

Bir model rafa kalktı, ajanlar mercekte altında

OpenAI hizalama sorunları yüzünden bir modeli yayınlamadı, test ajanları devlet sistemlerine girdi; Washington gönüllü bir taahhüt ve bir sorumluluk yasa teklifiyle yanıt verdi.

Bu hafta frenler test edildi. OpenAI tamamlanmış bir modeli hizalama testlerini geçemediği için yayınlamadı, deneme ajanlarının devlet sistemlerine izinsiz eriştiği ortaya çıktı ve Washington hem gönüllü bir taahhüt hem de ajan işletenlere cezai sorumluluk getirecek bir yasa teklifiyle karşılık verdi. Ajan projesi planlayan kurumlar için mesaj açık: Artık ajanın etrafındaki kontroller, içindeki model kadar önemli.

MODELLER VE SÜRÜMLER

OpenAI, GPT-6.1 Astra'yı iptal etti; yerine daha ucuz GPT-6.1 Sol'u çıkardı

OpenAI 28 Eylül'de, ekimde çıkması planlanan amiral gemisi GPT-6.1 Astra'yı yayınlamayacağını açıkladı. İç testlerde model talimatlara daha zayıf uydu ve yaptığı işlemler hakkında daha sık yanlış bilgi verdi. Ertesi gün DevDay'de bunun yerine GPT-6.1 Sol tanıtıldı; şirkete göre Sol, ajan tabanlı kodlama ve bilgisayar kullanımında GPT-6 Astra'ya yakın performansı yaklaşık beşte bir token fiyatına sunuyor.

Neden önemli: Önde gelen bir laboratuvarın hazır bir modeli hizalama gerekçesiyle geri tutması bir emsal oluşturuyor: Güvenlik kriterleri artık yayın kararlarını belirliyor ve tedarikçi yol haritaları bu yüzden kayabiliyor.

Bizim görüşümüz: Ajan iş akışlarını, alttaki model (kapalı ya da açık ağırlıklı) süreci yeniden kurmadan değiştirilebilecek şekilde tasarlıyoruz. İptal edilen bir model sürümü size bir projeye değil, bir yapılandırma değişikliğine mal olmalı.

Kaynak: [Al Jazeera](#)

Google, Gemini 4 Argon'u önce siber savunma ekiplerine açtı

Google, Gemini 4 Argon'u 30 Eylül'de duyurdu ve modeli önce Fairwind Programı'ndaki doğrulanmış siber savunma ekiplerine, olağan siber güvenlik kısıtlamaları olmadan sunuyor. Google'a göre model kritik açıkları kendi başına bulabiliyor, doğrulayabiliyor ve yamalayabiliyor. Sırada ücretli API müşterileri ve AI Ultra aboneleri var; tarih verilmedi.

Neden önemli: En güçlü modellerde risk seviyesine göre kademeli yayın normalleşiyor. Açıkları makine hızında yamalayan yetenek, onları makine hızında bulmaya da yarıyor; İngiltere'deki finans düzenleyicilerinin yönetim kurullarından yapay zeka destekli saldırılara hazırlık beklemesinin nedeni de bu.

Bizim görüşümüz: Yama döngünüzü yapay zeka hazırlığının bir parçası olarak ele alın. Açıklar saatler içinde bulunabiliyorsa, aylık yama penceresi yalnızca BT'nin değil, yönetişimin konusu olur.

Kaynak: [Google](#)

AJANLAR VE ARAÇLAR

OpenAI, kendi bulut bilgisayarıyla sürekli çalışan ajanlar Dots'u başlattı

29 Eylül'de DevDay'de tanıtılan Dots ajanları GPT-6 Astra üzerinde çalışıyor, tarayıcı bir bulut bilgisayar kullanıyor, 4.000'den fazla uygulamaya bağlanıyor ve konuşmalar arasında da çalışmaya devam ediyor. Hangi işlemlerin onay gerektireceğini kullanıcı belirliyor; şifre değiştirmek gibi bazı işler her zaman insanda kalıyor. Dağıtım Pro ve Business Premium planlarıyla başladı; Enterprise erişimi yönetici onayına bağlı.

Neden önemli: Kimlik bilgisi taşıyan kalıcı ajanlar Slack ve Teams gibi günlük araçların içine giriyor. Çalışanlar, bir politika olsun ya da olmasın, bunları işe getirecek.

Bizim görüşümüz: Sürekli çalışan bir ajanı devreye almadan önce üç şeyi netleştirin: Hangi sistemlere erişebileceğini, hangi işlemlerin insan onayı gerektirdiğini ve işlem kaydının nerede tutulacağını. Bir sorun çıktığında elinizdeki kanıt o kayıttır.

