

YAPAY ZEKA ETİĞİ: MÜHENDİSLİK PERSPEKTİFİNDEN KAPSAMLI RİSK YÖNETİMİ, VAKA ANALİZLERİ VE YÖNETİŞİM ÇERÇEVESİ

1. Giriş

1.1. Bilgi Teknolojilerinde Etik Paradigmasının Dönüşümü ve Kuramsal Temeller

Bilgi Teknolojileri (BT) etiği, disiplinin doğuşundan bu yana insan odaklı bir sorumluluk anlayışı üzerine inşa edilmiştir. Norbert Wiener'in sibernetik üzerine yaptığı erken dönem çalışmalarından James Moor'un "bilgisayar etiği" kavramsallaştırmasına kadar uzanan süreçte, etik daima insan failin eylemlerini "ne yapılabilir" ve "ne yapılmalıdır" ekseninde inceleyen normatif bir denetim mekanizması olarak kurgulanmıştır. Geleneksel BT etiği, verinin gizliliği, fikri mülkiyet hakları ve yetkisiz erişim gibi konuları ele alırken, teknolojiyi büyük ölçüde pasif bir araç (instrument) olarak konumlandırmıştır.¹ Ancak 21. yüzyılın ilk çeyreğinde, özellikle derin öğrenme (Deep Learning) ve üretken yapay zeka (Generative AI) modellerinin endüstriyel ölçekte yaygınlaşmasıyla birlikte, bu disiplin radikal bir ontolojik dönüşüm geçirmiştir.

Günümüz Yapay Zeka (YZ) etiği, failin (agent) artık sadece insan olmadığı, algoritmik sistemlerin otonom kararlar alabildiği hibrit bir ekosistemde ortaya çıkan "sorumluluk boşluklarını" (responsibility gaps) doldurmaya odaklanmaktadır. Teknolojinin sosyal, ekonomik ve politik sonuçlar doğuran bir aktör (actor) konumuna yükselmesi, mühendislik etiğinin sınırlarını zorlamaktadır.² Akademik literatürde bu durum, teknolojik determinizm ile sosyal inşacılık arasında bir köprü kurmayı gerektirir; zira teknoloji nötr değildir. Algoritmalar, tasarımcıların değer yargılarını, eğitim veri setlerinin tarihsel eşitsizliklerini ve geliştirildiği kurumların politik hedeflerini bünyesinde barındıran sosyo-teknik yapılardır.

Bu bağlamda BT etiği, artık sadece teorik bir felsefi tartışma alanı değil, sistem mimarisine gömülü (embedded) bir risk yönetimi protokolü ve mühendislik gerekliliğidir. "Ethics by Design" (Tasarım Yoluyla Etik) yaklaşımı, etik değerlerin kodun yazılmasından önce, sistemin tasarım aşamasında belirlenmesini ve teknik gereksinimler (requirements engineering) listesine eklenmesini zorunlu kılar.³ Bu rapor, YZ sistemlerinin teknik karmaşıklığı ile toplumsal değerler arasındaki gerilimi analiz ederek, mühendisler ve karar vericiler için kapsamlı bir yol haritası sunmayı amaçlamaktadır.

1.2. Yapay Zekanın BT Etiği Açısından Kritik Önemi ve Risk Yüzeyi

Yapay zekanın BT etiği içindeki merkezi ve kritik konumu, onun "karar verme" ve "öngörü" yetisinden kaynaklanır. Geleneksel yazılımlar deterministiktir; belirli bir girdi (input) her zaman aynı çıktıyı (output) üretir ve hata ayıklama (debugging) süreçleri lineer bir mantık izler. Ancak stokastik (olasılıksal) bir yapıya sahip olan makine öğrenmesi modelleri, eğitim verilerinden öğrendiği örüntüleri (patterns) genelleştirerek, daha önce hiç karşılaşmadığı durumlarda otonom kararlar verebilir. Bu durum, BT altyapılarında daha önce görülmemiş bir "öngörülemezlik" ve "kontrol kaybı" riskini doğurur.²

Bu risk yüzeyi, üç ana ekseninde genişleyerek klasik BT risk yönetimini yetersiz bırakmaktadır:

1. **Ölçeklenebilirlik (Scalability) ve Hız:** Bir insan operatörün hatası genellikle lokal bir etki yaratırken, bir yapay zeka algoritmasındaki hata (örneğin, ırksal önyargı içeren bir kredi skorum sistemi veya hatalı bir tıbbi tanı algoritması) milyonlarca kullanıcıyı aynı anda, sistematik olarak ve milisaniyeler içinde etkileyebilir. Bu durum, etik ihlallerin yayılma hızını ve etki alanını geometrik olarak artırır.⁵
2. **Opaklık (Opacity) ve Kara Kutu Problemi:** Derin sinir ağlarının (DNN) milyonlarca parametresi (weights & biases) arasında bir kararın tam olarak nasıl oluştuğunun izlenememesi, "hesap verebilirlik" ilkesini teknik olarak zorlaştırmaktadır. Bir otonom aracın neden fren yapmadığını veya bir İK algoritmasının neden belirli bir adayı elelediğini açıklayamamak, sadece teknik bir hata değil, hukuki ve etik bir krizdir.¹
3. **Otonomi (Autonomy) ve Ahlaki Eylemlilik:** İnsan müdahalesi olmadan (human-out-of-the-loop) çalışan sistemlerin, fiziksel dünyada (örneğin otonom araçlar, robotik cerrahi veya otonom silah sistemleri) doğrudan zarar verme potansiyeli, "zarar vermeme" (non-maleficence) ilkesini tehdit etmektedir. Bu sistemlerin ahlaki birer ajan (moral agent) sayılıp sayılamayacağı tartışması, hukuki sorumluluk rejimlerini (liability regimes) kökünden sarsmaktadır.⁶

