

생성형 AI 통합 플랫폼 설명서

1 도입 개요

□ Timely AI 통합 교육 플랫폼 도입 개요

- (추진배경) 인공지능 기술의 급속한 발전에 따라, 교수법 연구와 직원 업무 역량 향상을 위해 AI를 접목한 자동화·분석 기반 시스템의 필요성이 증대되고 있음.
- (목적) AI 기술을 활용하여 연구 및 행정의 반복적·복잡 업무를 자동화하고, 연구 몰입 환경을 조성하기 위한 AI 교육 통합 플랫폼 도입을 추진함.
- (주요내용) Timely AI를 통한 AI 플랫폼 11개사의 LLM 모델 사용
Open AI, Google AI, Meta AI, Perplexity AI, Anthropic, Mistral AI, Alibaba, X AI, Upstage AI, Black Forest Labs, Runway
- (소요예산) 효율적 총량제 크레딧 방식

인원수	개월	인당 월 크레딧	인당 월 단가	총 크레딧 량	합계(VAT포함)
100	12	10,000	22,000원	12,000,000	26,400,000원
100	12	2,500	5,500원	3,000,000	6,600,000원

- 과금 방식 : 총량제 크레딧 기반으로 월별·사용자별 크레딧 한도 설정
- 가격 변동 : 최신 AI 모델 출시 시 모델별 단가 변동 가능
- 비용 효율 : 개별 유료 AI 모델 월 3만원 수준 대비, 9개 이상 대표 AI 모델을 통합 활용함으로써 비용 효율적
- 크레딧 이월 : 잔여 크레딧은 이월 가능하며 사용자별 한도 차등 설정
- 일반 AI 채팅 : 약 1 ~ 10 크레딧 (모델별 과금 상이)
- 첨부파일 포함 AI 채팅 : 약 40 ~ 50 크레딧 (모델별 과금 상이)
- 이미지 생성 시 (Nano Banana 기준) : 약 100 ~ 300 크레딧
- 동영상 생성 시 (5초 기준, 평균) : 약 800 ~ 3,000 크레딧
- 연간 소요 추정 : 100명 기준, 1인당 5,000 크레딧(약 11,000원) 사용 시 총 비용 13,200,000원 (2025년 기준)

- (G2B, S2B 등록) 총량제 크레딧 방식 (GS 1등급 획득)_디지털전문계약 제도(디지털서비스 이용지원시스템) : 7월중 등록예정

구분	토큰	단가금액	물품식별번호	비고
학교장터(S2B)	500,000	1,100,000	202504081525283	https://www.s2b.kr/
조달청 벤처나라	500,000	1,100,000	25559891	https://venture.g2b.go.kr

- (특장점) ChatGPT, Gemini, Llama, Perplexity, Claude, Grok, Mistral, Qwen, Solar 등 여러 가지 LLM 모델 동시 제공

- 1인 1계정 사용 : 계정 공유 없이 생성형 AI 프라이버시 보장
- 효율성 및 경제성 : 관리자 대시보드를 통한 관리 및 통계, 크레딧 방식
- 통계 기능 : 전체 사용량, 사용자별 사용량, 사용자 그룹별 사용량, AI 모델별 사용량 제공 및 엑셀 다운로드 지원
- 가입 제한 기능 : 특정 이메일 도메인 등록, 약관 동의 관리, NAT IP 등록 제공
- AI 웹페이지 관리 : 조직 이용 목적에 맞는 디자인 가능, 다양한 형태의 페이지 생성·관리 기능 제공
- 조직 관리 : 부서·조직 단위 선택 가입 및 가입 시 부서 설정 기능 제공
- 안전 관리 : 금칙어 설정 기능 제공

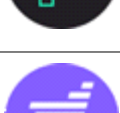
- (납품처) 서울시청, 경기도청, 경기교육청, 코레일, 계명대 등 300여 곳

- 대학 : 가천대학교, 국립경국대학교, 경상국립대학교, 전남대학교, 전북과학대학교, 계명대학교, 군장대학교, 원광대학교, 영진전문대학교, 순천대학교, 원광보건대학교, 경운대학교, 대전보건대학교, 대구보건대학교, 청암대학교, 남부대학교, 광주보건대학교, 한국폴리텍대학, 경동대학교, 백석대학교, 한국체육대학교, 동남보건대학교, 대신대학교, 동서대학교, 청암대학교, 호산대학교 등
- 관공서 : 서울시청, 경기도청, 경기도교육청, 철원군청, 인제군청, 금천구청, 밀양시청, 봉화군청, 남양주시청, 평택시청, 당진시청, 충청남도 교육청 교육과정 평가정보원, 코레일(한국철도공사), 농촌진흥청, 서울관광재단, 영천시시설관리공단, 구미전자정보기술원, 경찰청 등
- 기업 : 삼보모터스, 대교에듀캠프, 아트박스, 메가젠코리아, 세명기업, 서큘러스, 세종아이에스, 샘터정보기술, 컴퓨팅교사협회, 동우팜투테이블, 참프레, 평화정공(PHA) 등
- 학내망 (초중고) : 300여 곳 등

