Maxence Foulon

무료 자료 · 체크리스트

AI 플랜 B: 서비스 중단, 가격 인상, 시장 폭락에도 AI 프로젝트가 버티는 8가지 안전장치

15분이면 AI 공급업체가 내일 당장 서비스를 끊는다면 회사에서 무엇이 멈추는지 파악할 수 있습니다. 하루면 그런 일이 생겨도 회사가 멈추지 않을 계획을 세울 수 있습니다.

읽는 시간 10분 검증된 출처 16개 2026년 11월 7일 업데이트

30초 요약

- Anthropic은 9개월 동안 1,040억 달러의 손실을 냈습니다. 대부분은 회계상 손실이지만, 일부 투자자들은 11월 상장이 성사될지 의문을 품고 있습니다. 11월 5일 Nasdaq은 2025년 4월 이후 최악의 거래일을 기록했습니다.[1, 2, 3]
- 고객사 입장에서 실제 위험은 의존도입니다. 2026년 6월, 미국 정부의 결정에 따라 모든 고객을 대상으로 Claude Fable 5 제공이 3주 가까이 중단되었습니다.[4, 5, 6]
- 해법은 8가지 안전장치로 정리됩니다. 첫 번째인 AI 사용 현황 파악은 15분이면 끝납니다. 백업용으로는 Mistral이 10월 29일 가중치를 공개한 유럽산 모델 Mistral Large 4가 있습니다. 최상위 모델에는 한참 못 미치지만, 직접 내려받을 수 있습니다.[7, 8]

배경

'버블'이라는 말의 실체

2026년 10월, 세 가지 신호가 한꺼번에 겹쳤습니다. GPU 임대 업체(이른바 '네오클라우드') 가운데 가장 잘 알려진 CoreWeave의 채권은 시장에서 15%에 가까운 수익률로 거래됐고, 주가는 10월 1일부터 11월 5일 사이에 절반으로 떨어졌습니다.[9, 10] 일론 머스크는 신용 경색으로 200곳이 넘는 네오클라우드 대부분이 사라지거나 인수될 것이라고 예측했습니다.[11] 그리고 Anthropic은 6개월 동안 807억 달러의 순손실을 공개하며 IPO(기업공개)를 신청했습니다.[4]

이 마지막 숫자는 무섭지만, 잘못 읽기 쉽습니다. 이 가운데 724억 달러는 전환사채를 중심으로 한 금융 상품의 재평가에서 나왔습니다. 회사 가치가 오를수록 이 상품들이 장부상 부채로 더 크게 잡히지만, 실제로 빠져나가는 돈은 1달러도 없습니다. 영업손실은 6억 9,300만 달러였고, 매출은 163억 달러였습니다.[4]

11월 초에는 긴장이 한 단계 더 높아졌습니다. Anthropic이 제출한 증권신고서 정정본(S-1/A)에는 3분기 순손실 233억 달러가 추가되었습니다. 이번에도 대부분 회계상 손실(재평가 219억 달러)이며, 매출은 170억 달러였습니다.[1] 미국 중간선거 이틀 뒤인 11월 5일, Nasdaq은 2025년 4월 이후 최악의 거래일(−5.5%)을 기록했고, 잠재 투자자들은 Reuters에 Anthropic의 11월 IPO가 성사될지 의문이라고 말했습니다.[2, 3]

진짜 부담

바로 컴퓨팅 약정입니다. Anthropic은 약 8,950억 달러 규모의 구매·임대 계약을 맺었고, 거의 전부가 컴퓨팅과 인프라입니다. 이 가운데 4,475억 달러가 2030년 말 이전에 해당하고, 나머지는 2045년까지 이어집니다.[1] 지금 걸려 있는 것은 이 베팅, 그리고 GPU에 자금을 대는 공급망 전체의 베팅입니다.

