

2 오픈웨이트 AI, 가중치 공개에서 생태계 주도권 경쟁으로

→ 기업 채택 확대가 모델 밖의 수익화와 이용 경로 경쟁을 촉발

- 성능·비용·데이터 통제 이점을 바탕으로 기업의 실질적 선택지로 부상
 - 오픈웨이트는 AI 모델이 학습한 가중치를 내려받아 실행·수정할 수 있도록 공개한 방식으로, 폐쇄형보다 이용자 권한이 넓지만, 소스코드와 학습 데이터 등 개발 과정의 공개 범위는 모델마다 차이
 - '26년 들어 중국 개발사의 오픈웨이트 모델이 코딩과 에이전트 등 일부 업무에서 주요 폐쇄형 모델과의 성능 격차를 줄이면서, 기업도 연구용을 넘어 실제 업무에 적용할 대안으로 검토되기 시작
 - 낮은 이용료에 더해 실행 환경과 데이터 처리 방식을 기업이 직접 결정할 수 있다는 이점까지 더해지면서, 오픈웨이트는 연구와 실험을 넘어 실제 업무에 활용할 수 있는 선택지로 확대
- 모델 이용이 늘면서 수익과 통제권을 둘러싼 생태계 경쟁 본격화
 - 오픈웨이트는 이용자가 공개된 가중치를 직접 실행할 수 있어, 수익의 발생 지점도 가중치 자체에서 공식 추론 서비스와 모델 중개, 실행 인프라와 산업별 응용서비스로 확대
 - 실제로 '26년 8~9월 시장에서는 모델의 거래와 유통 관문을 확보하려는 플랫폼 인수가 이어진 가운데, 국가 차원에서도 특정 모델의 조달과 거래를 제한하려는 규제 검토가 가시화
 - 민간기업은 모델의 주요 유통·실행 경로를 확보하고 국가는 특정국 모델의 시장 진입을 제한하면서, 경쟁의 중심도 개별 모델의 성능에서 이용자 접근과 시장 접근의 결정권으로 이동

→ 오픈웨이트 확산이 바꾸는 AI 산업의 수익 구조와 경쟁 지형

1. (확산 배경) 성능과 비용 및 데이터 통제 이점으로 기업의 실질적 선택지로 진입

- 성능 격차 축소와 낮은 이용료로 적용 가능한 업무 확대
 - 일부 오픈웨이트 모델이 주요 폐쇄형과의 성능 격차를 줄이고 이용료도 낮아지면서, 기업이 요구 성능과 비용을 함께 고려해 업무 적용 가능성을 검토할 수 있는 여건 마련
 - 실제로 미국 AI 표준혁신센터(CAISI)는 DeepSeek V4 Pro와 자국 선도 모델의 성능 격차를 약 8개월로 평가해, 선도 오픈웨이트 모델이 폐쇄형 모델을 수개월의 시차로 추격하고 있다고 분석

