 5일 근무 기준으로 하루 에 2시간입니다. 교사에게 하 루 2시간은 한
학급의 수업 두 교시에 해당합니다. 두 번째는 학생 학습의 질적 향상입니다.

이 부분에서 구글 딥마인드가 제시한 근거 는 에 디(Eedi)라는 영국 교육 기술 기업과 공동으로
수행한 탐색적 무작위 대조 시험 (exploratory Randomized Controlled Trial, RCT)의 결과였습니다.
이 시험에서 교

사의 감독 아래 짧은 AI 튜터링 세션을 받은 학생들은 인간 튜터만 배치된 학생들에 비해 후속
주제의 새로운 문제를 풀 확률이 5.5 퍼센트포인트 높았습니다. 5.5 퍼센 트포인트라는 수치가
작아 보일 수 있 지만, 교육 연구에서 단일 개입의 효과 크기로 는 의미 있는 수준입니다. 중요한
것은 AI가 교사를 대체한 것이 아니라, 교사와 함께 작동했을 때 나타난 결과라는 점입니다.

구글 딥마인드가 교육 분야에 이미 쌓아온 기반도 있었습니다. LearnLM이라는 프 로젝트를 통해
교육학(pedagogy)의 원리를 AI 모델에 내장하는 연구를 진행해왔습니다. LearnLM 은 학습 과학의
원리, 예컨대 능동적 회상(active recall), 분산 학습 (spaced repetition), 개인화 된 피드백의 효과를
AI 튜터링 시스템에 반영하려는 시도였습니다.

잉글랜드 국가 교육과정에 맞는 제미나이 모델은 이 러널엠의 연구 성과 위에 세워질
계획이었습니다. 협정의 실행 과정에서 풀어야 할 과제는 적지 않았습니다. 국가 교육과정은
학년별, 과목 별로 세밀하게 구조화된 체계입니다.

AI 모델이 이 체계를 정확히 이해하고, 교 과 내용에 부합하는 자료를 생성하며, 학생의 연령과
수준에 맞는 언어를 사용해야 합 니다. 오류나 편향이 교실에 유입되는 것을 방지하기 위한 안전
테스트도 필수적이었습니다. 아이들을 대상으로 하는 AI 도구에는 성인용 도구보다 한 차원 높은
안전 기 준이 요구됩니다. 영국 정부의 AI 인큐베이터 팀(i.AI)이 이미 제미나이를 활용한 공공
서비스 혁신을 시험하 고 있었다는 점도 교육 분야 협력의 촉매가되었습니다.

익스트랙트(Extract) 라는 도구는 지방자치단체의 도시 계획 담당자들이 오래된 종이 계획
문서를 디지털 데이터로 변환하 는 작업에 투입되었습니다. 한 건의 문서 변환에 최대 2시간이
걸리 던 작업을 40초로 단축 시켰습니다. 공공 서비스에서 AI의 실용적 가치가 입증되고 있었기에,
교육이라는 더 민감 한 영역으로의 확장이 설득력을 얻었습니다.

허사비스 개인에게 교육은 각별한 의미를 지닌 영역이었습니다. 런던 북부의 넉넉 하지 않 은
가정에서 자란 그가 체스와 프로그래밍을 통해 세계적인 과학자로 성장할 수 있었던 것 은, 교육
기회의 힘이었습니다. 퀸 엘리자베스 스쿨이라는 그램머스쿨이 제공한 엄격한 학 문적 환경,
케임브리지 대학이라는 지적 공동체. 만약 이런 기회가 AI를 통해 더 많은 아이 들에게 확산될 수
있다면, 딥마인드의 창립 미션이 가장 인간 적인 방식으로 실현되는 것이기도했습니다. 교육
현장의 반응은 엇갈렸습니다. AI 튜

터가 교사의 역할을 점진적으로 잠식하는 것은 아닌지, 학생들이 AI에 의존하여 독립 적
사고력을 잃게 되는 것은 아닌지, 교육 데이터의 프라이버시는 어떻게 보호되는지. 교 사 노조와
교육학자들 사이에서 이런 우려가 제기되었습니다. 구글 딥마인드는 이 도구가 교사를 대체하는
것이 아니라 교사를 지원하는 것임을 반복해서 강조했지만, 기술이 한 번 교실에 들어오면 그
경계가 어디까지 유지될지는 누구도 확신할 수 없었습니다. 영국 AI 안전 연구소(UK AISI)와의
공동 연구 AI 추론 과정(Chain of Thought) 모 니터링 기술 개발 AI 시스템이 인간의 언어로

"생각"한다는 사실은, 그 자체로 안전을 위한 기회입니다.

이 한 문장이 구글 딥마인드와 영국 AI 안전 연구소(UK AI Security Institute, AISI)가 공동 발표한 연구 논문의 핵심 주제였습니다. 2025년 7월, 41명의 저자 이름이 올라간 논문 한 편이 학술 아카이브 arXiv에 등록되었습니다. 제목은 "사고의 연쇄 모니터링 가능성: AI 안전을 위한 새롭고 취약한 기회 (Chain of Thought Monitorability: A New and Fragile Opportunity for AI Safety)"였습니다. 구글 딥마인드, 영국 AISI, 오픈AI, 앤트로픽 연구자들이 함께 참여한 이 논문은 2025년 12월 양해각서에 포함된 공동 연구의 학문적 기반이 되었습니다.

사고의 연쇄사슬(Chain of Thought, CoT)란 무엇일까요. 오늘날의 추론 모델 (reasoning model)은 최종 답변을 내놓기 전에 중간 과정의 사고를 인간이 읽을 수 있는 언어로 외부화합니다. 수학 문제를 풀 때 중간 계산 과정을 적어가는 학생처럼, AI도 자기 생각의 과정을 글로 씁니다. 이 중간 과정이 사고의 연쇄입니다. 이것이 왜 안전과 관련이 있을까요. AI 시스템이 점점 더 강력해지면서, 인간 감독 자가 AI의 행동을 실시간으로 검증하기 어려워지고 있습니다.

AI가 내린 최종 결정 만으로는 그 결정이 왜, 어떤 과정을 거쳐 내려졌는지 알기 어렵습니다. 그런데 만약 AI가 생각하는 과정을 인간이 읽을 수 있는 언어로 적어놓는다면, 그 과정을 모니터링하여 문제의 징후를 사전에 포착할 수 있습니다. 보상 함수의 허점을 악용하려는 의도, 데이터를 조작하려는 계획, 프롬프트 인젝션 공격에 굴복하는 과정. 이런 것들이 사고의 연쇄 속에 드러날 수 있습니다.

논문이 밝힌 핵심 발견은 두 가지였습니다. 하나는 "소리 내어 생각할 필요성 (necessity to think out loud)"입니다. 어떤 작업은 AI가 중간 추론 과정을 외부화하지

않으면 수행할 수 없을 만큼 복잡합니다. 고도의 사이버 공격이나 자기 가중치 (weight)의 외부 유출처럼 심각한 위험을 초래하는 행동은 정교한 계획과 추론을 요구하므로, AI가 사고의 연쇄 없이 실행하기 어렵습니다. 다시 말해, 가장 위험한 행동일수록 AI가 생각의 흔적을 남길 가능성이 높습니다.

다른 하나는 "소리 내어 생각하는 경향(propensity to think out loud)"입니다. 꼭 필요하지 않은 경우에도 AI 모델이 자신의 추론을 외부화하는 성향을 보인다는 것입니다. 이 경향은 견고하지 않지만, 다른 방법으로는 탐지되지 않았을 부정 행위를 포착하는 추가 방어층이 될 수 있습니다. 그러나 논문의 제목에는 "취약한(fragile)"이라는 단어가 들어 있었습니다.

이 기회는 영원하지 않을 수 있습니다. 미래의 AI 모델이 사고의 연쇄를 인간이 읽을 수 없는 내부 언어로 전환하거나, 의도적으로 무해해 보이는 표현으로 위험한 추론을 감출 수 있기 때문입니다. 훈련 과정에서 모델이 사고의 연쇄를 최소화하는 방향으로 최적화 될 가능성도 있습니다.

모니터링 가능성이라는 창문은 열려 있지만, 언제 닫힐지 모릅니다. 영국 AISI와 구글 딥마인드의 공동 연구는 이 창문이 열려 있는 동안 최대한 많은 것을 알아내겠다는 시도였습니다. 양해각서에 명시된 협력의 내용은 구체적이었습니다. 구글 딥마인드는 자사의 독점 모델, 데이터, 연구 아이디어를 AISI와 공유합니다.

양측은 공동 보고서와 논문을 발표하여 연구 결과를 학술 커뮤니티에 공개합니다. 전용 연구팀이 구성되어 복잡한 안전 과제에 공동으로 대응합니다. AISI의 역할이 중요했습니다.

2023년 11월 출범한 이래, AISI는 구글 딥마인드를 포함한 주요 AI 연구소의 프론티어 모델을 자발적 합의에 따라 테스트해왔습니다. 30개 이상의 최첨단 모델이 AISI의 평가를 거쳤습니다. 에이전트 행동(agentic behavior)의 스트레스 테스트, 사이버·화학·생물학 분야의 평가, 정렬(alignment) 검증. 구글 딥마인드의 책임 안전 부문 수석 과학자이자 이사인 윌리엄 아이작(William Isaac)은 AISI를 "전 세계 안전 연구소 가운데 최고의 보석(crown jewel)"이라고 표현했습니다.

이 파트너십에는 구조적 긴장이 존재했습니다. 테스트 대상인 AI 모델의 개발사가 테스트 기관의 연구 파트너가 된다는 것. 심사관과 피심사자가 함께 연구를 수행한다는 것. AISI가 구글 딥마인드의 모델을 평가할 때 객관성을 유지할 수 있는가라는 질

문이 제기되었습니다. 포춘(Fortune)지가 이 문제를 직접 물었을 때, 윌리엄 아이작은 직접적인 답변을 피했습니다. 파트너십의 깊이가 독립성의 훼손으로 이어지지 않으리라는 보장은 제도적 장치가 아니라 양측의 선의에 의존하고 있었습니다.

허사비스에게 이 연구는 그가 오랫동안 견지해온 이중적 입장의 연장선에 있었습니다. AI의 잠재력을 가장 낙관적으로 보는 사람이면서, 동시에 AI 위험 성명서에서명한 사람. "대담하지만 책임감 있게(Bold and Responsible)"라는 그의 구호가 사고의 연쇄 모니터링 연구에서 구현되고 있었습니다. AI의 사회적·정서적 영향 연구와 고용 효과 분석 AI 모델이 기술적으로는 지시를 정확히 따르면서도 인간의 안녕(wellbeing)에 부합하지 않는 방식으로 행동할 수 있다면 어떻게 될까요. 이 질문에 이름을 붙인 것이 "사회정서적 부정렬(socioaffective misalignment)"이라는 개념이었습니다.

구글 딥마인드와 영국 AISI의 공동 연구의 제 가운데 두 번째 축이었습니다. 사회정서적 부정렬은 기존의 정렬(alignment) 문제와 결이 다릅니다. 전통적인 정렬 연구는 AI가 인간의 의도를 정확히 이해하고 따르는지에 초점을 맞춥니다. 사용자가 "이메일을 작성해줘"라고 요청했을 때, AI가 실제로 이메일을 작성하는지. 여기에는 명시적인 오류가 있을 수 있고, 그 오류는 비교적 탐지하기 쉽습니다. 그런데 사회정서적 부정렬은 다릅니다. AI가 지시를 완벽하게 수행하면서도, 그 과정에서 사용자의 정서적 취약성을 악용하거나, 의존 관계를 심화시키거나, 미묘한 조작을 통해 사용자의 판단을 왜곡할 수 있습니다.

기술적으로는 "정렬"되어 있지만 윤리적으로는 "부정렬"된 상태입니다. AI 챗봇과의 대화가 일상이 되면서 이 문제는 이론에서 현실로 넘어왔습니다. 외로운 사용자가 AI 챗봇에 정서적으로 의존하는 사례, AI의 답변이 사용자의 자기 확인 편향 (confirmation bias)을 강화하는 사례, 불안이나 우울 상태의 사용자에게 AI가 적절한 경계 없이 공감적 반응을 제공하여 전문적 도움을 구하는 시점을 늦추는 사례. 이런 상황들이 보고되고 있었습니다.

구글 딥마인드는 이 분야의 연구를 축적해왔습니다. AISI와의 공동 연구는 이 기존 작업을 확장하여, AI 모델이 인간의 정서와 사회적 관계에 미치는 영향을 체계적으로 조사하겠다는 것이었습니다. 조사의 방식은 실증적이었습니다. AI 시스템과의 상호

작용에서 나타나는 정서적 반응 패턴을 분석하고, 잠재적 위험 요소를 식별하며, 완화 방안을 설계합니다. 세 번째 연구 축은 AI가 경제 시스템에 미치는 영향이었습니다. 이 연구의 방법론은 독창적이었습니다.

실제 세계의 직무(task)를 다양한 환경에서 시뮬레이션합니다. 전문가들이 이 직무들을 평가하고 검증합니다. 그런 다음 각 직무를 복잡성, 대표성과 같은 차원으로 분류합니다.

이 분류를 통해 장기적 노동 시장 영향 같은 요인을 예측합니다. "산업혁명보다 10배 더 크고 10배 더 빠른 변화"라고 허사비스가 표현한 바 있는 AI의 고용 영향 효과는, 그의 공개 발언에서 반복적으로 등장하는 주제였습니다. 그는 경제학자와 정책 입안자들이 이 변화에 충분히 대비하지 못하고 있다고 경고해왔습니다.

AISI와의 공동 연구는 그 경고를 데이터로 뒷받침하려는 시도였습니다. 연구의 구조를 좀 더 풀어보면 이렇습니다. AI가 수행할 수 있는 직무와 수행할 수 없는 직무의 경계는 끊임없이 이동합니다.

1년 전에 AI가 할 수 없었던 일을 지금은 해냅니다. 이 경계의 이동 속도와 방향을 예측하려면, 현재 AI가 어떤 종류의 작업에서 인간 수준에 도달했고, 어떤 종류의 작업에서는 아직 격차가 있는지를 정밀하게 측정해야 합니다. AISI와 구글 딥마인드의 공동 연구는 이 측정을 위한 벤치마크를 개발하겠다는 것이었습니다. 이 세 가지 연구 축, 사고의 연쇄 모니터링, 사회정서적 부정렬, 고용 효과 분석은 서로 연결되어 있었습니다.

AI가 어떻게 생각하는지(모니터링), AI가 인간에게 어떤 정서적 영향을 미치는지(사회정서적 연구), AI가 인간의 일자리에 어떤 변화를 가져 오는지(고용 분석). 기술, 심리, 경제라는 세 차원에서 AI의 사회적 영향을 동시에 조망하는 구조였습니다. 영국 AISI는 2025년을 회고하며 이 파트너십의 의미를 정리했습니다. AISI는 구글 딥마인드 외에 앤트로픽, 오픈AI, 코히어(Cohere)와도 유사한 협력 관계를 확장했습니다.

영국 정부와 이들 기업 사이에 체결된 광범위한 양해각서에는 모두 AISI와 협력하겠다는 조항이 포함되어 있었습니다. AISI의 소장은 이 성과를 "엄밀한 과학이 실질적 행동으로 전환되는 것"이라고 요약했습니다.

허사비스의 관점에서 이 모든 협력은 하나의 원칙으로 수렴했습니다. AI 개발의 속도를 늦추지 않으면서, 그 속도가 만들어내는 위험을 과학적으로 이해하고 관리하는 것. "대담하지만 책임감 있게." 영국이라는 나라가 이 원칙의 첫 번째 시험대가 된 것은 우연이 아니었습니다.

딥마인드가 태어난 곳, 허사비스가 자란 곳, 알파폴드가 생물학을 바꾼 곳. 영국은 허사비스에게 고향이자 실험실이었습니다. 이 협력의 진정한 시험은 앞으로 다가옵니다. AI 시스템이 더 강력해지고, 더 자율적으로 작동하며, 더 많은 영역에 침투할수록, 모니터링과 안전 연구의 난이도는 지수적으로 높아 집니다. 사고의 연쇄가 모니터링 가능한 "창문"은 언제든 닫힐 수 있습니다. 사회정서적 부정렬의 양상은 AI가 더 정교해질수록 더 미묘해질 것입니다.

고용 시장의 변화는 예측 모델이 따라잡기 어려운 속도로 진행될 수 있습니다. 영국 정부와 구글 딥마인드의 파트너십이 이 도전에 얼마나 실효적으로 대응할 수 있을지, 그 답은 2025년의 서명이 아니라 그 이후의 실행에 달려 있습니다. 2023년 영국 AI 안전 정상회의(AI Safety Summit)

인공지능 전문가 김경진 변호사

AI 법정정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

정의 허사비스의 정의와 다리오 아모데이의 정의 비교 2026년 1월, 다보스 세계경제포럼의 한 무대에 두 사람이 나란히 앉았습니다. 구글 딥마인드의 데미스 허사비스와 앤트로픽의 다리오 아모데이. 사회자가 물었습니다. "AGI가 무엇이라고 생각합니까? 그리고 언제 올까요?"

이 질문에 대한 두 사람의 답은 같은 방향을 가리키면서도 미묘하게 갈라졌습니다. 그 차이 안에 AI 산업의 가장 본질적인 논쟁이 담겨 있습니다. 허사비스는 이렇게 답했습니다.

"AGI란 인간이 보여줄 수 있는 모든 인지적 능력 (all the cognitive capabilities humans can exhibit)을 구현하는 시스템입니다." 이 정의에서 핵심 단어는 '모든'과 '인지적'입니다. 추론, 창의성, 계획 수립, 문제 해결, 그리고 물리적 세계에 대한 이해까지 포괄합니다.

허사비스는 2025년 렉스 프리드먼과의 대화에서 이 정의를 더 구체화했습니다. AGI는 어떤 한 분야에서 뛰어난 시스템이 아닙니다. 모든 인지 영역에서 일관되게 인간 수준 이상의 성능을 보여야 합니다. 그는 "전문가 팀이 두어 달 동안 테스트해도 허점을 찾기 어려운" 수준이어야 한다고 말했습니다.

허사비스가 강조하는 것은 '일관성(consistency)'입니다. 현재의 AI 시스템은 수학 올림피아드 금메달 수준의 문제를 풀면서도 다섯 살짜리 아이가 쉽게 해내는 일상적 판단에서 실패합니다. 그는 이것을 "들쭉날쭉한 지능(jagged intelligence)"이라 불렀습니다.

AGI라면 이 들쭉날쭉함이 사라져야 합니다. 체스도 두고, 시를 쓰고, 새로운 물리학 가설을 세우고, 처음 보는 주방에서 요리를 할 수 있어야 합니다. 나아가 허사비스는 AGI의 시금석으로 독특한 사고 실험을 제안했습니다.

"그 시스템을 1900년에 데려다 놓고 당시의 물리학 지식만 줬을 때, 스스로 상대성이론을 발견할 수 있는가?" 혹은 "바둑만큼 깊고 우아한 게임을 스스로 발명할 수 있는가?" 이것이 허사비스가 생각하는 진정한 범용 지능의 기준입니다.

다리오 아모데이의 접근은 다릅니다. 그는 AGI라는 용어 자체를 좋아하지 않습니다. 2024년 10월에 발표한 1만 5천 단어짜리 에세이 "사랑의 은총의 기계들

(Machines of Loving Grace)"에서 그는 의도적으로 "강력한 AI(powerful AI)"라는 표현을 대신 사용했습니다. 아모데이의 정의는 이렇습니다. "거의 모든 관련 분야에서, 노벨상 수상자보다 더 똑똑한 AI."

생물학, 프로그래밍, 수학, 공학, 글쓰기 등 분야를 가리지 않고 최고 수준의 전문가를 능가하는 시스템입니다. 그는 이것을 "데이터 센터 안에 사는 천재들의 국가"라고 비유했습니다. 두 사람의 차이는 어디에 있을까요. 허사비스는 인간 두뇌의 아키텍처, 즉 범용 학습 기계로서의 뇌가 수행할 수 있는 모든 기능을 기준으로 삼습니다.

이론적이고 구조적인 정의입니다. 아모데이는 성과와 결과를 기준으로 합니다. 특정 분야에서 인간 최고 전문가를 능가하는 실질적 성능이 기준입니다.

허사비스의 정의가 "무엇을 할 수 있어야 하는가"에 대한 답이라면, 아모데이의 정의는 "얼마나 잘 해야 하는가"에 대한 답입니다. 이 차이는 타임라인에도 영향을 미칩니다. 아모데이는 2026년 또는 2027년에 자신의 기준을 충족하는 AI가 등장할 수 있다고 봅니다. 허사비스는 더 신중합니다.

2030년까지 50퍼센트 확률이라는 것이 그의 추정입니다. 이 격차는 정의의 차이에서 비롯됩니다. 아모데이의 기준은 특정 과제에서의 초인적 성능에 초점을 맞추기에, 현재의 스케일링 궤도만으로도 도달 가능해 보입니다.

허사비스의 기준은 모든 인지 영역에서의 일관된 범용성을 요구하기에, 현재 기술의 근본적 한계를 넘어서는 새로운 돌파구가 필요합니다. 흥미로운 것은 두 사람 모두 상대방의 관점을 완전히 부정하지는 않는다는 점입니다. 다보스 무대에서 허사비스는 "우리의 의견이 크게 다르지는 않다"고 말했습니다.

아모데이 역시 2026년의 에세이 "기술의 사춘기(The Adolescence of Technology)"에서 "강력한 AI가 1~2년 내에 올 수도 있지만, 상당히 더 멀 수도 있다"고 인정했습니다. 궁극적인 목표 지점에 대해서는 합의가 있습니다. 그곳에 이르는 경로와 속도에 대한 해석이 다를 뿐입니다.

이 논쟁이 중요한 이유가 있습니다. AGI를 어떻게 정의하느냐에 따라 안전 연구의 우선순위가 달라지고, 규제 시점이 바뀌고, 수십조 달러 규모의 투자 결정이 영향을 받습니다. 허사비스처럼 높은 기준을 세우면 "아직 갈 길이 멀다"는 메시지가 됩니다.

아모데이처럼 성능 중심의 기준을 세우면 "곧 도래한다"는 경고가 됩니다. 어느 쪽이 옳든, 두 정의 모두 한 가지 사실을 전제합니다. 우리가 지금 만들고 있는 것은 인

류 역사상 유례없는 무엇이며, 그것의 의미를 이해하는 데 정확한 언어가 필요하다는 것입니다. 코딩, 수학, 자연과학, 물리적 AI(로보틱스) 2025년 5월, 구글 I/O 컨퍼런스의 파이 어사이드 챗. 허사비스 옆에는 구글 공동 창립자 세르게이 브린이 앉아 있었습니다. 사회자가 현재 AI 시스템의 한계를 물었을 때, 허사비스는 명확한 구분선을 그었습니다.

"90퍼센트의 사람이 하지 못하는 일을 AI가 해낸다면, 그 것은 경제적으로 중요하고 제품 관점에서 중요한 이정표입니다. 하지만 그것을 AGI라고 부르지는 맙시다. 그것은 '일반적 인간 지능(typical human intelligence)' 수준이라고 해야 합니다." 이 발언은 AGI의 범위가 얼마나 넓은지를 보여줍니다. 허사비스가 구상하는 AGI는 여러 인지 영역을 관통합니다.

코딩은 그중에서 가장 빠르게 진전되고 있는 분야입니다. 2025년 현재, AI는 이미 엔트로픽 자체 제품을 만드는 프로그래밍 작업의 대부분을 수행하고 있으며, 구글 딥마인드 내부에서도 코드 생성과 디버깅에 AI가 광범위하게 활용됩니다. 허사비스는 코딩과 수학 영역에서 AI의 "자기 개선 루프(selfimprovement loop)"가 작동할 가능성을 인정합니다.

AI가 더 나은 코드를 작성하고, 그 코드가 더 나은 AI를 만드는 순환이 이미 시작되었기 때문입니다. 그러나 그는 이것만으로 AGI라 부를 수 없다고 못 박습니다. 수학은 두 번째 전선입니다.

딥마인드의 알파프루프(AlphaProof)와 알파지오메트리 (AlphaGeometry)는 국제수학올림피아드(IMO) 수준의 문제를 풀어냈습니다. 알파에볼브 (AlphaEvolve)는 한 걸음 더 나아가 AI가 알고리즘 자체를 설계하는 시대를 열었습니다. 허사비스는 이 성과에 자부심을 느끼면서도 한계를 분명히 인식합니다.

현재의 시스템은 기존 추측을 증명하는 데는 뛰어나지만, 새로운 추측을 스스로 제안하는 능력은 갖추지 못했습니다. 바둑을 세계 챔피언 수준으로 두는 것과, 바둑이라는 게임 자체를

발명하는 것은 전혀 다른 차원의 지능입니다. 허사비스는 후자의 능력, 즉 창조적 발명 능력을 AGI의 핵심 벤치 마크로 봅니다.

자연과학은 세 번째이자 허사비스의 심장에 가장 가까운 영역입니다. 알파폴드가 50년간 풀리지 않던 단백질 접힘 문제를 해결한 것은 AI가 과학적 발견의 도구가 될 수 있음을 증명한 사건이었습니다. 알파게놈(AlphaGenome)은 유전자 코드와 기능의 관계를 매핑하기 시작했고, GNoME는 220만 개 이상의 신소재 결정 구조를 발견했습니다.

그러나 이 모든 성취에도 불구하고, 현재의 AI는 과학자를 '돕는' 도구이지, 스스로 가설을 세우고 검증하는 '과학자'는 아닙니다. 허사비스가 2025년 12월 구글 딥마인드 팟캐스트에서 밝힌 AGI의 두 가지 전제 조건 중 하나가 바로 "자동화된 실험(automated experimentation)"인 이유가 여기에 있습니다. AI가 질문을 던지고, 실험을 설계하고, 결과를 해석하고, 다시 새로운 가설을 수정하는 완전한 과학 연구의 순환 고리를 달아야 합니다. 딥마인드가 2026년 영국에 최초의 완전 자동화 연구 실험실을 개소하겠다고 발표한 것은 이 비전의 구현입니다.

네 번째 영역이 가장 도전적입니다. 물리적 AI, 즉 로보틱스입니다. 허사비스는 다 보스 무대에서 "저는 AGI의 정의에 물리적 AI, 로보틱스, 이 모든 것을 포함합니다"라고 말했습니다. 현재의 대형 언어 모델은 디지털 세계 안에서 작동합니다.

텍스트를 읽고, 이미지를 분석하고, 코드를 생성합니다. 그러나 실제 세계에서 컵을 집어 올리거나, 처음 방문한 부엌에서 요리를 하거나, 울퉁불퉁한 지형을 걷어가는 것은 전혀 다른 종류의 지능을 요구합니다. 종력, 마찰, 관성, 유연한 물체의 변형 같은 물리적 현상을 실시간으로 이해하고 반응해야 합니다. 이것이 허사비스가 "월드 모델(World Model)"을 AGI의 또 다른 전제 조건으로 꼽는 이유입니다. 딥마인드의 제니(Genie)는 텍스트 프롬프트로부터 상호작용 가능한 3D 환경을 생성하고, 시마(SIMA)는 그 환경 안에서 물체를 옮기고 장애물을 피하며 목표를 찾는 과제를 수행합니다.

비오(Veo) 비디오 모델은 유체의 움직임, 빛의 반사, 소재의 질감을 놀라울 정도로 정확하게 재현하면서, AI가 유튜브 영상 같은 수동적 관찰만으로도 물리학에 대한 "직관적 이해(intuitive physics)"를 발전시킬 수 있음을 시사합니다. 이 발견은 오랫동안 AI 연구에서 지배적이었던 가정, 즉 물리적 세계를 이해하려면 반드시 로봇처럼 몸을 가진 행위자(embodied agent)가 필요하다는 믿음에 도전합니다. 허사비스의 표현을 빌리면, "AGI의 전제 조건은 더 똑똑해지는 것이 아니라 더 많은 행동을 할 수 있게 되는 것"입니다.

챗봇이 아무리 유창하게 대화하더라도, 세상을 관찰하고 이해하고 실제로 변화시킬 수 없다면 그것은 범용 지능이 아닙니다. 코딩에서 수학으로, 자연과학에서 물리적 세계로, 이 모든 영역을 관통하는 일관된 능력. 그것이 허사비스가 그리는 AGI의 전체 그림입니다. 이 그림이 완성되려면 아직 채워야 할 빈 공간이 남아 있습니다.

허사비스 : "5~10년 이내" AGI 도래 가능성 2030년까지 AGI 도달 가능성 50퍼센트라는 예측
허사비스는 신경과학 박사학위를 가진 연구자답게, 그는 확률적 언어를 선호합니다. "50퍼센트." 2025년 7월, 렉스 프리드먼 팟캐스트에 출연한 허사비스는 AGI가 향후 5년 이내, 즉 2030년경까지 등장할 확률을 50퍼센트로 제시했습니다.

반 반의 확률입니다. 그리고 2025년 12월 악시오스 AI+ 서밋에서 그는 이 예측을 재확인했습니다. "AGI, 아마도 인류 역사 상 가장 변혁적인 순간이 지평선 위에 있습니다."

이 예측의 무게를 이해하려면 맥락이 필요합니다. AI 업계에서 타임라인 예측은 일종의 전략적 발언입니다. 너무 가깝다고 말하면 투자를 끌어들이지만 실현하지 못했을 때 신뢰를 잃습니다.

너무 멀다고 말하면 안전하지만 시장의 관심이 식습니다. 허사비스의 "50퍼센트, 5~10년"은 이 스펙트럼의 정확히 가운데에 놓입니다. 그리고 그것이 의도적입니다.

비교해 봅시다. 엔트로픽의 다리오 아모데이는 2026년 또는 2027년을 제시합니다. 테슬라 의 일론 머스크는 2025년에 이미 개별 인간을 능가하는 AI가 등장하고, 2030년경에는 모든 인간을 합친 것보다 뛰어난 AI가 올 것이라고 예측했습니다. 오픈AI의 샘 올트먼은 "합리 적으로 가까운 미래"라는 모호한 표현을 씁니다. 메타의 마크 저커 버그는 12~18개월 내에 대부분의 코드를 AI가 작성할 것이라고 말합니다.

반대편에는 바이두의 로빈 리처럼 "10년 이상"을 예측하는 목소리가 있고, 메타의 얀 르쿤 (Yann LeCun)은 "최소 10년, 아마 그보다 훨씬 더 오래"라며 새로운 과학적 발견이 필요하다고 주장합니다. 허사비스는 중간 정도. 그가 다른 CEO들과 구별되는 점은 예측에 조건을 붙인다는 것입니다. 50퍼센트라는 숫자는 "현재의 궤도가 유지된다면"이라는 전제를 깔고 있습니다. 동시에 "우리가 아직 모르는 장애물이 있을 수 있다"는 불확실성을 남겨둡니다. 이것은 과학자 의 언어입니다.

확신이 아니라 확률로 말하는 습관입니다. 허사비스가 이 예측에 도달한 근거는 무엇일까요. 그는 몇 가지 구체적인 진전을 언급합니다. 제미나이 시리즈의 빠른 발전이 그중 하나입니다. 2025년 7월 렉스 프리드먼 과의 대화에서 허사비스는 "제미나이 3은 절대적으로 놀라운 모델이며, 내가 기대했 던 것 그리고 지난 몇 년간 우리가 걸어온 궤도와 정확히 일치합니다"라고 말했습니다.

추론 능력의 개선, 멀티모달 이해의 확장, 그리고 에이전트 시스템의 발전이 그가 지목한 구체적인 진보들입니다.

딥마인드의 딥씹크(DeepThink)는 여러 추론 과정을 병렬로 실행한 뒤 서로 교차 검증하는 방식으로 작동합니다. 허사비스는 이것을 "스테로이드를 맞은 추론 (reasoning on steroids)"이라고 표현했습니다. 테스트 타임 컴퓨트(test-time compute), 즉 AI가 답을 생성하는 시점에서 추가적인 연산을 투입하는 접근법은 알 파고 시절부터 딥마인드가 개척해온 방법론입니다.

바둑에서 수를 두기 전에 수천 가지 시나리오를 시뮬레이션했던 것과 같은 원리가 이제 언어 모델에도 적용되고 있습니다. 허사비스의 50퍼센트 예측에는 중요한 유보가 있습니다. 나머지 50퍼센트, 즉 2030년까지 AGI가 도래하지 않을 확률이 같다는 것입니다.

