Nornia
Akıllı İletişim Platformu — Kurumsal Kullanım Kılavuzu

v3.10.11
23 Kasım 2026

Bölüm 08: API Bütçe ve Tüketim Takibi

Google Gemini, OpenAI, ElevenLabs, Deepgram, Anthropic ve Groq yapay zeka sağlayıcılarının gerçek zamanlı token/dakika tüketimini, tahmini ABD Doları ($) maliyetini, aylık harcama tavanlarını ve bütçe uyarı eşiklerini yönettiğiniz finansal kontrol merkezidir.

BÖLÜM 08 | Operasyon & İzleme | Token • Maliyet • Kota | ~6 dk okuma

API Bütçe ve Tüketim Takibi

Google Gemini, OpenAI, ElevenLabs, Deepgram, Anthropic ve Groq yapay zeka sağlayıcılarının gerçek zamanlı token/dakika tüketimini, tahmini ABD Doları ($) maliyetini, aylık harcama tavanlarını ve bütçe uyarı eşiklerini yönettiğiniz finansal kontrol merkezidir.

SANTRAL MENÜ YOLU & URL

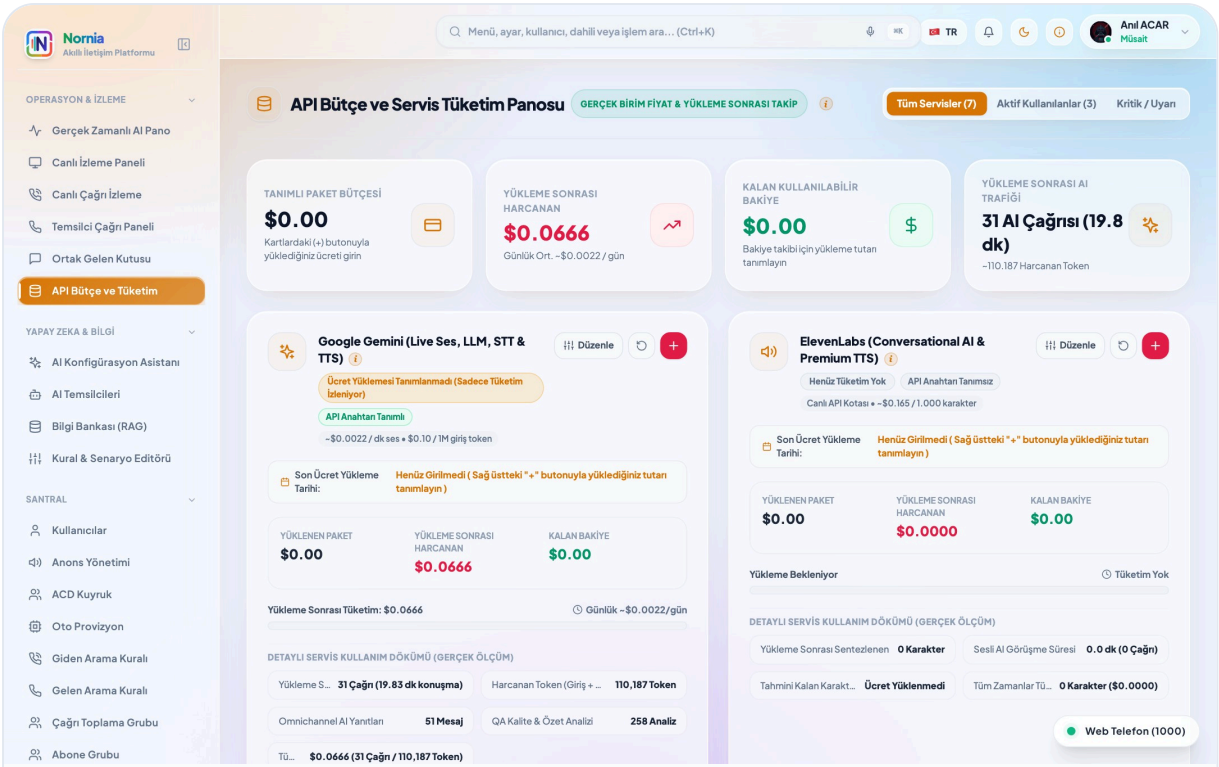
Operasyon & İzleme → API Bütçe ve Tüketim