- 오픈웨이트는 주요 폐쇄형 모델보다 이용료가 수십 분의 1 수준으로 낮으며, '26년 7월 AI 게이트웨이 전체 토큰의 36%를 처리하고도 지출 비중은 8.6%에 그쳐 높은 비용 효율을 확인 (Vercel, '26.8.)
- 자체 실행으로 데이터 통제권을 확보하되 운영 책임은 기업으로 이전
 - 기업은 오픈웨이트를 자체 관리 인프라에서 실행해 민감정보와 내부 업무 데이터를 외부 사업자에게 전송하지 않고, 저장 위치와 접근 권한, 모델의 실행·운영 방식까지 자체 기준에 따라 설정 가능
 - 산업별 데이터와 업무 규칙에 맞게 모델을 미세 조정하고 추론 방식을 변경할 수 있어, 자사 지식과 운영 경험을 외부 서비스에 종속되지 않는 내부 AI 자산으로 지속적 축적 가능
 - 다만 자체 배포 시 보안·접근통제·규제 준수와 결과물 오류의 책임이 도입 기업으로 이전되므로, 비용 절감뿐 아니라 이를 지속적으로 관리할 전문 인력과 운영체계를 함께 확보할 필요
- 업무 난이도와 위험 수준에 따라 오픈웨이트와 폐쇄형 모델을 조합
 - 데이터와 운영을 직접 결정할 수 있다는 이점에도 기업은 모든 업무를 오픈웨이트로 전환하지 않고, 업무 성격에 따라 두 모델을 나누어 쓰면서 기업 단위가 아닌 개별 업무별로 도입 여부를 판단
 - 오류 허용도가 높은 대량 반복 업무에는 요구 성능을 충족하는 저가 모델을 적용하고, 오류가 손실로 이어지는 코딩이나 백오피스 에이전트에는 비용이 높더라도 성능이 검증된 모델을 선택
 - 실제로 코딩 업무에서는 DeepSeek가 전체 토큰의 약 1/3을 처리한 반면, 관련 지출의 80% 이상은 Anthropic 모델에 집중돼 업무량과 지출을 서로 다른 모델군이 주도 (Vercel, '26.8)

2. (수익화) 오픈웨이트를 기반으로 모델 밖의 활용 계층에서 수익 창출

- 공식 서비스와 외부 유통을 통해 모델 개발사의 수익 기반 형성
 - 오픈웨이트 개발사는 가중치를 공개하는 대신 안정적인 추론과 지속적인 지원을 원하는 기업 수요를 공식 API·추론 서비스로 유도하고, 맞춤형과 기술지원 계약을 결합해 수익 창출
 - '26년 7월 오픈웨이트 모델 관련 지출은 전월보다 두 배 이상 증가, 증가분의 90% 이상이 Moonshot과 Z.ai가 직접 운영하는 API 서비스에서 발생해 이용 수요가 개발사가 직접 제공하는 서비스에 집중 (Vercel, '26.8.)



- 외부 클라우드 기업이 공개 모델을 자사 서버에서 운영해 수익을 내더라도 개발사에는 자동으로 배분되지 않아, 개발사는 라이선스와 매출 배분 계약을 통해 수익 일부를 확보하려는 움직임
- 모델 중개는 복수 모델의 연결·배분·관리를 통해 수익 창출
 - 기업의 복수 모델 이용이 늘면서 중개 플랫폼은 여러 개발사의 모델을 하나의 창구에서 비교·호출할 수 있도록 제공하고, 이용량 수수료와 결제·사용량 관리 기능을 결합해 수익 확보
 - 중개 플랫폼은 업무 난이도와 가격, 처리 속도, 서비스 가용성에 따라 기업의 요청을 적합한 모델과 추론 사업자에 배분하고, 장애가 발생하면 대체 모델로 자동 전환해 운영 부담을 완화
 - 이에 기업은 개발사별 계약과 기술 연동을 거치지 않고 여러 모델을 이용할 수 있어, 모델 중개가 개발사의 모델을 이용 기업에 연결하고 호출과 결제를 처리하는 독립적인 사업 계층으로 형성
- 클라우드·가속기는 오픈웨이트 실행 수요를 인프라 매출로 전환
 - 클라우드 기업은 오픈웨이트를 관리형 추론 서비스로 제공해 컴퓨팅 사용료를 확보하고, 반도체 기업은 모델 실행에 필요한 가속기와 소프트웨어를 결합해 오픈웨이트 확산을 인프라 매출로 전환
 - 이에 오픈웨이트 확산은 모델 개발사의 수익만 늘리는 것이 아니라, 이용자가 모델을 내려받아 조정·배포·실행하는 전 과정에서 클라우드와 가속기 사업자의 사업 기회를 함께 확대
 - 특히 공개 모델은 인프라 기업의 고객층을 소수 선도 개발사·대형 클라우드에서 다수의 기업·개발자로 확대해, 자체 칩을 개발하는 대형 고객에 편중된 매출 기반을 다변화하는 효과 (더밀크, '26.8.)
- 산업별 응용서비스는 범용 모델을 업무 성과로 전환하며 부가가치 확보
 - 응용서비스 기업은 공개된 가중치에 산업 데이터와 업무 절차, 사용자 인터페이스를 결합해 범용 모델을 특정 업무를 수행하는 서비스로 구현하고, 차별화된 부가가치를 창출
 - 무 처리 건수나 결과를 기준으로 요금을 받는 경우, 응용서비스 기업은 동일한 성과를 내는 더 저렴한 오픈웨이트 모델로 교체해 운영비를 줄이고 수익성을 높일 수 있음
 - 결국 오픈웨이트의 수익 구조는 모델 자체 판매에 의존하는 방식에서 벗어나, 공식 서비스·중개·인프라·응용 계층이 이용 과정의 서로 다른 지점에서 수익을 확보하는 다층 구조로 형성