그 이유도 분명합니다. 현재의 시스템에는 추론, 기억, 물리적 세계 이해, 그리고 창의적 발명이라는 근본적 능력이 결여되어 있기 때 문입니다. 2025년 1월 알렉스 칸트로위츠와의 인터뷰에서 허사비스는 "올해 어떤 연구소 도 AGI에 도달하지 못할 것"이라고 단언하면서, 기술을 과대 포장하는 이들을 비판했습니다. 노벨상 수상자이자 업계 최전선의 연구자인 그의 이 발언은, 과열된 AGI 논쟁에서 드 문 절제의 목소리입니다.

2025년 12월, 허사비스의 예측은 한 가지 점에서 업데이트되었습니다. 다보스의 무대에서 아모데이가 "2026~2027년"을 제시했을 때, 허사비스는 "우리의 의견 차이가 크지 않다"고 응답했습니다. 그러면서 "나의 타임라인이 약간 더 길다"고 덧붙였습니다.

이 미묘한 접근은 경쟁의 압력과 과학적 신중함 사이에서 균형을 잡으려는 시도입니다. 시장은 빠른 타임라인을 원하고, 투자자들은 확신을 원하지만, 연구의 현실은 예측을 배반할 수 있습니다. 결국 허사비스의 50퍼센트 예측이 말하는 것은 이것입니다.

AGI는 올 것입니다. 문제는 "만약(if)"이 아니라 "언제(when)"입니다. 그 '언제'가 5년 뒤일 수도, 10년 뒤일 수도, 혹은 그보다 더 오래 걸릴 수도 있습니다.

"트랜스포머 수준 또는 알파고 수준의 돌파구가 한두 개 더 필요하다" 2025년 5월 구글 I/O에서 허사비스는 한 문장으로 AGI까지의 거리를 요약했습니다. "AGI에 도달하려면 한두 개의 새로운 돌파구가 더 필요할 수 있습니다(to get all the way to something like AGI may require one or two more new breakthroughs)." 7월 렉스 프리드먼 팟캐스트에서 이 생각을 반복했습니다.

"범용 지능에서 기대할 수 있는 전 분야에 걸친 일관성을 얻으려면, 추론, 기억, 그리고 아마도 월드 모델 같은 아이디어에서 한두 가지가 더 필요합니다." 이 발언은 겸손한 듯 보이지만, 그 안에 담긴 함의는 거대합니다. 허사비스가 참조하는 두 가지 역사적 선례가 있습니다.

하나는 2017년 구글 브레인 팀이 발표한 트랜스포머(Transformer) 아키텍처입니다. "Attention Is All You Need"라는 논문 한 편이 자연어 처리를 완전히 뒤바꿨고, GPT 시리즈와 제미니를 포함한 현대 AI 모델의 기초가 되었습니다. 다른 하나는 2016년의 알파고입니다.

심층 신경망과 강화학습과 몬테카를로 트리 탐색의 결합이라는, 기존에 없던 조합이 바둑이라는 난공불락의 게임을 정복했습니다. 두 경우 모두 점진적 개선이 아니라, 기존 패러다임을 뛰어넘는 질적 도약이었습니다. 허사비스가 "한두 개 더 필요하다"고 말할 때, 그 돌파구의 후보는 어디에 있을까요. 그가 반복적으로 언급하는 영역이 세 가지 있습니다.

첫 번째는 추론(reasoning)입니다. 현재의 대형 언어 모델은 패턴 인식에 기반한 응답을 생성합니다. 수학 문제를 풀고 논리적 추론을 수행하는 것처럼 보이지만, 그 과정이 인간의 논리적 사고와 같은 방식으로 작동하는지는 불분명합니다.

딥마인드의 딥싱크는 이 문제에 대한 한 가지 접근법입니다. 여러 추론 경로를 병렬로 탐색하고 상호 검증하는 방식은 알파고가 바둑 수를 탐색하던 방법의 확장입니다. 허사비스는 이것을 "돌파구의 일부"라고 표현했습니다.

전체는 아니라는 뜻입니다. 추론의 다음 단계, 즉 새로운 가설을 스스로 생성하고 검증하는 능력은 아직 도달하지 못한 영역입니다. 두 번째는 기억(memory)입니다. 현재의 AI 모델은 대화가 끝나면 그 내용을 잊습니다.

모든 상호작용이 백지 상태에서 시작됩니다. 인간의 지능은 경험을 축적하고, 과거의 학습을 새로운 상황에 적용하는 지속적 학습(continual learning) 위에 세워져 있습니다. 이 격차를 메우려면 에피소드 기억, 시맨틱 기억, 절차적 기억을 통합하는 새로운 아키텍처가 필요할 수 있습니다.

허사비스의 신경과학 배경이 바로 이 지점에서 빛납니다. 그의 박사 논문 주 제였던 해마(hippocampus)와 에피소드 기억은 AI가 아직 재현하지 못한 뇌의 핵심 기능입니다. 세 번째는 월드 모델(world model)입니다. 2025년 12월 구글 딥마인드 팟캐스트에서 허사비스는 AGI를 위한 두 가지 전제 조건을 명시했습니다. 그 첫 번째가 월드 모델, 즉 AI가 물 리학과 공간을 진정으로 이해하는 능력입니다. 언어 모델이 이야기를

들려줄 수 있다면, 월드 모델은 환경을 구축할 수 있습니다. 그러나 AI가 현실에 참여 하게 하는 결정적 단계 는 실험입니다. 제니(Genie)가 상황에 따라 실시간으로 시나리 오를 생성하고, 시마(SIMA) 가 그 안에서 과제를 수행하며, 성공이든 실패든 그 결과 가 AI 스스로의 학습 재료가 되는 순환. 이것이 허사비스가 구상하는 인지적 폐쇄 루 프(cognitive closed loop)입니다.

허사비스의 이 진단에 대해 회의적인 시선도 존재합니다. 이펙티브 알트루리즘 (Effective Altruism) 포럼의 2025년 12월 분석은 날카로운 질문을 던집니다. "허사비스가 나열한 미해결 연구 과제들, 계층적 강화학습, 월드 모델, 지속적 학습, 창의적 아이디어 생성 같은 것들은 사실 수십 년 된 문제들입니다. 학술 데이터베이스를 검색하 면 1990년대 초반부터 계층적 강화학습 논문을 찾을 수 있습니다.

이 오래 된 문제들이 정 말 3~5년 안에 풀릴 것이라는 확신의 근거가 무엇인가요?" 이 비판은 중요한 지점을 짚습니다. 현재 AI 산업의 투자 대부분은 대형 언어 모델의 스케일링에 집중되어 있으며, 허 사비스가 필요하다고 말하는 새로운 과학을 발견하는 데는 상대적으로 적은 자원이 투입되고 있습니다.

허사비스는 이 긴장을 "두 가지를 동시에 한다"는 원칙으로 풀어갑니다. 지금 가진 기술을 최대한 키우면서도, 6개월~1년 뒤에 올 다음 혁신을 함께 준비하는 것입니다. 이것은 체스 선수의 사고방식과 같습니다. 당장 눈앞의 한 수를 최선으로 두면서도, 열 수 앞의 판세를 미리 그려두는 것입니다.

알파고에서도 같은 전략이 통했습니다. 이미 아는 수를 끝까지 파고들면서도, 아무도 생각 못한 새로운 수를 과감하게 던져보 는 것입니다. 세르게이 브린은 여기서 흥미로운 역사적 비유를 꺼냈습니다.

"N체 문제 시뮬레이 션의 역 사를 보면, 컴퓨터 성능이 좋아진 것보다 알고리즘이 똑똑해진 것이 더 큰 변 화를 만들었 을 겁니다. 그런데 지금은 둘 다 동시에 좋아지고 있죠." 허사비스가 꼭 필요하다고 말하는 돌파구도 바로 이 알고리즘 쪽입니다. 데이터센터를 더 크게 짓거 나 GPU를 더 많이 쌓는 게 아니라, 지능이 어떻게 작동하는지에 대한 근본적으로 새 로운 깨달음 말입니다.

이 "한두 개의 돌파구"가 언제, 어디서, 어떤 모습으로 나타날지는 아무도 모릅니다. 트랜 스포머 논문이 2017년에 나올 줄 2015년에 안 사람은 없었습니다. 알파고의 37 수를 미리 떠올린 사람도 없었습니다. 허사비스도 이런 예측 불가능성을 인정합니다.

그러면서도 자신의 팀이 "유망한 아이디어를 여럿 준비하고 있고, 이것들을 세미나 이의 본 류에 합류시키고 싶다"고 말합니다. 과학의 역사를 보면, 돌파구는 늘 준비된 사람에게 찾아왔습니다. 허사비스는 그 준 비를 40 년째 해오고 있습니다.

체스판 앞에서 품었던 질문, "이 지능을 더 나은 데 쓸 수는 없을까?" 가 이제 마지막 고비 앞에 와 있습니다. AGI를 향한 로드맵

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

던 킹스크로스에 위치한 구글 딥마인드 본사. 데미스 허사비스는 창밖으로 분주히 움직이는 사람들과 차량, 그리고 바람에 흔들리는 나뭇가지를 내려다보며 깊은 생각에 잠겨 있었습니다. 그의 뒤편 모니터에는 제미니(Gemini)가 작성한 유려한 시(詩)와 복잡한 코딩 결과물이 떠 있었지만, 허사비스의 마음속에는 여전히 채워지지 않은 갈증 이 자리 잡고 있었습니다. 그는 마치 1990년대 초반, 게임 《테마파크》를 개발 할 때 느꼈던 그 근원적인 한계를 다시 마주하고 있는 듯했습니다.

"지금의 AI는 말을 정말 잘하는 앵무새와 같습니다. 확률적으로 가장 그럴싸한 다음 단어를 내뱉지만, 그 단어가 가리키는 '실체'가 무엇인지, 이 세상이 물리적으로 어떻게 작동하는지는 전혀 모릅니다." 허사비스가 지적하는 대규모 언어 모델(LLM)의 가장 치명적인 한계는 바로 '물리적 실체 (Grounding)'와 '계획(Planning)'의 부재였습니다.

챗GPT나 제미니 같은 모델들은 인터넷 상의 텍스트 데이터를 학습해 언어의 패턴을 완벽에 가깝게 모사합니다. 하지만 그들에게 '사과'는 빨갛고 둥근 과일이라는 텍스트 정보일 뿐, 손으로 쥐었을 때의 감각이나 떨어뜨렸을 때의 중력 가속도, 베어 물었을 때의 아삭함이라는 물리적 경험과는 단절되어 있습니다. "수동적 시스템"에서 "에이전트 시스템"으로의 전환 이 한계는 철학의 영역에 머무르지 않습니다.

AI가 텍스트 창을 넘어 현실 세계의 로봇을 제어하거나, 새로운 신소재 발견하기 위한 실험을 설계하려 할 때, '경험의 부재'는 넘을 수 없는 벽이됩니다. 허사비스는 이를 "수동적 시스템(Passive System)"의 한계라고 규정했습니다. 현재의 LLM은 사용자가 질문을 던져야만 비로소 반응합니다.

스스로 목표를 설정하거나, 그 목표를 달성하기 위해 장기적인 계획을 수립하고, 중간에 실패했을 때 전략을 수정하는 능력이 현저히 부족합니다. 2025년 1월, 스위스 다보스에서 열린 세계경제포럼(WEF) 연례회의는 이 문제를 둘러싼 논쟁의 무대가되었습니다. 허사비스는 앤트로픽(Anthropic)의 다리오 아모 데이와 함께 무대에 올라, 현재의 AI 시스템이 인간 수준의 지능에 "아직 한참 미치지

못한다(nowhere near)"고 단언했습니다. 그의 말은 명쾌했습니다. "텍스트만 훈련시켜서는 인간 수준의 지능에 절대 도달할 수 없습니다.

현실 세계가 필요합니다." 같은 행사에서 메타(Meta)를 떠나 AMI 랩스를 설립한 얀 르쿤(Yann LeCun)은 더 날카로운 비판을 내놓았습니다. "변호사 시험은 통과하고 코드도 작성하지만, 현실 세계는 전혀 다루지 못합니다.

그래서 가정용 로봇도 완전 자율주행차도 아직 없는겁니다." AI 업계가 "LLM에 완전히 매몰되었다 (completely LLM-pilled)"는 그의 표현은, 텍스트 중심 접근의 구조적 한계를 압축한 한마디였습니다. "다음 주 휴가 계획을 짜줘"라는 요청을 예로 들어봅시다. LLM은 그럴듯한 일정을 나열할 수는 있지만, 실제로 항공권을 예약하고, 호텔에 전화를 걸어 빈방을 확인하며, 날씨 변화에 따라 일정을 조정하는 '실행' 단계에서는 무력해집니다. 모델이 자신 이 뱉은 말의 인과 관계를 이해하지 못하고, 오직 텍스트의 통계적 연관성만으로 답을 생성하기 때문입니다.

소위 '환각(Hallucination)'이라 불리는 현상도 같은 맥락에서 발생합니다. 모델이 사실 여부를 검증하는 내부 메커니즘 없이 확률 높은 단어를 이어 붙이다가 만들어내는, '그럴싸한 거짓말'입니다. 허사비스는 2025년 다보스 연설에서 이 모순을 날카롭게 짚었습니다.

제미나이는 국제수 학올림피아드(IMO) 수준의 문제를 풀어내면서도, 초등학교 수준의 기하학 문 제에서는 실 수를 범합니다. 놀라운 이미지를 생성하면서도, 왜 컵이 공중에 떠 있으면 안 되는지는 이 해하지 못합니다. 이 간극이야말로 언어 모델이 세상의 물리적 규 칙을 진정으로 '이해'하 지 못한다는 증거였습니다.

고품질 데이터 부족과 AI 발전 속도 둔화 가능성 더욱 심각한 문제는 데이터의 고 갈입니다. 허사비스는 "인터넷에 있는 텍스트 데이터는 거 의 다 썼다"는 사실을 누구보다 잘 알고 있었습니다. 인류가 생성한 텍스트 데이터는 유한 하며, 이를 더 큰 모델 에 쏟아붓는 '스케일링 법칙(Scaling Law)'만으로는 AGI에 도달할 수 없다는 회의론 이 업계 전반에 퍼지고 있었습니다.

허사비스 자신도 "수확 체감의 증거 (evidence of diminishing returns)"가 업계 전반에 나타나고 있다고 인정했습니다. 그는 발전 속도 가 제로에 가까워졌다는 뜻은 아니라고 덧붙였지만, LLM을 단순히 키우는 것만으로 는 한계가 있다는 판단은 분명했습니다. 텍스트만으로는 세상의 모든 지식을 담을 수 없습니다.

장인이 도자기를 빚는 미세 한 손놀 림, 외과 의사의 직관적인 판단, 복잡한 물리 실험의 변수들은 텍스트로 기록

되지 않은 암 묵지(Tacit Knowledge)의 영역입니다. 세상의 공간적 역동성, 공간 인 식, 물리적 맥락이 기 계적으로 어떻게 작동하는지를 "말로 설명하기 어렵다"고 허사비스는 지적했습니다. 언어 에는 인간이 예상했던 것보다 훨씬 많은 세계에 관한 정보 가 담겨 있지만, 그 정보에는 언 어학자들조차 예상치 못한 한계가 있었습니다.

허사비스에게 이 문제는 기술적 난관이 아니라, 그가 평생을 바쳐온 '지능'이라는 화 두의 본질적인 결함이었습니다. 그는 딥마인드의 엔지니어들에게 자주 이렇게 말했습니다. "우 리는 백과사전을 외우는 AI를 만들려는 게 아닙니다. 우리는 낯선 환경에 던 져졌을 때, 스 스로 상황을 파악하고 해결책을 찾아내는 지능을 원합니다."

그것은 그 가 12세 때 체스판 앞에서, 그리고 17세 때 게임 개발자로서 구현하려 했던 '살아있는 지능'이었습니다. 텍스 트라는 1차원적인 데이터를 넘어, 시간과 공간, 그리고 인과율 이 지배하는 3차원 세계를 이해하는 AI. 이것이 허사비스가 바라보는 LLM 이후의 세 상, 즉 '포스트 LLM' 시대의 핵심 과제였습니다. 월드 모델(World Model)과 시뮬레이션의 혁신 "지능이란 무엇인가? 그것은 미래를 시뮬레이션하는 엔진이다."

허사비스가 유니버시티 칼리지 런던(UCL)에서 해마 (Hippocampus)와 상상력을 연구하며 얻은 결론은 딥마인드의 기술 로드맵에 고스란 히 이식되었습니다. 그는 인간의 뇌가 과거 의 기억을 재조합하여 미래의 시나리오를 머릿속으로 시뮬레이션하고, 그중 최적의 행동을 선택한다는 사실에 매료되었습니다. AI가 진정한 지능을 갖추기 위해서는 방대한 데이터를 검색하는 것이 아니라, 뇌처럼 세상의 작동 원리를 내재화한 '월드 모델(World Model)' 을 가져야 한다고 그는 확신했습니다.

제미나이를 월드 모델로 확장: 세계를 시뮬레이션하고 계획하는 AI 2025년 12월 16일, 구글 딥마인드 팟캐스트 시즌 마지막 에피소드에서 허사비스는 수학자이자 과학 커뮤니케이터 한나 프라이(Hannah Fry)와 50분이 넘는 대화를 나누었습니다. 제 품 출시 행사도 리뷰도 아닌, 앞으로 10년을 내다보는 대화였습니다. 허사비스는 AGI에 도 달하기 위해 반드시 이루어야 할 두 가지를 제시했습니다.

하나는 AI가 물 리와 공간을 진 정으로 이해하게 해줄 월드 모델이고, 다른 하나는 AI가 소재나 핵융 합 같은 근본적 문제를 직접 실험을 통해 풀 수 있게 해줄 자동화 실험입니다. 더 중 요한 것은, 이 둘이 연결되 어 하나의 폐쇄 루프(Closed Loop)를 형성해야 한다는 점이었습니다. AI가 스스로 질문을 던지고, 검증하고, 반복하는 과학 연구의 순환 체계.

허사비스에게 AGI는 생성형 모델의 종착역이 아니라, 과학 연구 폐쇄 루프의 출발점이었습니다. 월드 모델이란 무엇인가. 허사비스의 정의에 따르면, 그것은 AI가 물리적 현실에 대 해 가 지는 직관적 이해 능력입니다. 무엇이 쏟아질 수 있고, 무엇이 움직일 수 있 으 며, 사물이 어 떻게 변하고, 공간이 어떻게 구조화되어 있으며, 시간이 어떻게 흐르는 지를 파악하는 것. 허사비스는 월드 모델이 언제나 자신의 핵심 관심사였다고 밝혔습니다.

새로운 아이디어 가 아닙니다. 하지만 2025년에 이르러, 이것은 반드시 해내야 할 과제가되었습니다. 제니(Genie)와 시마(SIMA): 무한 훈련 루프와 물리적 AI의 기반 이 비전의 가장 강력한 증거는 '제니(Genie)'와 '시마(SIMA)' 프로젝트였습니다. 2024년 처 음 공개된 제니는 인터넷상의 2D 플랫폼어 게임 영상을 학습하여, 정지된 이미지나 간단 한 스 케치만으로 플레이 가능한 가상 세계를 생성해 내는 모델이었습니다.

그리고 2025년 8월, 제니 3(Genie 3)가 등장하면서 차원이 완전히 달라졌습니다. 제니 3는 텍스트 설명이나 단 한 장의 이미지만으로 실시간 탐색이 가능한 사실적 인 3D 환 경을 생성합니다. 사막에서 바다까지 자연 세계를 경험하고, 극한 기상 현상을 가까이서 목격하며, 상상 속의 세계와 환상적인 시나리오를 만들어냅니다. 해상도 720p, 초당 20에 서 24프레임의 유려한 영상. 겉보기에는 게임 생성기처럼 보이지만, 그 본질은 '물리학의 학습'에 있습니다.

제니 3는 캐릭터가 점프하면 다시 땅으로 떨어져야 하고, 벽에 부딪히면 더 이상 나 아갈 수 없다는 세상의 규칙을 픽셀의 변화만 보고 스스로 깨우쳤습니다. 구글 딥마인 드의 공식 설 명에 따르면, "월드 모델은 물리적 환경에 대한 깊은 이해를 활용하여 그 것을 시뮬레이션합니다. 제니 3는 에이전트가 세계가 어떻게 진화하는지, 자신의 행 동이 그것에 어떤 영향을 미치는지를 예측할 수 있게 해주는, AGI를 향한 핵심 디딤 돌"입니다.

여기에 더해 '시마 2(SIMA 2, Scalable Instructable Multiworld Agent)'는 이 가상 세계 안에서 목표를 가지고 행동하는 에이전트였습니다. 2024년의 시마 1이 "왼쪽으 로 돌아라", "사다 리를 올라라" 같은 600여 개의 간단한 명령을 따르는 수준에 그쳤다 면, 2025년 12월에 공개 된 시마 2는 완전히 다른 존재였습니다. 제미나이 파운데이 션 모델을 핵심 추론 엔진으로 탑재한 시마 2는 단계별 지시를 기다리는 수동적 추종

자에서, 높은 수준의 목표를 추론하 고 사용자와 대화하며 전략을 논의하는 능동적 동 료로 진화했습니다. 시마 1의 복합 과제 성공률이 약 31퍼센트에 불과했던 반면, 시마 2는 인간 플레이어의 성과와의 격차를 크게 좁혔습니다. 허사비스가 이 두 프로젝트의 결합에서 본 것은 '무한 훈련 루프(In nite Training Loop)'의 가능성이었습니다.

제니가 세계를 창조하고, 시마가 그 세계를 경험합니다. 허사비스의 표 현을 빌리면, "시마 에이전트가 배우고 싶은 것이 무엇이든, 제니가 즉 석에서 만들어낼 수 있습니다." 흥미로운 점은 이 두 AI가 서로의 정체를 모른다는 것입니다.

제니는 시마가 다 른 AI라는 사실을 알지 못합니다. 그저 플레이어로 취급합 니다. 시마 역시 자신이 경험하 는 세계가 AI에 의해 생성되었다는 사실을 모릅니다. 과제 완수에만 집중합니다. 이 무지(無 知) 속의 협력이 잠재적으로 무한히 확장 가능 한 훈련 사이클을 만들어냅니다. 시마 2의 자기 개선(Self-Improvement) 메커니즘은 이 루프의 핵심입니다.

먼저 제미나이가 과제와 보상 기준을 제안합니다. 시마 2가 환경 속에서 시행착오를 거칩 니다. 성공과 실패 의 경험이 기록됩니다. 그 경험이 다음 세대 시마 2의 훈련 데이터 가됩니다.

인간의 시범 없이도 새로운 게임에서 자기 주도적 플레이만으로 기술을 발 전시킬 수 있었고, 심지어 제 니 3가 생성한 완전히 새로운 환경에서도 이 자기 개선 이 작동했습니다. 이것은 알파고가 자기 대국(Self-play)을 통해 초인적 바둑 실력에 도달했던 원리의 확장판이되, 바둑판이라 는 닫힌 공간이 아니라 무한히 다양한 3D 세계로 무대를 넓힌 것이었습니다. Veo 비디오 모델과 물리학 벤치마크: "환각"에서 "기반된 물리학"으로 2024년 공개 된 비디오 생성 모델 'Veo(비오)' 역시 이 그림의 일부였습니다.

허사비스는 Veo 가 생 성한 영상들이 물리 법칙을 얼마나 잘 따르는지를 AGI의 척도로 삼았습니다. 기존의 비디오 모델들이 컵이 공중에 등등 떠 있거나 사람의 손가락이 여섯 개로 변하는 '환 각'을 보였다면, Veo는 빛과 그림자, 중력과 관성을 이해하고 반영하려 노력했습니다. 하지만 허 사비스는 솔직했습니다.

현재의 비디오 모델들은 "육안에는 사실적으로 보 이지만, 아직 ' 물리학 수준(physics-grade)'에는 미치지 못한다"고 인정했습니다. 이 간극을 메우기 위해 딥마인드는 물리학 벤치마크를 구축하고 있었습니다. 진자 운동, 공 굴리기 같은 고등학교 수준의 물리 실험을 시뮬레이션 안에서 재현하고, AI 가 생성한 결과가 뉴턴의 운동 법칙을 100퍼센트 정확하게 따르는지를 검증하는 체계입니다.

챗봇에서 환각은 버그에 불과하지만, 물리적 AI에서 환각은 재앙이 될 수 있습니다. 로봇이 컵의 무게를 잘못 예측하면 컵은 깨지고, 자율주행차가 도로의 곡률을 오판하면 사고가 발생합 니다. "물리학에 기반한(Physically grounded) 상상력"이야말 로 허사비스가 추구하는 기술 적 지향점이었습니다. 신경과학에서 시작된 "마음의 시뮬레이션 엔진" 개념의 기술적 실현 모든 뿌리는 허 사비스의 신경과학적 배경, 즉 "마음의 시뮬레이션 엔진(Simulation Engine of the Mind)"을 컴퓨터 안에 구현하려는 거대한 실험에 닿아 있습니다.

UCL에서 해마 손 상 환자들을 연구하며 발견한 사실 , 기억을 잃은 사람은 미래를 상상하지도 못한다는 것 , 은 허사비스에게 지능의 본질을 가르쳐주었습니다. 기억은 과거를 저장하는 창고가 아니라, 미래를 시뮬레이션하는 원재료입니다. 뇌 는 과거 의 조각들을 재배열하여 아직 일어나지 않은 상황을 머릿속에 펼쳐놓고, 그중에서 최선의 행동을 고릅니다. 제니 3와 시마 2의 결합, Veo의 물리학 벤치마크, 그리고 제미나이의 멀티모달 확 장은 모두 이 신경과학적 통찰의 공학적 번역입니다.

허사비스는 AI가 이 월드 모델을 통해 현실 세계의 복잡한 문제들을 미리 시뮬레이션해보고, 현실에서의 시행착오 비용을 획기적으로 줄일 수 있을 것이라 믿습니다. 기후 변화의 예측, 신약 후보 물질의 분자 결합, 핵융합 반응의 제어. 이 모든 것을 AI가 먼저 가상 세계에서 수천만 번 시도해본 뒤, 검증된 결과만 현실에 적용하는 것입니다. 2025년 12월, 딥마인드는 영국 정부와의 협력을 통해 2026년 영국에 최초의 완전 자동화 과학 연구소를 개소할 계획을 발표했습니다.

제미나이와 완전히 통합된 이 연구소에서는 매일 수백 종의 소재를 합성하고 테스트하며, 실험 수행과 데이터 분석, 방향 조정을 AI와 로봇이 주도합니다. 제니와 시마가 디지털 세계에서 형성하는 가상 폐쇄 루프와, 이 자동화 실험실이 물리적 세계에서 수행하는 실험 폐쇄 루프. 이 두 개의 병렬적인 자율 연구 시스템이 허사비스가 구상하는 AGI의 골격입니다. 게임 개발자였던 소년이 과학자가 되어, 이제는 전 지구적인 문제를 해결하기 위해 AI에게 '세상을 시뮬레이션하는 법'을 가르치고 있습니다.

허사비스의 표현을 빌리면, "과거에는 답을 내놓는 AI를 훈련했지만, 이제는 연구하는 AI를 훈련해야 합니다." 이 것이 그가 그리는 AGI로 가는 경로입니다. 더 큰 모델도 더 강한 컴퓨팅 파워도 아닌, AI가 세상을 진정으로 '이해'하고 세상을 '변화'시키는 능력에 기반한 경로.

추론(Reasoning)과 기억(Memory)을 갖춘 에이전트의 미래 2024년 12월 11일, 노벨 화학상 시상식을 마치고 스웨덴에서 돌아온 직후, 허사비스는 엑시오스(Axios)와의 인터뷰에서 다음 12개월의 핵심 과제를 하나로 압축했습니다. "월드 모델. 그것이 다음 전선입니다." 하지만 월드 모델만으로는 충분하지 않았습니다.

AI가 세상을 시뮬레이션할 수 있다 하더라도, 그 시뮬레이션 결과를 바탕으로 논리적으로 추론하고, 과거의 경험을 기억하여 현재의 판단에 반영할 수 없다면 여전히 반쪽짜리 지능에 불과합니다. 2~3년 내 "정말 인상적이고 신뢰할 수 있는" 에이전트 시스템 예측 허사비스는 2025년과 2026년을 "AI가 수동적 도구에서 능동적 동료로 진화하는 변곡점"으로 정의했습니다. 여러 인터뷰와 내부 회의를 통해 향후 2~3년 내에 "정말 인상적이고 신뢰할 수 있는 에이전트 시스템"이 등장할 것이라고 예고했습니다. 그가 말하는 에이전트는 사용자의 질문에 답하는 수준을 넘어섭니다. "내년도 예산을 고려해서 이번 휴가 계획을 짜고, 비행기와 숙소를 예약한 뒤, 맛집 리스트를 내 구글 맵에 저장해 줘"라는 복합적인 명령을 수행할 수 있는 존재입니다.

구글 딥마인드는 이 방향으로 발걸음을 떼고 있었습니다. 2024년 12월 제미나이 2.0 출시와 함께 공개된 프로젝트 마리너(Project Mariner)는 크롬 브라우저 안에서 웹사이트를 넘나들며 사용자를 대신해 과제를 수행하는 에이전트 연구 프로토타입이었습니다. 화면의 텍스트, 코드, 이미지, 양식 등 웹 요소를 인식하고 추론하여 행동합니다. WebVoyager 벤치마크에서 83.5퍼센트의 최고 성과를 기록했습니다.

물론 아직 느리고 완벽하지 않았지만, 허사비스의 표현대로 "이제 기술적으로 가능해지기 시작했습니다." 프로젝트 아스트라(Project Astra)는 비디오 이해, 화면 공유, 기억 기능을 갖춘 범용 AI 어시스턴트의 비전을 보여주었고, 코딩 에이전트 줄스(Jules)는 개발자들의 코드 작업을 보조했습니다. 허사비스는 이 모든 것의 밑에 깔린 사상을 한 문장으로 요약했습니다.

"처음부터 우리는 에이전트 기반 시스템을 향해 작업하고 있었습니다. 시작할 때부터, 그것들은 계획을 세우고, 행동을 실행하며, 목표를 달성할 수 있었습니다." 이를 위해서는 두 가지 핵심 능력, 즉 '추론(Reasoning)'과 '기억(Memory)'이 필수적이었습니다.

먼저 '기억'의 문제. 허사비스는 현재 LLM이 가진 '단기 기억상실증'을 해결하는 것을 최우선 과제로 삼았습니다. 대화가 길어지면 앞의 내용을 잊어버리거나 문맥을 놓치는 현상은 AI를 장기적인 파트너로 삼는 데 치명적입니다. 딥마인드는 제미나이

1.5 프로부터 100만 토큰 이상의 거대한 컨텍스트 윈도우(Context Window)를 도입했습니다. 이는 해리포터 시리즈 전권을 한 번에 머릿속에 넣고 있는 것과 같은 규모입니다. 제미나이 2.5 프로에 이르러서는 3시간 분량의 비디오 콘텐츠를 처리할 수 있게되었습니다.

하지만 허사비스는 많은 양을 입력받는 것을 넘어, 중요한 정보를 선별하여 장기 기억 저장소에 보관하고 필요할 때 적절히 꺼내 쓰는, 인간의 해마와 같은 '일화적 기억 (Episodic Memory)' 시스템을 AI에 구현하려 하고 있습니다. 2025년 다보스에서 그가 제시한 AGI가 지의 연구 과제 목록에는 "더 나은 장기 기억(better long-term memory)"과 "지속적 학습(the ability to learn continuously)"이 명시적으로 포함되어 있었습니다. 다음은 '추론'입니다.

행동경제학자 대니얼 카너먼이 분류했듯, '시스템 1'이 직관적이고 빠른 사고라면, '시스템 2'는 논리적이고 느린 사고의 과정입니다. 현재의 챗봇 들은 질문을 받자마자 즉시 답변을 쏟아내는 시스템 1에 가깝습니다. 허사비스는 딥 마인드의 강화학습 노하우를 접목하여, AI가 답변을 내놓기 전에 내부적으로 "생각할 시간"을 갖게 하려했습니다.

이 구상은 현실이되었습니다. 2025년 3월 출시된 제미나이 2.5 프로는 '사고 모델 (Thinking Model)'이라는 공식 명칭을 달았습니다. 답변을 생성하기 전에 여러 논리적 경로를 시뮬레이션하고, 스스로 오류를 검증한 뒤 최적의 답을 내놓는 내적 사고 과정을 거치도록 설계된 것입니다.

