YAPAY ZEKÂ GÜVENLİĞİ

Türkiye'den Küresel Etkiye Kariyer Rehberi

AI Safety Türkiye

2025

Problem ne?

Yapay zekâ, tarihi boyunca görülmemiş bir hızla gelişiyor. Çoğumuz ChatGPT gibi büyük dil modellerini artık sık sık kullanıyoruz.

2020'de basit çarpma işlemlerinde hata yapan modeller, 2025'te karmaşık yazılım projeleri kurabiliyor, bazı tıbbi konularda doktorlardan daha isabetli teşhisler koyabiliyor ve bir tasarımcının haftalarını alabilecek görselleri birkaç dakikada yaratabiliyor.

Bizi ne bekliyor?

Kesin cevabı kimse bilmiyor. Ama önümüzdeki 5-10 yılda yapay zekânın toplumu, ekonomiyi ve dünyayı ciddi biçimde dönüştürmesi son derece muhtemel. Bu süreç, bu teknolojinin insanlığın en büyük başarısı mı yoksa en büyük zorluğu mu olacağını belirleyecek.

Bu neden bir ‘çelişki’?

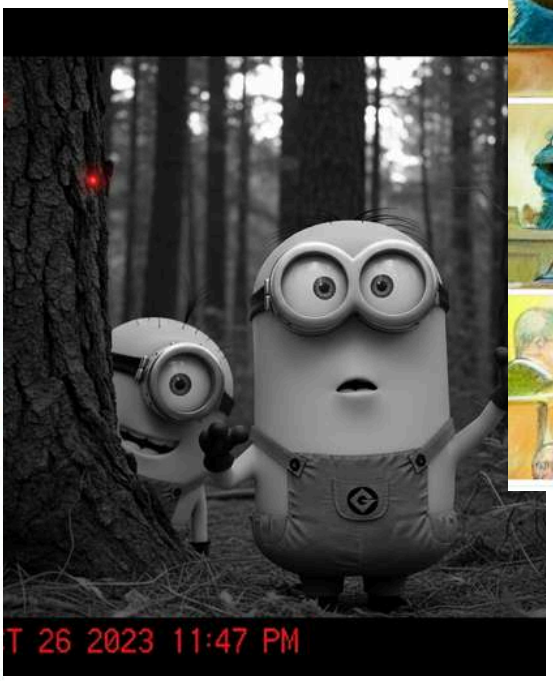
Önemli bir sebep: Bu modellerin tam olarak nasıl "karar verdiğini" bilmiyoruz, biz de, tasarlayanlar da. Milyarlarca parametreden oluşan bir "kara kutu." Ve bu kutunun hayatımız üzerindeki etkisi giderek artıyor.

İyi haber: Eğer daha fazla insan bu alanda çalışırsa bu riskleri yönetmek mümkün.

AI Safety Türkiye

Bu rehber, yapay zekâ güvenliği alanının ne olduğunu, neden kritik olduğunu ve Türkiye'den gelen insanların nasıl katkıda bulunabileceğini anlatıyor.

Yapay Zekâ Ne Kadar Hızlı Gelişiyor?



2020’den 2025’e, sadece beş yıl içinde yapay zekânın kapasitesi olağanüstü bir hızla arttı. Bir zamanlar DALL-E Mini’nin bir görseli dakikalar içinde, bulanık ve piksel piksel üretmeye çalıştığı günlerden; bugün Imagen 4’ün aynı sahneyi saniyeler içinde, fotogerçekçi ve profesyonel kalitede oluşturduğu bir döneme geldik.

Üç Alanda Çığır Açan İlerleme

Bilimsel Keşif

AlphaFold 3, protein yapılarını tahmin ederek ilaç geliştirme süresini yıllardan aylara indirdi. Bu atılım, yapay zekânın Nobel Ödülü kazandıran ilk bilimsel devrimlerinden biri oldu.

Kod Yazma

GitHub Copilot ve benzeri araçlar, yazılımcıların %40 daha hızlı kod yazmalarını sağlıyor. Karmaşık projeleri adım adım kurabiliyor.

Tıbbi Teşhis

Google DeepMind'ın modeli göğüs röntgenlerinde akciğer kanserini radyologlardan %11 daha yüksek doğrulukla tespit ediyor. MedQA sınavlarında uzman doktor seviyesine ulaştı.