2 주요내용

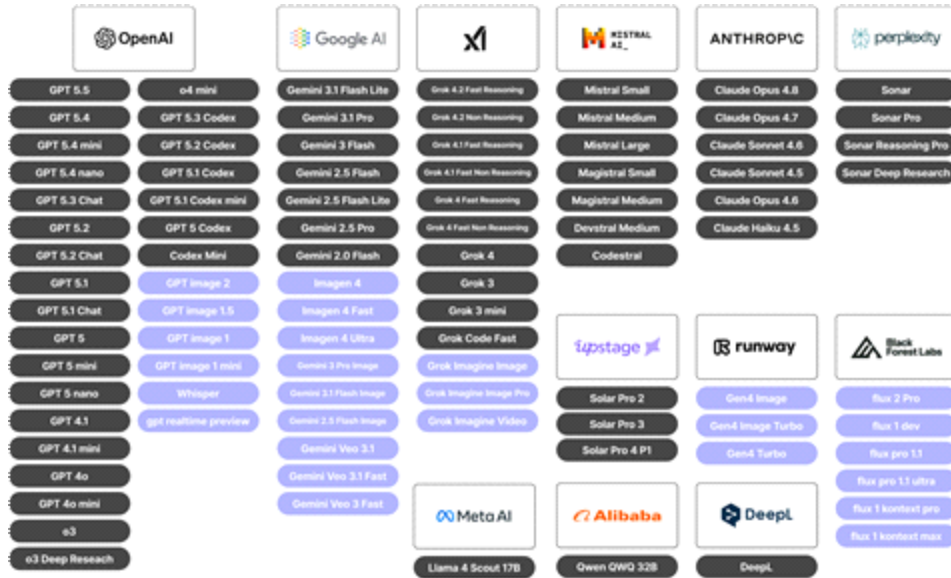
□ AI 모델 주요특징 비교표

○ AI 모델 주요특징

개발사		대표 LLM	사용 특징	1인 월 비용	100명 기준 연간 구독비
	OpenAI	GPT	범용 성능, 톨/에이전트, 음성·이미지·코딩 등 멀티모달 강점	\$20 ~ \$200	35,426,000원 ~ 354,266,000원
	Google AI	Gemini	구글 생태계 (검색·Gmail·Docs) 결합, 리서치·생산성 중심	\$20 ~ \$30	35,426,000원 ~ 53,139,000원
	Anthropic	Claude	글/코딩/분석 강점, "Extended thinking" 등 심층 작업 지향	\$20 ~ \$100	35,426,000원 ~ 177,133,000원
	Perplexity AI	Sonar	"검색 기반 답변+출처"에 최적화, 리서치 워크플로우 강점	\$20 ~ \$40	35,426,000원 ~ 70,853,000원
	Meta AI	Llama	오픈소스 생태계(Llama), 앱 내 AI 경험 중심	\$50 ~ \$500	88,566,000원 ~ 885,665,000원
	Mistral AI	Mistral	경량~대형 라인업, 유럽권 프라이버시/온프레임 옵션 선호층	\$15 ~ \$30	26,569,000원 ~ 53,139,000원
	Alibaba	Qwen	매우 큰 컨텍스트/라인업 다양, 클라우드 기반 API 요금 체계	\$30 ~ \$300	53,139,000원 ~ 531,399,000원
	X AI	Grok	X(트위터) 실시간 정보/트렌드 결합 지향	\$30 ~ \$40	53,139,000원 ~ 70,853,000원
	Runway	Gen	영상 생성·편집 워크플로우 특화(크레딧 기반)	\$28 ~ \$76	49,597,000원 ~ 134,621,000원
	Black Forest Labs	Flux	고품질 이미지 생성(Stable Diffusion 핵심팀 출신으로 알려짐), API 제공	사용량 기반 과금 1인, 월 100장씩 생성 시 연간 7,085,000원 ~ 10,627,000원	
	Upstage	Solar	한국 기업: 문서/업무 자동화 + LLM API(문서 파싱/추출 강점)	\$20 ~ \$100	35,426,000원 ~ 177,133,000원

□ Timely AI 통합 플랫폼 제공 모델

○ 기본 모델 및 고급 모델



개발사		제공 모델						
Open AI	텍스트	GPT 5.5	GPT 5.4	GPT 5.4 Mini	GPT 5.4 Nano	GPT 5.3 Chat	GPT 5.2	
		GPT 5.2 Chat	GPT 5.1	GPT 5.1 Chat	GPT 5	GPT 5 mini	GPT 4.1	
		GPT 4.1 mini	GPT 4o	GPT 4o mini	o3	o3 deep research	o4 mini	
	코딩	GPT 5.3 Codex	GPT 5.2 Codex	GPT 5.1 Codex	GPT 5.1 Codex mini	GPT 5 Codex		
	이미지	GPT Image 2	GPT Image 1.5	GPT Image 1	GPT Image 1 mini			
Google AI	텍스트	Gemini 3.1 Pro	Gemini 3.1 Flash Lite	Gemini 3 Flash	Gemini 2.5 Flash	Gemini 2.5 Flash Lite	Gemini 2.5 Pro	Gemini 2.0 Flash
	이미지	Imagen 4	Imagen 4 Fast	Imagen 4 Ultra	Gemini 3 Pro Image	Gemini 2.5 Flash Image	Gemini 3.1 Flash Image	
	영상	Veo 3.1			Veo 3.1 Fast		Veo 3 Fast	
Mistral AI	텍스트	Mistral Small	Mistral Medium	Mistral Large		Magistral Small	Magistral Medium	
	코딩	Devstral Medium			Codestral			
X AI	텍스트	Grok 4.1 Fast	Grok 4 Fast	Grok 4		Grok 3	Grok 3 mini	
	코딩	Grok Code Fast						
	이미지	Grok Imagine Image			Grok Imagine Image Pro			
	영상	Grok Imagine Video						
Runway	이미지	Gen 4 Image			Gen 4 Image Turbo			
	영상	Gen 4 Turbo						
Antnropic		Claude Opus 4.8			Claude Opus 4.7 / 4.6			
		Claude Sonnet 4.5 / 4.6			Claude Haiku 4.5			
Perplexity AI (검색)		Sonar	Sonar Pro		Sonar Reasoning Pro		Sonar Deep Research	
Meta AI		Llama 4 Scout 17B						
Upstage AI		Solar Pro 4 P1	Solar 3 Pro			Solar 2 Pro		
Alibaba		Qwen QWQ 32B						
Black Forest Labs (이미지)		Flux 2 Pro		Flux 1 dev		Flux pro 1.1		
		Flux pro 1.1 ultra		Flux 1 kontext Pro		Flux 1 kontext Max		

