

# SIMON AI

## MASTER PROJE YÖNETİM DOKÜMANI

Altyapıdan Ürüne - Eksiksiz Yol Haritası

**Versiyon: v5.0 FINAL**




Tarih: 02 Ocak 2026

**Proje Sahibi:** Ceyhun Bey

**Teknik Lider:** Yapay Zeka Agent Ekibi

**Durum:** Altyapı Tamamlandı  | Ürünleştirme Başlıyor 

## DOKÜMAN BİLGİLERİ

Doküman Adı	Simon AI - Master Proje Yönetim Dokümanı
Versiyon	v5.0 FINAL - Altyapıdan Ürüne Geçiş
Son Güncelleme	02 Ocak 2026, 21:30
Kapsam	Web Platform + Mobil App + Agent Studio
Proje Durumu	Altyapı: 100%    Ürün: Başlangıç 
Toplam Süre	Altyapı: 18 gün (  )   Ürünleştirme: 19-27 gün (tahmini)
Hedef Lansman	Q1 2026 (Ocak-Mart)

# 1. YÖNETİCİ ÖZETİ

## 1.1 Proje Vizyonu

Simon AI, global ölçekte erişilebilir, hibrit ekonomili (FREE + FREE+ + BYOK), çoklu model destekli ve otomatik failover'lı bir AI platformudur.

## 1.2 Mevcut Durum

### ✓ Agent Studio Altyapısı: 100% TAMAMLANDI (18 gün)

- 10/10 Container UP + 7/7 HEALTHY
- Observability Stack: Prometheus + Grafana + Loki
- Güvenlik: Egress Proxy + API Key İzolasyonu + Audit Logs
- Network: Tüm servisler simon-network'te (172.20.0.0/16)

### 🕒 Ürünleştirme: BAŞLIYOR (19-27 gün tahmini)

- FAZ 5: Web MVP-1 (Temel Chat) - 3-5 gün
- FAZ 6-8: Web Pro Features - 9-13 gün
- FAZ 9: Mobil MVP - 7-10 gün

## 1.3 Hedef Ürünler






- **Simon AI Web:** Tarayıcı tabanlı platform (MVP → Beta → v1.0)
- **Simon AI Mobil:** Flutter ile iOS/Android (MVP → Beta)

## 2. PROJE DURUM ANALİZİ

### 2.1 Tamamlanan Altyapı Fazları

Faz	Hedef	Süre	Teslim	Başarı
FAZ 0	Güvenlik + Altyapı	4 gün	30.12.25	100%
FAZ 1	Task Decomposition	4 gün	30.12.25	100%
FAZ 2	Browser Sandbox	4 gün	30.12.25	100%
FAZ 3	Observability	3 gün	31.12.25	85%
FAZ 4	Production Hardening	3 gün	01.01.26	100%
TOPLAM	Agent Studio	18 gün	-	97%

### 2.2 Ürünleştirme Fazları (PLANLANDI)

Faz	Hedef	Süre	Durum
FAZ 5	Web MVP-1: Temel Chat UI	3-5 gün	 Sonraki
FAZ 6	Web MVP-2: Layout Pro	3-4 gün	 Planlı
FAZ 7	Web MVP-3: Widgets	2-3 gün	 Planlı
FAZ 8	Web Beta: Feedback+Admin	4-5 gün	 Planlı
FAZ 9	Mobil MVP: Flutter	7-10 gün	 Planlı
TOPLAM	Web + Mobil	19-27 gün	-

## 2.3 Container Durumu

10/10 Container UP | 7/7 HEALTHY 

Container	Port	Rol	Durum
simon-api	8000	FastAPI Orchestrator	HEALTHY
simon-litellm	4000	LiteLLM Gateway	UP
simon-ollama	11434	Local Models	HEALTHY
simon-postgres	5432	Database	HEALTHY
simon-redis	6379	Cache + Jobs	HEALTHY
simon-celery-worker	-	Background	UP
simon-egress-proxy	3128	Squid Proxy	HEALTHY
simon-prometheus	9090	Metrics	HEALTHY
simon-grafana	3000	Dashboards	HEALTHY
simon-loki	3100	Logs	HEALTHY

## 3. MODEL KATALOĞU VE EKONOMİ

### 3.1 Key Mode Sistemi

Mode	Anahtar	Model Kaynağı	Hedef Kullanım	Limit/Politika
FREE	Yok	Ollama (lokal)	0 maliyet, temel kalite	Donanım/performans sınırlı
FREE+	Yok	Sponsorlu server key pool	Hızlı deneme, kısa görevler	Kota + abuse kontrol
BYOK	Kullanıcı	OpenAI/Anthropic/Google	Limitsiz/kaliteli	Kullanıcı limitine bağlı