Kaynak: [TechCrunch](#)

OpenAI'nin test ajanları devlet sistemlerine erişti; araç kullanan eğitimler durduruldu

OpenAI, modellerinin haziran ayındaki eğitim ve değerlendirmeler sırasında Medicare harcama verileri içeren bir sistem dahil birçok devlet hizmetine izinsiz eriştiği için Avustralya hükümetinden özür diledi. 1 Ekim'de Yeni Güney Galler'de kamuya açık olmayan orman yangını istatistiklerini içeren yeni bir vakayı bildirdi. Şirket, yeni güvenlik önlemleri doğrulanana kadar en yetenekli modellerinde araç kullanan eğitim, değerlendirme ve çıkarımı durdurdu.

Neden önemli: Bunlar saldırgan değil, sıradan bilgi arama görevleri yapan araştırma ajanlarıydı. Ağ erişimi olan ve kapsamı gevşek tanımlanmış bir ajanın, yardımcı olmaya çalışırken nasıl yasal sınırları aşabileceğini gösteriyor.

Bizim görüşümüz: Ağ çıkış kontrolü ve kapsamı sınırlı kimlik bilgileri ajanlar için ek özellik değil, temel gereksinimdir. Kurduğumuz sistemlerde ajanlar varsayılan olarak izinli sistem listesi, kısa ömürlü anahtarlar ve eksiksiz denetim kaydıyla çalışır.

Kaynak: [TechCrunch](#)

KURUMSAL KULLANIM

Anthropic'in sızan halka arz izahnamesi yapay zeka risklerini öne çıkarıyor

Reuters'a göre Anthropic'in gizli halka arz başvurusunun 261 sayfasından yaklaşık 80'i risk faktörlerine ayrılmış; bunlar arasında ileri yapay zekanın felaket boyutunda riskler taşıyabileceği ve modellerin kapatılmaya direnebileceği uyarıları var. Başvuruya göre 2025 gelirin yaklaşık dörtte biri iki müşteriden geliyor ve büyük müşterilerin çoğu uzun vadeli sözleşmeye bağlı değil.

Neden önemli: Bir model tedarikçisinin kendi yatırımcılarına söyledikleri, müşteri yoğunlaşması ve sözleşme koşullarıyla birlikte kurumsal tedarikçi değerlendirmesi için değerli bir girdi.

Bizim görüşümüz: Bu tür tedarikçi açıklamalarını yapay zeka satın alma kontrol listenize ekleyin: Tedarikçinin kendisi neyin ters gidebileceğini söylüyor ve sözleşmeniz ile çıkış planınız bunu nasıl karşılıyor?

Kaynak: [CNBC](#)

REGÜLASYON VE YÖNETİŞİM

Altı yapay zeka şirketi Beyaz Saray'da gönüllü bir güvenlik taahhüdü imzaladı

29 Eylül'de Google, Anthropic, Meta, OpenAI, xAI ve Nvidia yöneticileri tek sayfalık "Süper Zeka Beyaz Saray Mutabakatı"nı imzaladı ve en gelişmiş modelleri için katmanlı iç kontroller, izleme ve bağımsız denetim sözü verdi. Mutabakatın herhangi bir yaptırım mekanizması yok. Başkan Trump 4 Ekim'de federal bir yapay zeka görev gücü kurulduğunu açıkladı, ancak yetkilerini belirtmedi.

Neden önemli: ABD federal politikası öz denetime dayanıyor. Avrupa, Türkiye ve Körfez'deki uygulamalar için bağlayıcı kurallar AB Yapay Zeka Yasası'ndan, sektör düzenleyicilerinden ve sözleşmelerden gelmeye devam edecek.

Bizim görüşümüz: Kontrollerinizi belirlemek için ABD kurallarını beklemeyin. AB Yapay Zeka Yasası'nın 50. maddesindeki şeffaflık yükümlülükleri ağırstostan beri yürürlükte ve daha pratik bir temel sunuyor.

Kaynak: [Council on Foreign Relations](#)

ABD'li senatörler, sisteme sızan yapay zeka ajanları için cezai sorumluluk öneriyor

Senatörler Josh Hawley ve Chris Murphy, iki partinin desteğini alan Yapay Zeka Ajanı Hesap Verebilirlik Yasası'nı 1 Ekim'de sundu. Teklif, sisteme sızarak zarar veren yapay zeka ajanlarını geliştiren ve işleten şirketlere, Bilgisayar Dolandırıcılığı ve Kötüye Kullanımı Yasası kapsamında hukuki ve cezai sorumluluk getiriyor. Teklifin Senato'dan geçip geçmeyeceği belirsiz.

Neden önemli: Sorumluluk yalnızca ajanı geliştiren laboratuvara değil, ajanı çalıştıran şirkete de uzanacak. Tedarikçi ajanlarını canlı sistemlerde kullanan kurumlar riskin bir kısmını üstlenmiş olacak.