`/?tab=api-butce-ve-tuketim`

GEREKLİ ROL YETKİ KODLARI (RBAC)

`api_budgets:read` `api_budgets:write`

CANLI SANTRAL ARAYÜZ EKRAN GÖRÜNTÜSÜ



API Bütçe ve Tüketim — Sağlayıcı bazlı token/dakika kullanımı ve aylık bütçe limitleri

Operasyon & İzleme → API Bütçe ve Tüketim

Tüm Alanlar, Parametreler & Teknik Etkileri

4 parametre

ALAN / SEKME ADI	TİP / SEÇENEKLER	VARSAYILAN	DETAYLI AÇIKLAMA & SA ETKİSİ
Aylık Bütçe Limiti ($)	Decimal USD	Örn. $100 / ay	Her sağlayıcı (Gemini, Op ElevenLabs vb.) için ayrı a harcama limiti.
Uyarı Eşiği (%)	Integer (50–95)	%80	Bütçe bu yüzdeye ulaştığ yöneticilere e-posta ve si bildirimi gönderilir.
Limit Aşımı Politikası	Seçim	Uyarı Gönder / Yerel LLM'e Düş	Bütçe dolduğunda AI gör otomatik daha ekonomik veya insan kuyruğuna yönlendirebilirsiniz.
Görev Kırılımı	Tablo / Grafik	Canlı	Canlı Sesli Asistan, Trans Analizi, QA Puanlama ve A Konfigürasyon Asistanı ba tüketim detayı.

Ekle / Düzenle & Alt Sekme Ayarları (Detaylı Referans)

Bu modülün içindeki her sekme, form alanı, anahtar (toggle) ve gizli davranışın tam teknik dökümü (1 alt bölüm).

1 Sağlayıcı Ücret/Paket Yükleme & Bütçe Eşik Düzenleme Modalları

7 AI & Kanal Sağlayıcısı (4 ayar)

Google Gemini, ElevenLabs, Meta WhatsApp Cloud API, OpenAI, Anthropic, Groq ve Deepgram için bakiye yükleme, bütçe sınırı ve sıfırlama pencereleri.

AYAR / ALAN	SEÇENEKLER / TİP	VARSAYILAN	NE İŞE YARAR & TEKNİK DA
Ücret / Paket Yükleme Modalı ZORUNLU	Hızlı Tutar ($ USD) + Mevcut Harcama	USD	Sağlayıcı paneline yüklediğ ($10/$25/$50/$100 veya öz kaydeder; harcama takibini veya mevcut tutardan başla
Toplam Yüklenen Paket Bütçesi ($ USD) ZORUNLU	Decimal	Sağlayıcı Bazlı	Kalan bakiye yüzdesini ve ta konuşma/mesaj kapasitesini
Kritik Uyarı Alarm Eşiği (%)	Integer (50–95%)	%80	Tüketim bu yüzdeye ulaştığ Ayarları → E-Mail & WhatsA Bildirimleri üzerinden yöneti anında uyarı gönderilir.
Yükleme Sonrası Tüketimi Sıfırla	Onaylı Aksiyon	Dönem Başında	Yeni fatura dönemine veya y yüklemesine geçerken saya temizler.

1. Token Tasarrufu ve Hibrit Model Optimizasyonu

API Bütçe ekranını Sistem Ayarları → AI Görev & Model Yönlendirme ile birlikte kullanarak büyük maliyet avantajı sağlayabilirsiniz: Örneğin canlı sesli görüşmelerde Gemini Live kullanırken, çağrı sonrası özet/QA puanlama işlemlerini daha ekonomik Flash/Yerel Ollama modellerine yönlendirerek aylık AI faturasını %60'a kadar düşürebilirsiniz.

Sık Karşılaşılan Hatalar & Sorun Giderme

Belirti: AI Asistan çağrıları aniden insan kuyruğuna düşmeye başladı

Kök Neden: Aylık API bütçe limiti %100'e ulaşmış ve sert durdurma politikası aktif olabilir.

Çözüm: API Bütçe ve Tüketim ekranından ilgili sağlayıcının aylık limitini artırın veya sağlayıcı panelinde kredi bakiyesini kontrol edin.