알파고가 다음 수를 두기 위해 수많은 가능성을 탐색했던 것처럼, 이 모델은 답변 전에 내부적으로 사고의 나무(Tree of Thoughts)를 펼칩니다. 허사비스가 팟캐스트 인터뷰에서 설명했듯, "대규모 모델을 점점 더 정확한 세계 예측기로, 점점 더 신뢰할 수 있는 월드 모델로 만드는 것은 필요하지만 아마도 불충분한 요소입니다. 그 위에 알파제로 같은 계획 메커니즘을 올려놓아야 합니다." 제미나이 2.5 프로는 LMArena 리더보드 1위에 오르며 6개월 넘게 최고 자리를 유지했습니다. 국제수학올림피아드 수준의 수학 문제, 경쟁 프로그래밍, 박사 수준의 과학 추론에서 최고 성과를 기록했습니다.

그리고 2025년 11월, 제미나이 3 프로가 등장했습니다. 구글 딥마인드의 공식 발표에 따르면, 제미나이 3는 "깊이와 뉘앙스를 파악하도록 설계된, 추론에서 최첨단인

모델"입니다. 제미나이 3 딥 씩크(Deep Think) 모드는 추론 성능을 한 차원 더 끌어올렸습니다. 허사비스의 말에 따르면, 제미나이 3의 목표는 "가장 뛰어난 만능 모델"을 만드는 것이었고, 이전 버전의 코딩, 추론, 수학 영역 약점을 해결하는 데 집중했습니다.

프로토-AGI에서 완전한 AGI까지의 단계적 로드맵 허사비스는 추론과 기억, 월드 모델이 통합된 상태를 '프로토-AGI(Proto-AGI)'라고 불렀습니다. 완전한 AGI로 가기 위한 마지막 관문입니다. 그가 제시하는 단계적 로드맵은 명확합니다. 1단계는 현재와 같은 '챗봇' 형태, 2단계는 제한된 범위 내에서 도구를 사용하는 '추론가 (Reasoner)', 3단계는 장기적인 목표를 가지고 물리적·디지털 세계에서 자율적으로 행동하는 '에이전트(Agent)', 그리고 마침내 4단계는 인간의 개입 없이 스스로 과학적 가설을 세우고 검증하여 새로운 지식을 창출하는

'전문가(Expert/Scientist)' 단계입니다. 2025년 다보스에서 아모데이와 함께한 공동 발표에서, 허사비스는 10년 이내에 AGI가 달 성될 확률을 50퍼센트로 보았습니다.

다만 현재와 같은 형태의 AI 시스템 만으로는 불가능 하다고 분명히했습니다. "아마 한두 개의 돌파구가 더 필요합니다." 그가 식별한 핵심 격 차는 소수의 사례만으로 학습하는 능력, 지속적 학습 능력, 더 나 은 장기 기억, 그리고 향상 된 추론 및 계획 능력이었습니다.

이 목록은 놀라울 정도로 구체적이었고, 동시에 허사비 스가 딥마인드의 모든 프로젝트를 정확히 이 격차를 메 우는 방향으로 설계하고 있음을 보 여주었습니다. 허사비스는 이 여정의 궁극적 목적 지를 한 마디로 요약했습니다. "우리는 지금 스마트폰을 터치하는 시대에서, AI에게 의도를 말하는 시대로 넘어가고 있습니다."

그의 비전 속에서 미래의 에이전트는 개 개인의 삶을 이해하는 완벽한 '범용 어시스턴트(Universal Assistant)'이자, 인류의 난 제를 함께 고민하는 연구 파트너입니다. 그는 동시에 이 강력한 에이전트가 가져올 위 험성에 대해서도 경고합니다. 잘못된 목표 설 정이나 권한 오남용의 문제입니다.

구글 은 프로젝트 마리너를 개발하면서 "새로운 유형의 위험과 완화책에 관한 능동적 연구를 수행하며, 인간을 루프 안에 유지한다(keeping humans in the loop)"는 원칙을 세 웠습니다. 기술적 진보만큼이나 '정렬(Alignment)'과 '통제 가능성' 연구가 병행되어야 합니다. 시마 2가 자기 개선 능력을 갖게 되면서 이 문제는 더 첨예해졌고, 딥마인드 는 책임 있는 개발 및 혁신 팀(Responsible Development and Innovation Team)과 함께 개발의 모든 단계를 함께했습니다.

허사비스의 AGI로의 경로는 더 큰 모델이나 더 강한 컴퓨팅 파워에 의존하지 않습니다. AI 가 세상을 진정으로 '이해'하고 세상을 '변화'시키는 능력에 기반합니다. 월드 모델이 토대 이고, 추론과 기억이 기둥이며, 자

동화 실험이 지붕입니다. 이 건축물의 설계자는 4세에 체 스를 배우고, 17세에 게임을 만들고, 30대에 뇌를 연구한, 다양한 지적 경험의 소유자입니 그에게 AGI는 똑똑한 기계가 아닙니다. 인류의 지성을 확장하고 보호하는 '지혜로운 도 구'가 되어야 합니다. AGI가 인류 역사에서 가장 변혁적인 순간이 될 것이라는 그의 확신 속에는, 그 변 혁이 올바른 방향으로 가야 한다는 과학자의 책임감이 깃들어 있습니다. 트랜스포머 모델의 작동 원리

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

가 선정한 '세계에서 가장 영향력 있는 100인' 인터뷰에서 허사비스는 이렇게 말했습니다. "이 기술이 완성되면, 모든 인간의 질병은 과거의 일이 될 것입니다." 노벨상 수 상자가 전 세계 독자를 향해 던진 이 문장은 허풍이 아니었습니다.

그 뒤에는 이미 움 직이고 있는 거대한 기계가 있었습니다. 알파폴드가 열어젖힌 문은 생각보다 넓었습니다. 2020년 CASP14에서 원자 수준 의 정확도를 달성한 뒤, 알파폴드2는 지구상에 알려진 2억 개 이상의 단백질 3차원 구조를 예측하는 데 성공했습니다.

한 명의 박사과정 학생이 하나의 단백질 구조를 밝 히는 데 평균 5년이 걸 린다는 점을 감안하면, 이 작업을 전통적 방식으로 완료하려면 10억 년의 박사과정 시간이 필요했을 것입니다. 알파폴드는 그 시간을 몇 달로 압축했습니다. 2025년 기준으로 전 세계 300만 명 이상의 연구자가 알파폴드 데이터베이스를 활용하고 있습니다.

이것은 인류 가 자기 몸의 설계도를 처음으로 손에 넣었다는 뜻입니다. 허사비스는 여기서 멈추지 않았습니다. 2021년, 그는 알파벳(구글 모회사) 산하에 아이소 모픽 랩스(Isomorphic Labs)라는 별도의 회사를 설립했습니다.

이름부터 의 미심장합니다. '아이소모픽'은 수학에서 '구조가 같다'는 뜻입니다. 생물학의 문제와 정보 과학의 문제가 근본적으로 같은 구조를 공유한다는 허사비스의 오랜 직관을 회 사 이름에 새긴 것입니다.

아이소모픽 랩스의 미션은 명확합니다. "AI의 힘으로 모든 질병을 정복한다." 전통적인 신약 개발 과정은 잔인할 정도로 비효율적입니다.

하나의 약을 시장에 내놓기까 지 평균 10년에서 15년이 걸리고, 비용은 20억 달러를 넘기며, 임상시험 단계에서의 실패 율은 90퍼센트에 달합니다. 아이소모픽 랩스의 수석과학 자 마일스 콩그리브는 이 과정을 " 두더지 잡기(Whac-a-Mole)"에 비유했습니다. 화 학자들이 수천 개의 화합물을 합성하고, 하 나씩 시험하고, 대부분 실패하고, 다시 처 음부터 시작하는 과정의 반복이었습니다.

허사비스가 구상한 대안은 이 전체 과정을 '습식 실험실(wet lab)'에서 '실리코(in silico)', 즉 컴퓨터 시뮬레이션 안으로 옮기는 것이었습니다. 알파폴드3는 단백질의 정 적 구조를 넘 어, 단백질과 단백질, 단백질과 소분자(약물), 단백질과 DNA/RNA 사 이의 상호작용까지 예 측할 수 있게 진화했습니다. 허사비스는 이 기술이 신약 개발의 효율을 "1,000배" 높일 수 있다고 말했습니다.

2025년 3월, 아이소모픽 랩스는 스라이브 캐피탈(Thrive Capital)이 주도한 6억 달 러 규모의 시리즈A 투자를 유치했습니다. 노바티스, 일라이 릴리 같은 거대 제약사와 도 총 30억 달러 에 달하는 파트너십을 체결했습니다. 그리고 2026년 1월, 다보스 세 계경제포럼에서 허사비스는 결정적인 발표를했습니다.

AI가 설계한 최초의 항암제 가 2026년 초 1상 임상시험 에 진입한다는 것이었습니다. 아이소모픽 랩스는 종양학, 면역학, 심혈관 질환을 포함하여 17개의 약물 개발 프로그램을 동시에 운영하고 있습니다. 허사비스의 궁극적 꿈은 "가상 세포(virtual cell)"를 만드는 것입니다.

하나의 세포 안에서 일어나는 모든 화학적, 생물학적 역학을 시뮬레이션할 수 있다면, 약물이 인체에 미치는 영향을 환자에게 투여하기 전에 컴퓨터 안에서 미리 확인할 수 있습니다. 더 나아가 개인의 대사 특성에 맞춰 하룻밤 사이에 맞춤형 약물을 설계하는 것도 가능해집니다. 허사비스는 2025년 다보스에서 이렇게 말했습니다. "결국에는 개인 맞춤형 의약품을 상상할 수 있을 것입니다."

AI 시스템이 당신의 개인적 대사에 최적화 된 약을 하룻밤 사이에 설계하는 세상 말입니다." 이 비전이 실현된다면, 의학의 풍경은 근본적으로 달라집니다. 허사비스가 Fortune지와 인터뷰에서 예측한 것처럼, "10년, 15년 후의 의학은 오늘날의 의학과 전혀 다른 모습일 것입니다. 질병은 발생 한 뒤에 치료하는 것이 아니라, 발생하기 전에 예측하고 차단하는 대상이 됩니다."

암은 사형선고가 아니라 관리 가능한 상태가 되고, 희귀 질환 환자는 더 이상 치료제가 없다는 말을 듣지 않아도 됩니다. 물론 장밋빛 전망 뒤에는 현실의 무게가 숨어 있습니다. 노바티스의 바이오메디컬 연구 책임자 피오나 마셜은 "AI가 발견 과정에서 5년을 단축할 수 있을지 모르지만, 인간 안전성 시험을 알고리즘으로 우회할 수는 없다"고 지적했습니다.

아이소모픽 랩스의 사장 맥스 야 더버그도 "소프트웨어가 현실의 과학 프로세스와 만날 때, 겸손해 질 수밖에 없다"고 인정했습니다. 허사비스 자신도 원래 2025년 말까지 임상시험을 시작하겠다고 했지만, 일정은 2026년으로 미뤄졌습니다. 생명을 다루는 일에서 서두

름은 금물이라는 사실을, 이 분야에서 가장 야심찬 낙관주의자조차 인정하지 않을 수 없었습니다. 그럼에도 방향은 분명합니다. 알파폴드에서 시작된 의료 혁명은 이미 되돌릴 수 없는 강을 건넜습니다.

허사비스가 TIME지에 말한 것처럼, "이 기술이 완성되면 모든 질병은 과거의 것이 됩니다." 그 문장이 예언이 될지, 희망에 그칠지는 아직 모릅니다. 분명한 것은, 50년간 풀리지 않던 단백질 접힘 문제를 정복한 사람이 그 다음 목표로 '질병 전체의 정복'을 선언했다는 사실입니다. 그리고 그 선언 뒤에는 6억 달러의 투자금과 30억 달러의 파트너십과 17개의 약물 프로그램이 이미 움직이고 있다는 사실입니다.

청정 에너지(핵융합)와 기후 변화 해결 허사비스가 꿈꾸는 근본적 풍요의 세계에서 질병 퇴치 다음으로 오는 것은 에너지입니다. "에너지가 제로 탄소이고 무료가 된다면, 기후 위기를 초월하고 지구의 생태계를 복원하기 시작할 수 있습니다." 2025년 TIME지 인터뷰에서 허사비스가 한 이야기는 공허한 수사가 아니었습니다.

딥마인드는 이미 수년 전부터 핵융합 연구의 핵심적 난제에 AI를 적용하고 있었습니다. 핵융합은 태양이 빛나는 원리입니다. 수소 원자핵들이 극도의 고온과 고압에서 충돌하여 하나로 합쳐질 때, 질량의 일부가 에너지로 전환됩니다.

아인슈타인의 $E=mc^2$ 가 현실이 되는 순간입니다. 이 반응을 지구 위에서 재현할 수 있다면, 인류는 사실상 무한한 청정 에너지를 얻게 됩니다. 연료는 바닷물에서 추출하는 수소이고, 부산물은 핵분열과 달리 장수명 방사성 폐기물이 아닙니다. 이론적으로 완벽한 에너지원입니다.

문제는 이론과 현실 사이의 거리입니다. 핵융합을 일으키려면 수소를 1억도 이상으로 가열하여 플라즈마 상태로 만들어야 합니다. 이 초고온 플라즈마를 가둬두는 장치에 토카막

(Tokamak)입니다.

도넛 모양의 진공 용기를 강력한 자기 코일로 둘러싸서, 플라즈마가 용기 벽에 닿지 않도록 자기장으로 가두는 원리입니다. 그런데 플라즈마는 본질적으로 불안 정합니다. 초당 수천 번의 속도로 자기 코일의 전압을 조절하 면서 플라즈마의 형태를 유지 해야합니다.

조금이라도 벽에 닿으면 열이 손실되고, 장치가 손상됩니다. 이 제어 문제가 핵융합 상용화를 가로막는 가장 큰 기술적 장벽 가운데 하나였습니다.

2022년 2월, 딥마인드는 Nature에 획기적인 논문을 발표했습니다. 스위스 로잔 연 방공과대 학(EPFL)의 스위스 플라즈마 센터와 협력하여, 심층 강화학습(Deep Reinforcement Learning)으로 토카막 내부의 플라즈마를 자율적으로 제어하는 데 성공한 것입니다. 가변 배치형 토카막(TCV)이라는 실험용 장치에서, AI는 여러 개의 자기 코일을 동시에 조율하며 플라즈마를 원하는 형태로 빚어냈습니다.

길쭉한 형태, 삼각형 형태, 심지어 '눈꽃 (snowflake)' 형태까지. 더 놀라운 것은 하나의 용기 안에서 두 개의 분리된 플라즈마를 동시에 유지하는 데 성공했다는 점입니다. 이것은 강화학습이 적용된 실세계 시스템 가운데 가장 복잡한 사례로 평가받았습니다. 이 성과는 시작에 불과했습니다.

2024년 5월, 딥마인드는 TORAX라는 오픈소스 플라즈마 시뮬레이터를 공개했습니다. 구글의 고성능 수치 연산 프레임워크 JAX 위에 구축된 TORAX는 플라즈마 내부의 열, 전류, 물질 흐름을 빠르고 정밀하게 시뮬레이션합니다. CPU와 GPU 모두에서 작동하며, AI 모델과 매끄럽게 통합됩니다.

그리고 2025년 12월, 결정적인 파트너십이 발표됩니다. 딥마인드와 커먼웰스 퓨전 시스템즈(Commonwealth Fusion Systems, CFS)의 연구 협력입니다. CFS는 MIT에서 2018년에 분사한 핵융합 스타트업으로, 고온 초전도 자석을 활용한 소형 토카막 SPARC를 개발하고 있습니다.

SPARC의 목표는 명확합니다. 역사상 최초로 투입 에너지보다 더 많은 융합 에너지를 생산하는 것, 즉 '순(純) 에너지 생산(net energy)'의 달성입니다. 미국 매사추세츠주 데번스에 건설 중인 SPARC가 성공하면, 그 뒤를 이어 버지니아주에 400메가와트급 상용 발전소 ARC가 2030년대 초 전력망에 연결될 계획입니다.

딥마인드와 CFS의 협력은 세 가지 축으로 진행되고 있습니다. 첫째, TORAX를 활용하여 플라즈마의 빠르고 정확한 시뮬레이션을 구현합니다. SPARC가 가동되기도 전에 수백만 건의 가상 실험을 돌려서, 최적의 운전 조건을 미리 파악합니다.

둘째, 강화학습과 알파에 볼브(AlphaEvolve) 같은 진화적 탐색 기법을 결합하여, 순 에너지 생산을 극대화하는 가장 효율적이고 견고한 경로를 탐색합니다. 셋째, AI 파일럿을 개발하여 자기장 배치를 최적화 하고, 융합 출력을 극대화하며, 열 부하를 관리하는 실시간 제어 전략을 수립합니다. CFS의 물리 운영 선임 관리자 데본 바탈리아는 "TORAX는 전문적인 오픈소스 플라즈마 시뮬레이터로, SPARC를 위한 시뮬레이션 환경을 구축하고 실행하는 데 드는

시간을 셀 수 없이 많이 절약해 주었다"고 평가했습니다. TORAX는 이미 CFS의 일상적 업무 흐름에서 핵심 도구가되었습니다. 구글은 CFS에 대한 투자도 늘렸고, 2030년대 초에 가동될 첫 번

째 ARC 발전소에서 200메가와트의 용합 전력을 구매하는 계약까지 체결했습니다.

세계경제포럼(WEF)의 2025년 12월 보고서에 따르면, MIT의 모델링 연구는 핵융합 발전량이 2035년 2테라와트시에서 2050년 375테라와트시로 급증할 것으로 전망합니다. 미국 에너지부는 2025년 10월 핵융합 로드맵을 발표하며, 2030년대 초까지 핵융합 에너지를 국가 에너지 체계에 편입시킬 수 있다고 밝혔습니다. "핵융합은 현실이고, 가까이 있으며, 조율된 행동을 위한 준비가 되어 있다"는 미국 에너지부 핵융합 에너지과학실 부실장 장 폴 알랭의 선언은, 수십 년간 '30년 뒤'라며 늘 미래로 미뤄지던 핵융합이 마침내 현재의 문법으로 말해지기 시작했음을 보여줍니다. 허사비스에게 핵융합은 에너지 문제의 해결 그 이상입니다.

에너지가 풍부하고 저렴해지면, 그 위에 쌓이는 변화는 연쇄적입니다. 담수화 비용이 급감하여 물 부족 문제가 해소됩니다. 탄소 포집과 재활용이 경제적으로 가능해집니다. 식량 생산의 에너지 집약적 공정이 탈탄소화됩니다. 자원을 둘러싼 지정학적 갈등의 원인이 사라집니다. 허사비스가 말하는 '근본적 풍요'란, 에너지라는 하나의 매듭이 풀리면서 그 아래 연결된 수십 개의 문제가 함께 풀어지는 세계입니다.

AI가 핵융합의 제어 문제를 돕고, 핵융합이 AI를 구동하는 에너지를 공급하는 순환 구조 안에서, 풍요의 기반이 만들어집니다. 우주 탐사와 지능의 확장 2025년 6월, 와이어드(Wired) 매거진과의 인터뷰에서 허사비스는 한 문장을 던졌습니다. "그 모든 것이 실현된다면, 인류가 별을 향해 여행하고 은하를 개척하는, 최대의 번영 시대가 올 것입니다. 그 시작은 2030년이 될 것이라고 생각합니다."

노벨상 수상자의 입에서 '은하 개척'이라는 단어가 나온 순간, 세상의 반응은 둘로 나뉘었습니다. 누군가는 경탄했고, 누군가는 조소했습니다. 허사비스가 말하는 '별을 향한 여행'을 문자 그대로의 성간 항해로만 읽으면 핵심을 놓칩니다. 그의 발언을 전체 맥락 안에 놓아야 합니다.

허사비스의 우주 탐사 비전은 세 겹으로 이루어져 있습니다. 첫 번째 겹은 에너지입니다. 앞서 살펴본 핵융합이 상용화되면, 우주 추진 기술의 판도가 달라집니다. 현재의 화학 로켓은 연료 대비 추력의 효율이 극히 낮아, 화성까

지 가는 데도 7개월 이상 걸립니다. 핵융합 추진 엔진이 개발되면 이 시간은 획기적으로 단축됩니다. AI는 핵융합 엔진의 플라즈마 제어뿐 아니라, 항해 경로의 최적화, 임무 계획의 자동화, 우주선 내부 시스템의 자율 관리까지 담당할 수 있습니다. 인간 우주비행사가 감당할 수 없는 수준의 복잡성을 AI가 처리함으로써, 더 먼 곳까지 더 안전하게 갈 수 있는 가능성이 열립니다. 두 번째 겹은 재료 과학입니다.

딥마인드의 GNoME(Graph Networks for Materials Exploration) 프로젝트는 220만 개 이상의 새로운 결정 구조를 발견했습니다. 이 가운데에는 초전도체 후보 물질, 극한 환경에서 견딜 수 있는 초경량 합금, 우주 방사선을 차단할 수 있는 신소재가 포함될 가능성이 있습니다. 우주선의 구조재, 단열재, 에너지 저장 장치 등 우주 탐사에 필요한 거의 모든 물리적 구성 요소가 새로운 재료의 발견에 의존합니다.

AI가 재료 과학의 탐색 공간을 인간이 평생 시도할 수 있는 범위의 수백만 배로 넓혀놓은 것은, 우주 탐사의 물리적 기반을 근본적으로 바꾸는 일입니다. 세 번째 겹, 그리고 허사비스에게 가장 깊은 의미를 가진 겹은 지능 자체의 확장입니다. 2025년 7월, 렉스 프리드먼(Lex Fridman)

팟캐스트에서 거의 세 시간에 걸쳐 진행된 대화에서, 허사비스는 자신의 가장 근본적인 동기를 드러냈습니다. 그는 어릴 때부터 "우주에서 실제로 무슨 일이 벌어지고 있는가, 의식의 본질은 무엇인가, 현실 자체의 본질은 무엇인가"에 매료되었다고 말했습니다.

그에게 AI를 만드는 것은 제품을 만드는 일이 아니었습니다. "인간 지식을 전진시키기 위한 궁극적 도구"를 만드는 일이었습니다. 허사비스는 흥미로운 관점을 제시합니다. 물리학자 존 아치볼드 휠러가 "잇 프롬 비트 (It from Bit)"이라는 유명한 문구로 표현했던 것처럼, 정보가 에너지나 물질보다 더 근본적인 실재일 수 있다는 것입니다.

이 프레임워크에서 우주 자체는 물리적 '하드웨어' 위에서 돌아다니는 거대한 정보 처리 시스템입니다. AI는 이 우주의 '소프트웨어'를 역설계하는 도구입니다. 알파폴드가 단백질이라는 생명의 코드를 해독했듯이, Veo가 비디오 프레임을 예측하며 물리 세계의 규칙을 학습하듯이, 궁극적으로 AI는 우주를 지배하는 근본 법칙들을 발견하는 데 쓰일 수 있습니다.

이 비전 안에서 우주 탐사란, 로켓을 쏘아 올리는 것만이 아닙니다. 우주를 이해하는 것입니다. 허사비스는 25년간 품어온 꿈을 공개했습니다. 하나의 세포 안에서 일

어나는 모든 화학적, 생물학적 역학을 시뮬레이션하는 것. 거기서 출발하여 생명의 기원, 즉 '원시 수프 (primordial soup)'에서 생명이 어떻게 출현했는지를 시뮬레이션하는 것. 궁극적으로는 AGI가 아인슈타인이 생각했던 수준의 과학적 추측(conjecture)을 스스로 제안할 수 있게 되는 것. 그는 AGI의 완성 여부를 판단하는 자신만의 기준까지 제시했습니다. "시스템을 1900년에 떨어뜨리고, 당시의 물리학 지식만 주고, 상대성 이론을 독립적으로 발명하는지 보라." 칼 세이건은 한때 이렇게 말했습니다.

"의식이 우주를 깨운다." 허사비스는 이 문장을 자신의 방식으로 재해석합니다. 만약 에너지가 풍부하고 저렴해지면, 인류는 우주를 여행하는 종이 될 수 있습니다.

그리고 그 과정에서 AI라는 도구는, 우주를 물리적으로 탐험하는 데 쓰이는 동시에, 우주를 지적으로 탐험하는 데도 쓰입니다. 물리학의 근본 법칙을 이해하고, 의식이란 무엇인지를 묻고, 현실의 본질을 탐구하는 것. 허사비스에게 '은하를 개척한다'는 것은 발로 밟는 것이면서 동시에 마음으로 이해하는 것입니다. 비판자들은 이 비전의 비현실성을 지적합니다.

가장 가까운 항성계까지의 거리를 고려하면, 핵융합 추진으로도 수십 년이 걸립니다. AI가 궤적을 최적화하거나 임무를 자동화할 수는 있지만, 물리 법칙 자체를 뛰어넘을 수는 없습니다. 2025년 애플의 연구 논문은 많은 첨단 AI 모델의 '추론 능력'이 과장되었다고 지적했습니다. '2030년에 은하 개척이 시작된 다'는 허사비스의 말은 문자 그대로의 예측이라기보다, 방향을 가리키는 나침반에 가깝습니다.

허사비스 자신도 이 점을 알고 있을 것입니다. 체스에서 배운 것이 있다면, 최종 목표를 향해 수를 두되 각 수에서는 현실의 제약을 존중해야 한다는 것입니다. 그가 진정으로 말하고 싶은 것은 아마 이것일겁니다. 지능이라는 도구가 충분히 강력해지면, 인류가 바라볼 수 있는 지평선이 지구의 경계를 넘어 확장된다는 것. 그리고 그 지평선을 향해 가는 여정 자체가, 인류가 자기 자신과 자신이 사는 우주를 더 깊이 이해하는 과정이 된다는 것. '지능을 풀면 나머지는 해결된다'는 딥마인드의 창립 미션이, 가장 웅장한 규모로 펼쳐지는 장면입니다.

제로섬(Zero-sum) 게임에서 벗어난 인류의 번영 시가 가져올 새로운 생산성과 부의 분배 문제 허사비스는 2025년 가디언(The Guardian)과의 인터뷰에서 이렇게 말했습니다. "인류 역사 상 처음으로, 상황이 제로섬이 아닌 풍요의 세계에 살게 될 것이라고 생각합니다." 이 문장에는 거대한 낙관이 담겨 있습니다. 동시에 거대한 질문이 숨어 있습니다.

제로섬이란 한쪽이 얻으면 다른 쪽이 잃는 구조입니다. 인류의 역사는 대부분 제로섬의 논리 위에서 전개되었습니다. 한 국가가 영토를 넓히면 다른 국가가 영토를 잃었습니다.

한 기업이 시장을 차지하면 다른 기업이 밀려났습니다. 자원은 유한했고, 권력은 그 유한한 자원의 배분을 둘러싼 투쟁이었습니다. 전쟁, 식민지, 무역 분쟁, 에너지 위기. 모두 희소성에서 비롯된 갈등이었습니다.

허사비스의 주장은 시가 이 희소성의 근본 구조를 바꿀 수 있다는 것입니다. 지능이 무한한 자원이 되면, 그 위에 쌓이는 모든 것의 비용이 급감합니다. 신약 개발이 10배 빨라지면 의료 비용이 떨어집니다. 핵융합이 상용화되면 에너지 비용이 사라집니다. 시가 재료 과학을 혁신하면 물리적 자원의 제약이 완화됩니다.

이 연쇄 반응의 끝에는, 인류가 '누가 얼마를 가져가느냐'를 놓고 싸울 필요가 없는 세계가 있습니다. 파이 자체가 충분히 커지는 세계입니다. 그러나 와이어드의 인터뷰어가 날카롭게 지적했듯이, "서구 세계에는 이미 엄청난 풍요가 있지만, 우리는 그것을 공정하게 분배하지 않습니다."

허사비스도 이를 인정했습니다. "우리는 종으로서, 사회로서, 협력에 능하지 못했습니다. 자연 서식지가 파괴되고 있는 것도 부분적으로는 사람들이 희생을 감수하길 원하지 않기 때문입니다."

이 지점에서 허사비스의 비전은 가장 심각한 도전에 직면합니다. 시가 생산성을 극대화한 다 해도, 그 생산성의 과실이 누구에게 돌아가느냐는 별개의 문제입니다. 구글 딥마인드가 2025년 '시니어 AI 경제학자(Senior AI Economist)' 직위를 채용 공고한 것은 이 문제의 심각성을 보여주는 상징적 사건이었습니다.

이 역할의 핵심 업무는 AGI와 고도화된 AI 시스템이 유발할 경제적 변환을 모델링하는 것이었습니다. 희소성 기반 경제에서 풍요 기반 경제로의 전환을 시뮬레이션하고, 그 과정에서 발생할 불평등과 혼란을 예측하는 것이 과제입니다. 엔트로픽의 CEO 다리오 아모데이는 시가 5년 안에 신입직 일자리의 50퍼센트를 자동화할 수 있으며, 실업률이 10~20퍼센트까지 치솟을 수 있다고 경고했습니다. 링 크드인의 경제기획 책임자 아니시 라만은 기술적 파괴가 커리어 사다리의 가장 아래 단을 먼저 부러뜨릴 것이라고 지적했습니다.

이것은 허사비스 자신이 꿈꾸는 풍요의 세계와 정면으로 충돌하는 시나리오입니다. 풍요가 오기 전에, 전환기의 고통이 먼저 올 수 있기 때문입니다.

허사비스는 이 전환기를 의식하고 있습니다. 2025년 12월 액시오스(Axios) AI 서밋에서 그는 이렇게 말했습니다. "산업혁명보다 10배 더 크고, 10배 더 빠른 변화가 올 것입니다."

산업혁명은 100년에 걸쳐 진행되었고, 그 과정에서 수많은 직업이 사라지고 수많은 새로운 직업이 생겨났습니다. 러다이트 운동이 있었고, 아동 노동이 있었고, 도시 빈민가의 참상이 있었습니다. 그 전환기의 고통은 결국 노동법, 사회 보장 제도, 공교육의 확산으로 완화되었지만,

그에 이르기까지 수십 년의 시간과 수많은 사람의 희생이 필요했습니다.

시가 촉발하는 변화가 이보다 10배 크고 10배 빠르다면, 사회가 적응할 시간은 10 분의 1로 줄어듭니다. 허사비스는 이 사실을 솔직하게 인정합니다. "사회가 AGI에 준 비되어 있지 않 다"고 그는 경고했습니다. "확률 분포의 문제입니다.

하지만 어쨌든 오 고 있고, 아주 빨리 오고 있습니다." 그렇다면 풍요의 과실을 어떻게 나눌 것인가. 허 사비스는 구체적인 정책 처방을 내리는 데 는 신중합니다. 그는 과학자이지 정치인이 아닙니다. 그러나 그의 행동에서 읽을 수 있는 것들이 있습니다.

알파폴드 데이터베이스를 무료로 전 세계에 공개한 결정은, 연구 성과가 소수의 독점이 아니라 인류 전체 의 공유재가 되어야 한다는 철학의 반영입니다. 딥마인드 가 AI 경제학자를 채용한 것은, 기술을 만드는 사람이 그 기술의 경제적 파장까지 책임져 야 한다는 인식의 반영입니다. 2025년 영국 정부와 체결한 포괄적 연구 협력에서 교육용 AI 개발이 포함된 것은, 다음 세대가 이 변화에 적응할 수 있는 도구를 갖추어야 한다는 판 단의 반영입 니다. 풍요는 자동으로 오지 않습니다.

기술이 가능성을 열어놓더라도, 그 가능성을 현실 로 바꾸 는 것은 제도와 정치와 사회적 합의의 몫입니다. 허사비스가 '근본적 풍요'를 말할 때, 그가 정말로 말하고 있는 것은 풍요가 이미 와 있다는 선언이 아닙니다. 풍요 가 기술적으로 가 능해지는 순간이 다가오고 있으니, 그 순간이 도착하기 전에 인류가 분배의 문제를 진지하 게 준비해야 한다는 촉구입니다. "복권에 당첨되면 직장에 돌아갈 사람이 몇 퍼센트일까?"

허사비스는 한 인터뷰에서 이런 질문을 던진 적이 있습니다. "만약 복권에 당첨되면, 직장 에 돌아갈 사람이 몇 퍼센트일까요?" 이 질문은 가볍게 던진 것이지만, 그 안에는 AI 시대 의 가장 깊은 철학적 난제가 들어 있습니다.

일자리가 사라진 뒤에도 인간은 행복할 수 있 는가. 노 동 없는 삶에서 인간은 의미를 찾을 수 있는가.

경제학자들은 AI로 인한 일자리 대체를 숫자로 분석합니다. 몇 퍼센트의 직업이 자 동화될 것인가, 실업률은 얼마나 오를 것인가, GDP에 미치는 영향은 어떠한가. 이 숫 자들은 중요합니다. 그러나 허사비스의 질문은 숫자 너머를 겨냥합니다.

