

AI 및 디지털 산업에 대한 경쟁법적 분석 ②

AI 칩 생태계의 시장집중과 고착화*

법무법인(유) 세종 변호사 || 이창훈,** 박규태, 김태석

I. 서론

인공지능(Artificial Intelligence, AI) 분야에서의 경쟁법 이슈들에 관한 연작 중 지난 제1부¹⁾에서는 AI 공급망의 전반적 구조와 그 구조적 특성, 그리고 각 단계에서 제기되는 경쟁법적 이슈의 유형을 개관하였다. 제2부에서는 그중 AI 공급망의 최상류에 위치하는 AI 칩 영역을 대상으로, 그 영역의 구조적 특성과 이를 배경으로 논의되고 있는 경쟁법적 이슈들을 보다 구체적으로 살펴본다.

II. AI 칩 영역의 구조적 특성

AI 칩은 AI 모델의 학습과 추론을 뒷받침하는 연산 자원의 핵심으로서, 그 공급 구조는 AI 공급망 전체의 경쟁 조건에 영향을 미칠 수 있다는 점에서 주목받고 있다. 제1부에서 살펴본 바와 같이, AI 공급망 전체의 구조적 특성은 시장집중과 진입장벽, 핵심 자원의 공급 부족, 다층적 록인(Lock-in) 효과, 수직결합과 파트너십(Partnership)으로 요약할 수 있다. 이하에서는 AI 칩 영역을 대상으로 이러한 특성들이 구체적으로 어떻게 나타나고 있는지 살펴본다.

* 본 기획 연재는 총 4부로 구성되어 있다. 제1부 총론에서는 AI 공급망의 구조 및 특성, 그로부터 제기되는 경쟁법적 이슈와 주요 경쟁당국의 동향을 개관하였고, 본고는 제2부로서 AI 칩 영역에서 논의되고 있는 경쟁법적 이슈들을 살펴본다. 향후 연재될 제3부에서는 클라우드 인프라 및 파운데이션 모델 영역을, 제4부에서는 AI 서비스 영역을 각각 검토 대상으로 삼을 예정이다. 아울러 본고는 필자들 개인의 견해이며 소속 법무법인의 입장과는 무관함을 밝힌다.

** 필자는 법무법인(유) 세종의 경영위원이자 AI·디지털경쟁법팀, 공정거래그룹 팀장을 맡고 있다.

1) 경쟁저널 제234호(2026. 6.) 42면 이하.

1. AI 칩 영역의 다층적·순차적 구조

AI 칩 영역은 AI 공급망의 여러 단계 중 최상단에 위치하여 있다. 그런데 AI 칩 영역을 더 자세히 들여다보면 다시 여러 층위로 세분화되어 있다. AI 칩 영역은 AI 칩의 생산 과정에 따라 아키텍처(Architecture) 및 설계 자산(Intellectual Property, IP), 설계(Design), 제조(Fabrication), 조립·검사·패키징(Assembly, Testing and Packaging, ATP) 등 여러 계층이 순차적으로 결합되어 있다. 각 단계는 AI 칩 영역 내에서 서로 영향을 미칠 뿐만 아니라, AI 칩 영역보다 하류에 있는 클라우드(Cloud), 모델, 서비스 영역에도 영향을 미칠 수 있다.

AI 칩은 특정한 아키텍처에 따라 설계된다. 여기서 ‘아키텍처’란 소프트웨어와 하드웨어 간 통신을 가능하게 하는 규약으로서, 칩을 설계하고 프로그램을 작성하는 언어에 해당한다. 중앙처리장치(Central Processing Unit, CPU)의 경우 대표적인 아키텍처로 x86 아키텍처와 암(Arm) 아키텍처가 있다. x86 아키텍처는 주로 PC와 데이터센터(Data Center) 서버 CPU의 설계에 사용되어 왔다. 반면, 암 아키텍처는 전력 효율성을 강점으로 모바일, 사물인터넷(Internet of Things, IoT), 자동차 등의 영역에서 높은 비중을 차지하여 왔는데, 최근에는 데이터센터 CPU 영역에서도 전력 효율성에 대한 요구가 커지면서 암 아키텍처의 비중이 빠르게 확대되고 있다. 한편, 그래픽처리장치(Graphics Processing Unit, GPU)의 경우 엔비디아(NVIDIA) 등 제조사가 독자적으로 개발한 아키텍처를 기반으로 설계도를 작성하며, CPU에서와 같이 사실상 표준화된 아키텍처가 존재하지는 않는 것으로 보인다.

칩 설계는 고도로 복잡한 과정으로²⁾ 소수의 대형 사업자들은 직접 설계를 수행하는 반면, 대다수의 칩 설계 사업자들은 사전에 개발된 설계 모듈인 IP 코어(IP Core)를 구매하여 자사 칩 설계에 통합한다. 이를 통해 설계의 복잡성을 줄이고 출시 기간을 단축할 수 있다. 이후 칩 설계 사업자들은 전자설계자동화(Electronic Design Automation, EDA) 소프트웨어를 통해 회로 레이아웃(Layout) 검증, 물리적 설계 최적화, 제조 전 성능 시뮬레이션(Simulation)을 수행함으로써 실제 제조가 이루어지기 전 설계상의 오류를 발견하고 수정한

2) 칩 설계는 수천 명의 엔지니어, 수년 간의 연구개발, 수억 달러의 투자가 수반되는 고도로 복잡하고도 다원적인 과정이며, 제조의 상류이자 기초 연구의 하류에 위치하는 고부가가치 활동이다. 칩 설계는 상당한 지식재산 가치를 부가하며, AI 시스템의 성능, 비용, 에너지 효율을 좌우할 수 있다. OECD, 「Competition in Artificial Intelligence Infrastructure」, OECD Roundtables on Competition Policy Papers, No. 330 (2025, 11.), 12면.

다. 이렇게 완성된 설계도를 바탕으로 파운드리(Foundry), 즉 위탁제조를 전문으로 하는 기업에서 물리적 칩이 제조된다.³⁾ 제조된 칩은 조립, 검사, 패키징 단계를 거치며 메모리 및 인터커넥트(interconnect), 즉 칩과 칩을 잇는 고속 연결 시스템과 결합되어 완성된다.

이러한 단계들은 순차적으로 연결되어 있어, 상류 단계의 생산량, 경쟁 상황, 병목이 하류 단계 전체에 영향을 미칠 수 있다. 결과적으로 특정 단계의 시장집중이나 공급 제약은 칩 영역 내부에 그치지 않고, 그 하류에 있는 클라우드, 모델, AI 서비스 영역에까지 파급되어 비용 구조와 진입 조건에 영향을 미칠 가능성이 있다.