남의 일이 아닌 이유

여러분은 CoreWeave나 Nvidia에 돈을 내지 않습니다. 하지만 여러분의 AI 공급업체는 바로 이 공급망에서 컴퓨팅을 빌려 쓰고, 미국 법의 적용을 받습니다. 2026년 6월, 미국 상무부는 Fable 5와 Mythos 5에 수출 제한 조치를 내렸습니다. 고객을 국적별로 가려낼 방법이 없었던 Anthropic은 Fable 5 출시 이틀 뒤인 6월 12일, 모든 고객을 대상으로 두 모델의 제공을 중단했습니다. 7월 1일부터 다시 이용할 수 있게 되었습니다.[4, 5, 6] 그보다 앞선 2월 27일에는 미국 대통령이 연방 정부에 Anthropic 모델의 사용 중단을 지시했습니다.[1] S-1(증권신고서)은 이런 수출 제한이 다시 일어나거나, 더 오래 이어지거나, 다른 모델로 확대될 수 있다고 경고합니다.[4]

01

AI 사용 현황 파악 (15분)

무엇을 고르기 전에, 먼저 어디에 의존하고 있는지 파악하세요. 팀에서 AI를 쓰는 곳을 하나도 빠짐없이 적으세요. 재무팀이 쓰는 브라우저 확장 프로그램, 영업 담당자의 AI 어시스턴트, 인턴이 짠 스크립트처럼 아무도 보고하지 않은 사용처도 포함해야 합니다.

- [ ] 사용처마다 누가, 어떤 업무에, 어떤 도구와 어떤 모델로 쓰는지 적으세요.
- [ ] 일주일 동안 쓸 수 없게 되면 어떻게 되는지 적으세요: 영향 없음, 불편, 업무 중단 중 하나로.

- [ ] 데이터가 어디로 가는지 확인하세요: 어느 공급업체, 어느 국가인지. 모르면 '문의 필요'라고 적어 두세요.

AI 의존도 점검표 (스프레드시트에 붙여 넣기)

```
용도 | 팀 | 도구 | 모델 | 공급업체 | 데이터 위치(국가) | 1주일 중단 시 | 백업 테스트 여부 | 담당자
예시: 고객 문의 답변 | 고객지원팀 | 티켓 관리 도구 + API | (모델) | (공급업체) | 문의 필요 | 업무 중단 | 안 함 | (이름)
```

02

백업 모델, 이름만 정하지 말고 테스트까지

Fable 5가 중단된 동안에도 Opus 4.8을 비롯한 Anthropic의 다른 모델은 계속 쓸 수 있었습니다.[5] 이미 다른 모델로 업무를 돌려 본 팀은 설정 하나만 바꿔 전환할 수 있었습니다. 그렇지 않은 팀은 실제 운영 환경에서야 차이를 발견했습니다. Fable 5는 출시된 지 얼마 안 된 모델이었지만, 이 원리는 어떤 모델에나 똑같이 적용됩니다.

- [ ] 백업을 두 개 정하세요. 하나는 같은 공급업체의 다른 모델(특정 모델만 중단될 때 대비), 다른 하나는 다른 공급업체의 모델(더 넓은 장애, 가격 인상, 정부 결정에 대비)입니다.
- [ ] 두 모델 모두에 테스트 세트(04 참고)를 돌리고 차이를 기록하세요.
- [ ] 민감한 데이터를 위해 국내나 유럽, 또는 자체 서버에서 운영하는 선택지를 하나 남겨 두세요.

03

코드 속 AI 호출 창구는 하나로

모델 이름이 40개 파일에 하드코딩되어 있으면 공급업체를 바꾸는 데 몇 주가 걸립니다. 모든 호출이 내부 함수 하나를 거치고 모델을 설정에서 지정한다면, 기술적인 전환은 한 시간이면 됩니다. 결과는 테스트 세트(04 참고)로 확인하세요.

- [ ] 모든 AI 호출이 내부 모듈 하나를 거치게 하세요.
- [ ] 주 모델과 백업 모델은 코드가 아니라 설정에 두세요.

- [ ] 프롬프트는 따로 저장해 버전을 관리하고, 어떤 모델로 테스트했는지 함께 기록하세요.
- [ ] 재배포 없이 전환할 수 있는 스위치(기능 플래그)를 마련하세요.
- [ ] 자동 전환이 데이터를 보내기 전에, 백업 공급업체가 우리 데이터를 다뤄도 되는지(계약, 국가, 보안) 미리 승인해 두세요.