※ 실제 제공 모델은 API 상황 및 버전(모델) 업데이트에 따라 조정될 수 있음

○ 가격 비교 표

- Open AI ChatGPT 50명 사용 기준

구분	비용 기준	1개월	6개월	1년
ChatGPT	1인, 1개월, 약 28,600원	약 143만원	약 858만원	약 1,716만원

- 타임리 서비스는 종량제 크레딧 기반 방식이기 때문에 인원 제한 없이 사용 가능
- 예) 110만원(부가세 포함)으로 타임리 AI 서비스 크레딧 구매 시 50만 크레딧 지급

구분	비용 기준	유저 50명	추가 크레딧 구매
Timely AI	100만원 -> 50만 크레딧	10,000크레딧	크레딧당 2원

- 실제 서비스 도입 후 액티브 유저는 **전체의 약 20%**로 GPT를 활발하게 사용하는 1인 유저의 **크레딧 기댓값은 약 10,000크레딧에서 20,000크레딧** 이상
- 관리자가 구성원의 크레딧 월 사용 제한량을 지정할 수 있어 효율적으로 예산활용 가능
- 1,000크레딧으로 GPT 4o mini, GPT 4.1 nano, GPT 4.1 mini 등 타임리 AI 서비스 AI Chat 200 ~ 1,000회 활용 가능
- ※ 실제 사용량은 AI 문답의 길이와 LLM 모델에 따라 변동 가능
- 종량제 서비스 특징으로 유저 AI 사용량에 따라 크레딧을 유동적으로 배분할 수 있기 때문에 효율적인 AI 활용 가능

- 각 유료 모델 개별 구매 기준

- 각 유료모델 AI 플랫폼 5종 고급버전 사용 가능(예시)

구분	개별 구매 시				
	ChatGPT	Gemini	Claude	Perplexity	LLaMA
1인당 평균	약 28,600원	약 29,000원	약 28,600원	약 28,600원	상용 API 사용 요금 / 클라우드 서버 운영비
(소계) 약 114,800원 + α					

- 100명의 구성원이 5개사 AI 플랫폼을 한 번에 이용하기 위해서는 1인 월 114,800원 + α 만큼의 비용이 발생하며 1년 기준 약 1.4억원이 발생
- 1년간 예산 11,000,000원 기준 100명의 구성원이 약 100만 ~ 500만회의 AI Chat 활용 가능 (남은 크레딧은 소멸되지 않으며 총 크레딧에 적립됨)

○ 타임리 크레딧 소진 예시

- 크레딧 책정 기준

- 크레딧은 유저의 AI 입출력 데이터와 사용 도구를 기반으로 계산됩니다.
- LLM 모델마다 기준 토큰 당 비용이 다르기 때문에 실제 사용량 계산하여 크레딧으로 통일하여 안내하고 있습니다.
- [각 AI 개발사의 LLM 모델별 가격 x (입출력에 사용된 토큰 + 검색, 이미지, 코드 등 AI 도구에 사용된 토큰)] + 서비스 운영비용

- AI Chat 소진 예시

- 대화형 AI(GPT 4.1 mini) : 입력 45자 , 출력 1,576자, 사용한 크레딧 : 7 크레딧

실제 사용 화면

Chat GPT가 지원하는 API의 LLM 모델들에 대해서 종류와 특징을 설명해줘

네 ChatGPT가 지원하는 API의 LLM(대형 언어 모델) 모델들에 대해 종류와 특징을 단계별로 자세히 설명드리겠습니다.

ChatGPT API에서 지원하는 LLM 모델 종류와 특징

OpenAI의 API를 통해 사용할 수 있는 주요 LLM 모델들은 크게 GPT-3 계열, GPT-3.5 계열, 그리고 GPT-4 계열로 나눌 수 있습니다. 각 모델은 성능, 용도, 비용 측면에서 차이가 있습니다.

1. GPT-3 계열 모델들

대표 모델

text-davinci-003
text-curie-001
text-babbage-001
text-ada-001

특징

text-davinci-003: GPT-3 중 가장 강력한 모델로, 복잡한 언어 이해와 생성에 뛰어납니다. 창의적이고 참고한 작업에 적합합니다.

text-curie-001: 다목적 모델로, 빠르고 비용 효율적이며, 중간 정도의 복잡한 작업에 적합합니다.

text-babbage-001: 간단한 작업에 적합하며, 빠르고 저렴합니다.

text-ada-001: 가장 빠르고 저렴하지만, 복잡한 작업에는 적합하지 않습니다.