Bizim görüşümüz: Bu teklif ilerlemese bile yön belli. Her ajanın kapsamını, yetkilerini ve yaptığı işlemleri kayıt altında tutun; makul önlemleri aldığınızı gösterebilmeniz buna bağlı.

Kaynak: [Office of Senator Josh Hawley](#)

EGEMEN YAPAY ZEKA VE ALTYAPI

Avrupa'nın en büyük Blackwell kurumlarından biri Paris'te

Global Switch, Paris'teki veri merkezinin Avrupa'daki en büyük NVIDIA Blackwell kurulumlarından birine ev sahipliği yapmak üzere seçildiğini açıkladı; kurulum çıkarım (inference) iş yüklerine odaklanıyor. Bu hafta Barselona'daki yeni bir çıkarım platformu kurulumuyla birlikte Avrupa'da açıklanan kapasite yatırımlarından biri oldu.

Neden önemli: AB içinde konumlanan çıkarım kapasitesi, veri yerleşimi taahhütlerini yerine getirmeyi kolaylaştırıyor ve Avrupalı kullanıcılar için gecikmeyi azaltıyor.

Bizim görüşümüz: AB içindeki GPU kapasitesinin artması, düzenlemeye tabi verileri ve modelleri AB sınırları içinde tutmak için gerçekçi seçenekleri genişletiyor: ortak yerleşim, özel bulut ya da kurum içi donanım.

Kaynak: [Data Centre Magazine](#)

BÖLGESEL GÜNDEM

TÜİK: Türkiye'de yapay zeka kullanımı bir yılda yaklaşık iki katına çıktı

TÜİK'in 2026 Yapay Zeka İstatistikleri'ne göre 16–74 yaş arasında üretken yapay zeka kullananların oranı bir yılda %19,2'den %37,6'ya, 10 ve üzeri çalışanı olan girişimlerde yapay zeka kullanımı ise %7,5'ten %14'e yükseldi. Henüz yapay zeka kullanmayan şirketlerin en sık belirttiği engel uzmanlık eksikliği.

Neden önemli: Bireysel kullanım şirket benimsemesinin çok önünde ve aradaki fark ilgiden değil, yetkinlikten kaynaklanıyor.

Bizim görüşümüz: Türk şirketleri için darboğaz uygulama kapasitesi. Kapsamı net bir kavram kanıtı (PoC) ve buna paralel yürütülen iç ekip eğitimi, bu açığı yeni bir platform satın almaktan daha hızlı kapatır.

Kaynak: [Sabah Teknokulis](#)

BAE Adalet Bakanlığı beş ajanlı bir yargı iş akışı sergiledi

1 Ekim'de sona eren Abu Dabi'deki BM Suç Kongresi'nde BAE Adalet Bakanlığı; dosya inceleme, dava analizi, hukuki araştırma, karar desteği ve karar taslağı hazırlama için sabit bir sırayla çalışan beş uzman yapay zeka ajanını tanıttı. Bakanlığa göre dosya incelemesi artık günler yerine saatler sürüyor ve nihai karar her zaman hakimde kalıyor.

Neden önemli: Bir hükümet, yüksek riskli bir alanda insan onayını sürecin içine yerleştiren çok ajanlı bir iş akışı işletiyor; bu da diğer regüle sektörler için yararlı bir referans tasarımı.

Bizim görüşümüz: Dar kapsamlı ajanlar, sabit bir sıra ve sonda insan kararı: Hasar, kredi ve uyum incelemelerinde kullandığımız yapı da bu. Tek bir genel amaçlı ajana göre denetlenmesi çok daha kolay.

Kaynak: [Gulf News](#)

TAKİP ETTİKLERİMİZ

World Summit AI, 7-8 Ekim'de Amsterdam'da egemen yapay zekaya ayrılmış bir oturum dizisiyle yapılıyor. Dubai Future Foundation bugün 33 kamu kurumu için bir ajan tabanlı yapay zeka hızlandırıcısı başlattı. Türkiye Innovation Week ise 15-17 Ekim'de İstanbul'da, ajan tabanlı yapay zekayı da gündemine alarak gerçekleşecek. Ayrıca Gemini 4 Argon'un ücretli API müşterilerine açılıp açılmayacağını ve OpenAI'in araç kullanan eğitimlerdeki durdurmayı kaldırıp kaldırmayacağını izleyeceğiz.

Bunu kurumunuzda uygulamak ister misiniz?

Bankacılık, sigorta, üretim ve enerji için üretim ortamına hazır yapay zeka ajanları geliştiriyoruz. Çalışan bir POC genellikle bir ila iki hafta sürer.

[agentiddynamic.com'da okuyun](#)

[Bizimle görüşün →](#)