2. AI 칩의 종류와 패러다임 전환

AI 칩은 AI 모델의 학습과 추론에 필요한 연산을 위해 설계된 반도체이다. AI 모델의 학습에 가장 널리 사용되어 온 것은 GPU이다. GPU는 AI 연산에 필요한 병렬 처리에 특화되었으면서도, 다양한 AI 모델에 적용할 수 있는 유연성을 갖추고 있어 학습 단계에서 각광을 받아왔다. 그러나 특정 작업에 더 최적화된 AI 칩에 대한 수요가 증가하면서, 특정 용도에 맞추어 제작하는 주문형 반도체(Application-Specific Integrated Circuit, ASIC)와 제조 후에도 용도에 맞게 회로를 다시 프로그래밍할 수 있는 반도체(Field-Programmable Gate Array, FPGA)의 비중이 점차 증가하고 있다.⁴⁾ 이러한 흐름 속에서 AI의 새로운 패러다임 전환이 CPU를 비롯한 다양한 칩의 역할을 재정립하고 있다.

첫째, AI가 ‘훈련(Training)’ 중심에서 ‘추론(Inference)’ 중심으로 전환되면서 속도, 비용, 전력 효율이 중시됨에 따라 ASIC 등 다양한 칩이 더 많이 사용될 것으로 전망된다. 2025년 기준 GPU와 ASIC의 AI 칩 시장 점유율은 각각 약 73%와 27%로,⁵⁾ 이미 ASIC이 GPU 점유율을 잠식하기 시작한 것으로 보인다. 2026년 추론 워크로드(Work Load)는 전체 AI 컴퓨팅의 약 2/3를 차지할 것으로 예상되고 있으며,⁶⁾ ASIC 등 다른 칩의 점유율이 더 높아질 것으

3) 이 과정에서 실리콘 웨이퍼(Silicon Wafer), 희토류 등 원재료와 극자외선(Extreme Ultraviolet, EUV) 노광장비 등 고도로 특화된 장비가 필요하다.

4) 포르투갈 경쟁당국(Autoridade da Concorrência, 이하 ‘AdC’라 한다), 「Competition and Generative AI: Access to AI Chips」(2026, 2.), 2면; OECD, 「Competition in Artificial Intelligence Infrastructure」(2025, 11.), 11면.

5) Computerworld, 「CES 2026: AI compute sees a shift from training to inference」(2026, 1.); Third Bridge, 「Wrapping up 2025: AI Chips」(2025, 12.).

6) Deloitte, 「More compute for AI, not less」(2025, 11.).

로 예상된다. CPU의 역할도 재조명되고 있는데, 추론 단계에서는 제어, 조정, 복잡한 의사결정이 동시에 이루어지기 때문에 CPU의 역할이 더욱 중요하다.⁷⁾

둘째, 스스로 계획하고, 도구를 호출하며, 다단계 작업을 자율적으로 수행하는 에이전틱 AI(Agentic AI)의 부상이 AI 칩 수요 구조에 영향을 미치고 있다. 에이전틱 AI 시스템은 복잡한 ‘오케스트레이션(orchestration)’, 즉 여러 작업의 스케줄링(Scheduling)과 자원 할당 등을 통합적으로 관리하는 작업을 수행한다. 이러한 작업은 GPU가 아닌 CPU에서 처리되기 때문에 에이전틱 AI의 확산은 곧 CPU 중심 워크로드의 증가를 의미한다. 에이전틱 AI 환경에서는 AI 데이터센터의 GPU 대비 CPU 비율이 기존 1:4~1:8에서 1:1~1:2로 전환될 것이라는 전망도 나오고 있다.⁸⁾ 이러한 변화 속에서 엔비디아는 2026년 3월 AI 에이전트 맞춤형으로 설계된 ‘Vera CPU’를 독립 제품으로 출시하였고, 같은 달 IP 라이선싱(Licensing) 중심의 기업이었던 암도 자체 CPU인 ‘Arm AGI CPU’를 발표하였다.

셋째, AI가 물리적 세계와 상호 작용하는 ‘피지컬 AI(Physical AI)’가 확산되면서 반도체 수요가 더욱 다각화되고 있다. ‘피지컬 AI’란 로봇(Robot), 자율주행차, 산업용 자동화 시스템 등 실제 환경을 인지하고 추론하며 행동할 수 있는 AI를 의미한다. 피지컬 AI 시스템은 기기 가까이에서, 즉 엣지(Edge)에서 실시간으로 인지, 의사결정, 물리적 동작을 수행해야 하므로, 데이터센터 기반의 대규모 GPU 클러스터(Cluster)와는 다른 저지연, 저전력, 고효율의 칩 아키텍처를 필요로 한다.

이처럼 추론 기능, 에이전틱 AI, 피지컬 AI의 확산은 AI 칩에 대한 수요가 GPU 중심에서 CPU, 엣지 AI 칩 등으로 다각화되고 다양한 반도체의 역할이 함께 중요해지는 ‘AI 칩의 다원화’를 이끌고 있다.

3. AI 칩 공급망의 시장집중

AI 밸류체인(Value Chain)은 전체적으로 집중도가 높은 것으로 평가되고 있고, AI 칩 영역 역시 마찬가지인 것으로 보인다. 설계 영역에서는 데이터센터용 고성능 GPU의 80~90%

7) AMD, 「Agentic AI Brings New Attention to CPUs in the AI Data Center」(2026. 3.).

8) TrendForce, 「The Great Rebalance: How Agentic AI Is Reshaping the CPU: GPU Ratio」(2026. 4.).

② AI 칩 생태계의 시장집중과 고착화

이상을 엔비디아가 공급하고 있고, IP 코어와 EDA 소프트웨어 영역도 소수의 사업자가 60~70% 이상을 점유하는 것으로 분석되고 있다.⁹⁾ AI 칩 제조(Fabrication) 단계에서는 티에스엠씨(TSMC)가 80% 이상을 점유하고 있고, 노광(Lithography) 장비는 에이에스엠엘(ASML)이 80% 이상을 점유하고 있다고 분석되고 있다.¹⁰⁾ 고대역폭 메모리(High Bandwidth Memory, HBM) 영역도 SK하이닉스, 삼성전자, 마이크론 테크놀로지(Micron Technology) 등 3개 사업자가 상당 부분을 점유하고 있는 것으로 언급된다.¹¹⁾

한편, 최근 중요도가 커지고 있는 CPU 영역에서는 가장 최상류에 아키텍처 라이선싱 단계가 위치해 있다는 점이 주목된다. CPU 설계 아키텍처 중 x86 아키텍처는 인텔(Intel)이 개발한 독자 아키텍처로서 외부 라이선싱(Licensing)이 이루어지지 않고, 인텔과 에이엠디(AMD)만 설계와 제조를 하고 있다. 오픈소스(Open Source) 대안으로 부상하고 있는 리스크-파이버(RISC-V)는 로열티(Royalty) 없이 사용할 수 있는 장점이 있으나, 소프트웨어 생태계의 미성숙, 성능 격차, 표준화 부족 등으로 데이터센터 CPU 영역에서 단기간 내 다른 아키텍처를 대체하기는 어려울 것으로 평가되고 있다.¹²⁾ 결국 외부 라이선스(Licence)가 가능한 고성능 CPU 아키텍처는 암 아키텍처가 사실상 유일한데, 현재 하이퍼스케일러(hyperscaler), 즉 상위 대형 데이터센터 사업자의 약 절반이 암 아키텍처 CPU에 해당하는 것으로 추정되고 있다.