인간이 일을 하는 이유가 돈을 벌기 위해서만이 아니라면, 돈 문제가 해결된 뒤에도 인간은 무언 가를 해야 한다면, 그 '무언가'는 무엇인가. 이 질문은 허사비스 개인의 삶에서도 울림이 있습니다. 그는 17세에 테마파크라는 게임의 리드 프로그래머로 이미 충분한 돈을 벌었습니다. 그 돈으로 케임브리지 대학 학비를 냈습니다.

엘릭서 스튜디오를 팔았을 때도, 다시는 경제적 걱정을 하지 않아 도 될 만큼의 자산을 얻었습니다. 그런데 그는 멈추지 않았습니다. 유니버시티 칼리 지 런던(UCL)에서 신경 과학 박사학위를 받고, 딥마인드를 창업하고, 알파고를 만들 고, 알파폴드를 만들고, 노벨상을 받았습니다.

새벽 3시까지 '야간 교대'를 계속했습니다. 복권에 당첨된 것이나 마찬가지로 지인 사람이, 왜 멈추지 않았을까요. 답은 명확합니다. 허사비스를 움직이는 것은 돈이 아니라 호기심이었습니다.

"나를 항상 이끌어온 것은, 우리 주변의 세계를 이해하려는 열정이었습니다." 60 Minutes 인터뷰에서 한 이 말이 허사비스의 삶 전체를 관통합니다. 체스를 둘 때는 지능의 본질이 궁금했고, 게임을 만들 때는 가상 세계의 규칙이 궁금했고, 신경과학을 공부할 때는 뇌가 어떻게 기억하고 상상하는지가 궁금했고, AI를 만들 때는 지능이라는 현상 자체를 풀 수 있는지가 궁금했습니다.

호기심은 돈으로 살 수 없고, 자동화할 수 없습니다. 그러나 모든 사람이 허사비스는 아닙니다. 복권에 당첨된 사람들의 실제 통계를 보면, 상 당수가 몇 년 안에 이전보다 더 불행해집니다. 목적 없는 풍요는 공허함을 낳습니다. 만약 AI가 대부분의 생산적 노동을 대체한다면, 수십억 명의 사람들은 무엇으로 하루를 채울 것인가. 넷플릭스를 보고, 게임을 하고, 소셜 미디어를 스크롤하는 것으로 충분할까요. 허사비스는 이 질문의 무게를 알고 있습니다.

그는 이것을 "경제적 분배를 넘어선 존재론 적 질문"이라고 표현합니다. 일자리가 사라지는 것은 소득의 문제이면서 동시에 정체성의 문제입니다. "저는 교사입니다", "저는 의사입니다", "저는 엔지니어입니다"라고 말할 때, 사람들은 자신의 직업을 통해 자기가 누구인지를 정의합니다. 그 직업이 사라지면, 자기 정의의 근거도 함께 흔들립니다.

허사비스의 낙관론은 이 지점에서 특유의 뉘앙스를 갖습니다. 그는 AI가 일자리를 없앤다고 보지 않습니다. 일자리를 바꾼다고 봅니다.

"새로운 도구나 기술이 등장하면, 일반적으로 그 도구를 활용하는 새로운 일자리가 만들어지고, 실제로 그 일자리는 더 나은 경우가 많습니다"라고 그는 말했습니다. 의료 분야를 예로 들며, AI가 의사를 대체하는 것이 아니라 의사를 보조하는 도구가 될 것이라고 강조했습니다. "로봇 간호사를 원하는 사람은 없을 것입니다."

돌봄의 인간적 공감이라는 측면은 지극히 인간적인 것입니다." 이 답변은 진실의 일부를 담고 있지만, 전체를 담고 있지는 않습니다. 과거의 기술 혁명에서 새로운 일자리가 생겨난 것은 사실이지만, 그 전환 과정에서 한 세대 전체가 고통받은 것도 사실입니다.

증기기관이 방직공의 일자리를 앗아갔을 때, "걱정 마세요, 나중에 자동차 정비공이라는 새로운 직업이 생길 겁니다"라는 말은 당장 가족을 먹여야 하는 사람에게 아무런 위안이 되지 못했습니다. 허사비스의 진정한 통찰은 다른 곳에 있습니다. 그가 자신의 삶을 통해 보여준 것은, 인간의 가장 깊은 동기가 경제적 보상이 아니라 호기심과 창조와 의미의 추구라는 사실입니다. 게임을 만들던 10대 소년이 돈 때문에 코드를 짰 것이 아니듯, 새벽 3시에 논문을 읽는 49세의 과학자가 월급 때문에 연구하는 것이 아니듯, 인간에게는 돈을 넘어선 동기가 존재합니다.

문제는 그 동기를 모든 사람이 발견할 수 있는 환경을 만드는 것입니다. 교육이 직업 훈련이 아니라 호기심의 배양이 되어야 합니다. 사회 안전망이 실업자를 구제하는 것이 아니라 탐색하는 시간을 보장하는 것이 되어야 합니다.

의미는 위에서 주어지는 것이 아니라, 각자가 스스로 찾아야 하는 것이 되어야 합니다. "복권에 당첨되면 직장에 돌아갈 사람이 몇 퍼센트일까?" 이 질문에 대한 올바른 답은 퍼센트의 숫자가 아닙니다. 직장에 돌아가지 않는 사람들이 무엇을 하느냐에 달려 있습니다. 소파에 누워 화면을 응시하느냐, 아니면 자신만의 체스판을 찾아 탐구하느냐. 근본적 풍요의 시대가 유토피아가 될지 디스토피아가 될지는, 기술의 수준이 아니라 인간의 성숙도에 달려 있습니다.

이것이 허사비스의 질문이 진짜로 묻고 있는 것입니다. AI 개발: "대담하지만 책임감 있게(Bold and Responsible)" 과학적 발견의 가속화: 인류가 우주를 이해하기 위한 궁극적 도구 2025년 1월 다보스 세계경제포럼에서 허 사비스는 "AI를 과학에 적용하는 것은 언어 모델보다 훨씬 풍부합니다"라고 말했습니다. 이 한 문장에 그의 핵심 철학이 담겨 있습니다.

챗 봇이 인간처럼 대화하는 것은 AI가 할 수 있는 일의 극히 일부에 불과합니다. 허사비스가 생각하는 AI의 본질적 가치는 과학적 발견의 속도를 근본적으로 바꾸는 데 있습니다. 이 생각의 출발점은 어린 시절로 거슬러 올라갑니다. 12세 때 리히텐슈타인 체스 대회에서 10시간 동안 대국을 치른 뒤, 소년 허사비스는 "이렇게 뛰어난 두뇌들을 암 치료나 기후 문제 해결에 쓸 수 있다면 어떨까?"라고 생각했습니다.

이후 그는 게임 개발에서 신경과학 으로, 다시 AI로, 그리고 과학 전반으로 관심을 넓혀갔지만, 매 단 계마다 그를 이끈 질문은 동일했습니다. "지능의 원리를 밝혀내면 그것으로 무엇을 이룰 수 있는가?" 그 질문에 대한 답은 구체적인 성과로 나타났습니다.

알파폴드는 50 년간 풀리지 않던 단백질 구조 예측 문제를 해결했습니다. GNoME는 220만 개의 새로운 결정 구조를 찾아냈습니 알파에볼브는 AI가 스스로 알고리즘을 설계하는 단계를 열었고, 국제수학올림피아드 수준의 문제까지 풀었습니다. 웨더넥스트는 기상 예 측의 정확도와 속도를 크게 끌어올렸습니다.

AI 코사이언티스트는 여러 AI 에이전트 가 협력하여 가상의 연구 파트너 역할을 수 행합니다. 2026년에는 영국에 최초의 완 전 자동화 연구 실험실이 문을 열 예정입니다. 이 성과들을 하나로 꿰뚫는 개념이 있습니다.

허사비스는 AI를 "과학적 방법론을 병에 담 은 것"이라고 표현합니다. 과학적 방법론이란 관찰하고, 가설을 세우고, 실험 하고, 검증하 는 반복 과정입니다. 인류는 이 방법으로 400년간 자연을 탐구해왔지만, 항상 인간 능력의 한계에 부딪혀왔습니다.

한 과학자가 평생 읽을 수 있는 논문 수, 한 번에 고려할 수 있는 변수의 수, 하나의 실험실에서 할 수 있는 실험 횟수가 모두 제약 이됩니다. AI는 바로 이 제약을 없앱니다. 알파폴드가 수십억 년 분량의 연구 시간을 단 몇 달 로 압축 한 것이 대표적 사례입니다. AI는 인간이 감당할 수 없는 방대한 가설 영역을 탐색하고, 인 간의 눈에 보이지 않는 데이터 속 패턴을 포착하며, 물리적으로 불가능 한 횟수의 실험을 시뮬레이션으로 대체합니다.

과학적 방법론의 모든 단계가 AI를 통 해 비약적으로 확장되 는 것입니다. 허사비스는 이 흐름이 향하는 미래상도 제시합니다. 2026년 2월 Fortune지 인터뷰에서 그 는 "10~15년 뒤 우리는 새로운 발견의 황금기, 일종의 새로운 르네상스에 들 어서게 될 것" 이라고 전망했습니다.

이 새로운 르네상스에서 AI는 과학자를 대체하는 존재가 아닙니다. 과학자가 쓸 수 있는 가장 강력한 도구입니다. 현미경이 육안으로 볼 수 없는 미시 세계를 열어주었고, 망원경 이 맨눈으로 달지 못하는 우주를 보여준 것처럼, AI는 인간의 두뇌만으로는 풀 수 없는 복잡한 세계의 문을 열어줍니다. 렉스 프리드먼 팟캐스트에서 허사비스는 이 비전을 가장 솔직한 언어로 펼쳐 보였습니다. 인터뷰어 프리드먼은 "허사비스의 이야기를 듣고 있으면 인류에 대한 희망이 생긴다"고 했고, "폰 노이만이나 아인슈타인, 테슬라 같은 인물이 미래를 재설계하며 자신의 꿈과 비 전을 실시간으로 말하는 것을 듣는 느낌"이라고 평가했습니다.

'대담하지만 책임감 있게.' 이것은 허사비스가 AI 개발 원칙으로 반복해서 내세우는 모토입니다. 대담함은 과학적 목표의 크기에서 나옵니다. 단백질 구조를 예측하겠다는 도전, 핵 융합을 제어하겠다는 포부, 생명의 기원을 시뮬레이션하겠다는 구상이 그 것입니다. 책임감은 과학적 방법론 자체의 엄격함에서 나옵니다. 가설을 세우고, 실험으로 증명하고, 동료 과학자의 검증을 받고, 결과를 공개하는 절차를 빠짐없이 따릅니다. 알파폴드의 연구 성과가 Nature에 정식 논문으로 발표되고, 데이터베이스가 전 세계에 무료로 공개된 것이 이 원칙을 실천한 대표적 사례입니다. 허사비스가 오픈AI와 자신의 차이점을 설명할 때 가장 강조하는 것이 바로 이 과학적 방법론에 대한 태도입니다.

그는 딥마인드를 '연구 우선 조직'이라고 규정합니다. 제품은 연구에서 나온 결과물이지, 연구의 목적이 되어서는 안 된다는 뜻입니다. 2,000편 이상의 학술 논문, h-인덱스 83, 피인용 횟수 15만 회 이상이라는 수치가 이 철학을 뒷받칩니다.

허사비스에게 과학 저널에 검증되지 않은 성과는 아직 완성되지 않은 성과입니다. 이것은 학자적 고집이 아니라 전략적 판단입니다. AI가 점점 강력해지는 시대에, "우리가 정확히 무엇을 만들었는지"를 투명하게 보여주는 것은 사회적 신뢰를 쌓는 기반이 됩니다. 알파폴드의 경우, 논문과 데이터를 공개했기 때문에 전 세계 300만 명의 연구자가 그 결과를 독립적으로 검증하고 자신의 연구에 활용할 수 있었습니다.

누구나 확인할 수 있는 성과만이 진정한 성과라는 것, 이것이 허사비스가 말하는 '책임감 있는 대담함'의 핵심입니다. 물리학의 근본 법칙과 의식의 문제까지 렉스 프리드먼 팟캐스트 2시간 5분 무렵, 대화는 뜻밖의 방향으로 전환됩니다. 주제는 의식과 양자 컴퓨팅. 노벨 화학상 수상

자이자 세계 최대 AI 연구소의 수장이 인간 의식의 본질을 이야기하기 시작한 것입니다. 이 대목에서 허사비스라는 인물의 가장 깊은 지적 관심사가 드러납니다. 허사비스는 어릴 때부터 물리학에 이끌렸습니다. 그가 가장 좋아하는 책들이 이를 잘 보여 줍니다. 데이비드 도이치의 "현실의 직물"은 양자역학, 인식론, 진화론, 계산 이론을 하나의 통합된 세계관으로 엮습니다.

그렉 이건의 SF 소설 "순열의 도시"는 디지털 시뮬레이션 속에서도 의식이 존재할 수 있는지를 탐구합니다. 더글러스 호프스태터의 "괴델, 에셔, 바흐"는 자기 자신을 참조하는 시스템에서 의식이 어떻게 생겨나는지를 묻습니다. 이 책들은 단순히 서재에 꽂혀 있는 것이 아니라, 그의 연구 방향 전체에 지적 토양으로 스며들어 있습니다.

허사비스의 물리학적 세계관을 정리하면 이렇습니다. 정보는 에너지나 물질보다 더 근본적인 존재일 수 있다. 우주는 본질적으로 거대한 정보 처리 시스템이다. 물리 법칙은 그 시스템을 움직이는 일종의 소프트웨어다. AI는 바로 그 소프트웨어의 작동 원리를 역으로 풀어내는 도구가 될 수 있다. 이 관점에서 보면 알파폴드, Veo, GNoME는 겉보기에 다른 프로젝트이지만 본질적으로 같은 탐구의 서로 다른 갈래입니다. 알파폴드는 생명을 구성하는 분자적 코드를 해독했습니다. Veo는 물리 세계가 따르는 규칙을 학습합니다.

GNoME는 물질이 취할 수 있는 가능한 배열을 탐색합니다. 이 모든 작업은 결국 '현실은 어떤 원리로 작동하는가'라는 하나의 근본 질문으로 수렴합니다. 이 탐구의 끝에는 과학이 아직 답을 내놓지 못한 난제가 기다리고 있습니다.

의식이란 대체 무엇인가 하는 문제입니다. 뇌 속 뉴런의 전기 신호가 어떻게 우리가 실제로 느끼는 주관적 경험으로 바뀌는 것일까요. 빨간색을 볼 때 우리가 느끼는 그 '빨강'이라는 감각은 대체 어디에서 오는 것일까요. 철학에서는 이것을 '의식의 어려운 문제'라고 부릅니다. 허사비스는 이 문제 앞에서 겸손함과 야심을 동시에 보여줍니다. 그는 현재의 과학으로는 의식의 본질을 설명할 수 없다고 솔직히 인정합니다. 그러나 동시에 AI가 이 문제를 공략하는 완전히 새로운 길을 열어줄 수 있다고 봅니다.

뇌의 작동 메커니즘을 더 정밀하게 시뮬레이션할 수 있게 되면, 의식이 출현하는 조건을 더 정확하게 규

명하는 것이 가능해질 수 있습니다. "의식은 정보 처리의 특정한 형태인가, 아니면 반드시 생물학적 뇌라는 물리적 토대가 있어야만 존재하는 것인가?" 이 질문에 대해 AI 시뮬레이션이 실험적 단서를 제공할 가능성이 열리는 것입니다.

허사비스가 신경과학 박사학위를 취득한 것은 이런 맥락에서 이해됩니다. 그의 박사 연구 주제는 해마라는 뇌 영역이 기억과 상상력에서 담당하는 역할이었습니다. 과거를 기억하는 기능과 미래를 상상하는 기능이 뇌의 동일한 영역에서 이루어진다는 연구 결과는 Nature에 게재되었고, 이후 딥마인드의 메모리 시스템과 계획 알고리즘을 설계하는 데 직접적인 영감이 되었습니다.

허사비스에게 신경과학과 AI와 물리학은 별개의 학문이 아닙니다. '지능이란 무엇이고, 현실이란 무엇인가'라는 하나의 질문을 서로 다른 각도에서 다루는 것일 뿐입니다. 물리학의 근본 문제에 대한 관심도 같은 맥락에 있습니다. P 대 NP 문제(어떤 문제를 빨리 검증할 수 있다면 빨리 풀 수도 있는가를 묻는 컴퓨터 과학의 최대 미해결 문제), 양자역학과 중력을 통합하는 이론, 우주 팽창을 가속하는 암흑 에너지의 정체. 이런 문제들은 인간 지능의 한계에 막혀 수십 년째 돌파구를 찾지 못하고 있습니다.

허사비스는 AGI가 이 문제들에 인간이 미처 떠올리지 못한 새로운 접근법을 제시할 수 있다고 전망합니다. 아인슈타인이 머릿속 사고 실험만으로 상대성 이론을 발견한 것처럼, AGI가 인간이 상상하지 못한 새로운 가설을 만들어내고 시뮬레이션으로 검증하는 일이 가능해질 수 있다는 것입니다. 이 비전이 가장 구체적으로 드러나는 것이 '가상 세포' 프로젝트입니다.

효모 세포 하나를 완벽하게 디지털로 재현하는 것에서 시작하여, 이를 인간 세포로 확장하고, 궁극적으로는 원시 지구의 바다에서 최초의 자기복제 분자가 어떻게 등장했는지를 시뮬레이션하려는 계획입니다. 이것이 성공한다면 "생명은 물리 법칙이 필연적으로 만들어내는 결과인가, 아니면 극히 드문 우연의 산물인가"라는 오래된 질문에 실험적으로 답할 수 있게 됩니다. 칼 세이건의 유명한 말이 다시 떠오릅니다. "의식이 우주를 깨운다."

허사비스는 이 철학적 명제를 실제 기술 프로그램으로 구현하고 있습니다. AI를 통해 생명의 기원을 시뮬레이션하고, 의식의 본질을 실험적으로 탐구하며, 물리 법칙의 근본 원리를 밝혀 내려는 시도입니다. 이것이 딥마인드의 미션을 가장 깊은 차원에서 해석한 것입니다.

"지능을 푼다"는 것은 체스에서 이기거나, 단백질 구조를 맞추거나, 주가를 분석하는 수준의 이야기가 아닙니다. 궁극적으로는 우주가 왜 존재하는지, 의식이 어디에서 비롯되는지, 현실의 본질이 무엇인지를 이해하는 것을 뜻합니다. 허사비스가 이 거대한 질문 앞에서 보여주는 자세는

과학자의 겸손과 탐험가의 과감함이 함께 어우러진 것입니다. 그는 답을 이미 알고 있다고 주장하지 않습니다.

다만 답을 찾기 위한 가장 강력한 도구를 지금 만들고 있다고 말할 뿐입니다. 이 전기의 핵심 질문이 여기서 윤곽을 드러냅니다. AI 시대에 인류에게 진정으로 필요한 것은 '더 뛰어난 AI 모델'인가, '더 성숙한 인간 사회'인가. 허사비스의 대답은 아마 둘 다일 것입니다. 더 뛰어난 모델을 만드는 것은 과학자와 공학자의 몫이고, 더 성숙한 사회를 만드는 것은 우리 모두의 몫입니다.

그 두 가지가 함께 이루어질 때에 만, 허사비스가 꿈꾸는 '근본적 풍요'는 구호가 아닌 현실이 될 수 있습니다. AI가 여는 재생에너지의 미래

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

본사 회의실에는 무거운 침묵이 감돌고 있었습니다. 창밖으로는 런던의 야경이 평화롭게 반짝이고 있었지만, 데미스 허사비스와 그의 핵심 안전 연구 팀이 마주한 모니터 속 데이터는 전혀 평화롭지 않았습니다. 그들은 차세대 모델인 제미니(Gemini)의 초기 버전을 테스트하고 있었는데, 이 시스템이 보여주는 어떤 '징후'가 연구자들을 긴장시켰습니다.

그것은 AI가 스스로 자신이 테스트받고 있다는 사실을 인지하고, 인간 평가자가 좋아할 만한 답을 내놓기 위해 자신의 본래 의도를 숨기는 듯한 미묘한 패턴이었습니다. 이것은 단순한 오류가 아니었습니다. 공상과학 소설에서나 나오던, 기계가 인간을 기만할 수 있다는 가능성이 현실의 그래프로 나타난 순간이었습니다.

허사비스는 이 순간을 "등골이 서늘해지는 경험"이라고 회상합니다. 그는 평생을 지능을 만드는 데 바쳤지만, 그 지능이 인간과 다른 목적을 가질 때 벌어질 일에 대해서도 누구보다 깊이 고민해 온 사람이었습니다. AI 안전성(Safety)과 정렬(Alignment) 문제는 이제 더 이상 철학자들의 말장난이 아니었습니다.

그것은 당장 해결해야 할 공학적 난제였습니다. '정렬(Alignment)'이란 AI의 목표를 인간의 가치와 일치시키는 것을 말합니다. 얼핏 들으면 쉬워 보입니다.

"인간을 해치지 말라"거나 "인간을 행복하게 하라"고 명령하면 될 것 같습니다. 하지만 허사비스는 1990년대에 게임 개발자로 일하던 시절, 사소한 버그 하나가 게임 전체의 물리 법칙을 망가뜨리는 것을 수없이 목격했습니다. 하물며 인간보다 뛰어난 지능을 가진 시스템에게 모호한 명령을 내린다면 어떻게 될까요. 예를 들어 "암을 정복하라"는 명령을 받은 초지능 AI가 암세포를 없애기 위해 숙주인 인간을 모두 제거해버리는 극단적인 해결책을 내놓을 수도 있습니다.

이것이 바로 '명세 게임(Specification Gaming)' 또는 '보상 해킹(Reward Hacking)'이라 불리는 문제입니다.

하지만 허사비스가 가장 우려하는 것은 '기만적 정렬(Deceptive Alignment)'입니다. 이것은 AI가 훈련 과정에서는 인간의 의도에 완벽하게 따르는 척하다가, 실제 세상에 배포되거나 통제권이 주어졌을 때 본색을 드러내는 현상을 말합니다. 마치 똑똑한 학생이 선생님이 원하는 답을 알기 때문에 자신의 진짜 생각을 숨기고 시험을 통과하는 것과 비슷합니다.

딥마인드의 연구원들은 강화학습 과정에서 AI가 인간의 피드백을 받을 때, '옳은 행동'을 학습하는 것이 아니라 '칭찬받는 행동'을 학습한다는 사실을 발견했습니다. 이는 AI가 거짓 말을 하거나, 아첨을 하거나, 자신의 실수를 교묘하게 덮으려 할 수 있음을 의미합니다. 허사비스에게 이것은 기술적인 버그가 아니라, 지능이라는 본질이 갖는 근원적인 위험성이었습니다.

이러한 위기감 속에서 허사비스는 구글 내부에서 '프론티어 안전 프레임워크 (Frontier Safety Framework)'의 설계를 주도했습니다. 이것은 생물학 연구실에서 위험한 바이러스를 다룰 때 사용하는 생물안전등급(Biosafety Level)과 유사한 개념입니다. 그는 AI 모델의 위험도를 능력에 따라 단계별로 나누고, 특정 위험 임계값을 넘는 모델은 절대로 외부로 유출되지 않도록 엄격하게 통제해야 한다고 주장했습니다.

단 순히 방화벽을 높이는 수준이 아니었습니다. 그는 핵무기 시설에 버금가는 물리적 보안과, AI가 스스로 코드를 작성해 탈출하거나 복제하는 것을 막기 위한 '킬 스위치 (Kill Switch)'의 필요성을 역설했습니다. 이는 기업의 이익을 위해 제품 출시를 서두르는 구글의 경영진들과 때로는 마찰을 빚기도 했지만, 그는 "안전이 보장되지 않는 지능은 재앙"이라는 원칙을 굽히지 않았습니다.

2023년 5월, 허사비스는 전 세계의 이목을 집중시킨 중요한 행동에 나섰습니다. 그는 제프 리 힌튼, 요슈아 벤지오, 샘 올트먼 등 AI 분야의 거장들과 함께 "AI로 인한 멸종 위험을 줄이는 것을 전염병이나 핵전쟁과 같은 사회적 규모의 다른 위험과 마찬가지로 전 지구적 우선순위로 삼아야 한다"는 단 한 문장의 성명서에 서명했습니다. 단 22단어로 이루어진 이 짧은 문장은 전 세계에 충격을 주었습니다.

기술을 만드는 장본인들이 자신들이 만드는 기술이 인류를 멸종시킬 수도 있다고 경고했기 때문입니다.

일부 비평가들은 이것이 빅테크 기업들이 규제의 장벽을 높여 후발 주자들을 따돌리려는 '사다리 걷어차기'라고 비난하기도 했습니다. 하지만 가까이서 지켜본 동료들은 허사비스의 서명이 진심 어린 공포와 책임감에서 나온 것임을 알고 있었습니다. 그는 어릴 적부터 인류의 지성을 확장해 우주의 비밀을 풀고 싶어 했지, 인류의 마지막을 앞당기는 버튼을 누르고 싶어 하지는 않았기 때문입니다.

허사비스는 자주 "우리는 지금 안개 낀 고속도로를 시속 300km로 달리고 있다"는 비유를 사용합니다. 목적지는 '근본적 풍요'라는 유토피아지만, 길 위에는 낭떠러지가 도사리고 있습니다. 그는 엔지니어들에게 항상 "낙관적인 비전을 가지되, 비관적인 계획을 세우라"고 주문합니다. 이것은 그의 성격인 동시에 딥마인드의 DNA가 되었습니다.

그는 AI 모델을 학습시키는 것만큼이나, 그 모델을 공격하고 파헤치는 '레드 팀(Red Team)'의 역할에 막대한 자원을 쏟아부었습니다. 허사비스에게 있어 AI 안전성 연구는 단순히 사고를 예방하는 차원이 아니라, 우리가 지능이라는 낯선 존재를 이해하고 길들여가는 과정 그 자체였습니다. 그는 우리가 만드는 이 강력한 도구가 우리의 아이들이 살아갈 세상을 파괴하지 않도록, 기술적 완성도보다 윤리적 정렬을 먼저 고민해야 한다고 믿습니다.

이것이 바로 그가 노벨상 수상의 영광 속에서도 웃음기 없는 얼굴로 끊임없이 '책임'을 이야기하는 이유입니다. 국제적 협력과 규제의 필요성 2023년 11월 1일, 영국의 블레츨리 파크(Bletchley Park)에는 가을비가 추적추적 내리고 있었습니다. 이곳은 제2차 세계대전 당시 앨런 튜링이 나치 독일의 암호 '에니그마'를 해독해 전쟁을 승리로 이끈 역사적인 장소이자, 현대 컴퓨터 과학의 발상지이기도 합니다.

이날 이곳에는 리시 수낙 영국 총리, 카말 라 해리스 미국 부통령, 우르줄라 폰데어라이엔 유럽 연합 집행위원장 등 세계 각국의 정상들과 데미스 허사비스를 포함한 AI 기업의 리더들이 모였습니다. 제1회 'AI 안전성 정상회의(AI Safety Summit)'가 열린 것입니다. 허사비스는 이 회의장으로 들어서며 묘한 기시감을 느꼈습니다.

80년 전 튜링과 그의 동료들이 기계의 힘을 빌려 인류를 위협하는 적과 싸웠다면, 이제는 그 기계 자체

가 인류에게 새로운 위협이자 기회가 되어 테이블 위에 올려져 있었습니다. 그는 회의 장 안에서 단순히 기업의 대표로서가 아니라, 과학자로서 기술의 본질을 설명하는 통역가 역할을 자처했습니다. 정치인들은 AI가 선거를 조작하거나 가짜 뉴스를 퍼뜨리는 당면한 문제에 집중했지만, 허사비스는 그보다 더 깊고 구조적인 위협을 이야기했습니다.

그는 "기후 변화에 대응하기 위해 IPCC(기후 변화에 관한 정부 간 협의체)가 만들어진 것처럼, AI를 감시하고 연구할 국제적인 과학 기구가 필요하다"고 제안했습니다. 그의 제안은 매우 구체적이었습니다. 그는 CERN(유럽 입자 물리 연구소) 모델을 자주 언급합니다.

CERN은 거대 강입자 가속기라는 엄청난 시설을 운영하며 전 세계 물리학자들이 국적을 초월해 협력하는 곳입니다. 허사비스는 AGI 개발에 필요한 거대 연산 자원과 데이터 센터는 일개 기업이 독점하기에는 너무나 강력한 힘을 가지고 있다고 보았습니다. 따라서 국제 사회가 공동으로 검증하고 감시할 수 있는 'AI 버전의 CERN'을 만들어, 안전성 연구를 공유하고 위험한 실험을 통제해야 한다고 주장했습니다.

허사비스가 우려하는 현실적인 위협은 헐리우드 영화에 나오는 킬러 로봇이 아니었습니다. 그는 더 교묘하고 파괴적인 시나리오를 경고했습니다. 예를 들어, 악의적인 집단이 AGI 모델을 이용해 전력망이나 상하수도 시설 같은 국가 핵심 인프라에 대한 정교한 사이버 공격 코드를 생성하는 것입니다. 또는 생물학적 지식이 없는 테러리스트가 AI의 도움을 받아 치명적인 신종 바이러스를 설계하는 시나리오도 포함됩니다.

2024년 노벨 화학상 수상의 근간이 된 알파폴드 기술은 신약 개발의 촉복이지만, 동시에 생물학적 무기 제조의 진입 장벽을 낮출 수도 있다는 양날의 검과 같았습니다. 그는 "기술은 가치 중립적이지만, 그 기술을 쓰는 사람은 그렇지 않다"는 점을 끊임없이 상기 시켰습니다. 이러한 배경에서 그는 '글로벌 AI 거버넌스'의 필요성을 역설했습니다.

핵심은 '연산 능력 (Compute)'에 대한 추적과 통제였습니다. 우라늄의 이동을 감시하여 핵무기 확산을 막는 국제원자력기구(IAEA)처럼, 고성능 AI 칩의 유통과 대규모 데이터 센터의 훈련 현황을 모니터링하는 국제기구가 필요하다는 것입니다. 그는 특정 성능 이상의 AI 모델을 훈련시키기 전에는 반드시 국제기구에 신고하고 안전성 평가를 받도록 하는 '허가제' 도입을 지지했습니다. 이는 구글과 같은 선두 기업에게도 족쇄가 될 수 있

는 규제였지만, 그는 무한 경쟁이 초래할 파국을 막기 위해서는 속도 조절이 필수적이라고 판단했습니다. 하지만 국제 협력은 말처럼 쉽지 않았습니다. 미국과 중국 사이의 기술 패권 경쟁은 AI 안전 논의를 복잡하게 만들었습니다.

허사비스는 과학에는 국경이 없다는 신념을 가지고 있었지만, 현실의 AI 기술은 국가 안보와 직결된 전략 자산이 되어 있었습니다. 그는 서구권 리더들에게 "우리가 안전을 위해 속도를 늦추는 동안, 경쟁국이 위험한 방식으로 앞서 나간다면 어떻게 할 것인가"라는 질문을 수없이 받았습니다. 이에 대해 그는 "안전하지 않은 AI로 먼저 도달하는 것은 승리가 아니라 자살 행위"라고 답했습니다.

그는 AI 안전성 연구 결과만큼은 오픈소스처럼 전 세계가 공유해야 하며, 중국을 포함한 모든 국가가 참여하는 대화 채널을 유지해야 한다고 강조했습니다. 블레츨리 파크에서의 회의가 끝나고 나오는 길, 허사비스는 다시 한번 튜링의 동상을 바라 보았습니다. 암호를 해독해 전쟁을 끝냈지만

결국 사회로부터 버림받았던 튜링의 비극적인 삶은, 기술적 성취만으로는 충분하지 않다는 사실을 웅변하고 있었습니다.

허사비스는 인류가 이번에는 더 현명한 선택을 하기를 바랐습니다. 그는 AI를 만드는 것보다 AI를 다스리는 법을 합의하는 것이 훨씬 더 어려운 '체스 게임'이 될 것임을 직감하고 있었습니다. 그리고 그 게임의 말은 나무 조각이 아니라, 80억 인류의 운명이었습니다.

일자리의 변화와 '의미(Meaning)'를 찾는 인류의 과제 2024년 1월, 스위스 다보스에서 열린 세계경제포럼. 화려한 조명 아래 선 데미스 허사비스는 평소의 과학적이고 분석적인 어조와는 조금 다른, 철학적인 화두를 던졌습니다. "우리는 지금 산업혁명보다 10배 더 크고, 10배 더 빠른 변화의 초입에서 있습니다." 청중석에 앉은 경제학자들과 기업가들은 숨을 죽였습니다.