설정 템플릿 (상황에 맞게 수정)

```
ai:
  primary:
    provider: "입력 필요"
    model: "입력 필요"
  backup_same_provider:
    model: "입력 필요"
  backup_other_provider:
    provider: "입력 필요"
    model: "입력 필요"
  failover:
    automatic_if: "10분 동안 오류율 5% 초과"
    failover_owner: "담당자 이름"
  budget:
    monthly_cap_usd: "입력 필요"
    alert_at: "80%"
```

04

우리 회사만의 테스트 세트

AI 업체는 자사 모델이 돋보이는 벤치마크를 골라 공개합니다. Kimi 개발사인 중국 Moonshot은 Kimi K3.1 Pro가 Terminal-Bench 4.0에서 Claude Opus 5.5에 0.7점(65.7 대 66.4)밖에 뒤지지 않으면서 가격은 절반이라고 소개합니다. 입력은 100만 토큰당 2달러 대 4달러, 출력은 10달러 대 20달러입니다.[12, 13, 14] 하지만 이 숫자는 여러분의 계약서, 문의 티켓, 브랜드 말투에 대해 아무것도 알려 주지 않습니다. 그건 실제 사례 20건이 알려 줍니다.

- [ ] 익명 처리한 실제 요청 20건을 기대하는 답변과 함께 모으세요.
- [ ] 함정 사례 5건을 추가하세요: 정보 누락, 범위를 벗어난 요청, 답변에 되풀이하면 안 되는 개인정보.

- [ ] 모델이나 가격이 바뀔 때마다 모든 후보 모델을 같은 기준표로 채점하세요.

프롬프트: 실제 사례를 테스트 세트로 바꾸기

```
아래는 우리 팀이 실제로 받은 요청 20건입니다. 익명 처리했고, 우리가 좋다고 판단한 답변을 함께 적었습니다.
요청마다:
1. 답변에 반드시 들어가야 할 내용을 나열하세요(확인 가능한 기준 최대 3개).
2. 어떤 경우 받아들일 수 없는 답변이 되는지 적으세요(오류, 말투, 지어낸 정보).
3. 기준별로 0~2점 배점을 제안하세요.
그다음 이 요청들을 바탕으로 함정 사례 5건을 추가하세요: 정보 누락, 범위를 벗어난 요청, 답변에 되풀이하면 안 되는 개인정보.
결과는 모두 표로 정리하세요: 사례 | 입력 | 기준 | 불합격 조건 | 배점.

[여기에 요청 20건을 붙여 넣으세요]
```

05

계약 전에 꼭 물어야 할 질문

가격 인상이나 모델 서비스 종료에 맞서 회사를 법적으로 지켜 줄 수 있는 것은 계약서뿐입니다. Anthropic도 S-1에서 성과 기반 요금이나 작업당 요금 같은 가격 체계 변경 가능성을 직접 언급합니다.[4] 저는 법률 전문가가 아니므로, 조항 검토는 담당 변호사에게 맡기세요. 아래는 공급업체에 보낼 질문입니다.

AI 공급업체에 보낼 이메일

```
안녕하세요, (회사명) (이름)입니다.

계약 체결이나 갱신에 앞서 아래 사항을 서면으로 확인해 주실 수 있을지 문의드립니다.
1. 가격 인상 시 사전 통지 기간과 인상 상한 여부
2. 당사가 사용 중인 모델의 서비스 종료 전 사전 통지 기간과 마이그레이션 기간
3. 당사 데이터, 프롬프트, 로그를 내보내는 방법(형식 및 소요 기간)
4. 당사 데이터가 처리·저장되는 국가와 해당 재위탁 업체
5. 수출 제한이나 정부 결정으로 귀사가 모델 제공을 중단해야 하는 경우 당사에 적용되는 조치
(사전 통지, 대체 모델, 환불)
6. 서비스 가용성 보장 수준과 미달 시 제공되는 보상

바쁘신 와중에 확인 부탁드립니다.
감사합니다.

(이름) 드림
```

06

가격 인상을 견딜 수 있는 예산

가격은 계속 바뀝니다. 2026년 9월, Anthropic은 Fable 5.1로 일반적인 작업 비용을 약 25%, Opus 5.5로 작업당 비용을 약 40% 낮췄습니다.[4] 하지만 이 인하는 공급업체가 수천억 달러 규모의 컴퓨팅을 약정하고, 이들에게 GPU를 빌려주는 업체는 높은 금리로 돈을 빌리는 시장에서 이뤄지고 있습니다.[1, 9] 인하가 계속된다는 보장은 없습니다.