4. AI 칩의 공급 부족과 병목

AI 칩 영역에서는 공급 부족 상황이 상당 기간 지속되고 있다. 고성능 GPU는 한때 신규 구매 대기 기간이 1년 가까이에 달하였다가 3~4개월 수준으로 완화되었으나, 높은 가격이 유지되고 있다. HBM 메모리 역시 유사한 공급 부족을 겪고 있으며, 첨단 패키징 설비도 자본집약적이어서 신속한 증설이 어려운 가운데 제약을 받고 있다고 분석된다. 제조 단계에서

9) OECD, 「Competition in Artificial Intelligence Infrastructure」 (2025, 11.), 13-15면.

10) '노광(Lithography)'이란 칩을 구성하는 정교한 회로 패턴을 만들어 내는 공정이다. 점점 더 작아지는 칩에 더 많은 트랜지스터(Transistor)를 집적하려면, 나노(Nano) 단위의 미세한 규모에서 패턴을 형성하기 위해 더욱 고도화된 실리콘 노광 기술이 요구된다. OECD, 「Competition in Artificial Intelligence Infrastructure」 (2025, 11.), 23면.

11) OECD, 「Competition in Artificial Intelligence Infrastructure」 (2025, 11.), 23면.

12) Data Center Dynamics, 「Can RISC-V thrive in the data center?」 (2024, 8, 13.).

도 파운드리 서비스에 대한 이용이 제한적이어서 자본력이 튼튼한 고객이 우선적으로 배정된다는 분석이 있다. 칩의 세대 교체 주기가 짧아지고 AI 칩 등 하드웨어의 감가상각이 점차 빨라지는 상황은 사업자들에게 부담으로 작용할 수 있다.¹³⁾

최근에는 에이전틱 AI 등의 확산으로 CPU 수요가 급증하면서, CPU 공급 부족도 새로운 병목으로 부상하고 있다. 2026년 데이터센터 서버 CPU 시장은 상당한 공급 부족을 겪을 것으로 전망되고 있으며, 인텔은 대규모 데이터센터 기업들의 수요를 완전히 충족하는 데 어려움을 겪고 있는 것으로 알려져 있고, 에이엠디의 CPU도 2026년 물량이 이미 거의 품절된 상황으로 알려져 있다.¹⁴⁾

5. 소프트웨어 결합과 록인 효과

AI 칩은 하드웨어만이 아니라 이를 구동하는 소프트웨어 환경과 긴밀히 결합되어 활용된다. GPU 영역에서 가장 대표적인 것이 엔비디아의 통합 개발 환경인 CUDA(Compute Unified Device Architecture)이다. CUDA 생태계는 500만 명 이상의 개발자와 3,700개 이상의 애플리케이션에서 사용되고 있어 AI 개발의 사실상 표준으로 기능하고 있다. 그런데 CUDA는 엔비디아 하드웨어에서만 작동하도록 설계되어 있다. 따라서 다른 칩으로의 전환은 CUDA를 포기하고 대체 소프트웨어 스택(Stack)을 채택해야 함을 의미하며, 이 경우 개발자들은 기존에 의존하던 최적화된 도구와 라이브러리(Library)를 다시 마련해야 한다. 또한, 다른 칩을 도입할 때 기존 코드(Code)가 그대로 작동한다는 보장이 없어 추가 엔지니어링(Engineering) 작업에도 상당한 시간이 소요될 수 있다.¹⁵⁾ 이로 인해 다른 칩으로 전환하기 어려운 기술적 록인(Lock-in)이 형성될 수 있고, 이는 개발자에게 상당한 전환비용을 발생시켜 AI 하드웨어 시장에 진입하거나 확장하려는 경쟁 칩 공급자에게 장벽을 높일 수 있다.¹⁶⁾ 한편, 소프트웨어 환경이 CUDA로 인해 엔비디아 GPU를 전제로 설계된 경우가 많다

13) AdC, 「Competition and Generative AI: Access to AI Chips」 (2026. 2.), 5-6면.

14) Fusion Worldwide, 「Server CPU Supply and Pricing in 2026」 (2026. 5.); TechPowerUp, 「Intel Reallocates PC Production Capacity to Server CPUs Amid Tight Wafer Supply」 (2026. 1.).

15) '파이토치(PyTorch)'나 '텐서플로(TensorFlow)'와 같이 하드웨어의 차이를 가려 주는 프레임워크(Framework)가 전환비용을 일부 완화할 수 있지만, 이는 주로 일반적인 용도와 초기 개발 단계에서 효과가 있고 성능 최적화나 대규모 추론 단계까지 해결해 주지는 못하는 것으로 보인다. AdC, 「Competition and Generative AI: Access to AI Chips」 (2026. 2.), 14면.

16) AdC, 「Competition and Generative AI: Access to AI Chips」 (2026. 2.), 13-16면.

는 점에서, AI 칩 시장이 학습 단계에서 추론 단계로 이행하더라도 여전히 엔비디아 GPU의 우위가 지속될 것이라는 의견도 제시된다.¹⁷⁾

암 아키텍처가 높은 비중을 차지하고 있는 CPU 영역에서도 유사한 소프트웨어 생태계 록인 효과가 논의된다. 하드웨어와 소프트웨어가 긴밀히 결합된 생태계에서는 더 많은 개발자와 기업이 특정 생태계를 채택할수록 이용할 수 있는 도구와 활용 사례가 늘어, 그 소프트웨어의 가치가 다른 이용자에게도 커지는 네트워크 효과가 나타난다.¹⁸⁾ 암의 경우 이러한 과정을 거쳐 운영체제, 컴파일러(Compiler), 개발도구, 응용 프로그램에 이르는 광범위한 소프트웨어 생태계가 형성되었고, 그 결과 반도체 산업 전반이 20년 이상 암의 기술에 의존해 온 것으로 평가된다. 이 경우 설령 경쟁사업자가 이론상 더 효율적인 아키텍처를 설계하더라도, 전세계 개발자 커뮤니티가 새로운 표준으로 응용 프로그램을 이식(Porting)하도록 하는 데에는 상당한 시간과 비용이 소요될 수 있다.