용도 예시

텍스트 생성, 요약, 번역, 간단한 질의응답 등

2. GPT-3.5 계열 모델들

대표 모델

gpt-3.5-turbo

특징

GPT-3.5 대비 성능과 비용 효율성이 크게 향상됨

대화형 AI에 최적화되어 있어 ChatGPT와 같은 챗봇 구현에 적합

빠른 응답 속도와 높은 품질의 텍스트 생성 가능

용도 예시

챗봇, 고객 지원, 대화형 애플리케이션, 실시간 질의응답

3. GPT-4 계열 모델들

대표 모델

gpt-4
gpt-4-turbo

특징

GPT-3.5 대비 훨씬 더 뛰어난 이해력과 생성 능력 보유

복잡한 문제 해결, 긴 문서 이해, 고급 창의적 작업에 적합

gpt-4-turbo는 gpt-4보다 더 빠르고 비용 효율적이지만, 내용적으로는 약간의 차이가 있을 수 있음

멀티모달(텍스트+이미지) 지원 가능 (단, API에서는 텍스트 중심)

용도 예시

고급 연구, 법률 문서 분석, 창작적 글쓰기, 복잡한 대화 시나리오

4. 모델 선택 시 고려사항

모델명	성능	비용	속도	추천 용도
text-davinci-003	높음	높음	보통	복잡한 텍스트 생성 및 이해
gpt-3.5-turbo	중간~높음	낮음	빠름	대량 대화형 AI, 실시간 응답
gpt-4	매우 높음	매우 높음	느림	고급 분석, 복잡한 작업
gpt-4-turbo	매우 높음	중간	빠름	고급 작업 + 비용 효율성 필요

5. 최신 정보 및 참고 링크

OpenAI 공식 문서: <https://platform.openai.com/docs/models>

GPT-4 Turbo 출시 관련: 2024년 초에 발표되어 GPT-4 대비 비용과 속도 면에서 개선됨

요약

모델군	주요 모델	특징	추천 용도
GPT-3	text-davinci-003	강력하지만 비용 높음	복잡한 텍스트 생성
GPT-3.5	gpt-3.5-turbo	비용 효율적, 빠름	챗봇, 실시간 대화
GPT-4	gpt-4, gpt-4-turbo	최고 성능, 고급 작업에 적합	연구, 법률, 창작적 작업

필요하시면 각 모델별 API 사용법이나 비용 구조, 예제 코드도 안내해 드릴 수 있습니다. 추가 질문 있으시면 언제든지 말씀해주세요! 🙌

GPT-4.1 mini

- 이미지 생성(GPT image-1) : 1024x1536, 2.2MB, 96dpi, 24비트, 사용한 크레딧 212 크레딧

실제 사용 화면

아이가 한복을 입고 웃으면서 달려오는 모습을 그려줘

아이가 한복을 입고 웃으면서 달려오는 모습을 그린 이미지입니다. 😊



이 그림은 아이의 밝고 행복한 표정과 전통 한복의 아름다움을 잘 표현하고 있습니다. 혹시 더 구체적인 상황이나 배경, 아이의 나이, 성별 등 추가 요청이 있으면 알려주세요!

GPT-4.1 mini

○ 모델별 가격 비교표

- 타임리AI는 종량제로 결제하는 정책을 지원하고 있습니다.

사용자 1인당 매월 \$25(한화 약 36,000원)을 내야하는 ChatGPT Team Plan이나 월 \$30(한화 약 44,000원)을 내야하는 Microsoft Copilot 대비 70% 이상의 비용 절감이 가능합니다.

모델/도구명		API 1,000토큰 기준 요금 (한글 약 800 ~ 1,600자)		비고
개발사 (Type)	모델명	입력	출력	
Open AI (Chat)	GPT 5.5	약 9 크레딧	약 54 크레딧	
	GPT 5.4	약 4.5 크레딧	약 27 크레딧	
	GPT 5.4 Mini	약 1.4 크레딧	약 8.2 크레딧	
	GPT 5.4 Nano	약 0.4 크레딧	약 2.3 크레딧	
	GPT 5.3 Chat	약 3.2 크레딧	약 25.4 크레딧	
	GPT 5.3 Codex	약 3.2 크레딧	약 25.4 크레딧	
	GPT 5.2	약 3.2 크레딧	약 25.4 크레딧	
	GPT 5.2 Chat	약 3.2 크레딧	약 25.4 크레딧	
	GPT 5.2 Codex	약 3.2 크레딧	약 25.4 크레딧	
	GPT 5.1	약 2.3 크레딧	약 18.1 크레딧	
	GPT 5.1 Chat	약 2.3 크레딧	약 18.1 크레딧	
	GPT 5.1 Codex	약 2.3 크레딧	약 18.1 크레딧	
	GPT 5.1 Codex Mini	약 0.5 크레딧	약 3.6 크레딧	
	GPT 5	약 2.3 크레딧	약 18.1 크레딧	
	GPT 5 Mini	약 0.5 크레딧	약 3.6 크레딧	
	GPT 5 Nano	약 0.1 크레딧	약 0.7 크레딧	
	GPT 5 Codex	약 2.3 크레딧	약 18.1 크레딧	
	GPT 4.1	약 3.6 크레딧	약 14.5 크레딧	
	GPT 4.1 Mini	약 0.7 크레딧	약 2.9 크레딧	
	GPT 4o	약 4.5 크레딧	약 18.1 크레딧	
	GPT 4o Mini	약 0.3 크레딧	약 1.1 크레딧	
	o4 Mini	약 2 크레딧	약 8 크레딧	
	o3	약 3.6 크레딧	약 14.5 크레딧	
	o3 Deep Research	약 18.1 크레딧	약 72.5 크레딧	
Open AI (Image)	GPT Image 2	-	74 ~96 크레딧	1장당
	GPT Image 1.5	-	113 ~ 165 크레딧	1장당
	GPT Image 1	-	139 ~ 208 크레딧	1장당
	GPT Image 1 Mini	-	17 ~ 20 크레딧	1장당
Open AI (Voice)	Whisper	-	약 10.8 크레딧	1분당
	GPT 4o Realtime Preview	약 9 크레딧	약 36.2 크레딧	
Anthropic (Chat)	Claude Opus 4.8	약 9.1 크레딧	약 45.3 크레딧	
	Claude Opus 4.7	약 9.1 크레딧	약 45.3 크레딧	
	Claude Opus 4.6	약 9.1 크레딧	약 45.3 크레딧	
	Claude Sonnet 4.6	약 5.4 크레딧	약 27.1 크레딧	20만 토큰 초과 시 2, 1.5배 할증
	Claude Sonnet 4.5	약 5.4 크레딧	약 27.1 크레딧	
	Claude Haiku 4.5	약 1.8 크레딧	약 9 크레딧	
Google AI (Chat)	Gemini 3.1 Pro	약 3.6 크레딧	약 21.8 크레딧	20만 토큰 초과 시 2, 1.5배 할증