3. (플랫폼화) 모델의 선택·유통·실행 경로를 장악하려는 경쟁 본격화

● (Stripe) OpenRouter 인수로 모델 배분·결제 관문 확보

- 모델 이용 과정의 수익 창출 지점을 확보하려는 경쟁이 본격화되면서, 美 결제 기업 Stripe는 '26년 8월 모델 중개 플랫폼 OpenRouter를 70억 달러 이상에 인수하기로 합의 (Fortune, '26.8.)
- Stripe는 OpenRouter의 모델 비교·배분 기능에 자사의 결제·비용 관리를 결합해 모델 선택과 호출, 대금 정산을 하나의 경로로 통합하고 모델별 거래와 이용량, 지출 데이터를 자산으로 축적
- 이번 인수로 Stripe는 특정 모델을 직접 소유하지 않고도 이용자 접점과 유통 경로를 확보해, 개별 모델의 성패와 관계없이 오픈웨이트 이용 확대를 지속적인 플랫폼 수익으로 전환 가능

● (NVIDIA) Hugging Face 인수로 개발자 접점과 모델 배포 생태계 확보

- 모델 배분·결제 경로에 이어, 미국 NVIDIA는 '26년 9월 세계 최대 규모의 오픈모델 허브 Hugging Face 인수 합의를 발표하며 경쟁 범위를 개발자 생태계로 확대 (NVIDIA, '26.9.)
- 개발자·연구자 등 1,800만 명이 이용하는 Hugging Face는 모델 탐색·시험·공유·배포를 지원, '26년 2월 llama.cpp 개발팀 합류로 개인 PC나 기업 내부 서버에서 모델을 직접 실행하는 기능까지 확대
- NVIDIA는 이번 인수를 통해 Hugging Face의 개발자 접점과 모델 유통망을 확보하고, 모델 선택·배포 과정에서 발생하는 연산 수요를 자사 컴퓨팅·소프트웨어 생태계와 연결할 기반 마련

● 플랫폼 결합은 이용 편의를 높이는 동시에 유통 관문의 집중 가능성 확대

- 모델 중개와 결제, 모델 허브와 컴퓨팅이 각각 결합되면 기업과 개발자의 모델 탐색, 연동, 정산 및 배포 부담은 줄어들지만, 주요 이용 경로가 소수 플랫폼에 집중될 가능성도 확대
- 이용자가 집중된 플랫폼은 모델 선택과 이용량, 실행 비용 데이터를 축적하고, 이를 유망 모델의 조기 발굴과 투자, 가격 책정 및 인프라 공급 전략에 활용해 영향력 강화 가능
- 따라서 기업은 단기적인 이용 편의뿐 아니라 데이터 이전과 다른 플랫폼과의 호환 여부, 계약 종료 후 이용·운영 기록의 회수 조건까지 점검해 특정 사업자에 대한 종속 위험을 관리할 필요