소프트웨어 결합에 따른 록인 효과의 정도는 AI 워크로드의 단계에 따라 차이가 있을 수 있다.¹⁹⁾ 학습 단계에서는 대규모 병렬 연산이 필수적이어서 엔비디아 GPU에 대한 의존이 강한 것으로 보인다. 반면, 추론 단계에서는 학습 단계만큼의 연산 용량이 필요하지 않고 빠른 처리와 저전력 소비가 중요하기 때문에 다양한 사업자들의 대안적 칩이 상대적으로 활발한 것으로 분석되어, 록인 효과가 단계마다 다르게 나타날 수 있는 것으로 해석되고 있다. 그러나 앞서 언급한 바와 같이, 생태계의 전환에는 매우 높은 전환비용이 따를 수 있다.

6. 수직결합

AI 칩 영역에서는 다양한 형태의 수직결합이 관찰된다. 먼저 하류 영역인 클라우드, 파운데이션 모델(Foundational Model), AI 서비스 영역과의 결합이 나타나고 있다. 대표적으로 AI 칩 사업자인 엔비디아는 클라우드 서비스(DGX Cloud, GeForce Now 등)와 파운데이션 모델(Nemotron 등) 등 하류 영역으로 사업을 확장하고 있고, 이는 해당 하류 영역의 사업자들의 경쟁구조 및 경쟁압력에 영향을 미칠 수 있다.²⁰⁾ 반대로, 클라우드 사업자들도 외부

17) 일본 공정취인위원회(Japan Fair Trade Commission, 이하 'JFTC'라 한다), 「Report Regarding Generative AI ver. 2.0」(2026. 4.), 8, 10면.

18) OECD, 「Competition in Artificial Intelligence Infrastructure」(2025. 11.), 28면.

19) OECD, 「Competition in Artificial Intelligence Infrastructure」(2025. 11.), 39면.

공급자에 대한 의존을 낮추기 위해 자체 AI 칩을 설계하는 경우가 관찰되는데, 이 경우 자체 AI 칩을 주로 자사 클라우드 환경에서 제공하고 있다.²¹⁾ 이러한 관계는 칩 공급자와 클라우드, 모델 개발사 사이의 ‘순환투자(circular investments)’, 즉 대형 사업자가 AI 개발사에 투자하고, 그 개발사가 다시 같은 사업자의 칩이나 클라우드를 이용하게 되는 형태로도 나타나고 있다. 이에 대해서는 클라우드 사업자가 칩 설계에 진입하고 칩 공급자가 클라우드 서비스로 확장함에 따라 하드웨어와 인프라에 대한 통제력이 더욱 집중될 수 있고, 이는 개발자를 록인하려는 생태계상의 유인을 강화하여 상호운용성, 자사우대, 봉쇄 위험에 관한 우려를 낳을 수 있다는 의견이 있다.²²⁾

AI 칩 영역 내부의 여러 층위 사이에서도 수직결합이 나타날 수 있다. AI 워크로드의 규모가 개별 가속기의 처리 용량을 넘어서면서 성능은 점점 더 칩 간 통신의 효율성에 좌우되고 있으며, 이러한 맥락에서 네트워킹 기술이 AI 컴퓨팅의 핵심 구성 요소가 되고 있다. 서버 내부에서 복수의 칩을 연결하는 인터커넥트는 각 칩 설계사 고유의 생태계에 속하며,²³⁾ 이 기술들은 각자의 생태계 안에서만 작동하기 때문에 서로 다른 AI 칩을 조합할 가능성을 제한한다. 이러한 상황에서 인터커넥트에 대한 통제는 AI 스택 전반으로 시장지배력을 확장시킬 수 있으며, 칩 설계사가 네트워킹 계층으로 수직통합하면서 기술적 통합 또는 계약적 관행을 통해 록인 효과를 강화할 수 있고, 이는 칩 단계의 지위를 AI 스택의 인접 부문으로 전이할 위험을 높일 수 있다.²⁴⁾

이와 관련하여, 특히 AI 공급망 최상단에 위치한 아키텍처 라이선싱 단계의 사업자가 칩 공급 단계를 겸하게 되는 경우 그 영향이 넓게 미칠 수 있다는 점도 지적된다. 2020년 엔비디아의 암 인수 시도가 그러한 수직 통합이 야기할 수 있는 경쟁제한 우려를 여러 경쟁당국이 제기하면서 무산되었는데, 최근에는 암이 자체 칩을 직접 공급하기 시작하면서 같은 우려에 대한 논의가 다시 제기되고 있다.

20) OECD, 「Competition in Artificial Intelligence Infrastructure」 (2025. 11.), 25면; 호주 경쟁소비자위원회(Australian Competition and Consumer Commission, 이하 ‘ACCC’라 한다), 「Digital Platform Services Inquiry – Final Report」 (2025. 3.), 259면.

21) 프랑스 경쟁당국(Autorité de la concurrence, 이하 ‘ADLC’라 한다), 「Opinion 24-A-05 of 28 June 2024 on the competitive functioning of the generative artificial intelligence sector」 (2024. 6.), 54면.

22) AdC, 「Competition and Generative AI: Access to AI Chips」 (2026. 2.), 12면.

23) NVIDIA의 ‘NVLink’와 AMD의 ‘Infinity Fabric’이 그 예이다.

24) AdC, 「Competition and Generative AI: Access to AI Chips」 (2026. 2.), 16–17면.

Ⅲ. AI 칩 영역의 경쟁법적 이슈

AI 칩 영역에서는 차별취급, 결합판매, 상호호환성 제한, 거래 거절 및 접근 제한 등 개별 사업자의 단독행위와 이를 구조적으로 강화할 수 있는 수직결합, 그리고 다수 사업자 간의 담합 등이 주요 경쟁법적 이슈로 논의되고 있다. 이하에서는 각 유형별로 논의되는 행위의 내용을 살펴본 뒤, 실제 사안들을 어느 경쟁당국이 어떠한 절차로 다루고 있는지를 중심으로 소개한다.²⁵⁾

1. 차별취급

‘차별취급’이란, 사업자가 거래상대방에 따라 가격, 공급 시기, 공급 물량 등 거래 조건을 합리적 이유 없이 달리하여 특정 사업자를 경쟁에서 불리하게 만드는 행위를 말한다. AI 칩 영역에서는, 특히 공급이 부족한 국면에서 칩 배분에 강한 영향력을 가진 사업자가 자사 제품을 배타적으로 구매하는 고객 등을 우대하는 반면, 경쟁사업자의 칩을 함께 구매하는 고객 등에게는 더 높은 가격을 제시하거나 공급 물량을 제한하는 방식이 논의된다.²⁶⁾

이와 관련하여 실제로 다루어지고 있는 대표적 사안은 미국 법무부(Department of Justice, 이하 ‘DOJ’라 한다)의 엔비디아에 대한 조사이다. 이 조사는 2024년경 경쟁사업자들의 문제 제기를 계기로 개시된 것으로 전해지며, 엔비디아가 경쟁사업자의 칩을 함께 구매할 고객에게 자사 GPU를 더 높은 가격에 공급하거나, 그 고객에게 판매하는 GPU 물량을 제한하고 있는지 여부를 살펴보고 있는 것으로 알려졌다.²⁷⁾