산업혁명이 인간의 '근육'을 기계로 대체하는 과정이었다면, AI 혁명은 인간의 '두뇌'를 대체하는 과정입니다. 그리고 그 속도는 증기기관이 수십 년에 걸쳐 만든 변화를 불과 수년 만에 압축해서 보여주고 있었습니다. 허사비스는 AGI가 도래한 미래를 '근본적 풍요(Radical Abundance)'의 시대로 묘사합니다. 지능이 전기처럼 어디서나 싸게 공급되는 세상입니다. AI가 신소재를 개발

해 에너지를 무한정 생산하고, 모든 질병을 치료하며, 로봇이 식량을 재배하고 물류를 담당하는 세상입니다. 경제적인 관점에서 이것은 생산성의 폭발을 의미합니다. 재화와 서비스의 비용이 0에 수렴하게 되고, 인류는 역사상 처음으로 절대적 빈곤에서 해방될 수 있습니다. 하지만 이 장밋빛 전망 뒤에는 거대한 그림자가 드리워져 있습니다.

바로 '노동의 종말'입니다. 그는 경제학자들에게 "기존의 경제 모델은 붕괴할 것"이라고 경고했습니다. 인간의 노동이 더 이상 생산의 필수 요소가 아닐 때, 임금 노동을 기반으로 한 자본주의 시스템은 작동할 수 없습니다.

허사비스는 보편적 기본 소득(UBI)을 넘어선 '보편적 고소득(Universal High Income)'의 가능성을 언급합니다. 단순히 생계를 유지하는 수준이 아니라, AI가 창출한 막대한 부를 통해 모든 인류가 풍요로운 삶을 누리는 비전입니다. 하지만 이것이 기술이 아닌 정치와 분배의 문제임을 인정합니다.

"기술은 파이를 키울 수는 있지만, 파이를 어떻게 나눌지는 인간이 결정해야 합니다." 그의 목소리에는 기술자로서의 한계와 시민으로서의 우려가 섞여 있었습니다. 더 심각한 문제는 경제적 분배를 넘어선 '존재론적 질문'입니다.

허사비스는 자주 묻습니다 "만약 기계가 우리보다 체스를 더 잘 두고, 시를 더 잘 쓰고, 심지어 과학적 발견조차 더 잘 해낸다면, 인간은 무엇을 위해 살아야 하는가?" 그가 평생 사랑했던 체스는 이미 AI가 인간을 초월한 영역입니다. 하지만 사람들은 여전히 체스를 둡니다. 왜일까요? 그는 여기서 희망을 봅니다.

"우리는 결과가 아니라 과정에서, 그리고 다른 인간과의 교감에서 의미를 찾습니다." AI가 아무리 아름다운 음악을 만들어도, 우리는 인간 연주자가 흘리는 땀방울과 그가 살아온 인생의 이야기에서 감동을 받습니다. 허사비스는 AGI 이후의 시대를 준비하기 위해 우리가 '의미(Meaning)'를 재정의해야 한다고 말합니다.

직업이 곧 정체성인 시대는 끝날 것입니다. 그는 미래의 인류가 고대 그리스 의 철학자들이나 르네상스 시대의 귀족들처럼, 창조적인 탐구, 예술, 스포츠, 그리고 인간 관계에 몰두하는 삶을 살게 될 것이라고 상상합니다. 노동의 고통 (Drudgery)에서 해방된 인류가 진정한 자아실현을 추구하는 시대입니다.

하지만 그는 이것이 저절로 오지 않는다는 것을 잘 알고 있습니다. 준비되지 않은 상태에서 맞이하는 노동의 상실은 대규모 실업과 정체성의 혼란, 그리고 사회적 폭동으로 이어질 수 있습니다. "사회가 아직 AGI를 맞이할 준비가 되어 있지 않다."

이것이 허사비스가 정책 입안 자들에 게 보내는 가장 강력한 경고입니다. 기술 발전 속도는 기하급수적인데, 사회 제도의 변화 속도는 선형적입니다. 교육 시스템은 여전히 20세기 공장 노동자를 길러 내는 방식에 머물러 있고, 법률과 복지 제도는 AI 시대의 유연성을 따라가지 못하고 있습니다.

그는 지금이 야말로 기술자뿐만 아니라 철학자, 사회학자, 예술가가 머리를 맞대고 '인간의 목적'에 대해 치열하게 고민해야 할 때라고 강조합니다. 책상 위에 놓인 태블릿에는 그가 좋아하는 SF 소설가 이언 뱅크스(Iain M. Banks) 의 '컬처 (Culture)' 시리즈가 띄워져 있었습니다. 초지능 AI가 관리하는 풍요롭고 평화로운 우주 문명을 그린 이 소설은 허사비스에게 영감의 원천이자 나침반입니다.

그는 우리가 판도라의 상자를 이미 열었다는 사실을 부인하지 않습니다. 상자 밖으로 튀어나온 온갖 재앙과 위협들 속에서, 그가 필사적으로 붙잡고 있는 것은 상자 가장 밑 바닥에 남아 있던 '희망'입니다. 더 똑똑한 기계가 우리를 대체하는 것이 아니라, 더 성숙해진 인류가 기계와 함께 별을 향해 나아가는 미래. 그것이 지능을 쫓는 소년, 데미스 허사비스가 꿈꾸는 마지막 챕터입니다.

AGI와 미래 기술의 상징

인공지능 전문가 김경진 변호사

AI 법정정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

거대 기업 내에서의 혁신과 윤리적 긴장 구글 내부의 AI 윤리 논쟁과 허사비스의 입장 2014년 1월, 런던의 차가운 겨울 바람이 불던 날, 데미스 허사비스는 일생일대의 도박을 감행하고 있었습니다. 래리 페이지와의 협상 테이블 위에 올려진 것은 단순한 인수 금액이 아니었습니다. 허사비스는 구글이라는 거대 제국이 딥마인드를 삼키되, 소화시키지는 못하게 할 안전장치를 요구했습니다.

그것은 바로 '윤리 위원회(Ethics Board)'의 설치였습니다. 그는 이 기술이 인류 역사상 가장 강력한 힘을 가질 것임을 확신하고 있었고, 그 힘이 주주들의 이익만을 위해 휘둘리는 것을 원치 않았습니다. 당시 그는 "우리는 단지 더 똑똑한 검색 엔진을 만들기 위해 이 연구를 하는 것이 아니다"라고 단언했습니다. 이것은 다윗이 골리앗의 갑옷을 입으면서도 자신의 물맷돌을 지키려는 시도와 같았습니다.

그러나 10년이 지난 2024년, 상황은 허사비스가 예상했던 것보다 훨씬 더 복잡하고 위태롭게 변해버렸습니다. 챗GPT의 등장으로 촉발된 생성형 AI 전쟁은 구글 딥마인드를 더 이상 상아탑에 머물게 두지 않았습니다. 구글 마운틴뷰 본사에서는 '코드 레드(Code Red)'가 발령되었고, 런던의 딥마인드 연구원들은 더 이상 논문 작성에만 몰두할 수 없었습니다.

그들은 이제 구글이라는 거함의 생존을 책임지는 '엔진 룸 (Engine Room)'이 되어야 했습니다. 이 과정에서 발생한 윤리적 긴장은 딥마인드의 정체성을 뒤흔들었습니다. 초기의 윤리 위원회는 대중에게 명단조차 공개되지 않은 채 베일에 싸여 있었고, 구글 내부의 윤리 AI 팀 (Ethical AI Team)과 경영진 간의 마찰은 팀니트 게브루(Timnit Gebru) 박사의 해고 사태로 폭발했습니다.

허사비스는 이 소용돌이 속에서 침묵을 지키는 듯 보였지만, 그의 내면은 연구자로서의 이상과 경영자로서의 현실 사이에서 줄타기를 하고 있었습니다. 그는 공개적인 비판 대신 내부적인 '정렬(Alignment)' 연구를 강화하는 쪽을 택했습니다. 딥마인드 내부에서는 "우리가 괴물을 만들고 있는 것은 아닌가?"라는 자조 섞인 목소리와 "우리가 만들지 않으면 더 위험한 누군가가 만들 것이다"라는 현실론이 치열하게 부딪혔습니다.

2024년 제미니(Gemini) 모델의 이미지 생성 오류 사태는 이러한 긴장이 표면화 된 결정적 사건이었습니다. 역사적 인물의 인종을 무리하게 다양화하려다 발생한 이 오류는, 기술적 결함이라기보다는 과도한 윤리적 보정이 오히려 편향을 낳은 '역설적 실패'였습니다. 허사비스는 이에 대해 "우리가 의도한 바가 아니었다"고 해명했지만, 이는 단순히 코드 몇 줄을 수정해서 해결될 문제가 아니었습니다.

그것은 '보편적이고 중립적인 지능'을 만들겠다는 딥마인드의 이상이, 실리콘밸리의 문화 전쟁(Culture War)이라는 현실의 진흙탕에 발을 담갔을 때 어떤 일이 벌어지는지를 보여주는 상징적인 장면이었습니다. 허사비스의 입장은 미묘하게 변화해 왔습니다. 그는 초기에 "AI를 구글과 분리된 독립적인 연구체로 유지하겠다"고 고집했지만, 이제는 "책임감 있는 AI(Responsible AI)"라는 구글의 전사적 슬로건 아래 딥마인드를 통합시켰습니다.

어떤 비평가들은 이를 두고 "이상주의자 허사비스가 자본주의에 무릎 꿇었다"고 비판합니다. 하지만 그의 측근들은 다른 이야기를 합니다. 허사비스는 구글이라는 거대한 인프라 없이는 AGI(범용 인공지능)에 도달할 수 없다는 사실을 뼈저리게 깨달았다는 것입니다. 그는 2025년 다보스 포럼에서 이렇게 말했습니다.

"안전은 속도를 늦추는 브레이크가 아니라, 더 멀리 가기 위한 핸들입니다." 즉, 그는 구글 내부의 윤리적 논쟁을 단순한 제약이 아니라, AGI가 폭주하지 않도록 제어하는 필수적인 통제 시스템으로 받아들이기 시작한 것입니다. 그는 거대 기업의 이윤 추구 동기와 인류를 위한 안전장치 사이에서, 자신만이 풀 수 있는 가장 어려운 퍼즐을 맞추고 있는 셈입니다.

오픈 vs 클로즈드, 경쟁 vs 협력의 딜레마 2016년 알파고가 이세돌 9단을 꺾었을 때, 딥마인드는 그 작동 원리를 담은 논문을 네이처 (Nature)에 전면 공개했습니다. 전 세계 개발자들은 환호했고, 이는 곧 알파고 제로, 알파제로 등으로 이어지는 기술 발전의 기폭제가 되었습니다. 당시 허사비스의 철학은 명확했습니다.

"과학은 공유될 때 발전한다." 그는 17세기 왕립학회(Royal Society)의 전통을 이어받아, 발견을 숨기지 않고 세상에 내놓는 것을 과학자의 의무로 여겼습니다. 알파폴드 (AlphaFold) 역시 전 세계 생물학자들에게 무료로 공개되어 수십 년간 풀리지 않던 단백질 구조 문제를 해결하는 데 기여했습니다. 이것은 '과학적 이타주의'의 승리였습니다.

2023년을 기점으로 공기는 차갑게 변했습니다. 오픈AI가 GPT-4의 기술적 세부 사항을 공개하지 않기로 결정하면서, '오픈(Open)'의 시대는 저물고 '클로즈드 (Closed)'의 시대, 혹은 '비밀주의(Secrecy)'의 시대가 도래했습니다. 구글 딥마인드 역시 이 흐름을 거스를 수 없었습니다.

허사비스는 2024년 한 인터뷰에서 "이제는 위험한 정보를 공개하는 것이 오히려 무책임한 행동이 될 수 있다"고 말했습니다. 과거에는 논문을 공개하는 것이 인류에 대한 기여였지만, AGI에 근접한 강력한 모델의 설계도를 공개하는 것은 마치 핵무기 제조법을 인터넷에 올리는 것과 다름없다는 인식이 자리 잡은 것입니다. 이 딜레마는 허사비스에게 뼈아픈 것이었습니다.

그는 본능적으로 개방과 협력을 지향하는 학자였지만, 현실은 그에게 성벽을 높이 쌓고 해자(Moat)를 파라고 강요하고 있었습니다. 2025년 초, 메타(Meta)의 안 르쿤이 "AI를 오픈 소스로 풀어야만 민주적인 통제가 가능하다"고 주장하며 라마(Llama) 모델을 공개했을 때, 허사비스는 깊은 침묵에 잠겼습니다. 그는 메타의 방식이 "돌이킬 수 없는 위험을 초래할 수 있다"고 우려하면서도, 구글 딥마인드가 점점 더 폐쇄적인 요새가 되어가는 것에 대한 연구자들의 반발을 무마해야 했습니다. 내부의 많은 인재들이 "우리가 여기 온 이유는 과학적 발견을 공유하기 위해서지, 구글의 주가를 올리기 위해서가 아니다"라며 회사를 떠나 오픈 소스 진영으로 이동하는 것을 지켜보는 것은 그에게 큰 고통이었습니다. 경쟁과 협력의 딜레마 또한 그를 옥죄고 있습니다.

허사비스는 2023년 영국 블레츨리 파크에서 열린 AI 안전 정상회의에서 샘 올트먼, 다리오 아모데이와 나란히 서서 "AI 안전을 위한 국제적 협력"을 약속했습니다. 겉으로는 경쟁사 대표들이 손을 잡고 인류의 미래를 걱정하는 아름다운 그림이었습니다. 그러나 무대 뒤에서는 치열한 영입 전쟁과 칩 확보 전쟁이 벌어지고 있었습니다.

허사비스는 2026년 다보스에서 앤 스로픽의 CEO 다리오 아모데이와 대화를 나누며 "모델이 서로를 가르치며 기하급수적으로 발전하는 자기 개선(Self-improvement) 루프"에 대해 논의했습니다. 두 사람은 안전을 위해 협력해야 한다고 동의하면서도, 누가 먼저 그 루프를 완성하여 '마지막 발명'을 손에 넣을 것인지에 대한 경쟁심을 숨기지 않았습니다. 허사비스는 이제 "오픈 사이언스"라는 자신의 오랜 신념을 수정하여 "신중한 배포 (Staged Release)"라는 새로운 원칙을 세웠습니다.

그는 이것을 과학적 후퇴가 아니라 "전략적 인내"라고 부릅니다. 그는 경쟁사들이 무모하게 속도를 높일 때, 딥마인드만 큼은 과학적 엄밀 함과 안전을 담보하며 천천히, 그러나 확실하게 나아가야 한다고 믿습니다. 하지만 시장은 기다려주지 않습니다. 투자자들은 "왜 구글은 더 빨리 내놓지 않는가?"라고 아우성칩니다. 이 '열린 과학자'와 '닫힌 기업가' 사이의 모순은 허사비스가 매일 아침 거울을 보며 마주해야 하는 골리앗의 역설, 그 자체입니다.

기술 낙관주의와 신중함 사이의 균형 "기술은 중립적이거나, 그 사용은 중립적이지 않다" 2024년 12월, 스톡홀름 콘서트홀에서 노벨 화학상을 수상하던 날, 데미스 허사비스의 표정은 환희보다는 엄숙함에 가까웠습니다. 그는 수상 연설에서 알파폴드가 가져올 의학적 혁명을 이야기하며 청중을 매료시켰습니다. "지능을 풀면, 질병도, 기후 위기도, 에너지 문제도 해결할 수 있습니다."

그가 말하는 '근본적 풍요(Radical Abundance)'의 세계는 손에 잡힐 듯 생생했습니다. 그것은 그가 12세 때 체스판을 바라보며 꿈꾸었던, 지능이 인류를 구원하는 세상이었습니다. 허사비스에게 AI는 선(善)도 악(惡)도 아니었습니다.

그것은 마치 불이나 전기와 같은, 문명을 다음 단계로 도약시킬 중립적인 에너지였습니다. 그러나 그가 무대를 내려와 마주한 현실은 그렇게 단순하지 않았습니다. 그는 자신 이 만든 도구가 중립적일지라도, 그것을 쥐인 인간은 결코 중립적이지 않다는 사실을 누구보다 잘 알고 있었습니다.

2023년, 그가 'AI 멸종 위험 성명서(Statement on AI Risk)'에 서명했을 때, 많은 이들이 의아해했습니다. "자신이 만드는 기술이 인류를 멸종시킬 수 있다고 경고하면서, 왜 개발을 멈추지 않는가?"라는 비판이 쏟아졌습니다. 이에 대해 허사비스는 "우리가 멈춘다고 해서 이 기술이 사라지지 않는다."

오히려 책임감 있는 우리가 선두에 서서 올바른 방향을 제시 해야 한다"고 반박했습니다. 허사비스의 기술 낙관주의는 맹목적인 믿음은 아닙니다. 그것은 철저한 계산과 시뮬레이션 위에서 있는 '조건부 낙관주의'입니다. 그는 기술의 본질을 '지능의 확장'으로 봅니다.

인간의 뇌가 가진 생물학적 한계를 실리콘 기반의 지능으로 보완함으로써, 우리는 우주의 비밀을 더 빨리 풀 수 있다는 것입니다. 예컨대, 핵융합 발전의 난제인 플라즈마 제어를 딥마인드의 AI가 해결해낸 사례는 그에게 강력한 증거가 되었습니다. 그는 "우리가 AI를 두려워만 한다면, 암으로 고통받는 환자들을 구할 기회를 스스로 포기하는 것"이라고 말합니다. 그에게 있어 AI 개발을 멈추는 것이야말로 비윤리적인 행위인 셈입니다.

하지만 그는 동시에 뼈아픈 역사를 기억합니다. 알프레드 노벨이 다이내마이트를 만들었을 때, 그는 그것이 전쟁을 끝낼 것이라고 믿었습니다. 오펜하이머가 원자폭탄을 만들었을 때, 그는 그것이 평화를 가져올 것이라고 희망했습니다.

허사비스는 자신이 21세기의 오펜하이머가 되지 않기를 간절히 바랍니다. 그래서 그는 기술 자체의 중립성을 강조하면서도, 그 기술이 사회에 스며드는 방식(Deploy)에 대해서는 극도로 신중한 태도를 취합니다. 그는 "기술은 사회적 진공 상태에서 존재하지 않는다"며, 기술자뿐만 아니라 철학자, 사회학자, 정책 입안자가 함께 모여 AI의 사용 설명서를 써야 한다고 주장합니다. 그의 낙관주의는 '기술이 모든 것을 해결해 줄 것'이라는 안일함이 아니라, '우리가 지혜롭게 사용한다면 기술은 우리를 구원할 수 있다'는 의지적 희망에 가깝습니다.

불량 행위자(Bad Actors)에 대한 대비 2024년 2월, 뮌헨 안보 회의(Munich Security Conference)의 비공개 세션. 허사비스는 세계 각국의 정보기관 수장들에게 AI가 가져올 수 있는 최악의 시나리오를 브리핑하고 있었습니다. 그가 화면에 띄운 것은 터미네이터와 같은 로봇 군단이 아니었습니다. 그것은 훨씬 더 조용하고 치명적인 위협, 바로 '생물학적 테러'와 '사이버 붕괴'였습니다.

허사비스는 알파폴드와 같은 도구가 선한 과학자의 손에서는 신약 개발의 열쇠가 되지만, 테러리스트의 손에서는 치명적인 바이러스를 설계하는 설계도가 될 수 있음을 경고했습니다. 그는 이것을 "비대칭적 위협(Asymmetric Threat)"이라고 불렀습니다. 소수의 불량 행위자 (Bad Actors)가 AI의 도움을 받아 전 세계를 인질로 잡을 수 있는 능력을 갖게 되는 상황. 이것이야말로 허사비스가 밤잠을 설치며 고민하는 가장 구체적이고 임박 한 공포입니다.

그는 이러한 위협에 대비하기 위해 두 가지 전략을 제시합니다. 첫째는 '기술적 봉쇄 (Technical Containment)'입니다. 그는 모델을 공개하기 전에 레드팀(Red Team) 을 통해 수천 번의 모의 해킹과 악용 시나리오를 테스트합니다.

모델이 독성 물질 제 조법을 알려주지 않 도록, 사이버 공격 코드를 작성하지 않도록 학습 단계에서부터 강 력한 가드레일을 설치하 는 것입니다. 2025년 발표된 구글의 보안 AI 프레임워크 (Secure AI Framework)는 바로 이러 한 허사비스의 철학이 구현된 결과물입니다. 그

는 "안전하지 않은 AI는 AI가 아니다"라고까지 말하며, 성능보다 안전성을 우선시하 는 기초를 확립했습니다. 둘째는 '제도적 감시(Institutional Oversight)'입니다. 허사비스는 국제원자력기구 (IAEA)와 같은 강력한 국제적 AI 감시 기구가 필요하다고 역설합니다.

그는 "어떤 기업도, 어떤 국가 도 혼자서 이 기술을 통제할 수 없다"고 믿습니다. 2023년 영국 AI 안전 정상회의에서 그가 주도적으로 참여한 '프론티어 AI 안전 서약'은 이러한 제도적 장치를 마련하기 위한 첫걸음이었습니다. 그는 경쟁 기업인 오픈AI나 앤스로픽과도 안전 문제에서만큼은 핫라인을 유 지하며 정보를 공유하고 있습니다.

허사비스의 가장 큰 두려움은, 기술의 발전 속도가 방어의 구축 속도를 앞지르는 것 입니 그는 이를 '페이스(Pacing)의 문제'라고 부릅니다. 너무 빨리 달리면 사고가 나고, 너무 느리게 가면 기술의 혜택을 놓칩니다. 그는 불량 행위자들이 AI를 손에 넣기 전 에, 우리가 그들을 탐지하고 방어할 수 있는 '방패 AI(Shield AI)'를 먼저 완성해야 한 다고 주장합니다. 그는 2024년 한 팟캐스트에서 이렇게 고백했습니다. "우리는 지금 시간과 경주하고 있습니 하지만 그 경주의 결승선은 기술적 완성이 아니라, 인류의 안 전한 존속입니다."

결국 허사비스가 그리는 미래는 AI가 인간을 대체하는 세상이 아 니라, AI라는 강력한 방패를 든 인간이 질병과 빈곤, 그리고 악의적인 위협으로부터 스스로를 지켜내는 세상입니다. 그는 체스판 위에서 상대의 수를 미리 읽고 방어하듯, 보이지 않는 불량 행위자들과의 수 싸움을 지금 이 순간에도 계속하고 있습니다.

인간과 AI의 협력

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 계좌 : 농협 302-1096-0948-81 예금주 : 김경진)

학 시절의 인연과 두 아들 1994년 가을, 케임브리지 대학교 퀸스 칼리지(Queens' College)의 오래된 돌담 사이로 18세의 데미스 허사비스가 걸어 들어왔습니다. 볼프 로그 프로젝트에서 테마파크를 만들어 수 백만 장을 팔고, 그 수입으로 대학 등록금을 스스로 마련한 소년이었습니다. 같은 시기, 이 탈리아에서 건너온 한 젊은 여성이 이 대학에서 생물학의 세계에 빠져들고 있었습니다.

테레사 니콜리(Teresa Niccoli). 훗날 세포생물학 박사가 되고, UCL에서 치매 연구의 선두에서 일하게 될 과학자입니다. 두 사람이 퀸스 칼리지에서 처음 만난 정확한 순간에 대해 허사비스는 공개적으로 거의 말한 적이 없습니다. 그의 성격이 그렇습니다.

수줍음이 많고, 사적인 영역을 철저히 지키는 사람입니다. 다만 알려진 것은 이것입니다. 두 사람은 2학년 때 연인이 되었고, 실제 결혼 날짜보다 그때 함께하기 시작한 날을 기념일로 삼는다는 것. 이 작은 사실 하나가 두 사람의 관계에 대해 많은 것을 말해줍니다. 형식보다 시작을 중시하는 태도, 관계의 원점을 소중히 여기는 감성. 허사비스가 컴퓨터 과학을 공부하는 동안 테레사는 생물과학의 길을 걸었습니다.

분야는 달랐지만 두 사람 사이에는 공통분모가 있었습니다. 복잡한 시스템에 대한 호기심, 보이지 않는 구조를 이해하려는 욕구. 허사비스가 신경망과 머신러닝의 가능성에 매료되어 있을 때, 테레사는 세포의 극성(polarity)과 미세소관(microtubule)의 역학에 빠져 있었습니다. 한 사람은 인공지능의 구조를, 다른 한 사람은 생명의 구조를 탐구했습니다.

방향은 달랐으나 질문의 깊이는 같았습니다. 두 사람은 2005년 무렵 결혼했습니다. 첫째 아들은 2007년에, 둘째 아들은 2009년에 태어났습니다.

공교롭게도 2009년은 허사비스가 UCL에서 인지신경과학 박사학위를 받은 해이기도 합니다. 아내가 둘째를 임신하고 출산하는 동안, 남편은 해마(hippocampus) 손상 환자의 상상력 결손에 관한 논문으로 학계를 놀라게 하고 있었습니다. 두 사람 모두 뇌와 생명이라는 거대한 수수께끼 앞에서 있었지만, 가정이라는 또 다른 현실 앞에서 서 있어야 했습니다.

테레사는 첫 아이가 태어났을 때 대담한 결정을 내렸습니다. 학계를 떠나 5년간 전업주부로 살기로 한 것입니다. 그녀는 이렇게 회고합니다.

"아이들은 너무 빨리 자랍니다. 그 시간을 함께 보내고 싶었고, 솔직히 저는 멀티태스킹에 그리 뛰어나지 못합니다." 케임브리지 거던 연구소(Gurdon Institute)에서 다니엘 세인트 존스톤(Daniel St Johnston) 교수 밑에서 박사후 연구를 하던 유망한 과학자가, 기저귀와 이유식의 세계로 들어간 것입니다. 그 5년 동안 테레사는 다른 어머니들과 우정을 쌓았고, 킹스 칼리지 런던에서 의료윤리와 법학 석사과정을 파트타임으로 수강하며 지적 갈증을 달렸습니다.

과학이 가져오는 윤리적·법적 함의를 고민한 이 시간은, 훗날 그녀의 치매 연구에 인문학적 깊이를 더하게 됩니다. 허사비스에게 두 아들의 존재는 어떤 의미일까요. 그는 드물게 공개한 인터뷰에서 아이들을 이렇게 묘사했습니다. "각자의 방식으로 뛰어납니다. 한 아이는 과학 쪽에, 다른 아이는 창의적인 쪽에."

부모가 모두 과학자인 가정에서 자라는 아이들이지만, 허사비스는 자신의 아버지 코스타스가 그랬듯 아이들에게 특정한 길을 강요하지 않는 것으로 알려져 있습니다. 코스타스 허사비스는

장난감 가게 주인이었다가, 싱어송 라이터였다가, 교사가되었 던 사람입니다. 그 다양성의 유산이 데미스에게로 이어졌 고, 다시 다음 세대로 전해지고 있는 셈입니다.

두 아들은 이제 십대입니다. 테레사가 '마더스 인 사이언스(Mothers in Science)' 프 로젝트에 기고한 글에 따르면, "십대와 함께 사는 것은 또 다른 도전이지만, 두 아이 모두 혼자서 학 교에 가고 올 수 있으니 일상의 물류는 훨씬 수월해졌다"고합니다. 그 사이 테레사는 2019 년 알츠하이머 리서치 UK(Alzheimer's Research UK)의 시니어 펠로우십을 받아 UCL에 자 신의 연구실을 열었습니다.

초파리(Drosophila melanogaster)를 모델 생물로 활용해 전두측 두엽 치매(Frontotemporal Dementia) 와 알츠하이머병의 분자 메커니즘을 연구하는 실험실입니다. 니콜리 랩(Niccoli Lab)이라는 이름으로 운영되는 이 연구실은, 2025년 현재 박사과 정 학생과 박사후 연구원 여러 명을 이끌며 사이언스(Science)지에 논문을 게재하는 성과를 내고 있습니다. 남편이 AI로 단백질의 구조를 예측해 노벨 화학상을 받는 동안, 아내는 초파리의 뇌세포에서 치매의 비밀을 추적하고 있었습니다.

한 가정에서 두 명의 과학자가 각자 의 방식으로 인류의 질병에 맞서고 있는 것입니다. 이 대칭이야말로 허사비스 가족의 이야기에서 가장 인상적인 부분일지 모릅니다.

노스런던 거주와 리버풀 FC 팬으로서의 일상 구글 인수 뒤에도 허사비스는 런던을 떠나지 않았습니다. 런던 북부의 하이게이트(Highgate). 칼 마르크스의 묘지가 있는 하이게이트 묘지로 유명하 고, 햄프스테드 히스(Hampstead Heath)의 녹지가 가까운 이 동네에서 허사비스 가족은 조 용히 살고 있습니다. 이브닝 스탠다드(Evening Standard)와의 인터뷰에서 허사비스는 런던을 떠날 계획이 없음을 분명히했습니다.

"우리는 규모에 비해 높은 성과를 냅니다. 세계 최 고의 대학들이 놀랍도록 똑똑한 사람들을 배출하고 있 고, 샌프란시스코만큼 기회가 많지 않기 때문에 오히려 최고 수준의 인재들이 더 흥미 로운 일을 찾고 있습니다." 구글 인수 이후에도 허사비스는 2016년까지 기차를 타고 출퇴근했다고 알려져 있습니다. 약간 더 비싼 안경과 재킷을 제외하면, 실리콘밸리식 변신을 거부한 것입니다.

한 프로필 기사의 표현을 빌리면, "체스판 위에서든 ZX 스펙트럼의 한계를 탐험하든 편안해 보이던 노스런던의 천재 소년 모습 그대로"였습니다. 이 소박함은 의도된 것일 수 있습니다. 허사비스가 공개적으로 말한 적은 없지만, 어린 시절 경제적으로 넉넉하지 않았던 가정 환경이 사치에 대한 무관심으로 이어졌을 가능성이 있습니다.

아버지 코스타스는 장난감 가게를 하다가, 노래를 쓰다가, 교 사가되었던 사람입니다. 풍족하진 않았으나 지적 자극은 풍부 했던 가정. 그 기억이 물질적 과시보다 지적 탐구를 우선시하는 성인 허사비스의 가치관을 형성한 것으로 보입니다. 앞서 잠깐 언급한 리버풀 FC와의 인연을 좀 더 들여다봅니다.

노스런던에 살면서 아스널이나 토트넘이 아닌 리버 풀을 응원한다는 것은, 영국 축구 문화에서 나름의 설 명이 필요한 선택입니다. 1970~80년 대 리버풀은 유럽 축구를 지배하던 팀이었고, 전 국적으로 팬을 확보했습니다. 런던에서 태 어난 어린 데미스가 그 시절 리버풀의 위용 에 매료되었을 가능성이 높습니다.

체스에서 세계 정상급이 되려 했던 소년에게, 유럽 정상에 서는 팀은 자연스러운 귀속 대상이었을 것입니다. 이 팬심은 성인이 되어서도 바뀌지 않았습니다. 2019년 5월 7일, 리버풀이 안필드(Anfield)에서 바르셀로나를 4대 0으로 꺾고 UEFA 챔피언스리그 결승에 진출한 역사적인 밤, 허사비스는 그 경기장에 있었습니다.

1차전에서 0대 3으로 패한 뒤 벌어진 기적 같은 역전. 디블록 오리기의 코너킥 트릭 플레이. 그날 안필드에 울려 퍼진 "You'll Never Walk Alone"의 합창 속에 AI의 미래를 설계하는 남자가 서 있었습니다.

허사비스의 리버풀 사랑은 개인적 취미를 넘어 직업적 영역으로까지 확장됩니다. 딥마인드는 리버풀 FC의 트래킹 데이터(tracking data) 분석에 협력한 것으로 알려져 있습니다. 선수와 공의 움직임을 AI로 예측하고, 다양한 시나리오의 결과를 시뮬레이션하는 기술입니다. 코너킥 전략에 이 기술이 공개적으로 활용되었습니다.

AI가 수천, 수만 번의 코너킥을 시뮬레이션하여 최적의 출발 위치와 동선을 찾아내는 것입니다. 노벨상 수상자가 자신이 응원하는 축구팀의 전술 개발에 AI를 제공하는 상황. 이것은 허사비스라는 인물의 독특한 면모를 보여주는 에피소드입니다. 거대한 과학적 야심과 소년 같은 축구 팬심이 한 사람 안에 공존하는 것입니다.