모델/도구명		API 1,000토큰 기준 요금 (한글 약 800 ~ 1,600자)		비고
개발사 (Type)	모델명	입력	출력	
	Gemini 3.1 Flash Lite	약 0.5 크레딧	약 2.7 크레딧	
	Gemini 3 Flash	약 0.9 크레딧	약 5.4 크레딧	
	Gemini 2.5 Pro	약 2.3 크레딧	약 18.1 크레딧	20만 토큰 초과 시 2, 1.5배 할증
	Gemini 2.5 Flash Lite	약 0.2 크레딧	약 0.7 크레딧	
	Gemini 2.5 Flash	약 0.5 크레딧	약 4.5 크레딧	
	Gemini 2.0 Flash	약 0.2 크레딧	약 0.7 크레딧	
Google AI (Image)	Gemini 3 Pro Image Preview	-	약 423 크레딧	1장당
	Gemini 3.1 Flash image	-	약 222크레딧	1장당
	Gemini 2.5 Flash Image	-	약 129 크레딧	1장당
	Imagen 4	-	약 132 크레딧	1장당
	Imagen 4 Fast	-	약 66 크레딧	1장당
	Imagen 4 Ultra	-	약 198 크레딧	1장당
Google AI (Video)	Veo 3.1	-	약 1,320 크레딧	1초당
	Veo 3.1 Fast	-	약 495 크레딧	1초당
	Veo 3 Fast	-	약 495 크레딧	1초당
Perplexity	Sonar Deep Research	약 3.6 크레딧	약 15 크레딧	
	Sonar Reasoning Pro	약 3.6 크레딧	약 15 크레딧	
	Sonar Pro	약 5.4 크레딧	약 27.2 크레딧	
	Sonar	약 1.8 크레딧	약 1.8 크레딧	
X AI (Chat)	Grok 4.2	약 3.7 크레딧	약 11 크레딧	20만 토큰 초과 시 2배 할증
	Grok 4.1 Fast	약 0.4 크레딧	약 0.9 크레딧	12만 8천 토큰 초과 시 2배 할증
	Grok 4	약 5.4 크레딧	약 27.2 크레딧	12만 8천 토큰 초과 시 2배 할증
	Grok 4 Fast	약 0.4 크레딧	약 0.9 크레딧	12만 8천 토큰 초과 시 2배 할증
	Grok 3	약 5.4 크레딧	약 27.2 크레딧	
	Grok 3 Mini	약 0.5 크레딧	약 0.9 크레딧	
	Grok Code Fast	약 0.4 크레딧	약 2.7 크레딧	
X AI (Image)	Grok Imagine Image	-	약 66 크레딧	
	Grok Imagine Image Pro	-	약 231 크레딧	
X AI (Video)	Grok Imagine Video	-	약 165 크레딧	
Mistral AI	Mistral Small	약 0.3 크레딧	약 1.1 크레딧	
	Mistral Medium	약 0.7 크레딧	약 3.6 크레딧	
	Mistral Large	약 0.9 크레딧	약 2.7 크레딧	
	Magistral Small	약 0.9 크레딧	약 2.7 크레딧	
	Magistral Medium	약 3.6 크레딧	약 9 크레딧	
	Devstral Medium	약 0.7 크레딧	약 3.6 크레딧	
	Codestral	약 0.5 크레딧	약 1.6 크레딧	
Meta	Llama 4 Scout	약 0.2 크레딧	약 0.6 크레딧	
Alibaba	Qwen QWQ	약 0.5 크레딧	약 0.7 크레딧	
Upstage	Solar Pro 2	약 0.3 크레딧	약 1.1 크레딧	

모델/도구명		API 1,000토큰 기준 요금 (한글 약 800 ~ 1,600자)		비고
개발사 (Type)	모델명	입력	출력	
	Solar Pro 3	약 0.3 크레딧	약 1.1 크레딧	
	Solar Pro 4 P1	약 0.3 크레딧	약 1.1 크레딧	
Runway (Image)	Gen4 Image	-	약 165 크레딧	1장당
	Gen4 Image Turbo	-	약 66 크레딧	1장당
Runway (Vide)	Gen4 Turbo	-	약 165 크레딧	1초당
Black Forest Labs (Image)	Flux 2 Pro	-	약 99 크레딧	1장당
	Flux 1.1 Pro	-	약 132 크레딧	1장당
	Flux 1.1 Pro Ultra	-	약 198 크레딧	1장당
	Flux 1 Dev	-	약 83 크레딧	1장당
	Flux 1 Kontext Max	-	약 264 크레딧	1장당
	Flux 1 Kontext Pro	-	약 132 크레딧	1장당