이는 공급이 부족한 상황에서 칩 배분에 영향력을 가진 사업자가 거래상대방의 구매 선택에 따라 거래 조건을 달리하는 것이 경쟁에 어떠한 영향을 미치는지의 문제로 논의될 수 있다. 공급 부족 상황 자체는 하루 영역의 AI 혁신을 지연시키는 한편, 시장력을 집중시켜 자원 선점과 우선적 배분의 위험을 낳을 수 있다. 배타적 계약이나 수직결합을 통해 우선적 접근을 확보한 사업자는 AI 개발을 가속할 수 있는 반면 그렇지 못한 사업자는 뒤처지게 되므로, 이러한

25) 이하에서 소개하는 사안들의 상당수는 조사 또는 문제 제기 단계에 있거나 경쟁당국의 연구보고서에서 언급되는 것들로서, 현재까지 경쟁법 위반 여부가 확정된 것은 아니라는 점을 먼저 밝혀둔다.

26) ACCC, 「Digital Platform Services Inquiry — Final Report」(2025. 3.), 316면.

27) OECD, 「Competition in Artificial Intelligence Infrastructure」(2025. 11.), 58면 각주 14; ACCC, 「Digital Platform Services Inquiry — Final Report」(2025. 3.), 316면; Data Center Dynamics, 「US DOJ opens two antitrust probes into Nvidia」(2024. 8.).

환경에서는 누구에게 공급할 것인지에 대한 통제가 전략적 수단이 된다. 그 결과, 지배적 지위에 있는 칩 제조사가 자사 내부 프로젝트 또는 선호하는 파트너에게 공급을 우선함으로써 사실상 경쟁구도를 좌우할 수 있다. 아울러 대형 고객이 물량 약정, 배타조건, 우대가격 등을 포함한 다년계약을 체결하는 관행도 단기적으로 고객의 전환 가능성을 제약할 수 있다.²⁸⁾

유럽 집행위원회(European Commission, 이하 ‘EC’라 한다)는 2024년 엔비디아의 런:AI(Run:ai) 인수를 심사하면서 엔비디아의 차별취급 우려 문제를 다루었다. 런:AI는 여러 개의 GPU에 연산 작업을 나누어 배분하는 이른바 ‘GPU 오케스트레이션 소프트웨어’를 공급하는 기업으로, 이러한 소프트웨어는 GPU와 서로 호환되어야 제 기능을 한다. EC는 심사 과정에서 엔비디아가 데이터센터용 독립형 GPU 세계 시장에서 지배적 지위를 보유할 가능성이 높다고 보면서,²⁹⁾ 결합 후 엔비디아가 런:AI 소프트웨어와 엔비디아 경쟁사업자의 GPU 사이의 호환성을 저해하여 경쟁 소프트웨어나 경쟁 GPU를 차별적으로 취급할 수 있는지를 검토하였다. 당시 EC는 호환성을 확보해 주는 도구가 이미 이용 가능하고 널리 사용되고 있어 엔비디아가 호환성을 저해할 기술적 능력이나 유인을 갖기 어렵다고 보아, 2024년 12월 이 거래를 승인하였다.³⁰⁾

2. 결합판매

‘결합판매(끼워팔기)’란, 서로 구별되는 둘 이상의 상품을 묶어 판매하면서 한 상품에 대한 시장 지위를 다른 상품 시장으로 확장하는 행위를 말한다. AI 칩 영역에서는 대표적으로 AI 칩을 자사 네트워킹 장비 등 다른 상품과 함께 구매하도록 유도하는 행위가 논의된다. AI 워크로드가 개별 가속기의 용량을 넘어 확장되면서 칩 사이의 통신을 담당하는 네트워킹의 중요성이 커지고 있는데, 칩 설계 사업자가 네트워킹 영역으로 사업을 확장하는 과정에서 칩 단계에서의 지위가 인접 영역으로 전이될 수 있다는 점이 지적된다.³¹⁾

28) OECD, 「Competition in Artificial Intelligence Infrastructure」 (2025. 11.), 28-29면.

29) 우리 공정거래위원회도 국내 AI 반도체 영역에서 엔비디아 제품에 대한 높은 수요와 선호도, 이를 바탕으로 한 엔비디아의 지위를 언급한 바 있다. 공정거래위원회, 「생성형 AI와 경쟁」 정책보고서 (2024. 12.), 문단 142.

30) EC, 「Commission approves acquisition of Run:ai by NVIDIA」, IP/24/6548 (2024. 12.); EC, Case M.11766 — NVIDIA/Run:AI (2024. 12.), 32면.

31) AdC, 「Competition and Generative AI: Access to AI Chips」 (2026. 2.), 17면.

앞서 본 DOJ의 엔비디아 조사에서 이러한 결합판매 측면도 함께 검토되고 있으며, EC 역시 이를 살펴보고 있는 것으로 전해진다.³²⁾ 경쟁사업자들은 엔비디아로부터 GPU와 네트워킹 케이블을 함께 구매한 고객이 각 제품을 더 낮은 가격에 구매하게 된다는 점을, DOJ에 문제로 제기하였다.

엔비디아의 네트워킹 사업은 2020년 네트워킹 장비 사업자인 멜라노스(Mellanox) 인수를 통해 확보된 것이다. EC는 2019년 12월 엔비디아의 멜라노스 인수 거래를 조건 없이 승인하였다. 이 거래는 데이터센터용 GPU 분야의 사업자와 데이터센터용 네트워크 인터커넥트 분야의 사업자 간 결합으로, 두 제품은 함께 사용되는 보완재라는 관점에서, 한 제품에서의 지위를 보완재 시장으로 확장할 수 있는지에 초점이 맞추어져 심사가 이루어졌다. EC는 엔비디아가 멜라노스 제품을 지렛대로 활용할 능력과 유인이 있더라도, 그 영향이 GPU 시장의 일부에 그칠 것이라고 보았다.³³⁾

반면, 중국 국가시장감독관리총국(State Administration for Market Regulation, 이하 ‘SAMR’이라 한다)은 약 1년에 걸친 2단계(Phase II) 심사³⁴⁾를 거쳐, 2020년 4월 조건부로 승인하였다. SAMR은 결합기업이 인접시장에서 높은 지위를 갖게 된다고 보아 끼워팔기, 결합판매, 봉쇄(Foreclosure)의 위험을 완화하기 위하여 FRAND, 즉 공정·합리적·비차별적 조건에 따른 공급, 상호운용성 유지, 제3자 기밀정보 보호 등의 행태적 시정조치를 부과하고, 아울러 GPU 팀과 인터커넥트 팀 사이의 조직적 분리 유지와 오픈소스 관련 약속의 준수를 요구하였다.³⁵⁾

3. 상호호환성 제한

AI 칩 영역에서 상호호환성 제한은, AI 개발에 필수적인 소프트웨어를 공급하면서 해당 소프트웨어가 자사 AI 칩에서만 작동하도록 설계하여, 다른 칩으로의 전환에 상당한 비용을

32) Data Center Dynamics, 「US DOJ opens two antitrust probes into Nvidia」 (2024. 8.); ACCC, 「Digital Platform Services Inquiry — Final Report」 (2025. 3.), 316면.