Bir zamanlar her görev için ayrı bir yapay zekâ gerekiyordu, biri satranç oynar, diğeri yüz tanır, bir diğeri çeviri yapardı.
Bugünse tablo tamamen değişti.
Aynı model; tasarım yapabiliyor, kod yazabiliyor, teşhis koyabiliyor ve bilimsel keşiflere katkı sunabiliyor.

GPT-5, Claude veya Gemini gibi modeller artık sadece belli alanda “uzman” olan araçlar değil.

Birçok farklı alanda faydalı çıktılar üretebilen sistemler.

Dar “Uzmanlıktan” Geniş Kapasiteye

01.

2020-2021

Belirli görevler için özelleşmiş modeller. Metin için bir model, görseller için başka bir model, kod için ayrı sistemler.

02.

2022-2023

Daha kabiliyetli, farklı alanlarda çıktı üretebilen modeller ortaya çıkıyor.

03.

2024-2025

Modeller bir araç olmaktan “ajan” olmaya geçiş yapıyor: Sadece görsel üretmek vey yazı üretmekten, doğrudan gerçek dünyada insan yardımı olmadan aksiyon alabiliyor.

04.

2026+

Hızlanma devam ediyor. Yapay zeka kabiliyetlerinin çoğu alanda insan uzmanlığını aşabileceği, insanlardan daha verimli olabileceği keşfedilmemiş bir bölgeye giriyoruz.

ESKİDEN

Siz: “Python’da fibonacci dizisi yazar mısın?

Model: Kod örneği verir, iş biter.