Yapay zeka, veri madenciliğinden bulut bilişime kadar tüm BT süreçlerini optimize ederken, aynı zamanda bu süreçlerin etik zafiyetlerini de derinleştirmektedir. Yanlış etiketlenmiş bir veri seti, artık sadece veritabanı yönetiminin bir sorunu değil, o veriyi kullanan modelin üreteceği ayrımcı kararların temel nedenidir. Dolayısıyla YZ etiği; veri kalitesi, siber güvenlik, yazılım mühendisliği ve hukuk disiplinlerinin kesişim noktasında yer alan, çok katmanlı ve disiplinler arası bir güvenlik protokolü olarak ele alınmalıdır.

2. Yapay Zeka Etiğine Derinlemesine Giriş

2.1. Yapay Zeka Taksonomisi ve Etik İmplikasyonları

Yapay zekanın etik analizi, sistemin yetenek düzeyine ve işleyiş mekanizmasına göre farklılaşan risk profillerinin anlaşılmasını gerektirir. Her bir YZ türü, kendine özgü bir "etik risk vektörü" taşıyor ve mühendislik yaklaşımı bu vektörlere göre özelleştirilmelidir.

2.1.1. Yetenek Düzeyine Göre Sınıflandırma ve Riskler

Sistemlerin bilişsel kapasitelerine göre yapılan bu sınıflandırma, riskin niteliğini belirler:

- **Dar Yapay Zeka (ANI - Artificial Narrow Intelligence):** Günümüzde kullanılan tüm YZ sistemlerini (ChatGPT, Tesla Autopilot, Siri, AlphaGo) kapsar. Bu sistemler, tek bir alanda veya görevde (örneğin satranç oynamak veya protein katlanmasını tahmin etmek) insanüstü performans gösterebilirken, bağlam dışına çıktıklarında başarısız olurlar.
 - *Etik Risk: "Kırılganlık" (Brittleness) ve Antropomorfizm.* ANI sistemleri, eğitildikleri alanın dışına çıkıldığında veya veride gürültü (noise) olduğunda mantıksız ve tehlikeli hatalar yapabilir. Ayrıca, kullanıcıların bu sistemlere olduğundan fazla zeka ve bilinç atfetmesi (antropomorfizm), güvenli kullanım sınırlarını ihlal etmelerine ve sisteme aşırı güven duymalarına neden olabilir.¹
- **Genel Yapay Zeka (AGI - Artificial General Intelligence):** İnsan benzeri bilişsel esnekliğe sahip, farklı alanlar arasında bilgi transferi yapabilen ve öğrenebilen teorik sistemlerdir. Henüz gerçekleşmemiş olsa da, GPT-4 gibi Büyük Dil Modelleri (LLM) bu yönde atılmış adımlar olarak görülmektedir.
 - *Etik Risk: "Değer Hizalaması" (Value Alignment).* Bir AGI'nın hedeflerinin insanlığın temel değerleriyle (yaşam hakkı, özgürlük, refah) nasıl örtüştürüleceği, "Kontrol Problemi" olarak bilinen varoluşsal bir risk tartışmasıdır.
- **Süper Yapay Zeka (ASI - Artificial Super Intelligence):** İnsan zekasını bilimsel, yaratıcı ve sosyal her alanda aşan sistemlerdir.
 - *Etik Risk: İnsanlığın kaderinin kontrolü ve teknolojik tekillik (singularity).*

2.1.2. İşleyiş Yöntemine Göre Risk Analizi

Sistemin mimarisi, hata türlerini ve açıklanabilirlik seviyesini belirler:

- **Kural Tabanlı Sistemler (Rule-Based / Expert Systems):** "Eğer-O Halde" (If-Then) mantığıyla çalışan deterministik sistemlerdir.
 - *Risk: Esneklik Eksikliği.* Tüm senaryoların önceden kodlanamaması, beklenmedik durumlarda sistemin kilitlenmesine veya yanlış kuralı uygulamasına yol açar.
- **Makine Öğrenimi (Machine Learning - ML):** İstatistiksel yöntemlerle veriden öğrenen sistemlerdir.
 - *Risk: "Veri Önyargısı" (Data Bias).* Veri seti geçmişteki toplumsal eşitsizlikleri içeriyorsa, model bu eşitsizlikleri matematiksel bir gerçeklik olarak öğrenir ve geleceğe taşır (Bias in, Bias out).
- **Derin Öğrenme (Deep Learning - DL):** İnsan beyninden esinlenen çok katmanlı yapay sinir ağları kullanır.
 - *Risk: "Açıklanabilirlik Eksikliği" (Black Box).* Sistemin içsel mantığının insan tarafından anlaşılamaması, hata ayıklamayı (debugging) ve yasal savunmayı imkansızlaştırır. Modelin neden belirli bir karar verdiğini matematiksel olarak izlemek son derece güçtür.¹
- **Üretken Yapay Zeka (Generative AI):** Mevcut verilerden yola çıkarak yeni içerik (metin, görsel, kod, ses) üreten sistemlerdir (ör. ChatGPT, Midjourney).
 - *Risk: "Halüsinasyon" (Hallucination) ve Telif Hakları.* Gerçek olmayan bilgiyi son