33) OECD, 「Competition in Artificial Intelligence Infrastructure」 (2025. 11.), 37면.

34) 1단계 예비심사에서 경쟁제한 우려가 해소되지 않아 추가로 진행되는 심층심사이다.

35) OECD, 「Competition in Artificial Intelligence Infrastructure」 (2025. 11.), 37면. 이후 SAMR은 2020년 승인 당시 부과된 조건의 이행 여부에 관한 조사를 개시하여, 2025년 9월 예비조사 결과 엔비디아가 승인 조건을 위반한 것으로 판단된다고 발표하였다. 이는 엔비디아가 미국 수출 통제를 준수하기 위해 중국 시장에 저사양 제품을 공급한 것이 중국 고객에 대한 차별에 해당하는지의 문제로서, 기업결합 승인 조건의 사후 이행에 관한 사안이다. CNBC, 「China: Nvidia violated anti-monopoly law, will continue investigation」 (2025. 9.); 엔비디아 2026 회계연도 분기보고서(Form 10-Q).

발생시키는 행위를 위주로 논의되고 있다.

대표적으로 거론되는 사례는, 고성능 GPU 영역에서 사실상 표준으로 기능하고 있는 CUDA이다. 앞서 살펴본 바와 같이, CUDA는 엔비디아 하드웨어에서만 작동하도록 설계되어 있어, 다른 칩으로 전환하려는 개발자는 기존에 축적한 코드와 라이브러리 자산을 상당 부분 새로 마련하고 소프트웨어를 재설계해야 한다. 이에 따라 이러한 전환비용이 경쟁 칩 사업자의 진입과 확장에 영향을 줄 수 있다는 점이 지적되고 있다.³⁶⁾

이러한 문제의식은 실제 경쟁당국의 조사로 이어지고 있는 것으로 파악된다. 프랑스 경쟁당국(ADLC)은 2023년 9월 26일 엔비디아의 파리 및 니스 사무소에 대하여 GPU 부문의 경쟁제한행위 혐의로 현장조사를 실시하였다.³⁷⁾ 이는 프랑스 클라우드 컴퓨팅(Cloud Computing) 부문에 관한 보다 넓은 조사의 일환으로 이루어진 것으로서, 이후 이 조사는 엔비디아 자체를 대상으로 하는 별도의 경쟁법 사건으로 전환되었다.³⁸⁾ 위 조사는 엔비디아의 GPU 시장지배력과 CUDA에 대한 조사를 포함한 것으로 알려졌다.³⁹⁾ CUDA와 관련하여 거론되는 행위 중 하나는, 엔비디아가 CUDA 라이선스 조건을 통해 번역계층(Translation Layer)을 이용하여 CUDA 기반 소프트웨어를 다른 하드웨어 플랫폼에서 실행하는 것을 금지하고 있다는 점이며, 프랑스 경쟁당국 조사의 쟁점이 이를 포함하고 있다는 관측도 제기된다.⁴⁰⁾ 이후 ADLC는 2024년 6월 생성형 AI 부문에 관한 의견서(Opinion 24-A-05)를 발표하며, 엔비디아가 개발한 CUDA가 자사 칩에서만 작동함에도 이 부문 사업자들이 가장 널리 사용하고 있다는 점, 그리고 이 부문이 가속연산에 필수적인 GPU와 100% 호환되는 유일한 소프트웨어인 CUDA에 의존하고 있다는 점에 대한 우려가 제기되었다고 밝혔다.⁴¹⁾ 2026년 7월, ADLC의 조사가 마무리 단계에 이르러 심사보고서(Statement of Objections) 발부 여부에 관한 판단이 임박한 것으로 보도되었다.⁴²⁾

36) AdC, 「Competition and Generative AI: Access to AI Chips」(2026, 2.), 15면.

37) TechRadar, 「Nvidia office reportedly raided over cloud dominance inquiry」(2023, 9.); Crypto Briefing, 「French Competition Authority nears completion of Nvidia probe」(2026, 7.).

38) The Next Web, 「France's antitrust probe into Nvidia is nearing its end, regulator says」(2026, 7.).

39) Techzine, 「French antitrust case against Nvidia to combat AI domination」(2024, 7.); Crypto Briefing, 「French Competition Authority nears completion of Nvidia probe」(2026, 7.).

40) Tom's Hardware, 「Nvidia bans using translation layers for CUDA software」(2024, 3.); Techzine, 「French antitrust case against Nvidia to combat AI domination」(2024, 7.).

41) ADLC, 「Opinion 24-A-05 of 28 June 2024 on the competitive functioning of the generative artificial intelligence sector」(2024, 6.), 문단 128, 242.

42) Reuters, 「French competition authority's Nvidia probe nearing end」(2026, 7, 9.); Global Competition Review, 「France signals possible chargesheet in Nvidia antitrust probe」(2026, 7, 9.).

한편, CUDA를 둘러싼 문제의식은 프랑스에 국한되지 않는다. 앞서 본 DOJ의 조사에 관해서도, CUDA가 AI 개발자로 하여금 엔비디아 GPU와 함께 다른 사업자의 GPU를 사용하는 것을 사실상 어렵게 하고 있다는 우려가 조사의 배경으로 보도된 바 있다.⁴³⁾

이러한 조사들이 공통적으로 향하고 있는 지점은, 소프트웨어와 하드웨어의 결합이 그 자체로 위법한 것은 아니지만, 그 소프트웨어가 사실상의 산업 표준으로 기능하는 국면에서는 경쟁에 대한 함의가 달라질 수 있다는 것이다. CUDA는 엔비디아의 강한 시장 지위를 배경으로 사실상 업계 표준이 되었고, 엔비디아의 최고경영자 스스로 이를 AI의 운영체제라고 표현한 바 있다.⁴⁴⁾ 이러한 지위에 있는 소프트웨어가 자사 하드웨어에서만 작동한다면 개발자에게는 높은 전환비용이 발생하고 경쟁 칩 공급자에게는 진입과 확장의 장벽이 높아질 수 있으며, 독자적 소프트웨어 프레임워크와 그 기반 하드웨어 사이의 긴밀한 결합이 칩 설계 부문의 진입·확장장벽을 강화하는 데 기여할 수 있다는 지적도 같은 취지로 이해된다.⁴⁵⁾ 향후 이와 관련하여 상호운용성을 실질적으로 어떻게 확보할 것인가가 논의의 중심이 되어야 할 것이다.