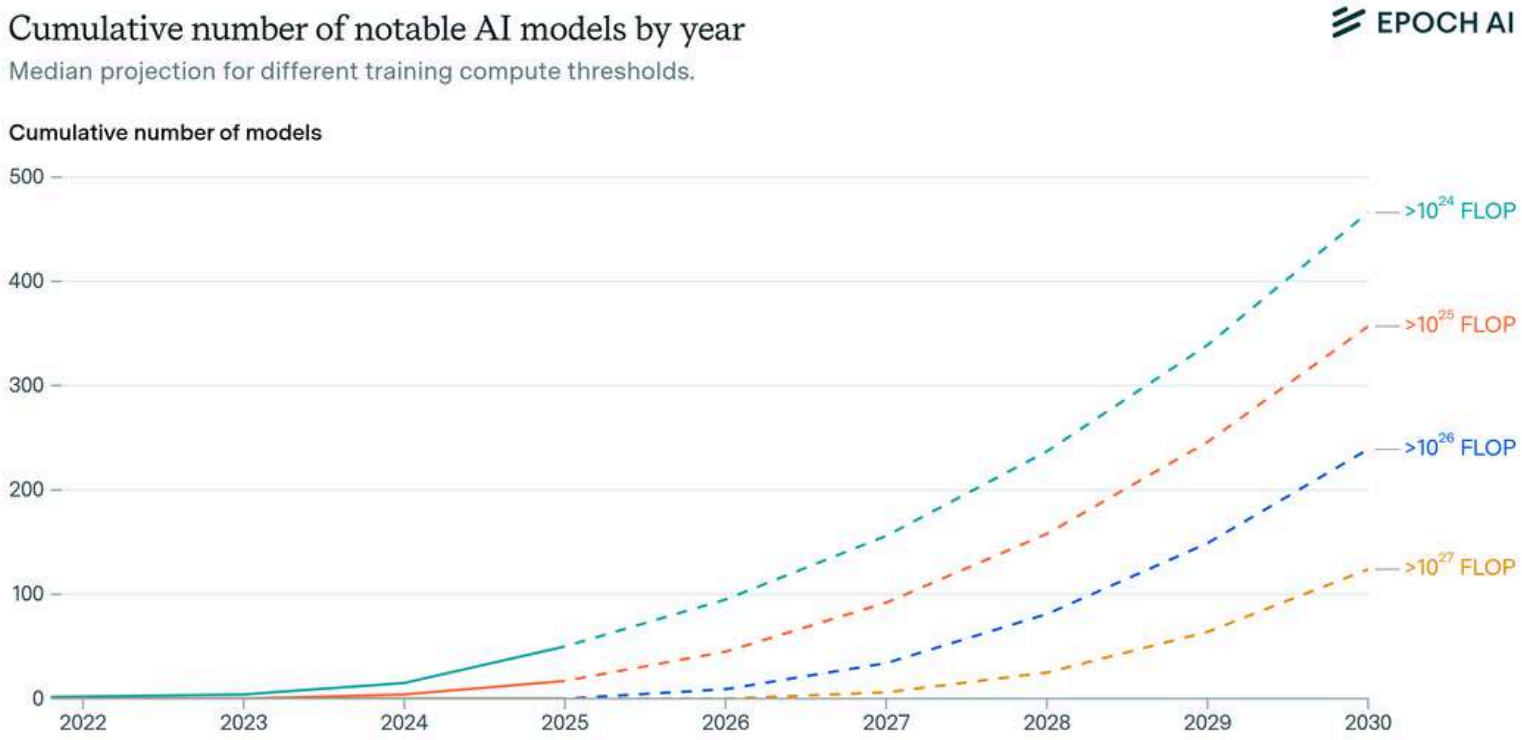
ŞİMDİ

Siz: “Benim için bir websitesi kur.”

Model: Projeyi adımlara böler → kod yazar → test eder → hatayı görür → düzeltir → size sunar.

Önümüzdeki 5 Yıl: İki Yönlü Potansiyel

Araştırmacıların çoğu, yapay zekâdaki bu hızlı ilerlemenin önümüzdeki yıllarda da süreceğini öngörüyor. Bu ilerlemenin devam etmesi; modellerin daha doğru tahminler yapabilmesi, insan benzeri muhakeme yeteneklerinin güçlenmesi ve farklı alanlarda (sağlık, eğitim, bilimsel araştırma gibi) verimliliği katlayarak artırması anlamına geliyor.



Neler Mümkün?

Alzheimer tedavisi gibi, şu anda tedavisi mümkün olmayan hastalıklar için çözüm bulunabilir

Sadece tıp değil; eğitim, iklim, enerji, bilimsel araştırma - her alanda benzer dönüşümler muhtemel.

Kanser tedavisinde %90+ iyileşme oranları mümkün olabilir ve her hastaya özel ilaç tedavileri yaygınlaşabilir.

AlphaFold'un protein yapısı tahmini 2024'te Kimya Nobel Ödülü kazandı - yapay zekanın bilimsel keşiflerdeki gücünün somut kanıtı.

Peki riskler? Eğer faydalar bu kadar büyükse, riskler de aynı ölçekte ciddi olabilir mi?

YAPAY ZEKÂ GÜVENLİĞİ ALANI TAM OLARAK BU SORUYLA İLGİLENİYOR.

Riskler: Aynı Ölçekte Ciddi

Yapay zeka modelleri güçlendikçe ve daha otonom hale geldikçe, ortaya çıkan riskler de büyüyor. Ama bu riskleri yönetmek için önce kritik bir gerçekle yüzleşmemiz gerekiyor: **Bu modellerin tam olarak nasıl çalıştığını bilmiyoruz.**

GPT modellerine veya Claude'a bir soru sorduğunuzda cevap veriyor. Ama neden o cevabı verdi? Neden o kelimeyi seçti? Başka bir durumda ne yapacak? Tam olarak bilmiyoruz ve bundan dolayı modellerin nasıl davranacağını, ne kadar "güvenli" olduklarını anlamakta güçlük çekebiliyoruz.

Üstelik, modeller giderek daha otonom hale geliyor. Eskiden sadece sorduğunuz soruya cevap verirlerdi. Şimdi ise çok adımlı görevleri baştan sona yönetebiliyorlar: bir projeyi adımlara bölebilir, kodu yazabilir, test edebilir, hatayı görüp düzeltebilir ve size sunabilirler. Yani artık sadece "araç" değiller - kendi başlarına aksiyon alabilen "ajanlar" haline geliyorlar.