- [ ] 월 청구액만 보지 말고 작업당 비용(처리한 티켓당, 작성한 문서당)을 측정하세요.
- [ ] 가격이 두 배가 되면 프로젝트가 어떻게 되는지 계산해 보세요. 버티지 못한다면, 업무를 감당할 수 있는 더 저렴한 모델을 찾아 두세요.
- [ ] 모든 공급업체에 지출 한도와 알림을 설정하세요.

07

오픈 모델은 라이선스부터 확인

직접 운영하거나 국내·유럽 호스팅 업체에 맡기는 오픈 웨이트(가중치 공개) 모델은 한 번 내려받고 나면 개발사의 결정에 좌우되지 않습니다. 하지만 '오픈'이 '아무 조건 없음'을 뜻하지는 않습니다. 예를 들어 Kimi K3.1 Pro 라이선스는 12개월 매출이 2,000만 달러를 넘는 규모로 이 모델을 서비스 형태로 재판매하는 경우 별도 계약을 요구합니다. 내부 사용은 예외입니다.[15]

유럽 쪽에서는 Mistral이 10월 29일 수정된 MIT 라이선스로 Mistral Large 4의 가중치를 공개했습니다. 상업적 사용이 허용되며, 모델 접근 권한을 재판매하는 사업을 운영할 때만 별도 라이선스가 필요합니다.[7, 16] 성능은 최상위 모델에 한참 못 미칩니다. 11월 7일 기준 독립 평가 기관 Artificial Analysis의 지수는 26점으로, Kimi K3.1 Pro(46점)나 미국 최상위 모델들(57~60점)과 차이가 큽니다.[8] 파라미터가 약 1조 개에 달해, 현실적으로는 중소기업 서버가 아니라 호스팅 업체에서 돌려야 합니다.[16] 백업 모델이 최고일 필요는 없습니다. 핵심 업무를 감당할 수 있으면 되고, 감당할 수 있는지는 테스트 세트로만 확인할 수 있습니다.

- [ ] 성능이 조금 떨어지더라도 핵심 업무를 감당할 수 있는 오픈 모델을 찾아 두세요.
- [ ] 라이선스를 읽어 보세요: 상업적 사용, 재판매, 필수 출처 표기.
- [ ] 테스트 세트로 돌려 보고 주 모델과의 차이를 기록하세요.

08

문서로 남기고 분기마다 테스트하는 중단 대응 계획

쓰던 모델이 사라지는 날, 누가 결정하는지 묻는 사람이 있어서는 안 됩니다. 한 페이지면 충분합니다.

AI 중단 대응 계획 (한 페이지)

```
발동 조건: 주 모델을 쓸 수 없거나, 사용이 제한되거나, 가격이 ___% 넘게 오른 경우.
결정권자: ___ (부재 시 대행: ___).
결정 시한: ___시간 이내.
전환: ___(같은 공급업체)로 전환하고, 필요하면 ___(다른 공급업체)로 전환.
전환 중 멈추는 업무: ___.
팀 공지: "___ 어시스턴트는 ___까지 다른 모델로 운영됩니다. 이상한 답변이 나오면 ___
에게 알려 주세요."
테스트: 분기에 한 번, 하루 동안(또는 트래픽 일부를) 백업 모델로 운영하고 결과를 비교.
```

- [ ] 이 한 페이지를 작성해 관련 팀과 공유하세요.
- [ ] 한 달 안에 첫 전환 테스트 일정을 잡으세요.