### 3.2 FREE: Ollama Kataloğu (Minimum 15 Model)

#### Hafif (Low RAM, Hızlı)

- qwen2.5:1.5b
- phi-4
- llama3.2:1b
- llama3.2:3b
- gemma2:2b

#### Genel Amaç (Daha İyi Kalite)

- qwen2.5:7b
- mistral:7b
- llama3.1:8b
- gemma2:9b
- deepseek-r1:7b (uygun sürüm)

#### Kod/Araç Odaklı

- qwen2.5-coder:7b
- deepseek-coder:6.7b
- codestral:22b (opsiyonel - donanıma bağlı)

## Özet/Akıl Yürütme (Opsiyonel)

- llama3.1:70b (cloud/self-host)
- mixtral:8x7b (cloud/self-host)

## 3.3 BYOK: En İyi 4 Model (Varsayılan)

- **OpenAI GPT-5.2:** Genel + agentic + coding
- **Anthropic Claude Opus 4.5:** Maksimum kalite
- **Anthropic Claude Sonnet 4.5:** Hız/kalite dengesi
- **Google Gemini 3 Pro:** Multimodal + uzun context + agentic

**Güncelleme Politikası:** Liste yılda en az 4 kez 'leaderboard + maliyet + stabilite' kriterleriyle güncellenir.

## 4. ÜRÜN ÖZELLİKLERİ (DETAYLI)

---

### 4.1 Ürün Vizyonu ve İlkeler

#### Vizyon

Global ölçekte erişilebilir, hibrit ekonomili, çoklu model destekli ve failover'lı bir AI platformu.

#### Temel İlkeler

- **Ultra Hızlı Algı:** İlk etkileşim < 10 saniye
- **Kesintisiz Deneyim:** Sağlayıcı hata/limit → otomatik geçiş
- **Güvenlik:** Anahtar izolasyonu, şifreli kasa, audit log
- **Sadelik:** Arayüz 'az ama etkili'; yoğun özellikler katlanabilir yüzeylerde

### 4.2 Web Platform - Arayüz Özellikleri

#### 4.2.1 Sol Sidebar (ChatGPT Benzeri)

- **Navigasyon:** Araçlar, proje, klasör, sohbet, ayarlar, profil, logo
- **Görünüm Modları:**
  - Açık mod (tam genişlik)
  - Icon-only kapalı mod (minimal)
- **Pin/Unpin:** Sık kullanılan öğeleri sabitleme
- **Hızlı Arama:** ⌘K / Ctrl+K kısayolu
- **Tema:** Dark mode + Light mode desteği

#### 4.2.2 Üst Bar (Top Bar)

- **Sabitlenebilir/Gizlenebilir:** Kullanıcı tercihi
- **İkon-only Mod:** Maksimum ekran alanı
- **Model & Key Mode Dropdown:**

- FREE: 15+ model görünür
- BYOK: 4 best model + kullanıcı custom
- Hızlı geçiş (tek tık)
- **Mini Dashboard:**
  - Token sayımı (anlık + toplam)
  - Kelime/karakter sayısı
  - Süre (session + toplam)
  - Proje/sohbet sayısı
  - Maliyet tracking (BYOK)

### 4.2.3 Sağ Panel (Widgets Dock)

- **Aç/Kapa İkonları:** Her widget için ayrı
- **Widget Kartları:**
  - Haberler (RSS feeds - allowlist)
  - Sosyal (Twitter/Reddit feeds)
  - Skor (Borsa, kripto, spor)
  - Özel linkler (kullanıcı ekler)
- **Sağ Alt Mini Slayt:**
  - Otomatik dönen kartlar (10s interval)
  - Seçili kaynağa göre içerik
  - Tıklanabilir (detay açar)

## 4.3 Chat Özellikleri

### 4.3.1 Temel Chat Yetenekleri

- **Streaming Yanıt:** Token by token (real-time)
- **Mesaj Düzenleme:** Geçmiş mesajları edit + yeniden gönder
- **Mesaj İşlemleri:**
  - Kopyala (copy to clipboard)
  - Sil (soft delete + undo 10s)

- Favori işaretleme (yıldız)
- Pin (üstte sabitle)
- **Code Highlighting:**
  - Syntax highlighting (100+ dil)
  - Tema seçimi (dark/light compatible)
  - Tek tık kopyala butonu
  - Çalıştır butonu (güvenli sandbox - opsiyonel)

### 4.3.2 İleri Rendering

- **Markdown:** Tam destek (GFM - GitHub Flavored Markdown)
- **LaTeX:** KaTeX ile matemat render
- **Mermaid:** Diagram çizimi
- **Tablo:** Responsive table render
- **Embed:** Video, iframe (allowlist)