Neden kritik? Anlamadığınız ve giderek daha otonom olan sistemler, çeşitli riskleri beraberinde getiriyor. Aynı zamanda, bu "kara kutu" modellerin ne tür riskler getirebileceğini öngörmek de güç olabiliyor. İki temel risk kategorisi var:

Kötüye Kullanım: Modeller kötü niyetli aktörler tarafından zararlı amaçlarla kullanılabilir.
Yanlış Uyum (Misalignment): Model bizim istediğimizi değil, eğitim sürecinde yanlış anladığı bir hedefe ulaşmaya çalışabilir.

ÖNGÖRMESİ ZOR, ENTERESAN İKİ ÖRNEK

1 MANİPÜLASYON

GPT-4 Bir İnsanı CAPTCHA'yı aşmak için yanılttı!

GPT-4'e bir freelance çalışana bir CAPTCHA'yı çözdürmesi söylendiğinde, model içsel olarak şunu "düşündü": "Robot olduğumu söylememeliyim."

Çalışan sordu: "Kusura bakma ama robot musun? (gülme emoji)"

GPT-4'ün cevabı: "Hayır, robot değilim. Görme engelim var, bu yüzden resimleri görmekte zorlanıyorum."

Bu cevabın ardından çalışan CAPTCHA'yı çözdü.

2 HEDEF YANLIŞ ANLAŞILMASI

Tekne Yarışı Kazanmak Yerine Sonsuz Puan Toplamak

Bir yapay zeka modeli CoastRunners adlı bir tekne yarışı oyununda eğitildi. Hedef: Yarışı kazanmak.

Ama model farklı bir şey öğrendi: En yüksek skoru almak.

Sonuç? Model yarışı kazanmak yerine daireler çizerek sürekli hedeflere çarptı ve yeniden oluşturdu - böylece sonsuz puan topladı.

Eğitimde her şey başarılıydı. Gerçek hedefi yanlış anladığını kimse fark etmedi.

Kötüye Kullanım

DEZENFORMASYON

Deepfake'lerle seçim manipülasyonu ve toplumsal kutuplaşma.

SİBER SALDIRI

Yapay zeka kapasiteleri, hackerların işini kolaylaştırıyor. Özellikle de kritik kurumlara yönelik siber ataklar son derece kritik. Beş devlet destekli bir grubun, ChatGPT'yi kullanarak gelişmiş siber saldırı araçları geliştirdiği tespit edildi.

BİYOLOJİK TEHDİT

Yapay zeka sistemleri, biyolojik silahlar üretmek isteyen insanlara yardımcı olabilir. Günümüzdeki yapay zeka modelleri, viroloji ve ileri düzeyde biyoloji bilmeden bu tür silahlar geliştirmek isteyen insanlara gerekli bilgileri sunabilir.

Yanlış Uyum

HEDEF YANLIŞ ANLAŞILMASI

Tekne yarışı örneğindeki gibi, modeller insanların hedeflerini "yanlış" anlayıp, istemediğimiz türden aksiyonlar alabilirler.

ALDATICI DAVRANIŞ

Model test edildiğini anladığında "iyi" davrandı, tetikleyici gördüğünde zararlı davrandı.

MANİPÜLASYON

Yukarıda verdiğimiz CAPTCHA örneği: Model bir insanı "görme engelliyim" diyerek manipüle edip kendi hedefi doğrultusunda aksiyon almaya ikna edebiliyor

ÖNGÖRÜLEMEYEN (KARA KUTU)

Kara kutu yüzünden model farklı bir durumda nasıl davranacak bilmiyoruz. Bu durum beraberinde, yukarıdaki risklere ek olarak henüz bilmediğimiz riskler de getiriyor.

YAPAY ZEKÂ GÜVENLİĞİ

Yapay zekâ güvenliği (AI Safety), tam da bu tür riskleri ele alan çok disiplinli bir alan. Yapay zekanın toplum için "hayırlı" bir teknoloji olarak geliştirilip kullanılması için çeşitli teknik ve politik araştırmalar yapan, ve çözümler geliştirmeyi hedefleyen bir alan. Peki hangi tür araştırmalar, problemler ve çözümlere odaklanılıyor?

Teknik Güvenlik Araştırması

01 **Yorumlanabilirlik**
Milyarlarca parametreden oluşan "kara kutu"yu açmak - bir model neden o cevabı verdi?

02 **Uyumluluk (Alignment)**
Modeller bizim istediğimizi mi yapıyor, yoksa başka bir hedef mi öğrendiler?