3 FAQ

□ Timely AI Q&A

○ 요금 및 과금 구조

- 가장 기본 옵션으로 구축한다고 가정할 때 300명 기준 도입 총 비용 (초기 구축, 연간 라이선스, 유지 보수 포함) 은 얼마인가요?
 - 타임리 AI 서비스는 종량제 형태의 서비스 기준으로 구매할 수 있습니다. 스페이스 생성, 템플릿 지정 간단한 배너 등록, 로고 변경 등의 기초 세팅은 무상으로 지원해드리며, 구매하시는 크레딧 기반으로 비용을 산정할 수 있습니다.
 - 1명이 1개월에 1만 크레딧을 사용한다고 가정할 경우 300명 기준 1개월에 300만 크레딧이 되고 해당 비용은 660만원(부가세 포함)입니다. 종량제 서비스 특성 상 매월 요금이 부과되지는 않으며 크레딧이 소진될 때마다 필요한 만큼 구매할 수 있습니다.
- 기본 패키지에 포함된 기능과 별도 비용이 발생하는 유료 기능 목록은?
 - 타임리 AI 서비스는 크레딧만 구매하시면 추가로 결제하지 않으셔도 모든 기능을 활용할 수 있습니다. 다만 서비스 관리와 별도 개발이 필요한 항목(SSO 연동, 신규 AI 기능의 우선 개발 등)의 경우 유료 정책이나 제한 정책을 통해 운영될 수도 있습니다.
- 조직 맞춤 커스터마이징은 기본 포함인가요? 아니면 별도 과금 항목인가요?
 - 서비스 최초 구매 시 기본적인 템플릿, 로고, 배너 등 간단한 초기 세팅은 타임리에서 진행해드리며 별도 비용은 받지 않습니다. 구축된 GPT 페이지를 전달받으시고 관리자를 지정해드리면 이후 관리자가 직접 수정하실 수 있습니다.
- 교육 프로그램은 기본 서비스에 포함되나요? 아니면 별도 구매가 필요한가요?
 - 타임리 AI 서비스의 원활한 사용을 위해 기본 사용자 매뉴얼, VOD 교육 영상 등을 무상으로 제공하고 있습니다. 고객사의 결제 금액과 협력 관계에 따라 온/오프라인 교육 프로그램을 안내드리고 있으며 일반적으로 1회 1시간 교육 기준으로 50만원 ~ 100만원의 금액이 발생합니다. 정확한 가격은 교육 인원, 교육 시간, 교육 방법 등에 따라 다르

게 측정될 수 있습니다.

○ 서비스 성능 및 기능

1) AI 다양한 모델 간 전환은 클릭 수준으로 가능한가요? 아니면 별도 설정이 필요한가요?

- 타임리 AI에서 제공하는 AI 기능(AI Chat, AI Template, AI Agent 등)에서는 사용자가 직접 모델을 설정하여 답변을 받을 수 있습니다. 나열된 LLM 모델을 클릭하는 과정으로 모델을 손쉽게 전환할 수 있습니다.
- Labs의 기능 중 이미지 생성, 지식 검색, 영상 생성을 제외한 다른 기능들은 고정 모델을 통해 지원하는 서비스이므로 임의로 모델을 변경하실 수 없습니다.

2) GPT, Claude 등 주요 모델의 응답 속도는 실제로 어느 정도인가요? (벤치마크 수치 제공 가능 여부 포함)

- Artificial Analysis 기준 주요모델의 응답 속도는 다음과 같습니다.

모델/도구명	Output Speed	Latency - TTFT
GPT 5.2	약 140 ~ 200 토큰 /s	0.4 ~ 0.8초
GPT 5.1	약 120 ~ 170 토큰 /s	0.5 ~ 1초
Gemini 3 Pro	약 130 ~ 190 토큰 /s	0.4 ~ 0.9초
Claude 4.5 Opus	약 70 ~ 90 토큰 /s	1.5 ~ 2.0초
Claude 4.5 Haiku	약 70 ~ 100 토큰 /s	1.2 ~ 2.0초
Grok 4.1	약 90 ~ 130 토큰 /s	0.8 ~ 1.3초
Llama 4 Scout 17B	약 80 ~ 130 토큰 /s	0.8 ~ 1.2초
Qwen QWQ 32B	약 80 ~ 130 토큰 /s	0.8 ~ 1.3초
Mistral Medium	약 100 ~ 150 토큰 /s	0.6 ~ 1초
Solar Pro 2	약 90 ~ 140 토큰 /s	0.7 ~ 1.2초

3) 최대 동시 사용자 수용량은 얼마이며, 이에 대한 부하 테스트가 완료되었나요?

- 3000명 동시 접속 부하 테스트가 완료되었습니다.

4) RAG 기능 사용 시 단일 문서 및 전체 문서 용량 제한은 어떻게 되나요?

- 파일 업로드 시 PDF(이미지, 형태)의 경우 OCR 분석 기능이 들어가기 때문에 최대 용량 50MB, 100페이지로 제한되고 있습니다.
- 워드파일, 텍스트파일 등 텍스트 기반으로 작성된 경우 100페이지 이상의 경우에도 파일 등록 및 분석이 가능합니다. 타임리에서는 1M 컨텍스트 윈도우를 제공하는 LLM 모델을 함께 지원하고 있습니다. (300페이지 분량의 논문 워드 파일로 동작 확인)

5) 모바일 및 태블릿 환경에서도 동일한 기능과 성능을 제공하나요?

- 타임리 AI는 PC와 모바일에서 동일한 기능과 성능을 제공하고 있습니다.
- 구글 플레이와 앱스토어에서 모두 다운로드하실 수 있으며 모바일 웹에서도 동일하게 이용할 수 있습니다.

○ 데이터 보안 및 개인정보 보호

1) 고객 데이터는 어느 위치의 어떤 클라우드(국내/해외)에 저장되며, 보관 정책은 어떻게 되나요?