출처

1. SEC EDGAR, Anthropic, PBC, Form S-1/A https://www.sec.gov/Archives/edgar/data/2130357/000162828026068676/anthropicpbc-sx1ax1.htm, 2026년 11월 2일.
2. Reuters, via Investing.com, S&P 500 futures steady after midterm rally; Nasdaq lags on chip weakness https://www.investing.com/news/stock-market-news/sp-500-futures-steady-after-midterm-rally-nasdaq-lags-on-chip-weakness-5003237, 2026년 11월 5일.
3. Reuters, via Investing.com, Exclusive: AI backlash could threaten Anthropic IPO timing, potential investors say https://www.investing.com/news/stock-market-news/exclusiveai-backlash-could-threaten-anthropic-ipo-timing-potential-investors-say-5006110, 2026년 11월 5일.
4. SEC EDGAR, Anthropic, PBC, Form S-1 https://www.sec.gov/Archives/edgar/data/2130357/000162828026067168/anthropicpbc-sx1.htm, 2026년 10월 26일.
5. Greenberg Traurig, AI Company Anthropic Suspends Access to Claude Fable 5, Claude Mythos 5 Following US Export Control Directive https://www.gtlaw.com/it/insights/2026/6/ai-company-anthropic-suspends-access-to-claude-fable-5-claude-mythos-5-following-us-export-control-directive, 2026년 6월 17일.
6. Engadget, US government allows Anthropic to redeploy its Mythos and Fable AI models https://www.engadget.com/2205599/anthropic-redeploy-mythos-fable-ai-models/, 2026년 6월 30일.
7. Mistral AI, Mistral Large 4 (Le Chonk) is now open weight https://mistral.ai/news/mistral-large-4-oss/, 2026년 10월 29일.
8. Artificial Analysis, LLM Leaderboard, Intelligence Index v5.0.1 https://artificialanalysis.ai/leaderboards/models, 2026년 11월 7일 확인.
9. CNBC, CoreWeave shares tumble as bond yields near 15% on concern about AI borrowing binge https://www.cnbc.com/2026/10/14/coreweave-stock-bond-yields-ai-debt.html, 2026년 10월 14일.
10. StockAnalysis, CoreWeave (CRWV) stock price history https://stockanalysis.com/stocks/crwv/history/, 2026년 11월 6일 확인.

11. CNBC, Elon Musk says more than 200 neoclouds face 'guppy-level death' https://www.cnbc.com/2026/10/26/neocloud-stocks-today-coreweave-nebius-iren-elon-musk.html, 2026년 10월 26일.
12. Moonshot AI, Kimi K3.1 Pro https://www.kimi.ai/blog/kimi-k3-1-pro, 2026년 10월 23일.
13. Artificial Analysis, Kimi K3.1 Pro https://artificialanalysis.ai/models/kimi-k3-1-pro, 2026년 11월 6일 확인.
14. Anthropic, Pricing (Claude API) https://platform.claude.com/docs/en/about-claude/pricing, 2026년 11월 6일 확인.
15. Moonshot AI, Kimi K3.1-Pro License https://huggingface.co/moonshotai/Kimi-K3.1-Pro/raw/main/LICENSE, Hugging Face, 2026년 11월 6일 확인.
16. Hugging Face, Mistral AI, Mistral-Large-4-1T-A52B https://huggingface.co/mistralai/Mistral-Large-4-1T-A52B, 2026년 11월 7일 확인.

작성한 의존도 점검표, 외부 시선으로 점검받고 싶으신가요?

프리랜스 프로젝트를 맡고 있습니다. 팀이 실제로 쓰는 도구에 연결되는 AI 도구를 설계하고 출시합니다. 점검표를 보내 주시면 어디서 문제가 생길지 함께 살펴보겠습니다.

프로젝트 문의하기 **프로젝트 보기**

카카오톡, WhatsApp, 복사할 이메일 주소: maxence.pro/ko/#contact

다른 무료 자료: maxence.pro/ko/resources/

© 2026 Maxence Foulon · maxence.pro · 자유롭게 공유하셔도 됩니다. 출처를 꼭 밝혀 주세요.