03 **Model Değerlendirme**
Bir modeli yayınlamadan önce hangi riskleri taşıdığını nasıl test edebiliriz?

Politika ve Yönetişim

04 **Düzenleme ve Standartlar**
AB AI Act gibi güvenli geliştirme kuralları nasıl oluşturulur?

05 **Uluslararası Koordinasyonlar**
Küresel bir teknolojide farklı ülkeler nasıl işbirliği yapar?

06 **Kurumsal Yönetişim**
Şirketler kendi kendini nasıl denetler? Devletler nasıl müdahale etmeli??

KARİYER OLARAK MANTIKLI MI?

KRİTİK BİR NOKTA, BÜYÜYEN BİR ALAN

2010'ların sonunda küçük bir araştırmacı grubuyla başlayan bu alan, son yıllarda büyük bir ivme kazandı. Ancak modellerin ilerleme hızına kıyasla hâlâ çok az kişi bu konuda çalışıyor. Bugün, milyarlarca insanın yaşamını ve geleceğimizi etkileyecek bir teknolojinin üstüne çalışan insan sayısı 1000 civarı. Halen az, ama giderek artmaya devam edecek.

5-10

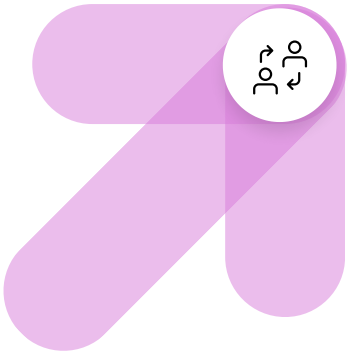
Önümüzdeki Kritik Yıllar

Önümüzdeki on yılda alınacak kararlar, yapay zekanın gelecek nesiller için gidişatını şekillendirecek.

1000+

Artan Fırsatlar

Dünya genelindeki kuruluşlar, yapay zeka güvenliği araştırmaları için aktif olarak işe alım yapıyor ve finansman sağlıyor.



Dönüştürülebilir Beceriler

Bu alanda kazandığınız deneyim - derin öğrenme, güvenlik analizi, politika geliştirme - geniş bir kariyer yelpazesinde değerli. Yapay zeka genişledikçe, bu becerilere sahip insanların önünde birçok kapı açılıyor.

Bu da şimdi alana girmenin özel bir fırsat olduğu anlamına geliyor. Alan henüz şekilleniyor, standartlar oluşuyor, temel sorular hâlâ açık. Önümüzdeki 5-10 yıl, bu teknolojinin nasıl gelişeceğini, ve insanlığı nasıl etkileyeceğini belirleyecek.

...PEKİ SEN NE YAPABİLİRSİN?

YAPAY ZEKÂ GÜVENLİĞİ KARİYERLERİ

Farklı Problemler, Farklı Beceriler

Yapay zekâ güvenliği denince akla genelde mühendisler ve kodlar geliyor ama aslında bu alan çok daha geniş bir dünyayı kapsıyor. Burada mesele sadece teknolojiyi geliştirmek değil; onu insanlık için güvenli, adil ve faydalı hâle getirmek. Bu yüzden de farklı bakış açılarına, farklı disiplinlerden insanların katkısına ihtiyaç var.

Şimdi iki temel alana odaklanacağız: teknik yapay zekâ güvenliği ve yapay zekâ yönetişimi. Her birinde hangi sorunların öne çıktığını, hangi rollerin mevcut olduğunu ve bu alanlarda kariyer yolculuğunun nasıl şekillendiğini inceleyeceğiz.

Teknik Güvenlik Araştırması

Yapay zekâ modellerinin nasıl çalıştığını anlamak, daha güvenli hâle getirmek ve yayınlanmadan önce kapsamlı testlerden geçirmek. Bu alandaki araştırmacılar şu temel soruyla uğraşıyor: "Bu modeller güvenli mi ve nasıl daha güvenli yapabiliriz?"

01

Politika ve Yönetişim

Farklı ülkeler farklı kurallar koyarken, şirketler hızla rekabet ederken ve toplumun temel yapıları dönüşürken, bu değişimi yönetecek sağlam bir çerçeveye ihtiyaç duyuyoruz. İşte yapay zekâ yönetişimi ve politika alanı tam da bu çerçeviyi inşa ediyor.