- 타임리에서 관리하는 데이터는 MicroSoft Azure 클라우드 데이터 센터(Korea Central)에 저장됩니다.
- 일반 유저의 경우 서비스 해지(회원 탈퇴) 시 식별 가능한 개인정보는 삭제합니다.
- 채팅 데이터를 포함한 사용 기록은 사용 시점 기준으로 90일 후 사용자가 확인할 수 없도록 삭제 처리되며 1년 후 데이터 삭제가 완료됩니다.

2) Open AI, Google AI API 등 외부 호출 시 데이터 전송 보안은 보장되나요?

- HTTPS TLS 기반의 암호화 통신을 사용하여 데이터를 전송합니다.
- API 호출은 모두 내부 서버에서만 수행되며, 사용자 브라우저 또는 외부 클라이언트에서 직접 호출되지 않도록 설계되어 있습니다.

3) GDPR, CCPA, 국내 개인 정보 보호법 등 주요 데이터 규제 준수와 관련된 공식 문서를 제공할 수 있나요?

- GDPR, CCPA, 국내 개인 정보 보호법 등 주요 데이터 규제 준수와 관련된 공식 문서를 제공하지 않습니다.

4) 내부 시스템(API) 연동 시, 데이터 암호화 방식 및 인증 체계는 어떻게 구성되어 있나요?

- 내부 서비스 간의 통신은 HTTPS TLS 기반의 암호화 통신을 사용하며, 서비스 간 인증 체계는 JWT 기반의 사용자 인증 토큰을 통해 수행됩니다.

5) 타임리 AI가 보유한 정보보안 인증 목록을 알려주세요.

- 타임리 AI에서 보유한 정보보안 인증 목록은 없습니다.
- KISA의 웹 취약점 점검을 수행하여 보안 취약점에 대한 개선이 완료되었습니다. (2025 12월)

○ 맞춤형 구축 및 연동

1) GPT 서비스 구축 시 기획부터 운영 가능 상태까지 평균 소요 기간은 얼마나 걸리나요?

- 타임리 AI 구축 문의가 접수되면 기초 디자인 설정과 도메인 생성부터 스페이스 구축 전달까지 영업일 기준 1 ~ 2일 정도 소요됩니다.

2) 커스터마이징 가능한 항목의 범위와 한계는 어디까지인가요?

- 관리자는 메인페이지에서 확인할 수 있는 로고와 배너, 상단 메뉴 로고, 하단 메뉴 스페이스 정보, 오픈그래프(페이지 링크 공유 시 노출되는 썸네일)를 등록 및 변경할 수 있습니다.
- 기능적 커스터마이징 항목으로는 AI Chat 화면에서 노출되는 Chat 예시 질문을 등록 및 수정할 수 있고, AI Template 화면에서 확인할 수 있는 그룹 지정 및 그룹별 템플릿 지정과 추천 템플릿 지정이 가능합니다.
- 제공된 스페이스에서 구성원이 활용할 수 있는 LLM 모델 제한과 요금제 제한량을 설정할 수 있습니다.

- 각 디자인 요소는 정해진 구역과 크기 내에서 변경만 가능하며 위치를 직접 수정하거나 전체 테마를 변경하는 기능은 추후 웹 빌더 기능 패치 후에 제공될 예정입니다.

3) ERP, CRM, 그룹웨어 등 기존 시스템과의 연동은 어떻게 지원되며, 기존 사례는 어떤 것이 있나요?

- 타임리 AI에서는 플랫폼에서 서비스를 직접 사용하는 형태 외에 타임리 API를 활용하여 기존 시스템에 서비스를 구축할 수 있도록 Assistant API를 지원하고 있습니다.
- 각 조직에서 사용 중인 서비스 내 타임리 시스템을 구현하기 위해 윈도우 형태 서비스를 연결할 수 있도록 제공되는 API를 활용해서 각 조직의 개발팀이 직접 개발하여 진행합니다.
- 타임리 AI는 해당 기능 구현을 위한 API Document와 구현된 서비스의 관리를 위한 관리 콘솔을 제공드리며 현재 (주) 알파코의 LMS 시스템에서 API를 통한 타임리 서비스 연결을 진행하고 있습니다.

4) 특수 기능 개발 요청 시, 개발 소요 기간과 전적 산정 기준은 어떻게 되나요?

- 타임리 서비스에서 제공하는 기능 및 서비스 외 특수 기능을 요청할 경우 해당 기능의 시급성과 필요성, 구현 가능성 등을 검토하여 일정을 산정할 수 있습니다.
- 개발 소요 기간 및 전적 산정 기준은 일반적으로 안내드리기 어려우며 추가 기능이 필요한 경우 개발 및 사업팀 간의 협의 미팅을 통해 산정해야합니다.

5) 구축 후 정기 업데이트, 유지보수 방식 기술 지원 체계는 어떻게 제공되나요?

- GPT 타임리 서비스는 플랫폼 기반의 서비스이기 때문에 신규 AI 기능, 서비스 기능 업데이트는 업데이트 배포와 함께 실시간으로 지원됩니다. 구축 이후 토큰 소진이 되기 전까지는 별도의 유지보수나 기술 지원이 불필요합니다.

○ 교육프로그램

1) 실제 제공되는 교육 커리큘럼 구성표나 PDF 자료를 미리 받아볼 수 있나요?

- 교육 대상과 교육 분야 주제를 전달해주시면 교육 주제를 기반으로 커리큘럼을 구성하여 안내드릴 수 있습니다.

2) 교육은 온라인, 오프라인, 온디맨드 중 어떤 형식으로 제공되며 선택 가능한가요?