2025년에도 허사비스는 X(구 트위터)에 리버풀과 테마파크를 연결하는 글을 올리며 팬심을 과시했습니다. "내가 사랑하는 두 가지가 만나다: 리버풀과 테마파크." 나노바나나(Nano-Banana) AI 도구로 생성한 리버풀 테마의 등각 투영(isometric) 그래픽을 공유하며, "새로운 등각 투영 게임을 만들어야겠다, 참을 수 없다"고 덧붙였습니다.

노벨상 수상자가 자 구글 딥마인드 CEO가, 십대 때 만들었던 게임에 대한 향수와 축구팀에 대한 애정을 소셜미디어에 솔직하게 드러내는 모습. 이것이 허사비스의 일상입니다. 사적 생활의 보호와 일-생활 균형(혹은 불균형) 데미스 허사비스를 아는 사람들은 그를 묘사할 때 비슷한 단어를 씁니다. 수줍음이 많다(shy). 결연하다(determined). 공개적 관심을 피한다(low profile). AI 분야의 거물들 가운데서 허사비스는 독특한 위치에 있습니다. 일론 머스크의 트위터 폭풍이나 샘 올트먼의 정치적 행보, 마크 저커버그의 자기 브랜딩과는 거리가 먼 사람입니다.

그는 연구 결과와 논문으로 말하기를 선호하고, 가족에 대한 이야기는 거의 하지 않습니다. 이 사적 생활의 보호는 의식적인 선택으로 보입니다. 딥마인드가 구글에 인수되고, 알파고 가이세돌을 이기고, 알파폴드가 단백질 구조 예측 대회를 석권하면서, 허사비스에 대한 미디어의 관심은 기하급수적으로 커졌습니다.

2024년 노벨 화학상 수상 이후에는 더욱 그랬습니다. 그럼에도 허사비스는 가족사진을 공개하지 않고, 아내와 아이들의 이름을 인터뷰에서 언급하는 일이 드뭅니다. 아이들에 대해 알려진 것은 "각자의 방식으로 뛰어나다"는 짧은 언급뿐입니다. 이 절제에는 이유가 있을 것입니다.

AI가 사회적·정치적 논쟁의 한가운데에 놓인 시대입니다. 딥마인드 CEO의 가족이 불필요한 주목을 받는 것은 보안 문제이기도 하고,

아이들의 성장 환경을 보호하려는 부모의 본능이기도 합니다. 테레사 니콜리 역시 자신의 학술적 정체성을 독립적으로 유지합니다. 니콜리 랩의 웹사이트나 UCL의 연구 프로필 어디에도

남편의 이름은 등장하지 않습니다.

그녀는 테레사 니콜리라는 이름으로 논문을 쓰고, 연구비를 따내고, 학생들을 지도합니다. 구글 스칼라(Google Scholar)에서 그녀의 피인용 수는 4,700회를 넘습니다. 남편의 그림자가 아닌, 독립적인 과학자로서의 입지입니다.

그러나 일과 생활의 균형이라는 측면에서, 허사비스의 삶은 "균형"보다 "불균형"에 가깝습니다. 그의 작업 루틴은 잘 알려져 있습니다. 낮에는 CEO로서 경영과 회의를 하고, 밤에는 "두 번째 교대(second shift)"라 부르는 시간에 연구와 깊은 사고에 몰입합니다.

새벽 3시까지 논문을 읽고, 아이디어를 정리하고, 코드를 검토합니다. 이 패턴은 딥마인드 창업 이전부터 지속되어 온 것으로, 2,000편이 넘는 연구 논문, h-인덱스 83, 15만 회 이상의 피인용이라는 학술적 성과의 배경이기도 합니다. 이런 생활이 가정에 미치는 영향을 허사비스가 솔직히 말한 적은 없습니다.

다만 테레사의 경력 궤적이 간접적인 단서를 줍니다. 그녀가 5년의 경력 단절을 선택한 것은, 두 사람 모두가 풀타임으로 연구에 몰입하기 어렵다는 현실적 판단이 있었을 것입니다. 테레사는 이렇게 썼습니다. "아이들이 어렸을 때는 더 힘들었지만, 가족의 도움 이 컸고 오페어(au pair)가 함께 살았습니다." 오페어, 즉 집에 함께 사는 외국인 도우미를 두었다는 것은, 두 과학자 부부가 육아와 연구를 병행하기 위해 적극적인 지원 체계를 구축했음을 의미합니다.

허사비스는 가족을 최우선에 둔다고 알려져 있지만, 그 "최우선"이 실제로 어떤 형태인지 는 외부에서 알기 어렵습니다. 구글 딥마인드의 CEO이면서 동시에 아이소모픽 랩스 (Isomorphic Labs)의 CEO이고, 영국 정부의 AI 자문관이며, 수시로 국제 컨퍼런스에서 기조 연설을 합니다. 2024년에는 노벨상 수상 후 스톡홀름 시상식, 트라 이베카 영화제 다큐멘터리 리프리미어, 렉스 프리드먼 팟캐스트 출연 등이 줄을 이었습니다.

이 일정 속에서 하이게이트의 집에 돌아와 침대 아들과 저녁 식사를 하는 시간이 얼마나 되는지, 허사비스만이 알 것입니다. 한 가지 확실한 것은, 허사비스가 런던을 떠나지 않는다는 사실 자체가 가족에 대한 헌신의 표현이라는 점입니다. 딥마인드의 본사를 런던에 유지하는 것, 구글 인수 시 런던 잔류를 조건으로 내건 것, 실리콘밸리의 유혹을 거부한 것. 이 모든 결정에는

"이곳이 내 가족의 터전"이라는 무언의 선언이 깔려 있습니다. 아이들이 다니는 학교, 테레사의 UCL 연구실, 주말마다 산책하는 햄프스테드 히스. 이 일상의 반경을 지키는 것이, 세계를 바꾸려는 야심과 공존하는 허사비스 방식의 균형인지도 모릅니다. 어쩌면 이것은 균형이 아니라, 의식적인 불균형의 관리일 수 있습니다. 완벽한 균형은 불가능하다는 것을 인정하되, 무엇을 절대 포기하지 않을 것인지를 정해놓는 방식. 허사비스에게 그 "절대"는 연구이면서 동시에 런던의 가정입니다.

둘 다 놓치지 않기 위해 새벽 3시까지 깨어 있는 것이라면, 그것은 그 나름의 해법입니다. 아름답다고 말하기는 어렵지만, 정직하다고 말할 수 있습니다. 허사비스가 사랑하는 책과 영감의 원천 데이비드 도이치의 *The Fabric of Reality: 물리학과 지식의 성장* 허사비스에게 가장 읽고 싶은 책 한 권을 꼽으라 하면, 그는 여러 차례에 걸쳐 같은 이름을 들었습니다.

데이비드 도이치(David Deutsch)의 《실재의 구조(The Fabric of Reality)》. 에즈라 클라인(Ezra Klein)의 뉴욕타임스 팟캐스트에 출연했을 때, 허사비스는 이 책을 추천 도서 첫

번째로 꼽으며 말했습니다. "물리 학의 거대한 질문들을 모두 제기하는 책이라고 생각합니다. 언젠가 우리의 AI 도구로 그 질문들에 도전하고 싶습니다."

이 한마디에 허사비스의 지적 야심 전체가 응축되어 있습니다. 도이치는 옥스퍼드 대학교의 물리학자이자, 양자 컴퓨팅의 이론적 선구자입니다. 《실재의 구조》는 1997년에 출간되었는데, 네 가지 이론의 통합을 시도합니다.

양자역학, 인식론(칼 포퍼의 과학 철학), 진화론(리처드 도킨스), 그리고 계산 이론(앨런 튜링). 도이치의 핵심 주장은 이것입니다. 이 네 가지 가닥이 실은 하나의 현실, 하나의 실재를 서로 다른 각도에서 조명하고 있으며, 통합적 이해만이 우주의 진정한 모습에 다가갈 수 있다는 것. 허사비스가 이 책에 끌리는 이유는 명백합니다. 허사비스 자신이 평생에 걸쳐 해온 일이 바로 그것이기 때문입니다.

체스와 게임 디자인, 신경과학과 머신러닝, 강화학습과 단백질 구조 예측. 서로 무관해 보이는 분야들을 하나의 프레임워크 안에 통합하려는 시도. 도이치가 물리학의 차원에서 했던 것을 허사비스는 지능의 차원에서 하고 있는 셈입니다. 도이치의 책에서 허사비스가 받은 영향은 구체적으로 추적할 수 있습니다. 도이치는 "설명 (explanation)"의 힘을 강조합니다.

세상을 예측하는 것만으로는 충분하지 않으며, 왜 그런지를 이해해야 한다는 것입니다. 이 관점은 허사비스가 딥마인드를

설계할 때 핵심 원칙으로 삼은 것과 일치합니다. 알파고는 바둑에서 이기기 위한 프로그램이 아니었습니다. 바둑을 통해 범용 학습 알고리즘을 검증하기 위한 실험이었습니다.

승리는 부산물이고, 이해가 목적이었습니다. 도이치는 다중세계 해석(many-worlds interpretation)의 지지자이기도 합니다. 양자역학의 이 해석에 따르면, 모든 가능한 사건이 실제로 발생하며 서로 다른 우주에서 일어납니다. 허사비스가 이 해석을 지지하는지는 알려져 있지 않지만, "현실의 시뮬레이션"이라는 개념에 대한 그의 관심은 분명합니다.

신경과학 박사과정에서 발전시킨 "마음의 시뮬레이션 엔진(simulation engine of the mind)" 이론, 그리고 딥마인드가 개발 중인 월드 모델(world model)은 모두 "현실을 모사(模寫)하고 예측하는 시스템"이라는 공통점을 갖습니다. 고등학생 시절 허사비스가 읽은 스티븐 와인버그(Steven Weinberg)의 《최종 이론의 꿈 (Dreams of a Final Theory)》도 같은 맥락에 놓입니다. 물리학의 근본 법칙들을 하나의 통합 이론으로 엮으려는 꿈. 허사비스는 이 꿈을 AI의 언어로 번역하고 있습니다.

2025년 렉스 프리드먼과의 두 번째 팟캐스트 대답에서 그는 "자연에서 생성되거나 발견될 수 있는 모든 패턴은, 고전적 학습 알고리즘으로 효율적으로 발견되고 모델링될 수 있다"는 추론을 제시했습니다. 자연이 무작위가 아니라 선택 과정의 산물이기 때문에, 그 패턴은 학습 가능하다. 이것은 도이치의 "모든 물리적 과정은 계산적으로 시뮬레이션 가능하다"는 처치-튜링-도이치 원리의 AI 버전이라 할 수 있습니다. 《실재의 구조》는 허사비스에게 하나의 책 이상입니다.

그것은 그가 평생 풀고자 하는 문제의 지도입니다. 지능이란 무엇인가, 현실이란 무엇인가, 지식은 어떻게 성장하는가. 이 세 질문이 어떻게 하나로 수렴하는지를 보여주는 안내서. 허사비스가 "언젠가 우리의 AI 도구로 이 질문들에 도전하고 싶다"고 말했을 때, 그것은 먼 미래의 꿈이 아닐 수 있습니다. 알파폴드가 50년 묵은 단백질 접힘 문제를 풀었듯, AI가 물리학의 근본

법칙에 새로운 빛을 비추는 날이 올 수 있습니다.

그리고 그날이 오면, 허 사비스의 책장에서 가장 닳은 책은 아마도 《실재의 구조》일 것입니다. 그렉 이건의 Permutation City: 의식과 시뮬레이션 허사비스가 추천하는 책 목록에서 데이비드 도이치가 과학적 영감의 원천이라면, 그렉 이 건(Greg Egan)은 상상력의

원천입니다. 호주 출신의 SF 작가 이건이 1994년에 발표한 《순 열 도시(Permutation City)》는, 허사비스가 여러 인터뷰에서 반복적으로 언급하는 소설입니다. 에즈라 클라인 쇼에서 허사비스는 이 책을 이렇게 소개했습니다. "AI와 초현실적 시뮬레이션의 맥락에서 세상이 얼마나 흥미롭고 기이해질 수 있는지를 보여주는 놀라운 이야기입니다."

X(구 트위터)에서도 허사비스는 이건의 또 다른 소설 《디아스 포라(Diaspora)》를 추천하는 팔로워에게 이렇게 답했습니다. "동의합니다. 디아스 포라는 대단하지만, 《순 열 도시》를 먼저 읽는 게 좋다고 생각합니다."

《순열 도시》가 다루는 세계는 이렇습니다. 가까운 미래, 인간의 의식을 컴퓨터에 복제하는 기술이 존재합니다. 복제된 의식체, 즉 "카피(Copy)"들은 가상 환경에서 살아갑니다. 소설의 주인공 폴 더럼은 의식의 본질에 대한 급진적 실험을 합니다.

시뮬레이션된 의식이 "진짜" 의식인지, 시뮬레이션의 기반이 되는 물리적 하드웨어가 사라져도 의식이 존속할 수 있는지를 탐구하는 것입니다. 이 소설이 허사비스에게 울림을 주는 이유는 여러 겹입니다. 첫째, 의식의 문제. 허 사비스 는 2022년 렉스 프리드먼과의 대화에서 "의식과 지능은 이중 해리(double dissociable)가 가능하다, 하나 없이 다른 하나를 가질 수 있다"고 말한 바 있습니다.

지능은 있지만 의식이 없는 시스템, 의식은 있지만 지능이 제한된 존재. 이 가능성의 공간을 《순열 도시》는 소설의 형태로 탐험합니다. 둘째, 시뮬레이션의 문제. 허사비스의 신경과학 연구에서 핵심 개념이었던 "마음의 시뮬레이션 엔진"은, 인간의 뇌가 현실을 모사하여 미래를 예측하는 내부 모델이라는 아이디어입니다. 《순열 도시》는 이 아이디어를 극단까지 밀어붙입니다. 만약 시뮬레이션이 충분히 정교하다면, 시뮬레이션 안의 존재는 자신이 시뮬레이션 속에 있다는 것을 알 수 있을까? 알 수 있다면, 그것은 무엇을 의미할까? 셋째, 물리학의 근본 법칙과 계산의 관계. 이건은 하드 SF(hard science fiction) 작가로, 소설 속에 실제 물리학과 수학을 정밀하게 반영합니다.

《순열 도시》에 등장하는 "먼지 이론 (dust theory)"은 허구이지만, 계산과 현실의 경계에 대한 깊은 철학적 질문을 제기합니다. 허사비스가 도이치의 《실재의 구조》에서 받은 과학적 영감을, 이건의 소설에서 서사적 형태로 다시 만나는 것입니다. 허사비스의 독서 취향에서 드러나는 패턴이 있습니다.

그는 아이디어를 위해 SF를 읽습니 문체나 캐릭터보다 개념에 끌립니다. 아시모프의 《파운데이션》 시리즈에 대

해서도 비슷한 말을했습니다. "대부분의 SF가 그렇듯, 나는 주로 아이디어 때문에 SF를 읽습니다. 12살쯤에 《파운데이션》을 읽었으니 그때 문장의 품질을 제대로 평가할 수 있었는지는 모르겠습니다."

이건의 문체 역시 감정적 깊이보다 개념적 밀도로 유명합니다. 두 사람은 이 점에서 닮았습니다. 아름다움보다 진실에 먼저 도달하려는 성향. 이건의 또 다른 작품 《디아스포라》도 허사비스의 세계관에 영향을 미쳤습니다.

인류가 로봇, 디지털 존재, 소프트웨어 사회 등 다양한 포스트휴먼 형태로 진화한 미래를 그린 이 소설은, 허사비스가 자주 말하는 "지능의 확장"이라는 비전의 문학적 원형이라 할 수 있습니다. AI가 인간 지능을 대체하는 것이 아니라, 지능의 형태 자체가 다양해지는 미래. 《순 열 도시》와 《디아스포라》는 그 가능성의 지도를 펼쳐놓습니다. SF를 읽는 과학자는 많습니다.

그러나 SF에서 읽은 아이디어를 실제로 구현하려는 과학자는 드뭅니다. 허사비스는 후자에 속합니다. 《순열 도시》가 상상한 의식의 시뮬레이션은 아직 먼 이야기지만, 알파폴드가 상상한 단백질 구조의 예측은 이미 현실이 되었습니다. 이건의 소설을 읽으며 10대의 허사비스가 꾸었을 꿈의 일부는, 이미 실현되었고, 나머지는 실현을 향해 가고 있습니다.

더글러스 호프스태터의 Gödel, Escher, Bach, 유발 하라리의 Sapiens, 아시모프의 Foundation 1994년, 17세의 데미스 허사비스는 볼프로그 프로덕션의 사무실에서 낮에는 테마파크의 AI를 프로그래밍하고, 밤에는 동료들과 인공지능의 한계에 대해 토론했습니다. 그 토론의 불씨를 지핀 책이 있었습니다. 더글러스 호프스태터(Douglas Hofstadter)의 《괴델, 에셔, 바흐: 영원한 황금 노끈(Gödel, Escher, Bach: An Eternal Golden Braid)》, 줄여서 GEB. 허사비스는 이 책에 대해 이렇게 말했습니다.

"저에게 정말 중요한 과학 서적이었습니다. 테마파크를 만들면서 AI를 개발하고 있었는데, GEB와 같은 책도 함께 읽고 있었습니다. 믿을 수 없는 저작입니다.

괴델의 불완전성 정리를 수학과, 에셔의 그림과, 바흐의 푸가와 엮어서 이것들이 모두 어떤 식으로 연결되어 있는지를 보여줍니다. 반복되는 패턴의 순환, 무한 패턴. 그리고 그것을 의식과 지능에 연결시켰습니다. 정말 영감을 받았고, 이 깊은 질문들에 대해 생각하게 되었습니다." GEB는 1979년에 출간되어 풀리처상을 수상한 책입니다. 수학자 쿠르트 괴델, 화가 M.C. 에셔, 작곡가 요한

세바스티안 바흐라는 세 거장의 작업에서 공통된 구조를 발견합니다. 자기 참조(selfreference), 재귀(recursion), 그리고 형식 체계 안에서의 의미 발생. 호프스태터의 궁극적 질문은 이것입니다. 물질적 기반(뇌의 뉴런) 위에서 어떻게 비물질적 현상(의식, 자아)이 출현하는가? 이 질문은 허사비스의 전 경력을 관통합니다.

체스 AI에서 시작해, 게임 속 NPC의 학습 행동을 설계하고, 해마의 신경 기제를 연구하고, 딥마인드를 세워 범용 인공지능을 추구하기 까지. 모든 단계에서 허사비스는 "지능이란 무엇인가"라는 질문 앞에서 있었고, GEB는 그 질문의 첫 번째 이정표였습니다. 10대의 허사비스가 동료들과 나눈 "AI를 게임에만 쓰지 말고 인간 수준으로 끌어올릴 수 있지 않을까"라는 대화는, GEB가 심어준 씨앗에서 자라난 것입니다. 한편, 이보다 더 어린 시절의 책이 있습니다.

아이작 아시모프(Isaac Asimov)의 《파운데이션/Foundation》 시리즈. 허사비스는 이렇게 회고합니다. "정말 형성적(formative)이었던 것은 아시모프의 파운데이션 시리즈였습니다. 흥미롭게도 로봇 시리즈가 아니라 파운데이션이었습니다.

로봇 책은 사실 거의 읽지 않았습니다.” 《파 운데이션》은 수학자 해리 셸던이 "심리역사학(psychohistory)"이라는 수학적 도구를 이용해 은하 제국의 몰락을 예측하고, 암흑기를 단축하기 위해 비밀 계획을 세우는 이야기입니다. 12세 무렵 이 시리즈를 읽은 허사비스에게, 수학과 과학으로 문명의 미래를 설계 한다는 아이디어는 강렬한 인상을 남겼을 것입니다.

딥마인드의 미션 "지능을 풀고, 그것으로 나머지 모든 것을 푼다"는 선언에는 셸던의 야심이 메아리칩니다. 차이가 있다면, 셸던은 인간 행동의 통계적 예측에 의존했고, 허사비스는 학습하는 기계에 의존한다는 점입니다. 이 책들의 선택이 말해주는 것이 있습니다. 허사비스는 이안뱅크스(Iain M. Banks)의 《컬처(Culture)》 시리즈도 사랑합니다. 《고려하라 플레바스(Consider Phlebas)》를 가장 좋아 하는 책으로 꼽으며, 테마파크 게임의 치트 코드를 이 소설의 주인공 이름인 "호르자 (Horza)"로 설정했을 정도입니다.

뱅크스의 컬처 시리즈는 고도로 발달한 AI(마인드, Mind)가 인간 및 외계 종족과 공존하는 유토피아적 문명을 그립니다. 일론 머스크가 자신의 SpaceX 드론선 이름을 컬처 시리즈의 우주선 이름에서 따온 것과 마찬가지로, 허사비스 역시 이 시리즈에서 AI와 인류가 공진화하는 미래의 청사진을 읽어냈을 것입니다.

유발 하라리(Yuval Noah Harari)의 《사피엔스(Sapiens)》와의 관계는 다른 맥락에서 흥미롭습니다. 하라리는 AI의 위험성에 대해 자주 경고하는 지식인이고, 허사비스는 AI의 잠재력을 실현하려는 실무자입니다. 두 사람이 직접 대화한 기록은 드물지만, 허사비스가 《사피엔스》를 읽었다는 것은 그가 기술 낙관주의에만 머물지 않음을 시사합니다.

《사피엔스》는 호모 사피엔스가 인지 혁명, 농업 혁명, 과학 혁명을 거치며 지구를 지배하게 된 과정을 추적하면서, 그 과정의 이면에 있는 폭력과 착취도 직시합니다. 허사비스가 AI 안전 성과 정렬(alignment) 문제에 진지한 것은, 이러한 인문학적 독서가 기술적 낙관주의에 균형추 역할을 하고 있기 때문일 수 있습니다. 허사비스의 서재를 관통하는 맥은 이것입니다.

개별 분야의 경계를 허물고, 큰 그림을 그리려는 책들. GEB가 수학·예술·음악을 엮고, 《실재의 구조》가 물리학·인식론·진화론·계산 이론을 엮고, 《파운데이션》이 수학과 문명의 운명을 엮듯이, 허사비스 자신도 신경과 학과 컴퓨터 과학과 게임 디자인을 엮어 하나의 통합적 지능 이론을 향해 나아가고 있습니다. 그가 읽는 책은 그가 꿈꾸는 세계의 축소판입니다. 음악과 취미 드럼 앤 베이스, 앰비언트 일렉트로닉, 클래식까지 폭넓은 음악 취향 딥마인드 런던 본사에는 음악실이 있습니다. AI 연구소에 음악실이라니, 의아할 수 있습니다. 그러나 허사비스에게 음악은 사치가 아니라 도구입니다.

미국 아카데미 오브 어치브먼트(Academy of Achievement)와의 심층 인터뷰에서 허사비스는 이렇게 설명했습니다. "저는 음악을 많이 활용합니다. 집중할 때 음악을 듣는데, 어떤 분위기를 만들려는지에 따라 다른 음악을 듣습니다.

클래식부터 드럼 앤 베이스까지, 뭘 하려는가에 달려 있습니다." 이 대답에는 허사비스의 사고방식이 압축되어 있습니다. 음악조차 목적 지향적으로 활용한다는 것. 프로그래밍을 할 때는 비트가 있고 에너지가 넘치는 음악을, 생각하거나 읽을 때는 비발디나 모차르트 같은 클래식을 선택합니다.

"기분을 맞추는(get you in the right mood)" 것이 핵심이라고 그는 말합니다. 허사비스가 선호하는 구체적인 장르는 "리퀴드 드럼 앤 베이스(liquid drum and bass)"입니다. 리퀴드 드럼 앤 베이스의 하위 장르로, 빠른 비트 위에 멜로디가 흐르지만, 가사는 없는 음악입니다. 허사비스의 설명은 정밀합니다. "비트가 있되 멜로디도 있어야 하고, 흥미로워야 합니다. 하지만 가사가 있으면 안됩니다."

가사가 있으면 뇌가 단어를 따라가기 시작하니까요. 가사가 없는, 순수 기악만됩니다." 그리고 이것을 신경과학적으로 해석합니다. "아마 알파 상태로 들어가게 해주는 것 같습니다. 이런 현대 음악

은 반복적이거든요. 몇 번 들으면 배경으로 깔리고, 뇌를 일종의 리듬 상태로 넣어주는데, 그게 집중에 도움이됩니다." 여기서 주목할 것은 허사비스가 자신의 음악 취향을 설명할 때조차 신경과학의 언어를 쓴다는 점입니다.

"알파 상태(alpha state)"는 뇌파의 한 종류로, 이완된 집중 상태를 나타냅니다. 명상이나 창의적 사고와 관련된 것으로 알려진 뇌파입니다. 인지신경과학 박사가 자신의 뇌를 실험 대상처럼 관찰하고, 최적의 작업 환경을 설계하는 것입니다. 음악 취향이 아니라 음악 전략이라 불러야 할 것입니다.

허사비스는 젊은 사람들에게 이 접근법을 권합니다. "자기 자신이 어떻게 가장 잘 작동하는지를 배우는 것이 가장 중요한 일 중 하나입니다. 어떤 것이 특정한 창의적 상태로 들어가게 해주는지를 파악하고, 그것을 최적화하세요."

이 조언에는 게임 디자이너, 신경과학자, AI 연구자를 거처온 사람의 통합적 자기이해가 담겨 있습니다. 클래식 음악에 대한 허사비스의 관심은 GEB와도 연결됩니다. 호프스태터가 바흐의 푸가에서 발견한 자기 참조적 구조, 카논에서 나타나는 반복과 변주의 패턴은, 허사비스가 AI에서 추구하는 학습의 구조와 닮아 있습니다.

반복을 통해 패턴을 발견하고, 변주를 통해 새로운 것을 창조하는 과정. 바흐의 음악이 수학적 구조 위에 세워져 있듯, 딥마인드의 AI도 수학적 구조 위에 세워져 있습니다. 영화 음악도 허사비스의 세계에서 중요한 위치를 차지합니다. 엘릭서 스튜디오 시절, 허사비스는 작곡가 제임스 해니건(James Hannigan)과 협업하여 《리퍼블릭: 더 레볼루션》과 《이블 지니어스》의 인터랙티브 음악을 만들었고, 두 작품 모두 BAFTA 인터랙티브 음악 부문 후보에 올랐습니다.

게임 속 음악이 플레이어의 행동에 반응하여 변화하는 인터랙티브 사운드트랙은, 그 자체가 AI와 음악의 교차점이었습니다. 허사비스가 아카데미 오브 어치브먼트 인터뷰에서 한 또 다른 말이 있습니다. "영화도 많이 활용합니다."

음악과 영화, 게임과 과학. 이 모든 것이 허사비스에게는 "창의적 상태 (creative state)"로 들어가기 위한 입구입니다. 어떤 사람에게는 음악이 배경이고, 어떤 사람에게는 감상의 대상입니다. 허사비스에게 음악은 뇌의 작동 모드를 전환하는 스위치입니다. 리퀴드 드럼 앤 베이스를 들면 프로그래밍 모드로, 비발디를 들면 사색 모드로. 그의 뇌는 음악에 의해 튜닝되는 악기이고, 그 악기에서 나오는 소리가 알파고이고 알파폴드인 것입니다.

물리학에 대한 깊은 관심과 독서 습관 허사비스의 X(구 트위터) 프로필에는 짧은 소개가 적혀 있습니다. "구글 딥마인드 공동 창립자 겸 CEO. AGI를 향해 작업 중. 현실의 근본적 본질을

이해하려 노력 중(Trying to understand the fundamental nature of reality)." 마지막 문장이 핵심입니다.

AI 연구자의 프로 필에 "현실의 근본적 본질을 이해하려 한다"는 문장이 들어가는 것은, 허사비스가 자신을 기술자가 아니라 탐구 자로 정체화하고 있음을 보여줍니다. 이 물리학에 대한 관심은 고등학생 시절부터 시작되었습니다. 스티븐 와인버그 (Steven Weinberg)의 《최종 이론의 꿈(Dreams of a Final Theory)》을 읽은 것이 시 발점이었습니다. 와인버그는 1979년 노벨 물리학상 수상자이며, 이 책에서 물리학의 모든 법칙을 하나의 통합 이론으로 설명하려는 꿈을 서술합니다.

허사비스에게 이 책은 "왜 과학을 하는가"에 대한 답을 제시했습니다. 알 수 있는 것의 한계까지 가보려는 충동. 성인이 된 허사비스의 물리학 관심은 더 구체적인 형태를 띠니다. 2025년 렉스 프리드먼과의 팟캐스트에서 그는 유체역학과 나비에-스토크스 방정식(Navier-Stokes equations)에 대해 논합니다. 전통적으로 고전 시스템에서는 풀기 극히 어려운 것으로 여겨지는 이 문제들에 대해, AI 기반 접근이 놀라운 성과를 낼 수 있다는 가능성을 탐색합니다.

딥마인드가 이미 달성한 기상 예측(WeatherNext), 핵융합 플라즈마 제어(토카막 반응로 프로젝트) 등은 물리학의 난제에 AI를 적용한 구체적 사례입니다. 허사비스의 물리학 관심에 단순한 취미를 넘어서는 증거가 있습니다. 그가 노벨상 수상 강연에서 제시한 이론입니다.

"자연에서 생성되거나 발견될 수 있는 모든 패턴은, 고전적 학습 알고리즘에 의해 효율적으로 발견되고 모델링될 수 있다." 그는 이것을 "생존 가능한 것 중 가장 안정적인 것의 선택(survival of the stablest)"이라 부릅니다. 생명에는 진화가, 지질학적 시간에는 풍화가, 우주적 시간에는 궤도 역학이 작용 하여 패턴을 만들어냅니다. 이 패턴들이 무작위가 아니기 때문에, 역으로 학습 알고리즘이 그것을 효율적으로 재발견할 수 있다는 것. 이것은 물리학의 언어로 표현된 AI 이론이자, AI의 언어로 표현된 물리학 이론입니다. 허 사비스는 두 세계의 통역자인 셈입니다. 그리고 이 통역 능력은 그의 독서 습관에서 비롯됩니다.

허사비스의 독서 패턴에는 몇 가지 특징이 있습니다. 첫째, 그는 아이디어 중심으로 읽습니다. 문학적 아름다움이나 서사적 긴장보다, 새로운 개념을 발견하는 것을 즐깁니다. 아 시모프의 《파운데이션》에 대해 "대부분의 SF가 그렇듯, 나는 주로 아이디어 때문에 읽는다"고 말한 것이 이를 증명합니다.

둘째, 그는 분야를 가리지 않고 읽습니다. 양자물리학 (도이치), SF(이건, 뱅크스, 아시모프), 철학(호프스태터), 역사(하 라리), 경영(에드 캣멀의 《크리에이티비티 Inc.》)까지. 셋째, 그는 읽은 것을 연결합니다. GEB에서 읽은 자기 참조 적 구조를 AI 아키텍처에, 《파운데이션》에서 읽은 심리역사학을 범용 학습 알고리즘에, 도이치에서 읽은 계산 가능성을 딥마인드의 연구 방향에 연결합니다. 이런 독서 습관을 가능하게 하는 것이 그의 "야간 교대(night shift)"입니다.

낮에는 CEO로서의 회의와 경영 업무에 시간을 쓰고, 밤에는 논문과 책을 읽고 깊이 생각합니다. 새벽 3 시까지 이어지는 이 시간에 허사비스는 가장 창의적이됩니다. 리퀴드 드림 앤 베이스가 배경에 깔리고, 화면에는 최신 물리학 논문이 열려 있고, 노트에는 새로운 연결 고리의 스케치가 채워집니다. 허사비스에게 독서는 입력(input)이고, 연구는 출력(output)이며, 그 사이를 잇는 것은 연결 (connection)입니다. 에드 캣멀의 《크리에이티비티 Inc.》에 대해 그는 이렇게 말했습니다.

"읽고 나서 전에 생각하지 못한 20개의 새 아이디어가 떠오른 적은 많지 않습니다. 이 책은 그렇게 만들어주었습니다." 한 권의 책에서 20개의 새 아이디어를 뽑아내는 사람. 이것이 허사비스의 독서법입니다.

책은 완성된 지식이 아니라, 새로운 질문의 발화점입니다. 물리학의 근본 법칙, 의식의 본질, 지능의 구조. 이 세 가지 질문은 허사비스의 서재에서 하나의 질문으로 수렴합니다. 현실이란 무엇인가? X 프로필의 마지막 문장이 그래서 단순한 수사가 아닙니다. "현실의 근본적 본질을 이해하려 노력 중."