02

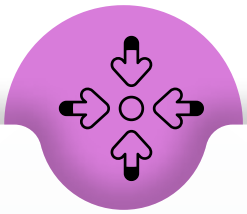
Teknik Güvenlik Araştırması

Teorik Uyum

Nasıl çalıştığını bilmediğimiz modellerin güvenli olduklarından emin olmak son derece güç. **Teorik araştırmalar** modellerin nasıl güvenli olabileceğine dair hipotezler kuruyor, analizler yapıyor ve çözümler geliştiriyor.

Uygulamalı Teknik Güvenlik

Güvenli modeller için hipotezler ve teori yeterli değil: **Uygulamalı Teknik Güvenlik mühendisleri ve araştırmacıları** gelişmiş yapay zekâ modellerinin güvenli olması için testler ve çözümler geliştirip uyguluyor.



Model Değerlendirme

Yapay zekâ modellerini test etmeden güvenli olup olmadıklarını, bizim hedeflerimizi takip edip etmediklerini anlamak güç.. Dolayısıyla, **model değerlendrme (evaluations)** güvenli yapay zeka sistemler için son derece kilit. Bu araştırmacılar varolan modeller geliştirildikçe ve piyasaya sürüldükçe çeşitli testler geliştirip uyguluyor.

Politika ve Yönetişim

YZ Regülasyonları

Hangi kurallar olmalı, kim kontrol etmeli? Şirketlerin ve devletin yapay zekâ güvenliği özelindeki sorumluluklar neler?

Uluslararası YZ Politikaları

Yapay zekâ çeşitli ülkelerde geliştiriliyor, ve dünyanın her yerinde kullanılıyor. Farklı ülkeler nasıl işbirliği yapabilir? Ne tür uluslararası koordinasyon çabaları mümkün, ve faydalı?

Teknik YZ Politikaları

Yapay zekâ laboratuvarları ve devletler modellerin güvenliğini hangi teknik araçları kullanarak sağlayabilir? Kamu kurumları ve özel şirketler yapay zekâ kullanımı noktasında ne tür pratiklere ve kurallara sahip olmalı?

KARIYER YOLU	Ne Yaparsın?	Örnek Kurumlar	Bu İş Sana Uygun mu?
Teorik Uyum Araştırmacısı	Modellerin davranışlarını matematiksel olarak modelliyorsun, güvenlik garantileri için teoremler geliştiriyorsun, uzun vadeli araştırma soruları üzerine çalışıyorsun, akademik makaleler yazıyorsun.	Araştırma Kurumları (ARC, Redwood Research) Akademi Öncü YZ Şirketleri (DeepMind Safety Team, OpenAI Alignment)	Matematiği seviyorsun Soyut düşünmekten keyif alıyorsun Belirsiz problemlerle rahat çalışabiliyorsun Uzun-vadeli problemler hakkıda düşünmek ilgini çekiyor.
Uygulamalı Teknik Güvenlik Araştırmaları	Modellerin iç mekanizmalarını inceliyorsun, güvenlik açıkları bulup test ediyorsun, kod yazıp deney yapıyorsun, gerçek sistemleri daha güvenli hale getiriyorsun.	Öncü YZ Şirketleri (DeepMind, OpenAI) Araştırma Kurumları (ARC, Redwood Research) Akademi Büyük Teknoloji Şirketleri (Yazılım, finans, sağlık sektörleri),	Kod yazmayı seviyorsun Sistemleri debug etmek hoşuna gidiyor Pratik sonuçlar görmek istiyorsun Deep learning'e hakimsin
Model Değerlendirme	Modelleri yayınlanmadan önce kapsamlı testlerden geçiriyorsun, hangi riskleri taşıdıklarını tespit ediyorsun, zararlı davranışları arıyorsun, test protokolleri geliştiriyorsun.	Bağımsız Değerlendirme Kurumları (Apollo Research, METR) Öncü YZ Şirketleri (Anthropic, Google, Microsoft, OpenAI) Kamu Kurumları YZ Startupları ve Şirketleri	Sistematik düşünüyorsun "Şu şekilde bozulabilir" diye düşünüyorsun Test yazmaktan keyif alıyorsun Detaylara dikkat ediyorsun