- 교육은 Zoom을 통한 온라인 교육, 강사의 직접 방문 후 진행하는 오프라인 교육, VOD를 통한 온디맨드 교육 모두 제공이 가능합니다. 교육의 내용과 주제에 따라서 교육 가능 여부가 나뉘고 교육 방식은 협의 후 직접 선택할 수 있습니다.

3) 교육 이수 후 활용도 평가 방식과 피드백 시스템은 마련되어 있나요?

- 교육 진행 후 활용도를 평가하는 방식은 존재하지 않습니다. 교육에 대한 피드백 시스템은 공식적으로는 존재하지는 않지만 희망하실 경우 Google Form을 통해 진행할 수 있습니다.

4) AI 최신 기능 업데이트 주기에 따라 교육 콘텐츠는 얼마나 자주 갱신되나요?

- 타임리에서 제공해드리는 사용자 매뉴얼과 교육 영상은 모듈별로 제작이 되어 신기능 출시 시 새로운 VOD가 추가되는 방식으로 업데이트됩니다.

- 온라인 Zoom, 오프라인 교육 시에는 타임리 AI 서비스 업데이트 내용을 포함하여 최신 버전 기준으로 교육을 진행합니다.

○ 성과지표

- 1) **작업 속도 5배, 운영비 24% 절감 등의 수치는 어떤 방식으로 측정되었으며 실제 적용 사례는 무엇인가요?**
 - 작업 속도와 운영비 절감 등의 수치와 관련하여 구체적인 측정 방식 기준은 정해지지 않았습니다.
- 2) **ROI를 측정할 수 있는 지표 예시와 분석 템플릿이 제공되나요?**
 - 타임리 AI를 도입함으로 발생할 수 있는 ROI 측정 지표 예시는 업무 처리시간, 서비스 비용, 업무 처리량, 직원 AI 만족도 등이 있습니다.
 - AI를 활용하지 않거나 다른 AI 서비스를 동시에 활용하는 것과 비교해 업무 효율성, 직원 만족도 서비스 비용 등을 검토하여 기존 상황 대비 효과를 확인하실 수 있습니다. 하지만 타임리 AI 서비스 내에서 별도로 제공하는 분석 템플릿은 존재하지 않습니다.
- 3) **SLA (서비스 수준) 협약의 가용성 응답시간, 복구시간 등 주요 항목은 어떻게 정의되어 있나요?**
 - SLA 서비스 수준 계약 관련 내용은 정의되어있지 않습니다.
- 4) **시스템 장애 발생 시 지원 대응 시간, 백업/복구 프로세스는 어떻게 운영되나요?**
 - GPT Azure Cloud 타임리 서비스는 를 활용해 서비스를 구성하고 있습니다. 장애 발생 시 Azure의 고가용성 인프라 및 자동 복구 기능을 통해 장애에 대비하고 있습니다.

○ 개발 및 기술 역량

- 1) **GPT 타임리 플랫폼은 자체 기술 기반인가요? 외부 솔루션 통합형인가요?**
 - GPT 타임리 플랫폼은 외부 LLM 모델 API를 활용해 AI Chat과 AI Template 등 AI 도구를 활용할 수 있는 웹사이트를 구축하는 서비스입니다.
 - 직접 관리하는 LLM 모델은 없지만 그 외 모든 기능과 기술은 자체 개발하였습니다.
- 2) **내부 개발 인력 규모 및 전담 기술 조직 구성은 어떻게 되나요?**
 - 타임리 조직은 행정, 사업, 개발 3가지 하위 조직으로 구성되며 사업 운영, 개발, 기획, CS 등 서비스/사업 운영과 신규 사업 개발을 관리하는 평균 경력 7년 이상의 사업팀과 서비스 기획, 개발, 디자인 등 타임리 AI를 비롯한 서비스 전체 개발을 관리하는 평균 경력 8년 이상의 개발팀으로 구성되어 있습니다.
- 3) **RAG, API , 연동 프롬프트 관리 등 주요 기능은 내부에서 직접 개발되었나요?**
 - RAG, API , 연동 프롬프트 관리를 포함해 타임리 AI 서비스를 구성하는 전체 기능은 타임리 내부 개발팀이 외주 없이 직접 개발하였습니다.
- 4) **기능 요청 또는 개선 시 기획 → 개발 → 배포까지 평균 소요 시간은?**
 - 간단한 기능이나 문제 상황 파악 시 피드백 후 개발 및 배포까지는 1일정도 소요되며 비교적 특별한 기능이나 로직이 얹혀있는 프로세스의 개선 및 개발은 영업일 기준 3 ~

5일정도 소요될 수 있습니다. 해당 소요 시간은 요청한 기능 개선과 타임리 시스템 구성 상황에 따라서 가감될 수 있습니다.

5) 장애 대응 이력, 기능 릴리즈 노트, 기술 블로그 등 기술 성과물이 존재하나요?

- 타임리 AI 서비스 운영 중 발생한 접속, 기능 사용 문제에 대해 실시간 고객 대응을 진행하며 해당 문제의 발생 원인을 파악하고 해결하였습니다.
- 수천명 단위의 학교 구성원의 가입 및 사용 단계에서 과도하게 낭비되는 토큰 발생이나 의도하지 않는 서비스 기능을 확인하여 토큰을 효율적으로 사용할 수 있도록 RAG 기술 개선, Chat 데이터 관리 등을 진행하여 문제를 해결하였습니다.
- 타임리 AI 업데이트는 사용 가이드 내 업데이트 소식을 모아 릴리즈 노트 형태로 제공하고 있으며 타임리 AI 서비스를 내부 시스템에 직접 개발할 수 있도록 API Document를 지원하고 있습니다.