fixcloud

SUPERINTELLIGENCE CLOUD · TÜRKİYE

# LLMaaS

# Tek API, Tüm Modeller

Büyük dil modellerini (LLM), tek bir API üzerinden dakikalar içinde uygulamalarınıza entegre edin.

| | |
|---|---|
| 10+ | Hazır Model |
| 256K | Bağlam |
| %100 | KVKK Uyumlu |

OpenAI · Qwen · deepseek · Mistral AI · Meta · Gemma 4 · NVIDIA

## LLMaaS NEDİR?

FixCloud LLMaaS; GPT-OSS, DeepSeek, Qwen, Llama ve Mistral dahil 10'dan fazla hazır modeli Türkiye'deki NVIDIA GPU altyapısı üzerinde hazır bir hizmet olarak sunar. API anahtarınızı oluşturun, dakikalar içinde kullanmaya başlayın. İhtiyacınıza uygun aylık paketi seçin ve kota aşımında tüketim bazlı ücretlendirmeden yararlanın.

## NEDEN FIXCLOUD?

- OpenAI uyumlu standart API
- %100 KVKK Uyumlu, Türkiye Veri Merkezi
- NVIDIA GH200 ile yüksek performans
- Dedicated endpoint seçeneği

### Kendi modeliniz mi var?

Özelleştirme desteği ve dedicated endpoint ile modelinizi size ayrılmış GPU'larda yayına alıyoruz.

## STANDART PAKETLER

### PRO

**$99** / ay

| | |
|---|---|
| Aylık token kotası | **50M token** |
| API anahtarı | ✓ |
| Tüm standart modeller | ✓ |
| Deep Search | ✓ |
| Research Jobs | **10 / ay** |
| Batch Jobs | ✓ |
| AI Studio | ✓ |
| Compare | ✓ |
| Hız limiti | **30 istek/dk** |
| Eşzamanlı istek | **4** |
| Aktivite kayıtları ve dışa aktarma | ✓ |
| Aylık harcama limiti | **Ayarlanabilir** |
| Destek | **Standart** |

EN POPÜLER

### BUSINESS

**$199** / ay

| | |
|---|---|
| Aylık token kotası | **100M token** |
| API anahtarı | ✓ |
| Tüm standart modeller | ✓ |
| Deep Search | ✓ |
| Research Jobs | **50 / ay** |
| Batch Jobs | **Öncelikli** |
| AI Studio | ✓ |
| Compare | ✓ |
| Hız limiti | **100 istek/dk** |
| Eşzamanlı istek | **16** |
| Aktivite kayıtları ve dışa aktarma | ✓ |
| Aylık harcama limiti | **Ayarlanabilir** |
| Destek | **Öncelikli** |

### ULTRA

**$499** / ay

| | |
|---|---|
| Aylık token kotası | **500M token** |
| API anahtarı | ✓ |
| Tüm standart modeller | ✓ |
| Deep Search | ✓ |
| Research Jobs | **Sınırsız** |
| Batch Jobs | **Öncelikli** |
| AI Studio | ✓ |
| Compare | ✓ |
| Hız limiti | **500 istek/dk** |
| Eşzamanlı istek | **64** |
| Aktivite kayıtları ve dışa aktarma | ✓ |
| Aylık harcama limiti | **Ayarlanabilir** |
| Destek | **Premium** |

## BİZE ULAŞIN


**TELEFON**
+90 212 344 00 44

**E-POSTA**
sales@fixcloud.ai

**WEB**
fixcloud.ai

**SELF SERVİS PORTAL**
console.fixcloud.ai

fixcloud

MODEL KATALOĞU VE ÖZELLİKLER

TÜRKİYE KURUMSAL BULUT PAZAR LİDERİ

**1000+**

kurumsal müşteri, 2016'dan beri

| | |
|---|---|
| **5** veri merkezi, 3 ülke | **60+ PB** kapasite |
| **50+** iş ortağı | **%100** yerli sermaye |

## Platform Avantajları

**OpenAI Uyumlu API**
Standart API, mevcut kodunuzla çalışır.

**Veri Egemenliği**
Verileriniz İstanbul Equinix IL4 veri merkezinde işlenir.

**Maliyet Kontrolü**
Harcama üst limitinizi tanımlayın.

**AI Studio ve Compare**
Modelleri web üzerinden test edin ve karşılaştırın.

## STANDART MODELLER

Kota üzerindeki kullanımda geçerli olan, 1M token başına girdi/çıktı fiyatlarıdır.

| MODEL | SAĞLAYICI | TÜR | BAĞLAM | GİRDİ $/1M | ÇIKTI $/1M |
|---|---|---|---|---|---|
| **GPT-OSS 120B** | OpenAI | Sohbet | 128K | **0,99** | **2,99** |
| **Qwen3.6 35B A3B** | Alibaba | Sohbet | 128K | **0,49** | **1,99** |
| **Gemma 4 26B A4B** | Google | Sohbet | 64K | **0,39** | **1,49** |
| **DeepSeek V4.1 Flash Max** | DeepSeek | Agent, Coding | 256K | **1,49** | **3,49** |
| **GLM-4.5-Air** | Z.ai | Sohbet | 64K | **0,99** | **3,49** |
| **Mistral Medium 3.5** | Mistral AI | Sohbet + Görsel | 64K | **1,49** | **3,99** |
| **Llama 3.3 70B Instruct** | Meta | Sohbet | 32K | **0,99** | **1,99** |
| **Whisper Large v3 Turbo** | OpenAI | Speech to Text | — | **1,49** | **2,49** |
| **Qwen3 Embedding** | Alibaba | Embedding | 32K | **0,39** | **0,99** |
| **BGE Reranker v2** | BAAI | Reranking | 8K | **0,19** | **0,49** |

**Token Ücretlendirmesi:** Paketinizdeki aylık token kotası dolduktan sonra kullanım durmaz; kota üzerindeki her 1 milyon token, kullandığınız modelin kendi girdi/çıktı fiyatı üzerinden ücretlendirilir ve dönem sonunda cüzdanınızdan tahsil edilir.

## FARKLI MODELLER, FARKLI YETENEKLER

**Chat**
Çok turlu sohbet için /v1/chat/completions.

**Agentic**
Çok adımlı ajan iş akışları için uygun.

**Coding**
Kod üretimi, açıklama ve düzeltme için optimize.

**Reasoning**
Yanıt öncesi düşünme adımları; reasoning alanında döner.

**Embedding**
Vektör gömme üretir (/v1/embeddings).

**Completion**
Ham metin tamamlama için /v1/completions.

**Vision**
Görselleri anlar ve açıklar.

**Streaming**
Token bazında akış (stream: true).

**Function Calling**
Fonksiyon çağrısı şemalarını destekler.

**JSON Schema**
response_format ile JSON şemasına uyan çıktı.

**Prefix Cache**
Tekrarlanan prompt önekleri önbellekten, daha hızlı ve ucuz.

**Batch Inference**
Asenkron toplu işler (Batch API).

**Image Input**
Mesajlarda görsel girdisi kabul eder.

**Tool Calling**
Tanımladığınız araçları OpenAI tools formatında çağırır.

**Structured Output**
Yapılandırılmış, ayrıştırılabilir çıktı üretir.

## BİZE ULAŞIN


**TELEFON**
+90 212 344 00 44

**E-POSTA**
sales@fixcloud.ai

**WEB**
fixcloud.ai

**SELF SERVİS PORTAL**
console.fixcloud.ai