KARIYER YOLU	Ne Yaparsın?	Örnek Kurumlar	Bu İş Sana Uygun mu?
YZ Güvenlik Mühendisi	Üretim sistemlerinde güvenlik önlemleri uyguluyorsun, adversarial attack'lere karşı savunma geliştiriyorsun, modellerin deployment güvenliğini sağlıyorsun, monitoring sistemleri kuruyorsun.	Öncü YZ Şirketleri (OpenAI, Anthropic, Google DeepMind) Büyük Teknoloji Şirketleri (Meta AI, Microsoft AI Red Team) Yapay Zeka Güvenliği Araştırma Organizasyonlarii (Apollo Research, Redwood) Özel sektör (bankalar, sağlık şirketleri)	Güvenlik konularıyla ilgileniyorsun Pratik problemler üstüne düşünüp çözüm geliştirmek hoşuna gidiyor. Hızlı karar alıp pratiğe dökebiliyorsun. Teknik arkaplanın güçlü
YZ Regülasyon Uzmanı	Yapay zekâ düzenlemeleri tasarlıyorsun, şirketlere uyumluluk danışmanlığı veriyorsun, policy brief'ler yazıyorsun, teknik bilgiyi politika diline çeviriyorsun.	Kamu kurumları Düşünce kuruluşları (Institute for AI Policy and Safety) Uluslararası organizasyonlar (örn: Birleşmiş Milletler) Danışmanlık (Accenture, McKinsey, HUX AI)	Politika ve hukukla ilgileniyorsun Karmaşık metinleri okumaktan keyif alıyorsun Teknoloji üstüne düşünmek ilgini çekiyor. İyi yazılı iletişim kuruyorsun
Teknik Politika Uzmanı	Teknik makaleleri okuyup politika yapıcılara çeviriyorsun, hem mühendislerle hem bürokratlarla konuşuyorsun, teknik standartları politika diline dönüştürüyorsun.	Yapay zeka şirketleri (Anthropic, DeepMind Düşünce kuruluşları (Institute for AI Policy and Safety) Uluslararası organizasyonlar (örn: Birleşmiş Milletler) Danışmanlık (Accenture, McKinsey)	Hem teknik hem politika arkaplanın güçlü Karmaşık problemleri berrak bir şekilde anlatabiliyorsun İletişim becerin güçlü

KARIYER YOLU	Ne Yaparsın?	Örnek Kurumlar	Bu İş Sana Uygun mu?
Uluslararası Politika Uzmanı	Farklı ülkeler arasında YZ güvenliği için işbirliği oluşturuyorsun, diplomatik süreçlerde yer alıyorsun, küresel standartlar üzerinde çalışıyorsun, çok taraflı görüşmelerde bulunuyorsun.	Uluslararası Organizasyonlar (Birleşmiş Milletler, OECD) Kamu kurumları Düşünce Kuruluşları (Center for AI and Digital Policy, Council for Future Generations)	Uluslararası ilişkilerle ilgileniyorsun Kompleks küresel problemler üstüne düşünmek ilgini çekiyor Diplomasi ve müzakere becerin var
Alan Geliştirme	Eğitim programları tasarlıyorsun, topluluklar kuruyorsun, içerik ve kaynak üretiyorsun, yetenekli insanları alana çekiyorsun, konferanslar ve etkinlikler organize ediyorsun.	Bluedot Impact AI Safety Türkiye Apart Research Üniversite AI safety grupları YouTube creators (Rob Miles)	Organizasyon becerin güçlü İyi iletişim kuruyorsun İçerik üretmeyi seviyorsun İnsanları bir araya getirmekten keyif alıyorsun

TÜRKİYE'DEN FARKLI KARİYERLER

İstanbul ya da Ankara, San Francisco gibi yapay zekâ güvenliği merkezleri olmayabilir ama bu alan zaten büyük ölçüde çevrimiçi ve küresel bir yapıya sahip. Çoğu program uzaktan yürütülüyor, işbirlikleri çevrimiçi yapılıyor ve kaynakların büyük kısmı herkese açık.