4. (통제 경쟁) 라이선스와 국가 정책이 모델의 실질적 시장 접근을 결정

- 오픈웨이트 모델의 상업적 이용 범위는 개발사의 라이선스가 규정
 - 오픈웨이트는 가중치를 내려받을 수 있다는 기술적 공개를 의미할 뿐, 상업적 이용과 수정·재배포, 대규모 서비스 제공 권한까지 자동으로 보장하는 것은 아니며 허용 범위는 모델별로 차이
 - 일부 개발사는 기업의 내부 이용과 제품 탑재는 무상으로 허용하되, 일정 규모 이상의 모델 재판매·추론 서비스에는 별도 계약을 요구하는 방식으로 이용 형태별 조건을 구분
 - 이에 기업은 모델 성능과 가격뿐 아니라 이용자 규모와 허용 용도, 수정·재배포 조건, 위반 시 책임을 함께 검토하고, 라이선스 변경이 사업 지속성에 미치는 영향도 함께 관리할 필요
- 국가의 조달·거래·호스팅 제한이 특정 모델의 시장 진입을 통제
 - 미국은 자국 오픈웨이트 생태계에 대한 지원을 확대하는 한편, 중국산 모델의 정부 조달과 거래·호스팅 제한을 검토하며 개발국과 보안 위험을 시장접근의 판단 기준으로 적용하려는 움직임
 - 국가 차원의 제한은 모델의 다운로드 가능 여부와 관계없이 공공 조달과 클라우드 제공, 기업 거래를 차단할 수 있어, 기술적으로 공개된 모델의 실제 이용 범위를 축소하는 통제 수단으로 작용
 - 이와 같은 제한은 국가별로 이용 가능한 모델과 클라우드 제공 경로를 달리해 오픈웨이트 시장을 규제권역별로 분절시키고, 개발사·플랫폼의 유통 전략에도 영향을 미칠 전망
- 기술적 공개와 실제 이용 가능성 간의 차이를 기준으로 도입 판단
 - 동일한 오픈웨이트 모델이라도 기업의 이용 목적과 서비스 규모, 적용 국가에 따라 허용 범위가 달라질 수 있어, 가중치 공개만으로 장기적인 활용 가능성을 판단하기는 어려움
 - 기업은 도입 당시의 조건을 확인하는 데 그치지 않고, 개발사의 라이선스 개정과 국가별 규제 변화가 기존 서비스의 운영·확장·해외 진출에 미치는 영향을 지속적으로 관리할 필요
 - 따라서 오픈웨이트의 실질적 개방성은 다운로드 가능 여부보다 기업이 모델을 안정적으로 이용·수정·배포하고 사업을 지속할 수 있는 권리가 보장되는지를 기준으로 평가할 필요