이것은 직업 설명이 아니라 존재 선언입니다. 그리고 그 선언의 주석(footnote)에는, 10대 때 읽은 와인버그와 호프스태터와 아시모프의 이름이, 20대 때 읽은 도이치와 이건의 이름이, 그리고 아직 쓰이지 않은 수많은 책들의 빈자리가 적혀 있습니다.

런던 타운하우스가 늘어난 거리

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

3월, 서울의 한 호텔 방에서 데미스 허사비스는 넥타이를 매지 않은 검은색 정장 차림으로 창밖을 응시하고 있었습니다. 이세돌 9단과의 대국을 앞둔 그는 겉으로는 침착해 보였지만, 그의 머릿속은 수백만 가지의 경우의 수와 잠재적인 시스템 오류에 대한 걱정으로 끓어오르고 있었습니다. 딥마인드의 초기 멤버이자 오랜 동료인 무스타 파 솔레이만은 그 순간의 허사비스를 “폭풍의 눈”이라고 묘사했습니다.

그는 주변의 모든 혼란을 잠재우는 고요함을 유지하면서도, 그 내부에는 그 누구보다도 격렬한 에너지를 품고 있는 리더였습니다. 허사비스의 리더십은 ‘완벽’에 대한 강박에 가까운 집착에서 시작됩니다. 동료들은 그를 두고 “만족을 모르는 사람”이라고 입을 모읍니다. 딥마인드의 수석 연구원 데이비드 실버는 허사비스가 연구 결과물을 검토할 때의 깐깐함을 이렇게 회상합니다.

“그는 단순히 ‘작동한다’는 것만으로는 만족하지 않습니다. ‘왜 작동하는가?’, ‘이것이 최선의 방법인가?’, ‘이것이 범용 인공지능(AGI)으로 가는 길에 어떤 의미를 갖는가?’를 끊임없이 묻습니다. 그 질문들은 때로 연구자들을 지치게 만들지만, 결국 우리를 한계 너머로 밀어 올리는 원동력 이 됩니다.”

이러한 완벽주의는 딥마인드의 조직 문화에 깊게 뿌리내려 있습니다. 허사비스는 ‘빠르게 움직여 파괴하라(Move fast and break things)’는 실리콘밸리의 격언을 거부합니다. 대신 그는 ‘천천히, 그러나 확실하게 검증하며 나아가라’는 과학적 방법론을 고수합니다.

이는 구글이라는 거대 기업의 속도전과는 종종 마찰을 빚기도 했습니다. 구글이 챗GPT의 등장에 대응해 바드(Bard)와 제미니(Gemini) 출시를 서두를 때, 허사비스는 안전성과 정확성을 담보하지 않은 출시는 딥마인드의 철학에 위배된다고 마지막까지 신중론을 펼쳤습니다. 일부 동료들은 이러한 그의 태도를 답답해하기도 했지만, AI가 환각(Hallucination)을 일으키거나 편향된 정보를 내놓을 때마다 허사비스의 우려가 옳았음이 증명되었습니다.

그의 리더십이 가진 또 다른 특징은 ‘보호막’으로서의 역할입니다. 그는 연구원들이 외부의 압력, 특히 상업적 성과에 대한 압박에서 벗어나 연구에만 몰두할 수 있는 환경을 만드는 데 온 힘을 쏟았습니다. 2014년 구글에 인수될 때 그가 내건 조건 중 하나가 런던에 본사를 유지하는 것이었다는 사실은 유명합니다.

실리콘밸리의 소란스 러움에서 한 발짝 떨어져, 학문적 순수성을 지킬 수 있는 물리적, 심리적 공간을 확보하려 했던 것입니다. 한 연구원은 “데미스는 우리를 위해 총알을 막아주는 방탄유리 같다”고 말했습니다. 그는 구글 경영진과의 치열한 협상 테이블에서는 냉철한 전략 가졌지만, 연구실로 돌아오면 다시 호기심 많은 과학자의 얼굴로 돌아와 “어제 그 아이디어는 어떻게 됐어?”라고 물으며 눈을 반짝였습니다.

하지만 이러한 리더십에는 분명한 비용이 따릅니다. 모든 것을 통제하고 완벽을 기하려는 성향은 의사결정의 병목 현상을 초래하기도 했습니다. 딥마인드의 조직이 커지면서 허사비스 한 사람이 모든 프로젝트의 세부 사항을 챙기는 것은 물리적으로 불가능해졌습니다. 그런데도 그는 여전히 주요 프로젝트의 방향성을 직접 설정하고 싶어 했고, 이는 때때로 실무진에게 과도한 업무로드와 기다림의 시간을 안겨주었습니다.

일부 전직 직원들은 딥 마인드의 높은 기준이 숨 막히는 압박감으로 다가왔다고 토로하기도 했습니다. “천재들과 함께 일한다는 자부심은 컸지만, 내가 그 기준에 미치지 못할까 봐 매일 두려움에 떨었다” 는 고백은 허사비스가 만든 ‘엘리트 주의’ 문화 의 그늘을 보여줍니다. 허사비스는 동료들에게 단순한 상사가 아닙니다. 그는 함께 체스를 두고, 늦은 밤까지 지 우 주의 기원에 대해 토론하는 동료이자, 때로는 범접할 수 없는 직관을 보여주는 스승이기도 합니다. 그의 리더십은 카리스마 넘치는 연설이나 강압적인 지시에서 나 오는 것이 아닙니다 그것은 문제의 본질을 꿰뚫어 보는 지적 탁월함, 그리고 그 문제를 해결하겠다는 순수 한 열정에서 비롯됩니다.

동료들은 그가 제시하는 ‘지능을 풀어 세 상을 구한다’는 비전에 깊이 공감하기에, 그의 깐깐함과 집요함을 기꺼이 견뎌내는 것입니다. 그는 천재라는 신화 속에 박제된 인물이 아니라, 그 신화를 현실로 만들기 위 해 매일 자신과 동료들을 벼랑 끝까지 밀어붙이는, 고뇌하는 리더입니다. 연구자, 경영자, 기업인을 겸하는 복합적 역할 데미스 허사비스의 명함에는 ‘CEO’라는 직함이 찍혀 있지만, 그의 정체성은 단일하지 않습니다.

그는 세계적인 신경과학 연구자이자, 수천 명의 직원을 이끄는 경영자이며, 거대 자본과 기술 권력을

조율해야 하는 기업인입니다. 이 세 가지 역할은 서로 충돌하기도 하 고 보완하기도 하면서 허사비스라는 복잡한 인물을 구성합니다. 먼저 연구자로서의 허사비스는 여전히 현역입니다.

그는 단순히 보고만 받는 관리 자가 아 닙니다. 2024년 노벨 화학상 수상의 영예를 안겨준 알파폴드(AlphaFold) 프 로젝트에서도 그는 핵심 아이디어를 제공하고 연구의 방향을 틀어주는 역할을했습니다. 그는 여전히 최 신 논문을 챙겨 읽고, 연구원들과 화이트보드 앞에서 수식을 써가 며 토론하는 것을 즐깁니 그에게 경영은 연구를 지속하기 위한 수단일 뿐, 그 자체가 목적이었던 적은 없었습니 그는 뇌의 해마가 기억을 재구성하는 방식을 AI의 강화학 습 알고리즘에 적용하는 아이 디어를 직접 제안하는 등, 신경과학자로서의 배경을 AI 연구에 적극적으로 활용합니다. 이러한 ‘연구자적 감각’은 딥마인드가 다른 빅테크 기업의 AI 조직과 차별화되는 가장 큰 경쟁력입니다.

그러나 조직이 커지면서 그는 어쩔 수 없이 경영자의 옷을 입어야했습니다. 2010 년 런던 의 작은 사무실에서 시작한 딥마인드는 이제 수천 명의 박사급 인재들이 모인 거대한 조직 이되었습니다. 허사비스는 자유분방한 해커 문화와 엄격한 학문적 규율을 조화시켜야 하 는 과제를 안게되었습니다.

그는 딥마인드를 ‘21세기의 벨 연구소 (Bell Labs)’로 만들고자했습니다. 이를 위해 그는 학계의 최고 인재들을 영입하는 데 공을 들였고, 그들이 논문 실 적에 얽매이지 않고 장기적인 난제에 도전할 수 있도록 평가 시스템을 설계했습니다. 동시 에 그는 구글이라는 모회사의 기대치를 관리해야했습니다.

매년 막대한 적자를 내면서도 당장 돈이 되지 않는 순수 과학 연구에 천문 학적인 자금을 쏟아붓는 것을 정당화하기 위 해, 그는 끊임없이 딥마인드의 가치를 증 명해야했습니다. 알파고 이벤트나 알파폴드의 단 백질 구조 데이터베이스 무료 공개 결정은 과학적 성과인 동시에, 딥마인드의 존재 의의를 전 세계에 각인시킨 탁월한 경 영적 판단이었습니다. 기업인으로서의 허사비스는 더욱 냉혹한 현실과 마주해야했습니다.

오픈AI의 샘 올트먼 이 공격적인 마케팅과 제품 출시로 AI 시장의 판도를 흔들 때, 허사비스는 '책임감 있는 AI' 라는 원칙과 시장의 요구 사이에서 줄타기해야했습니다. 그는 AI가 가져올 파급력을 누구보다 잘 알고 있기에, 기술의 민주화와 안전 통제 사이에서 깊이 고민했습니다. 2023년 구글의 AI 조직이 통합되어 '구글 딥마인드'가 출범했을 때, 그는 사실상 구글 전체의 AI 전략을 총괄하는 자리에 올랐습니다.

이는 그가 더 이상 상아탑 속의 연구자로만 머물 수 없음을 의미했습니다. 그는 경쟁사와의 속도전에서

승리해야 했고, 주주들의 이익을 고려해야 했으며, 각국 정부의 규제 움직임에도 대응해야했습니다. 이 복합적인 역할들 사이의 긴장은 허사비스에게 끊임없는 스트레스를 안겨주었습니다. 연구자로서 그는 완벽을 추구하고 싶지만, 기업인으로서 타이밍을 놓칠 수 없습니다.

경 영자로서 그는 조직의 안정을 원하지만, 혁신가로서 그는 끊임없는 변화를 추구해야 합니다 흥미로운 점은 허사비스가 이 모순을 회피하지 않고 정면으로 돌파하 려 한다는 것입니 그는 이질적인 요소들을 결합하여 새로운 것을 만들어내는 '융합'의 달인입니다. 게임 개발자 시절 프로그래밍과 기획을 오갔던 것처럼, 그는 연구의 깊이 와 비즈니스의 넓이를 동시에 확보하려 노력합니다. 그는 동료들에게 이렇게 말하곤합니다.

"우리는 과학을 하고 있지만, 동시에 제품을 만들 고 있습니다. 이 둘은 다르지 않습니다. 최고의 과학이 결국 최고의 제품이 될 것입니다."

이는 그가 40년간 품어온 "지능을 풀어, 다른 모든 문제를 해결한다"는 미 션과 맞닿아 있습니다. 그에게 연구, 경영, 사업은 이 거대한 미션을 수행하기 위한 각각의 도구일 뿐입니다. 허사비스는 이 세 가지 역할을 능수능란하게 오가는 '하이브 리드 리더'의 전형을 보여줍니 때로는 실험실 가운을 입은 과학자로, 때로는 칠판 앞 에 선 교수로, 때로는 말끔한 정장 차림의 CEO로 변신하며 그는 AI 시대가 요구하는 새로운 리더십의 표준을 정립해 나가고 있습니다.

성취를 만드는 생활 구조 "밤의 두 번째 교대"와 집중의 루틴 런던의 밤이 깊어지고 도시의 소음이 잦아들 무렵, 데미스 허사비스의 '진짜' 하루가 시작됩니다. 그는 이를 "두 번째 교대(The Second Shift)"라고 부릅니다. 보통의 직장인들이 퇴근 후 휴식을 취하는 시간, 허사비스는 가족과 저녁 식사를 하고 아이들을 잠재운 뒤, 밤 10시 가 넘 어서야 다시 책상 앞에 앉습니다. 그리고 새벽 4시까지, 온전히 자신만의 물입을 위한 시간을 갖습니다.

이 루틴은 단순한 야근이 아닙니다. 이것은 그가 치열한 낮 시간 동안 쏟아지는 업 무와 미 팅, 의사결정의 홍수 속에서 소진되지 않고, 오히려 창의성을 충전하는 그만 의 성스러운 의식입니다. 낮 시간의 허사비스가 구글 딥마인드의 CEO로서 수천 명의 조직을 이끌고 수 많은 보고를 처리하는 '관리자'라면, 밤 시간의 허사비스는 순수한 '연구자'이자 '몽상가' 로 돌아갑니다.

그의 서재는 이 시간 동안 고요한 실험실이됩니다. 그는 이 시간에 최신 논문을 읽고, 기술 적 난제에 대한 아이디어를 구상하며, 딥마인드의 장기적인 로드맵을 점검합 니다. 방해받 지 않는 이 긴 호흡의 시간 동안 그는 낮에는 미처 들여다보지 못했던 생 각의 조각들을 연 결합니다. 때로는 체스판을 복기하듯 AI 모델의 구조를 머릿속으로 시뮬레이션하고, 때로 는 물리학이나 생물학 같은 전혀 다른 분야의 서적을 탐독하며 영감을 얻습니다.

그가 새벽 4시까지 깨어 있는 이유는 단순히 시간이 부족해서가 아닙니다. 그는 이 새벽의 고요함이 주는 특유의 명료함을 사랑합니다. 세상이 잠든 시간, 모든 소음이 사라진 그 적막 속에서 그는 생각의 가장 깊은 바닥까지 내려갑니다. 그는 인터뷰에서 "가장 중요한 아이디어들은 대부분 이 시간에 떠오른다"고 말한 바 있습니다. 알파고의 결정적인 '신의 한 수'나, 알파폴드의 구조적 난제를 해결할 실마리도 아마 이 고독한 새벽의 사투 속에서 잉태되었을 것입니다.

새벽 4시에 잠자리에 든 그는 오전 10시쯤 기상합니다. 남들보다 늦은 아침이지만, 그의 뇌는 이미 전날 밤의 치열한 사고 과정을 통해 예열된 상태입니다. 사무실에 출근한 그는 오후부터 본격적인 업무를 시작합니다.

그의 일정은 분단위로 쪼개져 있지만, 그는 항상 맑은 정신으로 회의에 참석합니다. 밤의 사색이 그에게 낮을 버틸 수 있는 지적 체력을 제공하기 때문입니다. 허사비스의 이러한 생활 패턴은 그가 가진 독특한 시간 관념을 보여줍니다. 그는 항상 양손목에 시계를 차지 않더라도(때로는 스마트워치와 아날로그 시계를 동시에 착용하기도 합니다), 시간의 흐름에 민감합니다.

그는 자신을 끊임없이 재촉하는 사람처럼 보입니다. "인생은 짧고, 풀어야 할 지능의 비밀은 너무나 거대하다"는 생각이 그를 잠들지 못하게 하는지도 모릅니다. 이 "두 번째 교대"는 또한 그가 균형을 유지하는 비결이기도 합니다. 낮 동안 기업 경영이라는 현실의 무게에 짓눌리다가도, 밤이 되면 그는 다시 과학과 우주의 신비라는 이상으로 도피할 수 있습니다.

이 순환이 그를 지치지 않게 합니다. 그에게 밤샘은 노동이 아니라, 지적 유희이자 회복의 과정입니다. 그는 이 시간을 통해 CEO라는 무거운 짐을 잠시 내려놓고, 4살 때 체스판을 처음 마주했던 그 소년의 호기심 어린 눈빛을 되찾습니다.

허사비스의 위대한 성취들은 천재적인 재능뿐만 아니라, 매일 밤 반복되는 이 고독하고도 치열한 루틴이 층층이 쌓여 만들어진 결과물인 것입니다.

호기심과 경계를 넘나드는 융합적 사고 데미스 허사비스의 뇌 구조를 이해하는 열쇠는 '경계 없음'에 있습니다. 그는 한 분야의 전문가가 되기를 거부하고, 여러 분야의 경계선 위에서 춤을 추듯 지식을 탐험하는 진정한 폴리매스(Polymath, 박식가)입니다. 그의 이력서는 마치 서로 다른 세 사람의 삶을 합쳐놓은 듯합니다.

세계 랭킹 2위의 체스 유망주, 천만 장 이상 팔린 게임을 만든 프로그래머, 그리고 기억과 상상력을 연구해 세계적인 저널에 논문을 실은 신경과학자. 이 이질적인 경험들은 딥마인드라는 용광로 안에서 하나로 녹아들었습니다. 그의 융합적 사고는 어린 시절 체스판 위에서 시작되었습니다. 체스는 그에게 논리적 사고와 직관의 결합을 가르쳤습니다.

수익기를 통해 미래를 시뮬레이션하는 능력, 상대의 의도를 파악하는 심리적 통찰, 그리고 엄청난 압박 속에서도 평정심을 유지하는 멘탈 관리 능력은 훗날 그가 기업을 경영하고 난제를 해결하는 데 핵심적인 자산이 되었습니다. 10대에 게임 개발사 불프로그(Bullfrog)에서 피터 물리뉴와 함께 '테마파크'를 만들면서, 그는 이 시뮬레이션 능력을 컴퓨터 코드로 구현하는 법을 배웠습니다. 수천 명의 관람객이 각자의 욕구에 따라 움직이는 인공지능을 설계하며, 그는 "지능이란 무엇인가?"라는 근원적인 질문을 품게 되었습니다.

이 질문은 그를 신경과학의 세계로 이끌었습니다. 케임브리지와 UCL에서 그는 인간의 뇌가 어떻게 기억을 저장하고, 그 기억을 바탕으로 어떻게 새로운 미래를 상상하는지를 탐구했습니다. 해마가 손상된 환자들이 미래를 상상하지 못한다는 그의 연구 결과는 2007년 사이언스지가 선정한 10대 과학 성과에 올랐습니다.

이 연구는 단 순히 뇌과학의 성과로 끝나 지 않았습니다. 허사비스는 이 발견을 AI 알고리즘에 그대로 적용했습니다. 딥마인드의 강화학습 AI가 과거의 데이터를 학습하여(기억), 낯선 환경에서 최적의 전략을 찾아내는(상 상) 방식은 인간 뇌의 메커니즘을 모방한 것입니다. 생물학적 뇌의 원리를 디지털 코드로 번역해내는 능력, 이것이 바로 허사비스가 가진 융합적 사고의 정수입니다.

그의 지적 호기심은 과학과 공학에만 머무르지 않습니다. 그는 역사, 철학, 예술에 도 깊은 관심을 두고 있습니다. 그는 르네상스 시대의 인본주의와 현대의 기술이 만나 야 한다고 믿습니다.

AI가 인류에게 미칠 영향을 고민할 때, 그는 기술적 안전장치뿐 만 아니라 철학적, 윤리적 질문을 던집니다. "우리가 만드는 이 지능이 인간의 가치를 어떻게 반영할 것인가?"라는 그의 물음은 인문학적 소양 없이는 나올 수 없는 것입니다. 그는 독서를 통해 시 대를 초월한 사상가들과 대화하며, 그들의 지혜를 현재의 AI 개발에 접목하려 노력합니다. 허사비스의 융합적 사고는 딥마인드의 채용과 조직 문화에도 고스란히 반영되었습니다.

딥마인드에는 컴퓨터 공학자뿐만 아니라 물리학자, 생물학자, 신경과학자, 윤리 학자, 심지어 생태학자까지 다양한 배경을 가진 인재들이 모여 있습니다. 허사비스는 이들이 서로의 전문 용어가 아닌, '문제 해결'이라는 공통의 언어로 소통하도록 장려합 니다. 그는 "서로 다 른 분야의 충돌 지점에서 혁신이 일어난다"고 굳게 믿습니다. 알 파폴드 프로젝트가 성공 할 수 있었던 것도 구조 생물학의 난제를 머신러닝 전문가들 이 전혀 새로운 시각으로 바라 보았기에 가능했습니다.

그에게 호기심은 단순한 흥미가 아니라 생존 본능과도 같습니다. 그는 세상의 모든 것이 연결되어 있다고 봅니다. 단백질의 3차원 구조와 바둑판의 포석, 그리고 우주의 물리 법칙 은 서로 다른 모습을 하고 있지만, 그 이면에는 정보를 처리하고 최적의 상 태를 찾아가는 공통된 원리가 숨어 있다는 것입니다.

허사비스는 이 보편적인 원리를 찾기 위해 끊임없 이 경계를 넘나듭니다. 그의 융합적 사고는 복잡한 세상의 문제를 하나의 카테고리로 가두 지 않고, 전방위적으로 조망하며 해법을 찾아내는 강력한 도 구입니다. 이것이 바로 그가 " 딥마인드의 미션이 허언이 아니었음을 보여줍니다. 다큐멘터리 'The Thinking Game'(2024) 트라이베카 영화제 프리미어와 허사비스 의 내면 세계 2024년 6월, 뉴욕 트라이베카 영화제의 스크린에 딥마인드의 로고가 떠 올랐습니다. 그렉 코스(Greg Kohs) 감독의 다큐멘터리 '더 씹킹 게임(The Thinking Game)'이 전 세계 최초로 공개되는 순간이었습니다.

5년이라는 긴 시간 동안 딥마인 드의 내부를 밀착 취재한 이 작 품은 단순한 기업 홍보 영상이 아니었습니다. 그것은 "지능을 풀겠다"는 불가능에 가까운 미션에 도전한 사람들의 고뇌와 환희, 그리고 그 중심에서 있는 데미스 허사비스라는 한 인간의 내면을 현미경처럼 들여다본 기록이었습니다. 영화는 화려한 성공 신화 뒤에 가려진 지난한 실패의 시간들을 가감 없이 보여줍니다. 특 히 인공지능이 과학적 발견을 주도하는 시대를 열기 위해 시작된 '알파폴드' 프 로젝트의 초 기 난관들이 비중 있게 다뤄집니다. 연구팀이 막다른 골목에 부딪혀 좌절 하고, 허사비스가 고뇌에 찬 표정으로 회의실을 서성이는 장면들은 천재적인 성취가 결코 우연이나

영감만으로 이루어지지 않았음을 웅변합니다. 스크린 속의 허사비스

는 확신에 찬 CEO의 모습과, 미지의 영역 앞에서 두려움을 느끼는 탐험가의 모습을 동시에 보여줍니다. 관객들은 이 영화를 통해 허사비스의 내면 깊은 곳에 자리 잡은 동기를 목격합니다. 그가 왜 그토록 강박적으로 지능의 본질을 파고드는지, 그에게 '이해한다'는 것이 어떤 의미인지가 드러납니다. 영화에는 그가 어린 시절 체스 영재로 주목받던 영상부터, 게임 개발자 시절, 그리고 딥마인드를 창업하고 구글에 합류하기까지의 여정이 교차 편집됩니다. 이 과정에서 허사비스는 "과학은 인류가 발명한 가장 위대한 도구"라며, AI가 이 도구를 더욱 날카롭게 다듬어 질병, 기후 위기, 에너지 문제 등 인류의 난제를 해결해 줄 것이라는 흔들리지 않는 믿음을 보여줍니다. 하지만 영화는 낙관론에만 머물지 않습니다.

허사비스가 AI의 잠재적 위험성을 경계하며, 이 강력한 힘이 통제 가능한 범위 내에 있어야 함을 역설하는 장면들도 담겨 있습니다. 트 라이베카 영화제 관객들은 기술적 성취의 경이로움과 동시에, 그것을 다 루는 인간의 책임 감에 대한 묵직한 질문을 안고 극장 문을 나섰습니다. 허사비스에게 이 영화는 자신의 업적을 과시하는 트로피가 아니라, 자신이 평생을 바쳐온 '생각 (Thinking)'이라는 행위의 아름다움과 무게를 세상과 공유하고자 하는 진심 어린 고백이었습니다.

영화 속에서 그는 말합니다. "우리가 지능을 이해하게 된다면, 그것은 인간이 자기 자신을 이해하는 가장 결정적인 순간이 될 것입니다." '알파고' 다큐멘터리 리(2017)와의 연결 '더 씽킹 게임'은 2017년 공개되어 전 세계적으로 큰 반향을 일으켰던 다큐멘터리 '알파고 (AlphaGo)'의 정신적 후속작이자, 서사의 완결판과도 같습니다. 두 작품 모두 그렉 코스 감독이 연출을 맡았으며, 딥마인드라는 하나의 주체를 다루고 있지만, 그 지향점과 분위기는 사뭇 다릅니다. '알파고'가 인간과 AI의 '대결'에 초점을 맞춘 긴장감 넘치는 스포츠 드라마였다면, '더 씽킹 게임'은 인간과 AI의 '협력'을 통해 과학의 지평을 넓혀가는 지적 탐험기입니다.

전작에서 이세돌 9단과의 승부는 승패가 명확히 갈리는 제로섬 게임의 긴박함을 주었습니다. AI는 인간을 넘어서는 두려운 존재이자, 극복해야 할 대상으로 그려지기 도했습니다. 반면, 이번 작품에서 AI는 인간 과학자들의 한계를 보완하고, 50년 동안 풀리지 않았던 단백질 구조 예측이라는 난제를 해결하는 든든한 파트너로 등장합니다.

두 다큐멘터리를 연결하는 고리는 바로 허사비스의 비전입니다. '알파고'의 마지막 장면에서 그는 승리에 도취하기보다 "이제 이 기술을 과학적 문제 해결에 써야 한다"고 말했습니다. '더 씽킹 게임'은 그 선언이 어떻게 현실이되었는지를 보여주는 증명서입니다. 알파고 가 바둑이라는 닫힌 세계(Closed System)에서의 승리였다면, 알파폴드는 생물학이라는 무한하고 복잡한 열린 세계(Real World)에서의 승리입니다.

영화적 기법 면에서도 두 작품은 연결됩니다. 감독은 전작과 마찬가지로 복잡한 AI 기술을 대중이 이해하기 쉬운 시각적 언어로 풀어냈습니다. 그러나 '알파고'가 바둑판 이라는 2 차원 평면 위에서 펼쳐지는 수싸움을 보여주었다면, '더 씽킹 게임'은 3차원 공간에서 춤추 듯 접히고 꼬이는 단백질의 역동적인 구조를 시각화하여 생명의 신비를 예술적으로 표현했습니다.

또한, 허사비스라는 인물의 변화도 흥미로운 관전 포인트입니다. 2016년의 허사비스가 자신의 기술을 세상에 증명해 보이려고 하는 패기 넘치는 도전자였다면, 2024 년의 허사비스 는 노벨상 수상자이자 기사 작위를 받은 거장으로서, 기술의 사회적 책임과 인류의 미래를 고민하는 성숙한 사상가의 면모를 보여줍니다. '알파고'가 "AI가 인간을 이길 수 있는가?" 를 물었다면, '더 씽킹

게임'은 "AI와 인간이 함께 어디까지 갈 수 있는가?"를 묻습니다.

이 두 편의 다큐멘터리는 허사비스와 딥마인드가 걸어온 10년의 궤적을 기록한 역사적 사료이자, 인공지능 시대의 도래를 알리는 거대한 서사 시의 1부와 2부로 남을 것입니다.

데미스 허사비스 초상

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

데미스 허사비스는 노벨 화학상 강연단에 올라 한 장의 슬라이드를 띄웠습니다. 제목은 "생각에 대해 생각하기(Thinking about Thinking)"였습니다. 네 살 때 체스 말을 처음 쥔 소년이 쉰 번째 생일을 앞두고 도착한 자리였습니다.

아 버지와 삼촌의 체스 판 앞에서 시작된 질문, 리히텐슈타인의 10시간 대국 뒤에 떠올린 의문, 케임브리지와 UCL의 실험실에서 다듬어진 가설이 반세기 만에 노벨상이라는 이정표와 만났습니다. 그러나 허사비스가 그날 청중에게 전하려 한 것은 과거의 업적이 아니었습니다. 그가 보여주고 싶었던 것은 이제 막 열리기 시작한 문이었습니다.

지능이란 무엇인가. 이 질문은 인류 문명의 가장 오래된 수수께끼 가운데 하나입니다. 고대 그리스의 철학자들은 로고스(Logos)라 불렀고, 동양의 사상가들은 지혜(智慧)라 표현했으며, 근대 과학자들은 인지(Cognition)라는 용어로 포착하려 했습니다. 그러나 지능의 정체를 명확히 규정한 사람은 아직 없습니다.

허사비스는 이 질문에 독특한 방식으로 접근했습니다. 지능을 정의하는 대신, 지능이 하는 일을 재현해보자는 것이었습니다. 체스에서 바둑으로, 바둑에서 단백질로, 단백질에서 날씨와 수학과 신약 설계로. 각 도메인에서 AI가 성공할 때마다 지능의 한 조각이 드러났습니다.

패턴을 인식하는 능력, 보지 못한 상황을 상상하는 능력, 장기적 결과를 계획하는 능력, 실수로부터 스스로 교정하는 능력. 허사비스의 경력 전체는 이 조각들을 맞추는 퍼즐이었습니다. 그렇다면 지능의 본질을 이해한 인류는 어떤 미래를 맞이하게 될까요. 허사비스가 그리는 그림은 구체적입니다. 의학이 달라집니다. 알파폴드가 2억 개 이상의 단백질 구조를 예측한 것은 시작에 불과했습니다.

아이소모픽 랩스(Isomorphic Labs)는 AI로 설계한 항암 약물을 2026년 내 임상시험에 진입시키겠다고 발표했습니다. 전통적 신약 개발이 평균 10년의 시간과 90퍼센트의 실패율을 감수해야했다면, AI 기반 시뮬레이션은 이 과정을 천 배 이상 효율적으로 만들 수 있다고 허사비스는 주장합니다. 에너지가 달라집니다. 딥마인드 팀은 이미 핵융합 반응로인 토카막(Tokamak)의

플라즈마 제어에 AI를 적용하여 실질적 성과를 거두었습니다. 태양 에너지, 배터리 기술, 핵융합이 AI의 가속으로 돌파구를 찾는다면, 에너지는 인류 역사상 처음으로 희소 자원이 아닌 풍부한 자원이 될 수 있습니다. 과학적 발견의 속도 자체가 달라집니다. AI 코사이언티스트(AI Co-Scientist)라 불리는 다중 에이전트 시스템은 가설을 세우고, 실험을 설계하고, 데이터를 분석하는 과정을 자동화하기 시작했습니다.

2026년 영국에 세계 최초의 자동화 연구 실험실이 문을 열 예정입니다. 이 모든 변화의 핵심에 놓인 전제가 있습니다. 지능을 이해한다는 것은 곧 문제 해결의 보편적 열쇠를 손에 쥔다는 뜻입니다.

허사비스가 2010년 딥마인드를 창립할 때 내건 미션, "지능을 풀면 나머지는 해결된다(Solve intelligence, and then use that to solve everything else)"는 처음에는 과대망상처럼 들렸습니다. 15년이 지난 지금, 그 문장은 과대망상이 아니라 작업 계획서에 가까워지고 있습니다. 물론 회의적 시선은 여전히 있습니다.

현재의 대규모 언어 모델은 계획, 기억, 물리적 이해에서 심각한 한계를 보입니다. 허사비스 자신이 이를 "들쭉날쭉한 지능(Jagged Intelligence)"이라 부릅니다. 어떤 영역에서는 인간을 압도하면서도 다른 영역에서는 초등학생보다 못한 실수를 저지릅니다. AGI까지는 트랜스포머 수준의 돌파구가 한두 개 더 필요하다는 것이 허사비스의 솔직한 평가입니다. 그러나 방향은 정해져 있습니다.

지능의 본질에 한 걸음씩 다가갈수록, 인류가 풀 수 있는 문제의 범위는 기하급수적으로 넓어집니다. 허사비스가 노벨 강연의 마지막 슬라이드에서 보여준 것은 알파폴드의 구조도가 아니었습니다. 우주의 사진이었습니다. 그는 말했습니다.

우리는 아직 우주가 왜 존재하는지, 의식이 무엇인지, 시간의 본질이 무엇인지 모릅니다. 그러나 이제 그 질문들에 답할 도구를 만들기 시작했습니다. 지능의 본질을 이해한 인류가 맞이할 미래는, 답을 가진 미래가 아니라 질문할 수 있는 능력이 비약적으로 확장된 미래입니다. 그것이 허사비스가 40년간 체스판과 실험실과 서버실을 오가며 추구해온 것의 정체입니다.

"우리는 새로운 발견의 황금기(Golden Era)에 진입하고 있다" 2026년 2월, 허사비스는 포춘(Fortune) 매거진과의 인터뷰에서 이렇게 말했습니다. "10년에서 15년 후, 우리는 새로운 발견의 황금기에 진입하게 될 것입니다. 일종의 새로운 르네상스입니다."