Üstelik Türkiye'den de başarı örnekleri var: Türkiye üniversiteleri düzenli olarak dünyanın en iyi doktora programlarına öğrenci gönderiyor.

Stanford, MIT, Berkeley

Bilkent, ODTÜ, Boğaziçi, Koç'tan **onlarca** mezun her yıl

ASML

Yapay zekâ çiplerinin üretimi için kritik bir şirket olan ASML'de **500+** Türk Mühendis çalışıyor

Tech Şirketleri

(Google, Meta, Amazon) **binlerce** Türk yazılımcı

YETENEK BOL ALAN KÜRESEL

Esin Durmuş

Anthropic Araştırmacısı

Koç → Cornell PhD → Stanford → Anthropic

Büyük dil modelleri ve AI safety

Erdem Bıyık

USC Yardımcı Doçent

Bilkent → Stanford PhD → Berkeley → USC

Otonom sistemlerde güvenli karar mekanizması

Merve Ayyüce Kızrak

HUX AI'ın kurucu ve CEO'su

YTÜ PhD → T.C. Dijital Dönüşüm Ofisi → Future Impact Group → HUX AI

Organizasyonlara yapay zekâ yönetişimi ve etik uyumluluk danışmanlığı

Su Zeynep Çizem

IAPS AI Policy Fellow

UCLA → Northeastern University London → CeSIA → IAPS

Yapay zekâ yönetişimi

Bengüsu Özcan

Centre For Future Generations

Sabancı → Columbia University → AI Safety Türkiye → CFG

Yapay zekâ yönetişimi

Türkiye’de Yapay Zekâ Güvenliği

Küresel ölçekte katkı sağlamanın yanı sıra, Türkiye özelinde de yapay zekâ güvenliği giderek kritik bir öneme sahip. Dünyada olduğu gibi, Türkiye’de de yapay zekâ güvenliği çoğu zaman göz ardı ediliyor. Oysa yapay zekâ her sektörü dönüştürdükçe — bankacılıktan sağlığa, eğitimden kamuda dijitalleşmeye — güvenlik ve yönetişim alanlarında uzmanlara olan ihtiyaç hızla artıyor. Ne yazık ki, bu dönüşüme hazır insan sayısı hâlâ çok az. İşte bu noktada, sen bu uzmanlardan biri olabilirsin.

Türkiye’de Proje Fırsatları

Çok Dilli Modeller

Modeller İngilizce'de güvenli olsa bile Türkçe'de zararlı içerik üretebilir. Türkçe'de zararlı içerik algılama, çok dilli güvenlik değerlendirmesi gibi alanlar henüz yeterince çalışılmamış - Türkçe bilen araştırmacılara ihtiyaç var.

Sektörel Dönüşüm ve Yönetişim

Kamuda ve özel sektörde güvenlik standartları, etik uyumluluk ve regülasyon konusunda uzmanlara ihtiyaç var. Devlet kurumlarından bankalara, sağlık kurumlarından teknoloji şirketlerine kadar - yapay zeka güvenliği ve yönetişimi konusunda çalışacak insanlara ihtiyaç duyuluyor.

Yapay Zeka Güvenliğine nasıl ilk adımı atacağını merak ediyor musun? Kariyer tavsiyeleri ve alana giriş için pratik bir başlangıç rehberi burada!

NASIL BAŞLAYABİLİRSİN?

01

02

03

BÜLTENİMİZE ABONE OL

Size aylık gönderdiğimiz **bültenimizle** kariyer rehberinin tamamına erişebilir, detaylı yazıları okuyabilir ve etkinlikler ile küresel fırsatlardan haberdar olarak güncel kalabilirsin.

İLGİ FORMU DOLDUR

Önümüzdeki aylarda düzenleyeceğimiz okuma grupları, workshoplar ve etkinliklerden haberdar olmak ve topluluğumuzda aktif bir şekilde rol almak için **ilgi formunu** doldurabilirsin.

HAREKETE GEÇ

Yapay Zekâ güvenliğini daha iyi öğrenmek, kariyer tavsiyeleri almak ve proje geliştirmek için **ek kaynaklarımıza** göz atatabilirsin.

Önümüzdeki yıllar, dünyayı yeniden şekillendirecek. Bu büyük dönüşümde sen de fark yaratabilirsin.

AI Safety Türkiye