### 4.3.3 Dosya İşlemleri

- **Upload:** Drag & drop, multi-file, progress bar
- **Desteklenen Formatlar:**
  - Doküman: PDF, DOCX, TXT, MD
  - Kod: PY, JS, TS, JAVA, CPP, vb.
  - Görsel: PNG, JPG, GIF, WEBP
  - Data: CSV, JSON, XML, XLSX
- **Download:** Sohbet export (PDF, MD, JSON)

## 4.4 Proje ve Sohbet Yönetimi

### 4.4.1 Proje Hiyerarşisi

- **Klasör/Proje:** İç içe klasör desteği (sınırsız derinlik)
- **Sohbet Tagging:** Çoklu tag, renk kodlama
- **Arama + Filtreleme:**

- Tarih aralığı
- Model bazlı
- Tag bazlı
- Full-text search
- **Sıralama:** Tarih, isim, favori, kullanım sıklığı

#### 4.4.2 Senkronizasyon

- **Çoklu Cihaz:** Web + Mobil + Desktop senkron
- **Offline Destek:** Local cache, otomatik sync
- **Conflict Resolution:** Otomatik merge + manuel çözüm

#### 4.5 UI Kalite Kriterleri

- **Hız:**
  - İlk yükleme < 3 saniye
  - Navigasyon < 100ms
  - 60 FPS scroll/animasyon
- **Responsive:**
  - Desktop (1920x1080 optimal)
  - Tablet (landscape/portrait)
  - Mobile (360px+ genişlik)
- **Accessibility:**
  - WCAG 2.1 AA uyumluluk
  - Klavye navigasyon
  - Screen reader uyumlu
- **Renk Paleti:** Kurumsal (profesyonel, sade, şık)

## 5. KRİTİK ÜRÜN YETENEKLERİ

---

### 5.1 Geri Bildirim Sistemi + AI Triage

#### 5.1.1 Kullanıcı Tarafı

- **Tek Form:** Hata / İyileştirme / Talep (dropdown)
- **Otomatik Bağlam:**
  - Tarayıcı (user agent)
  - Ekran çözünürlüğü
  - App versiyon
  - Seçili model + key mode
  - Son 50 log satırı (PII maskeleye ile)
- **Repro Adımları:** Otomatik taslak (AI ile)
- **Screenshot:** Opsiyonel (otomatik blur sensitive data)

#### 5.1.2 AI Triage Agent

##### Akış:

1. Feedback alınır → DB'ye kaydedilir (status: NEW)
2. Feedback Agent analiz eder:
  - Problem özeti
  - Kök neden
  - Çözüm planı
  - Risk skoru (LOW/MEDIUM/HIGH)
  - Test planı
3. Sistem admin'e: "Bu fix'i uygulayayım mı?" sorusu
4. **Admin ONAY:** Otomatik PR/patch + CI test + deploy
5. **Admin RED:** Issue kapanır veya manuel aksiyon

**Güvenlik:** Asla doğrudan prod'da değişiklik yok; her şey PR + CI + tag + rollback ile.

## 5.2 Admin Panel

### 5.2.1 Model Debate Arena (Model Tartışırma)

- **Paralel Çalıştırma:** Aynı prompt'u 2-5 modele gönder
- **Karşılaştırma:**
  - Argüman haritası (side-by-side)
  - Konsensus analizi (ortak noktalar)
  - Farklılık vurgulama
- **Hakem Modu:** Seçilen modelin diğer çıktıları değerlendirip sentez üretmesi
- **Export:** Karşılaştırma raporu (PDF/MD)

### 5.2.2 Agentic Factory (Tek Tuş Ürün Üretimi)

- **Uygulama Üretimi:**
  - Repo scaffold (folder structure)
  - UI (component library)
  - API (endpoint + database schema)
  - Test suite
  - Deployment config
- **Tasarım Üretimi:**
  - Tema (color palette + typography)
  - Component library (React/Vue)
  - Design system dokümanı
- **Doküman Üretimi:**
  - PDF raporlar (profesyonel)
  - DOCX dokümanlar
  - Markdown + HTML
- **Web Sitesi Üretimi:**

- Landing page
- İçerik (SEO optimized)
- Blog template






**Güvenlik:** Bu modül güvenli sandbox içinde çalışır ve admin onayı olmadan dış sistemlerde işlem yapmaz.

## 5.3 Failover + Model Routing

- **Otomatik Failover:** Sağlayıcı hata/limit → alternatif model
- **Maksimum Deneme:** 2 failover (sonsuz döngü yok)
- **Kullanıcıya Bildirim:** "Model değiştirildi (Sebebi: timeout/limit)"
- **Öncelik Sırası:**
  1. Seçili model
  2. Aynı sağlayıcı küçük model
  3. Alternatif sağlayıcı eşdeğer model
- **Maliyet Farkındalığı:** Kullanıcıya failover maliyeti göster (BYOK)