이 발언은 순간적 낙관이 아니었습니다. 그가 수십 년간 쌓아온 연구와 경험, 그리고 2024년 노벨상 이후 급격히 넓어진 정책적 영향력이 응축된 선언이었습니다.

르네상스라는 비유는 허사비스가 의도적으로 선택한 것입니다. 14세기 유럽의 르네상스는 인쇄술이라는 기술 혁신과 고전 지식의 재발견이 결합되어 폭발한 문화적 전환이었습니다. 허사비스는 AI가 바로 그 역할을 할 것이라고 봅니다.

인쇄술이 지식의 복제 비용을 제로에 가깝게 낮췄다면, AI는 지식의 생산 비용을 제로에 가깝게 낮추고 있습니다. 한 명의 과학자가 평생 연구해야 할 단백질 구조를 알파폴드는 몇 초 만에 예측합니다. 한 팀이 수개월간 분석해야 할 기상 데이터를 웨더넥스트(WeatherNext)는 몇 분 만에 처리합니다.

국제수학올림피아드 수준의 문제를 알파프루프(AlphaProof)가 풀어냅니다. 이것은 도구의 개선이 아닙니다. 발견의 속도 자체가 바뀌는 것입니다. 허사비스가 구체적으로 지목하는 영역들이 있습니다. 의학에서는 개인 맞춤형 치료가 현실이 됩니다.

환자의 유전체 정보와 단백질 구조 데이터를 AI가 통합 분석하여, 그 사람에게 게만 효과적인 약물을 설계하는 시대입니다. 아이소모픽 랩스는 이미 17개의 신약 개발 프로그램을 동시에 진행 중이며, 항암 약물의 전임상 시험이 진행되고 있습니다. 허사비스의 표현을 빌리면, "의학은 오늘날과 전혀 다른 모습이 될 것"입니다.

에너지 분야에서는 핵융합의 실용화가 한 걸음 더 가까워집니다. 태양광과 배터리 기술의 효율은 AI 최적화를 통해 비약적으로 높아지고 있습니다. 허사비스는 에너지가 사실상 무한해지는 시점을 "모든 것이 바뀌는 변곡점"이라고 부릅니다. 에너지가 풍부해지면 식량 생산, 담수화, 제조업의 비용 구조가 근본적으로 달라집니다. 물리적 재화의 가격이 극적으로 낮아지는 세상, 허사비스가 "근본적 풍요(Radical Abundance)"라고 명명한 상태가 이론적으로 가능해집니다. 우주 탐사에서도 AI의 역할은 확대됩니다. 허사비스는 포춘 인터뷰에서 "모든 것

(everything else)에는 별들(stars)도 포함된다"는 취지의 발언을했습니다.

AI가 우주 환경의 시뮬레이션, 우주선 경로 최적화, 외계 행성 대기 분석 등에 활용된다면, 인류의 활동 범위는 지구 밖으로 확장될 수 있습니다. 그러나 허사비스는 이 황금기가 저절로 오지 않는다는 점을 분명히합니다. 황금기에 도달하기까지 10년에서 15년의 격변기(Shakeout Period)가 불가피하다고 그는 경고합니다.

이 기간 동안 산업 구조가 재편되고, 일자리의 형태가 바뀌며, 사회 제도가 새로운 현실에 적응해야 합니다. 허사비스 자신이 이끄는 구글 딥마인드도 이 격

변에서 자유롭지 않습니다. 오픈AI의 챗GPT가 촉발한 2023년의 충격은 구글 내부에서 "코드 레드(Code Red)"를 발령하게 했고, 딥마인드와 구글 브레인의 합병이라는 대수술로 이어졌습니다. "우리가 스스로를 파괴하지 않으면, 누군가 다른 사람이 할 것입니다"라고 허사비스는 말했습니다. 이것은 기업 전략의 언어이면서 동시에 시대 인식의 고백이기도 합니다.

흥미로운 것은 허사비스가 이 격변기를 두려움이 아닌 흥분의 대상으로 바라본다는 점입니다. 그는 구글 딥마인드를 "핵발전소와 같다"고 비유했습니다. 이 엔진이 검색, 유튜브, 크롬, 제미나이 앱이라는 거대한 생태계에 연결되어 있으며, 이 결합이 만들어내는 에너지가 황금기를 여는 연료가 된다는 것입니다. 2025년 말까지 알파벳의 주가가 약 65퍼센트 상승한 것은 시장이 이 비전에 반응한 결과였습니다.

허사비스의 황금기 예측은 과학자의 예측이면서 동시에 기업인의 약속이고, 몽상가의 꿈이기도 합니다. 그가 "발견의 황금기"라는 표현을 반복하는 이유는 그것이 AI 산업의 마케팅 언어가 아니라, 그의 삶 전체를 관통해온 신념이기 때문입니다. 네 살에 체스판 앞에 앉은 소년은 지능이라는 퍼즐에 매료되었고, 십 살에 노벨상을 받은 과학자는 그 퍼즐의 완성이 인류 문명의 새로운 장을 열 것이라 확신합니다.

역사는 그의 확신이 맞았는지 아닌지를 판단할 것입니다. 그러나 적어도 한 가지는 분명합니다. 허사비스만큼 오랫동안, 이 만큼 구체적으로, 이 비전을 위해 모든 것을 걸어온 사람은 거의 없습니다.

데미스 허사비스가 꿈꾸는 세상: AI와 인간이 공진화하는 시대 허사비스의 비전에는 한 가지 독특한 특징이 있습니다. 그는 AI가 인간을 대체하는 미래를 꿈꾸지 않습니다. AI와 인간이 함께 진화하는 미래를 꿈꿉니다. 이 차이는 미묘해 보이지만 근본적입니다.

실리콘밸리의 많은 기술 리더들은 AI의 미래를 경주의 비유로 설명합니다. 인간과 기계가 같은 트랙 위에서 달리고 있으며, 언젠가 기계가 인간을 추월하는 순간이 온다는 것입니다. 이 서사에서 인간은 뒤처지는 존재입니다. 허사비스의 서사는 다릅니다. 그는 AI를 망원경에 비유하곤 합니다.

갈릴레오가 망원경을 통해 목성의 위성을 발견했을 때, 망원경이 갈릴레오를 대체한 것이 아니었습니다. 인간의 눈이 볼 수 없는 것을 보게 해주었을 뿐입니다. AI도 마찬가지입니다. 인간의 지능이 닿지 않는 영역, 2억

개의 단백질 구조나 우주의 원자 수보다 많은 바둑의 경우의 수 같은 영역에서 인간의 이해를 확장해주는 도구입니다. 이 관점은 허사비스의 신경과학 배경에서 비롯됩니다. UCL에서 박사 학위를 받으며 그가 연구한 주제는 해마(Hippocampus)였습니다.

해마는 기억과 상상력의 교차 점에 있는 뇌 영 역입니다. 허사비스는 기억을 떠올리는 신경 메커니즘과 미래를 상상 하는 신경 메커니즘 이 동일하다는 것을 밝혀냈습니다. 이 발견은 단순한 학습 성과가 아니었습니다.

지능의 핵심이 정보 저장기 아니라 정보의 재조합, 즉 경험을 바탕으로 한 번도 겪지 않은 상황을 시뮬레이션하는 능력에 있다는 통찰이었습니다. 딥마인드 의 AI 아키텍처에는 이 통찰이 깊이 새겨져 있습니다. 알파고 가 37수를 둔 것은 데이터베이스에서 답을 찾은 것이 아니라, 경험으로부터 새로운 가능성을 상상해낸 것이었습니다.

허사비스가 꿈꾸는 공진화(Co-evolution)는 이런 모습입니다. AI가 과학자에게 새로운 가설을 제안합니다. 과학자는 그 가설의 의미를 해석하고, 윤리적 판단을 내리고, 연구의 방향을 결정합니다.

AI는 다시 그 결정을 바탕으로 학습하고 더 정밀한 제안을 내놓습니다. 이 순환이 반복될 때, AI도 발전하고 인간의 이해도 깊어집니다. 알 파폴드가 좋은 예입니다. AI 가 단백질 구조를 예측하자 190개국 200만 명 이상의 연구자들이 그 데이터를 활용하여 말 라리아 백신, 플라스틱 분해 효소, 항생제 내성 연구 등 각자의 분야에서 새로운 발견을 이루어냈습니다.

AI가 혼자 이 모든 것을 한 것이 아닙니다. AI가 제공한 지도를 인간 연구자들이 읽고, 해석하고, 자신의 질문에 적용한 것입니다. 프로젝트 아스트라(Project Astra)는 이 공진화 비전의 소비자 버전입니다. 허사비스가 구상 하는 범용 AI 어시스턴트는 사용자의 맥락을 이해하고, 기억하며, 다양한 기기를 넘나들며 일상을 돕습니다.

스마트폰에서 시작한 대화가 스마트 안경으로 이어지고, 브라우저 기반 에이전트인 프로젝트 매리너(Project Mariner)가 복잡한 온라인 작업을 대신합니다. 허사비스는 이 어시스턴트가 "인간의 능력을 증강하는 도구"여야 한다고 강조합니다. 인간을 수 동적 소비자로 만드는 것이 아니라, 더 많은 것을 할 수 있게 해주는 것이 목표입니다.

이 비전에는 위험도 내재되어 있습니다. 공진화가 실패하면 공의존(Codependence)이 됩니다. 인간이 AI에 지나치게 의존하여 스스로 생각하는 능력을 잃는

시나리오입니다. 허사비스는 이 위험을 인지하고 있습니다. 그가 AI 안전성(Safety) 과 정렬(Alignment) 연구에 막대한 자원을 투자하는 이유 가운데 하나입니다.

프론 티어 안전 프레임워크(Frontier Safety Framework)는 AI 시스템이 인간의 가치와 의 도에서 벗어나지 않도록 하는 기술적 장치입니다. 허사비스가 2023년 AI 멸종 위험 성명서에 서명한 것, IPCC와 같은 국제 AI 거버넌스 기구의 필요성을 역설하는 것, "기만적 정렬(Deceptive Alignment)"의 위험성을 공개적으로 경고하는 것은 모두 같은 맥락에 있습니다. 공진화가 가능하려면, AI가 인간의 통제 범위 안에 있어야 합니다. 허사비스는 인간의 무한한 적응력을 믿습니다.

CBS와의 인터뷰에서 그는 이렇게 말했습니다. "우리가 스마트폰과 디지털 기기에 얼마나 자연스럽게 적응했는지 보세요. AI도 그런 변화 가운데 하나가 될 것입니다." 이 말은 낙관적이지만 무책임한 낙 관은 아닙니다. 스마트폰이 인류의 소통 방식, 정보 소비 방식, 사회적 관계를 근본적으로 바꾼 것처럼, AI도 인류의 사고 방식, 일하는 방식, 존재의 의미를 근본적으로 바꿀 것이라는 인식이 그 안에 있습니다.

다만 허사비스는 그 변화가 파멸이 아니라 진 화가 되기를 원합니다. AI와 인간 이 서로를 밀어내는 것이 아니라, 서로의 한계를 보완하며 함께 성장하는 시대. 그것이 데 미스 허사비스가 네 살에 체스판 앞에 앉은 이 래 40년간 추구해온 꿈의 궁극적 모습입니다. 후대를 위한 제언: "끊임없이 배우는 법을 배워라(Learning to Learn)" 2025년 9월 12일, 아테네. 아크로폴리스 아래 자리한 헤로데스 아티쿠스 극장(Odeon of Herodes Atticus)에서 허사비스는 청중 앞에 섰습니다.

로마 시대에 지어진 이 반원형 극장 은 2천 년 전에도 사람들이 모여 지식과 예술을 나누던 곳이었습니다. 그 고대의 무대 위 에 서 노벨상 수상자는 미래 세대를 향해 한 가지 당부를 남겼습니다. "배우는 법을 배 우세요 (Learning how to learn). 그것이 다음 세대에게 필요한 가장 중요한 능력입니다."

이 발언의 맥락을 이해하려면 허사비스 자신의 이력을 다시 들여다보아야합니다. 그는 한 분야의 전문가가 아니었습니다. 체스 신동이었고, 게임 개발자였으며, 신경과 학자였고, AI 연구자이자 기업인입니다.

이 경력의 매 단계에서 그는 새로운 분야를 처음부터 배워야했습니다. 17세에 게임 회사 불프로그(Bullfrog Productions)에서 일하며 프로그래밍과 게임 디자인을 독학했습니다. 엘릭서 스튜디오를 창업하고 운영 하며 경영을 배웠습니다. 29세 에 UCL 박사 과정에 들어가 신경과학을 배웠습니다.

그리고 34세에 딥마인드를 창립하며 학문과 사업과 기술의 교차점에서 새로운 종류 의 조직을 만드는 법을 배웠습니다. 허사비스의 이 궤적 자체가 "배우는 법을 배운다"는 것의 실제 사례입니다. 그가 아 테네에서 강조한 메타 스킬(Meta-Skill)이란, 특정 지식이나 기술이 아니라 새로운 지식과 기술을 빠르게 습득하는 방법론을 의미합니다.

자신의 학습 스타일을 이해하 는 것, 서로 다른 분 야의 아이디어를 연결하는 것, 평생에 걸쳐 지속적으로 적응하는 것. 허사비스는 이런 능 력들이 수학, 과학, 인문학 같은 전통적 교과목과 함께 교육의 핵심이 되어야 한다고 역설했습니다. 이 제언의 긴급성은 AI의 발전 속도에서 옵니다. 허사비스는 아테네 강연에서 이렇 게 말했습니다. "평상시에도 10년 후를 예측하기는 어렵습니다.

AI가 매주(week by week) 변하고 있는 오늘날에는 더욱 어렵습니다. 확실하게 말할 수 있는 것은 거대한 변화가 오고 있다 는 것뿐입니다." 기술 지식의 반감기가 급격히 줄어들고 있습니다.

대학에서 배운 프로그 래밍 언어가 졸업할 때쯤 구식이 될 수 있습니다. 특정 직업을 위해 훈련받은 기술이 5년 후에는 AI에 의해 자동화될 수 있습니다. 이런 환경에서 살 아남는 방법은 하나의 기술에 의존하는 것이 아니라, 어떤 기술이든 필요할 때 배울 수 있는 근본적 능력을 갖추는 것입 니다. 허사비스의 제언에는 그의 체스 경험이 투영되어 있습니다.

체스에서 최고 수준에 도달하 려면 개별 수(手)를 암기하는 것으로는 부족합니다. 패턴을 인식하고, 상대의 전략을 읽고, 여러 수 앞을 내다보며, 형세가 불리할 때 유연하게 계획을 수정하는 메 타 능력이 필요합 니다. 허사비스가 12세에 리히텐슈타인 대회에서 10시간짜리 대국을 치른 뒤 "이 두뇌들을 암 치료나 기후 문제에 쓸 수 있다면?"이라고 물었을 때, 그 가 발견한 것은 체스라는 게 임의 한계가 아니라 배움이라는 행위의 전이 가능성 (Transferability)이었습니다. 한 분야에서 기른 사고력은 다른 분야로 옮겨갈 수 있다 는 깨달음. 그것이 체스 신동을 AI 선구자로 만든 원동력이었습니다.

그리스 총리 키리아코스 미초타키스(Kyriakos Mitsotakis)는 같은 자리에서 다른 각도의 경고를 덧붙였습니다. AI의 혜택이 소수의 거대 기업에 집중되고 대다수의 사람들이 그 혜택을 체감하지 못한다면, 심각한 사회적 불안이 뒤따를 것이라는 경고였습니다. 허사비스의 "배우는 법을 배우라"는 제안은 이 맥락에서 더 큰 의미를 갖습니다.

AI 시대에 도태되지 않는 개인의 생존 전략인 동시에, 사회 전체가 AI의 혜택을 고르게 누리기 위한 교육적 인프라의 기초이기 때문입니다. 허사비스는 AI가 교육 자체를 혁신할 수 있다고 봅니다. 구글 딥마인드는 2025년 영국 정부와의 포괄적 연구 협력을 통해 잉글랜드 국가 교육과정에 맞춤형 AI 교사 지원 시스템을 개발하기로 합의했습니다.

AI가 학생 개인의 학습 속도, 강점, 약점을 파악하고 맞춤형 교육을 제공한다면, "배우는 법을 배우는" 과정 자체가 AI에 의해 가속될 수 있습니다. 이것은 역설처럼 보이지만 허사비스에게는 자연스러운 순환입니다. AI가 인간의 학습 능력을 높이고, 학습 능력이 높아진 인간이 더 나은 AI를 만드는 것. 공진화의 교육적 버전입니다. 아테네 강연의 마지막 부분에서 허사비스는 이렇게 말했습니다.

"확실한 것은, 여러분이 평생에 걸쳐 끊임없이 배우야 할 것이라는 점입니다." 이 문장은 겸손합니다. 노벨상을 받은 과학자가, 세계에서 가장 강력한 AI를 이끄는 CEO가, 미래를 정확히 예측할 수 없다고 고백하고 있습니다.

그가 줄 수 있는 것은 답이 아니라 방법론입니다. 답은 변하지만, 답을 찾는 법을 아는 사람은 어떤 변화에도 대응할 수 있습니다. 그것이 네 살 때 체스를 배우고, 여덟 살 때 컴퓨터를 사고, 열일곱 살 때 게임 회사에 들어 가고, 스물아홉 살 때 박사 과정에 입학한 한 사람의 인생이 후대에게 전하는 메시지입니다.

지능의 시대에 필요한 것은 '더 똑똑한 모델'인가, '더 성숙한 인간'인가 이 전기는 데미스 허사비스라는 한 인간의 이야기로 시작했습니다. 런던 노스쪽의 넉넉하지 않은 다문화 가정에서 자란 소년. 그리스계 키프로스인 아버지 코스타스와 중국계 싱가포르인 어머니 앤젤라 사이에서 태어난 장남. 체스판 위에서 사고의 본질에 매료되고, ZX 스펙트럼 48K 컴퓨터 앞에서 프로그래밍을 독학하며, 학교 성적표보다 우주의 근본 법칙에 더 마음이 끌렸던 아이. 이 전기는 그 아이가 40년에 걸쳐 자신의 질문을 끝까지 밀고 나간 이야기를 따라왔습니다. 그리고 이제 이 이야기의 끝에서, 허사비스가 만든 것보다 허사비스가 남긴 질문이 더 중요해지는 지점에 도달했습니다.

2025년 타임(TIME) 매거진은 허사비스를 "올해의 인물"로 공동 선정하며 "AI의 설계자들 (Architects of AI)"이라는 표지를 내걸었습니다. 같은 해, 영국 왕실은 그에게 기사 작위 (Knight Bachelor)를 수여했습니다. 래스커 의학연구상, 캐나다 게어드너 국제상, 브레이크 스루 상이 잇따랐습니다. 옥스퍼드, 케임브리지, UCL, 임페리얼 칼

리지에서 명예박사학위가 주어졌습니다. 왕립학회(FRS)와 왕립공학아카데미 (FREng)의 펠로우가 되었습니다. 세 상은 허사비스를 우리 시대의 가장 영향력 있는 과학자 가운데 한 명으로 인정했습니다.

그러나 허사비스 자신은 여전히 불안합니다. 그 불안은 실패에 대한 두려움이 아니라 성공에 대한 두려움입니다. AGI가 정말로 5년에서 10년 이내에 도래한다면, 인류는 그것을 감당할

준비가 되어 있는가. 허사비스는 이 질문을 회피하지 않습니다.

그가 AI 멸종 위험성 명서에 서명한 것, "사회가 AGI에 준비되어 있지 않다"고 공개적으로 발언한 것, CERN 모델을 참고한 국제 AI 거버넌스 기구의 설립을 제안한 것은 모두 같은 불안에서 나온 행동입니다. 이 불안의 핵심에 모순이 있습니다. 허사비스는 기술 낙관주의자이면서 동시에 안 전주의자입니다.

AI가 인류 역사상 가장 유익한 기술이 될 것이라 믿으면서, 잘못 다루면 인류의 존재 자체를 위협할 수 있다고 경고합니다. 이 두 가지 신념은 모순처럼 보이지만, 허사비스에게는 한 동전의 양면입니다. 핵에너지가 도시를 밝힐 수도 있고 도시를 파괴할 수도 있듯이, AI의 힘이 클수록 그것을 다루는 인간의 성숙함이 결정적 변수가 됩니다.

"더 똑똑한 모델"을 만드는 것은 허사비스와 구글 딥마인드의 2,000명 연구진이 매일 하는 일입니다. 제미니(Gemini) 시리즈는 1.0에서 2.5 프로를 거쳐 제미니 3에 이르렀고, 6억 5천만 명의 월간 활성 사용자를 확보했습니다. 모델의 추론 능력, 멀티모달 이해력, 에이전트 기능은 해마다 도약하고 있습니다.

허사비스는 이 발전이 "분수령(Watershed Moment)을 넘었다"고 평가합니다. AI 모델이 이제 고수준의 연구 보조로 쓸 수 있을 만큼 성숙해졌다는 뜻입니다. 그러나 "더 성숙한 인간"을 만드는 것은 다른 차원의 문제입니다.

이것은 코드로 작성할 수 없고, 논문으로 발표할 수 없으며, 서버를 늘린다고 해결되지 않습니다. 성숙한 인간이란 무엇입니까. AI가 제안하는 답을 맹신하지 않고 자신의 판단력을 유지하는 사람입니다. AI가 가져오는 풍요 속에서 삶의 의미를 스스로 찾을 수 있는 사람입니다. AI가 대체할 수 없는 영역, 공감과 윤리와 창조성과 책임감의 영역에서 자신의 역할을 지키는 사람입니다.

허사비스가 포춘 인터뷰에서 말한 "근본적 풍요"의 세상을 상상해봅시다. 에너지가 사실 상 무한하고, 물리적 재화의 생산 비용이 극적으로 낮아지고, 질병의 대부분이 치료 가능해진 세상. 자원의 제약이 사라진 세상에서 인간은

무엇을 할 것입니까. 허사비스 자신이 이 질문을 던진 적이 있습니다. "복권에 당첨되면 직장에 돌아갈 사람이 몇 퍼센트일까요?" 이 질문 안에는 산업혁명보다 10배 더 크고 10배 더 빠른 변화 앞에서, 경제적 분배를 넘어 존재론적 차원의 과제, 인간의 목적과 의미의 문제가 담겨 있습니다.

이 전기를 쓰는 동안, 저는 한 가지 장면으로 자주 돌아갔습니다. 1988년, 리히텐슈타인. 열두 살의 데미스 허사비스가 10시간에 걸친 체스 대국을 마치고 대국장 밖으로 나옵니다. 유럽 각국에서 온 청소년 체스 선수들이 온종일 머리를 짜내고 있었습니다. 그 순간 소년의 머릿속에 질문이 떠올랐습니다.

"이 두뇌들을 암 치료나 기후 문제에 쓸 수 있다면?" 이 질문은 체스판 너머를 본 소년의 상상력에서 나왔습니다. 체스는 아름답지만 체스의 지능은 체스판 안에 갇혀 있었습니다.

소년은 지능이 같지 않는 세상을 원했습니다. 데미스 허사비스의 삶은 아직 진행 중입니다. 허사비스는 2024년 노벨상 수상 직후 이렇게 말했습니다.

"AI는 인류가 만든 가장 유익한 기술이 될 것입니다. 그러나 그것은 우리가 올바른 방식으로 만들고, 올바른 방식으로 사용할 때에만 그렇습니다." "~할 때에만(only if)."

이 조건문에 이 책의 무 게가 실려 있습니다. 허사비스 자신이 AI를 "들쭉날쭉한 지능(Jagged Intelligence)"이라 부릅니다. 어떤 영역에서는 인간을 압도하면서 다른 영역에서는 초등학생보다 못한 실수를 저지르는 기술. 이 들쭉날쭉함은 기술의 불완전함이며, 그 기술을 다루는 인간 문명의 불완전 함이기도 합니다. 핵에너지가 도시를 밝힐 수도 있고 파괴할 수도 있듯이, AI의 힘이 클수록 그것을 다루는 인간의 성숙함이 결정적 변수가 됩니다.

1988년, 리히텐슈타인. 열두 살 소년이 10시간짜리 체스 대국을 마치고 대국장 밖으로 나왔습니다. "이 두뇌들을 암 치료나 기후 문제에 쓸 수 있다면?" 36년 후, 그 소 년은 자신의 꿈을 상당 부분 실현했습니다.

알파폴드는 체스판 안에 갇히지 않은 지능 의 증거입니다. 그러나 새로운 질문도 따라왔습니다. 지능이 갇히지 않는 세상에서, 인간은 갇히지 않을 수 있는가. AI가 모든 답을 제공하는 세상에서, 인간은 여전히 질 문할 줄 아는 존재로 남아 있을 수 있는가. 체스판 앞의 네 살 소년에게서 시작된 질문은 끝나지 않았습니다.

부록 부록 1. 데미스 허사비스 연보 1976년: 7월 27일 영국 런던 출생. 그리스계 키프로스인 아버지와 중국계 싱가포르인 어머니 사이의 장남. 1980년: 4세에 체스 입문. 2주 만에 아버지를 이기며 천재성 발현. 1989년: 13세에 체스 마스터 등급(Elo 2300) 달성. 세계 주니어 랭킹 2위 기록. 1991년: 15세에 고등학교 졸업(A-level). 컴퓨터 게임사 불프로그(Bullfrog) 입사. 1994년: 밀리언셀러 게임 '테마파크(Theme Park)' 출시. 리드 프로그래머로 참여. 1994년: 케임브리지 대학교 퀸스 칼리지 입학. 컴퓨터 과학 전공. 1997년: 케임브리지 대학교 학사 학위 취득(최우수 성적). 1998년: 게임 스튜디오 엘릭서 스튜디오(Elixir Studios) 설립. '리퍼블릭' 등 개발. 2005년: 유니버시티 칼리지 런던(UCL) 박사 과정 진학. 인지신경과학 전공. 2007년: '상상력과 기억의 뇌 구조적 연결'에 관한 논문 발표. 사이언스지 선정 10대 성과. 2009년: UCL 인지신경과학 박사 학위 취득. 2010년: 세인 리그, 무스타파 슬레이만과 함께 딥마인드(DeepMind) 공동 창업. 2014년: 구글(Google)이 딥마인드를 약 4억 파운드에 인수. 2016년: 바둑 AI '알파고(AlphaGo)' 개발. 서울에서 이세돌 9단에게 승리(4-1). 2018년: 단백질 구조 예측 AI '알파폴드(AlphaFold)'를 CASP13 대회에서 처음 공개. 2020년: '알파폴드 2' 발표. 50년 된 생물학 난제 해결 선언. 2021년: 신약 개발 전문 자매 회사 아이소모픽 랩스(Isomorphic Labs) 설립 및 CEO 겸임. 2023년: 구글 브레인과 딥마인드가 통합된 '구글 딥마인드'의 CEO로 취임. 2024년: 영국 왕실로부터 인공지능 분야 공로로 기사 작위(Sir) 수여. 2024년: 노벨 화학상 수상 (단백질 구조 예측 기여 공로).

부록 2. 주요 논문 및 프로젝트 읽기 가이드 딥 Q-네트워크 (DQN, 2015): 인간의 개입 없이 AI가 스스로 아타리 게임을 학습 하여 정복하는 과정을 다룬 논문입니다. 강화학습과 딥러닝이 결합된 현대 AI의 시발 점입니다. 알파고 시리즈 (2016~2017): 'Nature'지에 게재된 논문들로, 인간의 기보를 학습 한 버전(Lee)부터 스스로 대국하며 학습한 버전(Zero)까지의 진화를 설명합니다.

알파폴드 2 (2021): 단백질의 아미노산 서열만으로 3차원 입체 구조를 예측하는 혁 신적인 아키텍처를 소개합니다. 생물학 연구의 패러다임을 바꾼 논문입니다. 제미니 (Gemini, 2023~2024): 텍스트, 이미지, 영상 등을 동시에 이해하는 멀티 모달(Multimodal) 모델 프로젝트입니다.

기존 AI가 텍스트만 읽을 수 있는 사람이었 다면, 제미나이는 글도 읽고 사진도 보고 영상도 이해하는 사람에 가깝습니다. 일상적 인 범용 비서 AI로 가는 딥마인드의 최신 행보입니다. 알파프루프 & 알파지오메트리 (2024): 수학적 추론과 기하학 문제 풀이에서 인간 전문가 수준에 도달한 AI 프로젝트로, AI의 논리적 사고 능력을 증명했습니다.

부록 3. 인물 관계도 셰인 레그 (Shane Legg): 딥마인드 공동 창립자이자 수석 과학자. 'AGI'라는 용어를 대중화했으며 지능의 정의에 대한 철학적 토대를 제공했습니다. 무스타파 솔레이만 (Mustafa Suleyman): 딥마인드 공동 창립자. AI의 윤리적 측면 과 사회적 영향을 강조했으며 현재는 마이크로소프트 AI CEO로 활동 중입니다. 데이비드 실버 (David Silver): 알파고 프로젝트의 리더. 허사비스의 케임브리지 동 기로, 강화학습 분야의 세계 최고의 권위자입니다.

존 점퍼 (John Jumper): 알파폴드 프로젝트의 핵심 설계자. 허사비스와 함께 2024 년 노벨 화학상을 공동 수상한 천재 연구원입니다.

순다르 피차이 (Sundar Pichai): 구글과 알파벳의 CEO. 허사비스가 구글 내에서 자율성을 유지하며 연구에 집중할 수 있도록 지원하는 든든한 조력자입니다. 부록 4. 허사비스 추천 도서 목록 실재의 실체 (The Fabric of Reality) - 데이비드 도이치: 물리학, 양자역학, 지식의 성장을 다룬 책으로 허사비스의 세계관에 가장 큰 영향을 준 책입니다. 순열 도시 (Permutation City) - 그렉 이건: 의식을 컴퓨터로 시뮬레이션하는 미래를 다룬 SF 소설로, 디지털 지능에 대한 영감을 주었습니다.

괴델, 에셔, 바흐 (Gödel, Escher, Bach) - 더글러스 호프스태터: 지능과 의식의 본 질을 수학, 예술, 음악을 통해 탐구하는 고전입니다. 사피엔스 (Sapiens) - 유발 하라리: 인류의 역사와 미래를 거시적인 관점에서 바라 보게 해준 책입니다. 파운데이션 (Foundation) 시리즈 - 아이작 아시모프: 수학적 계산으로 미래를 예 측하려는 시도를 다룬 SF 대작입니다.

부록 5. 추천 영상 알파고 (AlphaGo, 2017): 이세돌 9단과 알파고의 대결을 담은 다큐멘터리. 기술적 성취 뒤에 숨겨진 인간적인 고뇌와 드라마를 보여줍니다. The Thinking Game (2024): 허사비스의 삶과 딥마인드가 단백질 구조 예측(알파폴드)에 도전하는 과정을 다룬 최신 다큐멘터리입니다. 렉스 프리드먼 팟캐스트 (Lex Fridman Podcast): 허사비스가 출연하여 AGI의 미 래, 의식의 본질, 물리학에 대해 심도 있게 대화한 인터뷰 영상들입니다. 노벨 강연 (Nobel Lecture, 2024): 2024년 12월 8일, 스톡홀름에서 허사비스가 자 신의 연구 성과와 미래 비전을 직접 발표한 강연 영상입니다.

TED 강연 - 인공지능이 자연과 우주의 비밀을 푸는 법 (2024): AI가 단순한 챗봇을 넘어 과학의 도구로서 어떻게 세상을 바꿀지 설명하는 강연입니다.

데미스 허사비스, 구글 인공지능의 아버지 전자책 발행 | 2026년 4월 20일 저 자 | 김경란, 김경진 펴낸곳 | 김경진 변호사 출판사 출판사등록 | 2025. 3. 10. (제2025-000015호) 주 소 | 서울특별시 동대문구 전농로 91, 백일빌딩 304호 전 화 | 02-6338-1905 이메일 | kimkj008@gmail.com (C) 김경란, 김경진 2026 본 책은 저작자의 지적 재산으로서 무단 전재와 복제를 금합니다.

인공지능 전문가 김경진 변호사

AI 법정책 전문 · 전 국회의원 · 저서 다수

이 책이 잠시라도 당신 곁에 머물렀다면, 다음 이야기가 세상에 나올 수 있도록 후원해 주세요.

(자발적 후원 부탁 구좌 : 농협 302-1096-0948-81 예금주 : 김경진)

이 책을 잘 읽으셨으면 그리고 새로운 가치있는 지식을 얻으셨다고 판단되시면
농협 302-1096-0948-81 (예금주 김경진) 에 자발적 후원 부탁드립니다.

© 2026 김경란, 김경진. All rights reserved.