4. 거래 거절 및 접근 제한

AI 칩 영역에서 거래 거절 및 접근 제한은, 칩 개발에 필요한 요소에 대한 거래, 이용, 접근을 거절하거나 그 조건을 제한하여 하류 영역의 경쟁에 영향을 미치는 행위가 주로 논의되고 있다. 칩 개발에 필요하며 달리 이를 대체할 수 없는 요소를 통제하고 있는 사업자가 하류 영역의 사업자로 하여금 이에 대한 접근을 하지 못하도록 한다면, 하류 영역의 경쟁구조에 직접적으로 영향을 미칠 수 있다.

AI 관련 칩 설계에는 특정 명령어 집합구조에 따라 직접 칩을 설계하거나 사전 설계된 IP 코어를 활용하는 것이 일반적이는데, 이러한 핵심 요소에 대한 라이선스를 거절하거나 그 조건을 제한하는 경우가 논의된다. ‘명령어 집합구조(Instruction Set Architecture, ISA)’란, 소프트웨어와 프로세서가 서로 소통할 수 있게 하는 기본 규약으로, 반도체 산업 전반이 20

43) Tech Monitor, 「Nvidia faces French antitrust probe」 (2024. 7.).

44) OECD, 「Competition in Artificial Intelligence Infrastructure」 (2025. 11.), 13면.

45) AdC, 「Competition and Generative AI: Access to AI Chips」 (2026. 2.), 15-16면.

년 이상 의존해 온 칩 설계의 토대에 해당한다. 이러한 아키텍처를 공급하는 대표적 사업자가 애플인데, 애플이 종래의 개방적 라이선스 정책에서 자사 칩 제품을 우대하는 보다 제한적인 방식으로 전환하고 있다는 점이 거론된다. 실제로, 미국 연방거래위원회(Federal Trade Commission, 이하 'FTC'라 한다) 등이 애플의 라이선스 모델에 관하여 검토하고 있는 것으로 알려져 있으며, 한국 공정거래위원회도 2025년 11월 애플의 서울사무소를 현장조사한 것으로 보도되었다.⁴⁶⁾

이러한 접근 제한에 관한 우려는 기업결합 심사에서 이미 다루어진 바 있다. 엔비디아는 2020년 9월 소프트뱅크(SoftBank)로부터 애플을 약 400억 달러에 인수하겠다고 발표하였다. 애플은 엔비디아의 경쟁사들을 포함하여 기술산업 전반이 널리 사용하는 칩 아키텍처의 공급자였으므로, 그동안 중립적 위치에 있던 아키텍처 공급자가 칩 사업자의 소유가 될 경우 어떠한 결과가 생기느냐가 심사의 핵심이었다. 경쟁당국들이 제기한 우려의 핵심은, 결합 후 엔비디아가 자사 경쟁사들이 의존하는 애플의 CPU 지식재산에 대한 접근을 제한하거나 관련 제품 간 상호운용성을 저해하는 방식으로 경쟁사업자에게 영향을 줄 수 있고, 특히 AI 인프라와 관련하여 CPU, 네트워크 인터페이스 컨트롤러(Network Interface Controller), GPU에 대한 접근이 제한될 경우 데이터 전송 효율과 서버 성능에 영향을 미칠 수 있다는 것이었다.⁴⁷⁾ 이에 따라 FTC는 2021년 12월 해당 거래를 저지하기 위한 소송을 제기하였고, EC는 2021년 10월 2단계 심층조사에 착수하였으며, 영국에서는 디지털·문화·미디어·스포츠 부장관이 경쟁 및 국가안보 문제를 이유로 이 거래를 영국 경쟁시장청(Competition and Markets Authority, 이하 'CMA'라 한다)에 회부하여 2단계 심층심사가 이루어졌다. 결국, 엔비디아와 소프트뱅크는 2022년 2월 해당 거래를 포기하였다.⁴⁸⁾

5. 수직결합

AI 칩 영역에서는 칩 설계 사업자가 클라우드, 모델 영역으로 확장하거나 클라우드 사업자가 자체 칩을 설계하는 등 사업자가 공급망의 여러 단계를 겸하는 수직결합이 두드러지며,

46) 연합뉴스, “공정위, 英 반도체 설계회사 Arm 서울사무소 현장조사” (2025. 11.).

47) OECD, 「Competition in Artificial Intelligence Infrastructure」 (2025. 11.), 37면.

AI 칩 영역 내의 여러 층위 간에서도 결합이 이루어지고 있다.

높은 시장점유율과 시장력을 보유한 사업자의 수직통합은 진입장벽을 높일 수 있고, 지배적 사업자가 가치사슬의 여러 단계에 걸친 자신의 지위를 이용하여 경쟁을 자신에게 유리하게 기울이고 다른 사업자의 경쟁을 더 어렵게 만들 위험을 높일 수 있다. AI 인프라와 같은 동태적 시장에서 수직적 봉쇄는 최적 전략이 될 수 있는데, 이는 단기적으로 이윤 손실을 초래할 수 있음에도 불구하고 장래의 시장구조에 영향을 미쳐 통합사업자의 장래 이윤을 증가시키기 때문이다.⁴⁹⁾ 이러한 봉쇄는 앞서 살펴본 차별취급, 결합판매, 접근 제한 등 여러 방식으로 발생할 수 있다.

아키텍처 라이선싱 사업자인 암의 자체 칩 진출이 대표적인 사례이다. 암은 2026년 3월 AI 데이터센터용 자체 칩인 ‘Arm AGI CPU’를 공식 발표하면서, IP 라이선싱을 넘어 완성된 칩을 직접 공급하는 단계로 사업을 확장하였다. 암은 그동안 20년 넘게 여러 반도체 기업에 중립적으로 설계를 공급하여 왔으나, 이제는 직접 칩을 공급함으로써 자사 설계를 라이선스 받아온 기존 고객들과 같은 시장에서 경쟁하게 되었다. 이러한 수직통합에 더하여, 기존 고객들인 라이선시(Licensee)에게 차별취급, 결합판매, 접근 제한 등 봉쇄행위를 실시하고 자사 칩 제품을 우대한다면, AI 칩 영역은 물론 AI 공급망 전체에 광범위한 영향을 미칠 가능성이 있다. 실제로 FTC가 2026년 5월경, 암이 고객의 자체 CPU 개발에 필요한 라이선스를 거절하거나 그 품질을 낮추는 방식으로 관련 반도체 시장을 독점화하려 하는지를 살펴보고 있는 것으로 보도된 바 있다.⁵⁰⁾

한편, 클라우드 인프라를 제공하는 하이퍼스케일러가 자체 칩을 설계하면서 자사 칩을 우대하는 자사우대(self-preferencing) 가능성도 거론된다. 예를 들어, 아마존웹서비스(AWS)는 학습용 ‘트레이니움(Trainium)’과 추론용 ‘인퍼런시아(Inferentia)’를, 구글은 텐서 처리장치(Tensor Processing Unit, TPU)를 각각 자체 설계하여 자사 클라우드에서 제공하고 있는데, 이러한 자체 칩은 대체로 외부 시장에서 별도로 판매되지 않고 해당 사업자의 클라우

48) FTC, ‘FTC Sues to Block \$40 Billion Semiconductor Chip Merger’ (2021, 12.); EC, ‘Mergers: Commission opens in-depth investigation into proposed acquisition of Arm by NVIDIA’, IP/21/5624 (2021, 10.); CMA, NVIDIA/Arm merger inquiry 사건 페이지, <https://www.gov.uk/cma-cases/nvidia-slash-arm-merger-inquiry>.