5. (국내 과제) 모델 개발을 넘어 오픈웨이트 생태계의 실질적 선택권 확보

- 국내외 모델의 라이선스·성능·보안을 검증하는 신뢰 기반 마련
 - 라이선스와 국가 정책이 모델의 실제 이용 범위를 좌우하는 만큼 국내에서도 기업이 공개 모델의 도입 가능성을 판단할 근거가 필요하며, 자체 구축 확대로 관련 검증 수요도 증가 전망 (삼성SDS, '26.5.)
 - 국내에서는 AI안전연구소의 안전성 평가와 보안 당국의 지침이 각각 운영되고 있으나, 기업 도입 관점에서 라이선스·성능·보안을 함께 판별할 수 있는 통합 검증 체계는 아직 미비
 - 이에 국내외 공개 모델을 공통 기준에 따라 등재·검증하고 결과를 기업의 모델 비교·선택에도 활용하도록 지원해, 특정 개발사나 해외 플랫폼에 의존하지 않는 신뢰 기반을 마련할 필요
- 주요 오픈웨이트 모델을 국산 AI 반도체에서 실행하는 최적화 기술 강화
 - 모델 종속이 플랫폼과 실행 경로로 옮겨가고 中 모델이 자국 칩 중심으로 최적화되는 흐름에서, 주요 공개 모델을 국산 가속기에서 효율적으로 실행하는 기술은 선택권 확보의 핵심 기반
 - 국내에서는 '26년 7월 국산 NPU 서버 한 대로 500B급 국산 모델을 구동한 실증과 포털 실서비스에 국산 모델·NPU 결합을 적용한 사례가 나와, 국산 모델 실행 기반은 형성 단계
 - 향후 국산 모델 중심의 실증을 넘어, 기업이 실제 널리 쓰는 해외 공개 모델까지 국산 가속기에서 실행하는 변환·최적화 기술과 실증 사례를 늘리는 것이 실행 경로 확보의 남은 과제
- 공개 건수보다 산업별 적용·기업 채택·운영비 절감 등 실질적 활용 성과를 중시
 - Hugging Face의 관심도와 다운로드 수 상위 25개 모델 중 중복 모델은 1개에 그쳤으며, 미국 기업의 공개 모델 플랫폼 이용률도 6.1%로 나타나 공개 건수·순위와 실제 활용에는 차이 (HF, '26.8.)
 - 국내 기업이 자체 구축을 선택한 이유도 데이터 유출 방지(41%)와 역량 내재화(40%)로 조사되어, 활용 성과는 데이터 통제·운영비·산업 적용으로 측정할 필요 (삼성SDS, '26.5.)
 - 결국 국내 대응은 모델 추가 개발에 더해 검증·최적화·배포·사업화 역량으로 범위를 넓히고, 성과를 공개 건수가 아닌 산업별 적용과 기업 채택, 운영비 절감으로 측정하는 것이 필요한 시점



출처 : Hugging Face 외(2026.8.)

<https://huggingface.co/blog/state-of-open-models-summer-2026>
<https://www.cooley.com/news/insight/2026/2026-08-12-unlocking-the-weight-s-what-enterprises-should-know-before-deploying-open-weight-ai-models>
<https://www.linkedin.com/pulse/what-open-weight-ai-can-learn-from-open-source-jim-whitehurst-dqcec>
<https://thecuberresearch.com/state-of-open-models-in-2026/>
<https://www.reuters.com/world/middle-east/abu-dhabi-ai-institute-releases-fully-open-source-models-with-training-data-code-2026-09-03/>
<https://www.csis.org/analysis/what-know-about-chinese-ai-models>
<https://www.brookings.edu/articles/why-open-weight-models-are-crucial-for-american-ai-leadership/>
<https://www.microsoft.com/en-us/corporate-responsibility/topics/open-weight/>
<https://www.weforum.org/stories/artificial-intelligence/open-weight-ai-the-difference-between-renting-and-owning-tech-sovereignty/>
<https://www.themiilk.com/articles/a8cc74b7b?u=9a5578ca&t=a082ad39d&from=https://vercel.com/blog/ai-gateway-production-index-july-2026>
<https://vercel.com/blog/deepseek-overtakes-google-on-volume-cost-per-token-falls>
<https://www.reuters.com/technology/artificial-intelligence/american-ai-model-makers-smell-an-opportunity-2026-08-12/>
<https://techcrunch.com/2026/08/28/open-weight-ai-companies-are-the-valleys-hottest-acquisition-targets/>
<https://www.reuters.com/business/retail-consumer/chinas-moonshot-talks-with-microsoft-amazon-google-over-k3-revenue-sharing-2026-08-26/>
<https://stripe.com/en-pl/newsroom/news/stripe-agrees-to-acquire-openrouter>
<https://fortune.com/2026/08/16/stripe-7-billion-deal-ai-firm-openrouter-acquisition/>
<https://blogs.nvidia.com/blog/nvidia-to-acquire-hugging-face/>