## 6. TEKNİK MİMARİ

### 6.1 Backend Stack (MEVCUT )

Katman	Teknoloji	Durum
Orchestrator	FastAPI + PostgreSQL	 Çalışıyor
AI Gateway	LiteLLM + Redis	 Çalışıyor
UI Runner	Playwright + Celery	 Çalışıyor
Network Security	Squid Proxy	 Çalışıyor
Observability	Prometheus + Grafana + Loki	 Çalışıyor

### 6.2 Frontend Stack (GELECEK)

- **Framework:** Next.js 14+ (App Router)
- **UI Library:** shadcn/ui + Tailwind CSS
- **State Management:** Zustand + React Query
- **Real-time:** WebSocket (streaming chat)
- **Hosting:** Vercel / Cloudflare Pages
- **CDN:** Cloudflare (global edge caching)

### 6.3 Mobil Stack (GELECEK)

- **Framework:** Flutter 3.x
- **State Management:** Riverpod / Bloc
- **Local Storage:** Hive / Drift (offline)
- **Push Notifications:** Firebase Cloud Messaging

- **Deployment:** App Store + Google Play

## 7. GÜVENLİK VE COMPLIANCE

---

### 7.1 API Key İzolasyonu

- API anahtarları **SADECE** Orchestrator + LiteLLM'de
- UI Runner: **SIFIR** credential erişimi
- Secrets: Environment variables (prod: HashiCorp Vault)
- Otomatik key rotation: 90 günde bir

### 7.2 Veri Gizliliği (KVKK/GDPR)

- Screenshot TTL: 30 gün otomatik purge
- PII detection: telefon, email, kredi kartı maskeleyme
- GDPR Article 17: Right to Erasure
- Encryption: AES-256 at rest, TLS 1.3 in transit

### 7.3 Network Güvenliği

- Egress Proxy: Domain allowlist enforcement
- DPI: Protocol validation
- Traffic logging: Forensic analysis
- Firewall: Strict inbound/outbound rules

## 8. MALİYET VE ROI

### 8.1 Maliyet Planı

Bileşen	İlk Ay	Normal	Notlar
Claude Sonnet 4.5	\$50	\$25	Prompt caching %90 tasarruf
OpenAI GPT-4o	\$10	\$8	Batch processing
Ollama (Local)	\$0	\$0	Self-hosted
Egress Proxy	\$4	\$4	Hetzner VPS
Hosting	\$4	\$4	Vercel + Railway
<b>TOPLAM</b>	<b>\$68</b>	<b>\$41</b>	<b>Hedef: ≤\$40/ay</b>






### 8.2 ROI Analizi

- **Baseline:** Geleneksel developer \$8,000/ay
- **Simon AI:** \$41/ay
- **Tasarruf:** \$7,959/ay (%99.5 maliyet azalışı)
- **Break-even:** 1 gün (ilk task)
- **Yıllık tasarruf:** \$95,508

## 9. SONRAKİ ADIMLAR

---

### 9.1 Acil Aksiyonlar (Hafta 1)

-  FAZ 4 tamamlandı (Production hardening)
-  **FAZ 5 başlat:** Web MVP-1 (Temel Chat UI)
-  Next.js proje scaffold
-  Model dropdown + Key Mode UI
-  Streaming chat implementasyonu

### 9.2 Orta Vadeli (Hafta 2-4)

- Layout Pro: Sol/üst/sağ paneller
- Widget ekosistemi
- Geri bildirim + AI triage
- Admin panel (debate + factory)

# EKLER

---

## A. Proje Konumu

- Ana Dizin: C:\Users\ceyhu\Desktop\simonai
- SSH: ceyhu@localhost (key-based)
- Reports: C:\Users\ceyhu\Desktop\simonai\reports

## B. Endpoint'ler

- API Health: http://localhost:8000/health
- API Metrics: http://localhost:8000/metrics
- API Docs: http://localhost:8000/docs
- Prometheus: http://localhost:9090
- Grafana: http://localhost:3000 (admin/SimonAI@2026!Secure)

## C. Referans Dokümanlar

- Simon\_AI\_v3\_1\_FINAL\_CORRECTED.pdf
- SimonAI\_Master\_Proje\_Dokumani\_v3\_1\_TR.pdf
- Yapay\_Zeka\_Talimatları\_v7.1.docx
- FAZ\_0-4 Raporları (reports/ klasörü)

---

— DOKÜMAN SONU —

Son Güncelleme: 02 Ocak 2026, 21:30