49) OECD, ‘Competition in Artificial Intelligence Infrastructure’, (2025, 11.), 25-26면.

50) Bloomberg, ‘Arm Holdings to Face US Antitrust Probe Over Chip Tech’ (2026, 5.); Tom’s Hardware, ‘US FTC reportedly launches antitrust probe into Arm following its launch of its own AGI CPU’ (2026, 5.).

드 인프라와 결합되어 제공된다. 하이퍼스케일러가 인프라 제공자이자 칩 설계자라는 이중적 지위를 가지게 되면, 가격구조, 할인 적용, 용량 배분, 자사 소프트웨어와의 통합 등을 통해 자사 칩 기반 서비스를 유리하게 취급할 유인을 가질 가능성이 있다는 분석이 있다.⁵¹⁾ 그 밖에 앞서 살펴본 바와 같이, 칩 공급자와 클라우드, 모델 개발사 사이에 자본과 공급계약이 순환적으로 엮히는 투자구조가 사업자 간 이해관계를 결속시켜 경쟁에 영향을 미칠 수 있다는 점도 수직결합과 관련하여 함께 거론된다.

6. 담합(공동행위)

AI 칩 영역과 관련하여 복수의 사업자가 가격, 산출량 등을 합의하는 담합(공동행위)의 가능성도 지적되고 있다. AI 공급망은 대체로 집중도가 높아 점유율이 비대칭적이며, 빠르게 변화하는 기술과 높은 수준의 혁신을 바탕으로 작동하기 때문에 담합의 위험이 상대적으로 낮다고 분석되나, 공급망 가운데 소수 사업자 사이에 과점적 경쟁이 이루어지는 영역에서는 담합의 위험이 더 높을 가능성이 있다고 평가된다.⁵²⁾ 예를 들어, AI 칩의 성능 확장에 필수적인 HBM 등 메모리 반도체 영역은 소수의 사업자가 공급하는 과점구조를 형성하고 있어 담합의 위험성이 거론된다.

실제로, 메모리 반도체 영역에서는 담합이 문제되어 관련 사업자들이 제재를 받은 사례가 있다. DRAM 카르텔은 1998년부터 2002년 사이 여러 DRAM 칩 제조사가 참여한 국제 가격담합 사건으로, DOJ는 2002년 「셔먼법(Sherman Antitrust Act)」에 근거하여 조사에 착수하였고, 2003년부터 2006년 사이 여러 임원과 기업이 유죄를 인정하였다. 예컨대, 인피니언(Infineon)은 2004년 1억 6,000만 달러, 하이닉스는 2005년 1억 8,500만 달러, 삼성전자는 2005년 3억 달러의 벌금을 각각 부과받았다. 2010년 EC도 카르텔 가담을 이유로 9개 기업에 총 3억 3,100만 유로의 과징금을 부과하였으며, 마이크론 테크놀로지는 자진신고로 면제를 받았다. DRAM 시장이 성숙하고 범용화되면서 칩이 상호 대체 가능해지고 산업 표준이 널리 준수된 가운데, 이러한 시장구조가 활발한 정보 공유와 결합되어 담합 위험을 높

51) AdC, 「Competition and Generative AI: Access to AI Chips」 (2026, 2.), 8면.

52) AdC, 「Competition and Generative AI: Access to AI Chips」 (2026, 2.), 4-5면; OECD, 「Competition in Artificial Intelligence Infrastructure」 (2025, 11.), 14, 23, 33면.

였다고 분석된다.⁵³⁾

2026년 6월 미국 캘리포니아 북부연방지방법원에는 삼성전자, SK하이닉스, 마이크론 테크놀로지를 상대로 한 집단소송이 제기되었다. 원고 측은 위 사업자들이 범용 DRAM(DDR3, DDR4 등)의 공급을 축소하여 4년 간 가격을 약 700% 상승시켰다고 주장한 것으로 보도된 바 있다.⁵⁴⁾ 다만, 과점 시장에서 사업자들의 행위가 나란히 나타난다는 사정만으로 담합의 합의가 인정되는 것은 아니다. 2018년 위 3개 사업자를 상대로 유사한 집단소송이 제기된 바 있으나 연방지방법원은 2020년 이를 기각하였고, 제9연방항소법원도 2022년 3개 사업자의 행위가 위법한 합의보다는 조율되지 않은 적법한 시장행위로 설명될 가능성이 더 높다고 보아 기각 판결을 유지하였다.⁵⁵⁾

IV. 결어

이상에서 살펴본 바와 같이, AI 칩 영역에서는 다층적 구조, 시장집중, 공급 부족, 록인 효과 등 구조적 특성을 배경으로 차별취급, 결합판매, 상호호환성 제한, 거래 거절 및 접근 제한 등 단독행위와 이러한 논의를 구조적으로 강화할 수 있는 수직결합 그리고 담합 등 다양한 경쟁법적 이슈가 논의되고 있다. 본고에서 소개한 사안들은 상당수가 아직 조사 또는 문제 제기 단계에 있으며, 앞으로도 AI 칩 영역에 대한 각국 경쟁당국의 연구와 집행은 계속 전개될 것으로 전망된다.

이어지는 제3부에서는 AI 공급망의 다음 단계인 클라우드 인프라(Cloud Infrastructure) 및 파운데이션 모델 영역에서 논의되는 경쟁법적 이슈들을 살펴본다. 

53) OECD, 「Competition in Artificial Intelligence Infrastructure」 (2025. 11.), 33-34면.

54) Tom's Hardware, 「Samsung, SK hynix, and Micron sued over alleged DRAM price fixing amid record memory costs」 (2026. 6.); Notebookcheck, 「Samsung, Micron, SK Hynix sued over DRAM price-fixing triggering a 'RAMpocalypse' as memory prices soar」 (2026. 7.).

55) In re Dynamic Random Access Memory (DRAM) Indirect Purchaser Antitrust Litigation, No. 4:18-cv-02518 (N.D. Cal. Dec. 21, 2020), aff'd, 28 F.4th 42 (9th Cir. 2022).