derece ikna edici ve otoriter bir dille sunma riski taşır. Ayrıca eğitim verisindeki eserlerin telif hakları konusunda büyük hukuki belirsizlikler yaratır.⁸

2.2. BT Etiği Açısından Sosyo-Teknik Yaklaşım

YZ sistemlerini yalnızca teknik bir kod bloğu veya matematiksel bir optimizasyon problemi olarak görmek, etik risklerin yönetiminde yetersiz kalır. Modern mühendislik etiği, YZ'yi "sosyo-teknik sistemler" (socio-technical systems) olarak ele alır. Bu yaklaşım, teknolojinin sadece teknik bileşenlerden (donanım, yazılım, veri) değil, aynı zamanda sosyal bileşenlerden (insanlar, kurumlar, yasalar, kültürel normlar) oluştuğunu kabul eder.

BT etiği bu noktada üç temel boyutta operasyonel hale gelir:

- Veri Etiği (Data Ethics):** Verinin soyut bir "varlık" değil, insanların dijital iz düşümü olduğu kabulüyle hareket eder. Veri toplama sürecindeki rıza mekanizmaları, veri temizliği, anonimleştirme teknikleri ve verinin bağlamı bu kapsamdadır.¹⁰
- Algoritmik Etik (Algorithmic Ethics):** Algoritmanın matematiksel optimizasyon fonksiyonunun, adalet ve eşitlik gibi sosyal değerlerle çelişmemesini hedefler. Örneğin, bir sigorta algoritmasının karı maksimize ederken, dezavantajlı grupları sistematik olarak sistem dışına itmemesi gerekir.
- Uygulama Etiği (Applied Ethics):** YZ'nin belirli bir bağlamda (örneğin eğitim, sağlık veya askeriye) kullanılmasının, o alanın yerleşik meslek etiği ilkeleriyle (örneğin tıpta Hipokrat yemini veya savaş hukukunda orantılılık ilkesi) uyumlu olup olmadığını sorgular.¹¹

2.3. YZ Yaşam Döngüsü Boyunca Etik Yönetişim

Etik, projenin sonunda yapılan bir "uyumluluk kontrolü" (checklist) değil, YZ yaşam döngüsünün (lifecycle) her aşamasına entegre edilmiş sürekli bir süreçtir:

Yaşam Döngüsü Aşaması	Temel Etik Görevler ve Sorular	İlgili Riskler
1. Tasarım ve Kapsam	"Bu sistemi inşa etmeli miyiz?" sorusu sorulmalıdır. Amaç meşru mu?	Amaç kayması, etik olmayan kullanım (ör. duygu tanıma).
2. Veri Hazırlama	Veri setindeki eksiklikler ve tarihsel önyargılar tespit edilmelidir.	Temsil önyargısı, etiketleme hataları, gizlilik ihlalleri.
3. Modelleme ve Eğitim	Modelin aşırı uyum (overfitting) sağlamaması ve karar sınırlarının adil	Model önyargısı, enerji tüketimi (çevresel etki).

	olması sağlanmalıdır.	
4. Test ve Doğrulama	Teknik doğruluk (accuracy) yanı sıra adalet metrikleri (fairness metrics) test edilmelidir.	Güvenlik açıkları, adversarial zafiyetler.
5. Dağıtım ve İzleme	Model drift (zamanla başarımların kaybı) ve beklenmedik yan etkiler izlenmelidir.	Performans düşüşü, bağlam değişikliği hataları.

Örneğin, duygu tanıyan (emotion recognition) YZ sistemlerinin işe alımlarda veya sınır güvenliğinde kullanılması, bilimsel temeli zayıf olduğu (duyguların evrenselliği tartışmalı olduğu için) ve manipülasyona açık olduğu gerekçesiyle Avrupa Birliği Yapay Zeka Yasası (AI Act) kapsamında "kabul edilemez risk" olarak değerlendirilip yasaklanma eğilimindedir.¹³

2.4. Veri Etiği: BT Yönetiminin Temel Taşı

Büyük veri (Big Data) çağında veri etiği, BT yönetiminin (IT Governance) en kritik bileşenidir. Veri etiği, yasal zorunlulukların (KVKK/GDPR) ötesinde, verinin **bağlamsal bütünlüğünü** (contextual integrity) korumayı amaçlar. Helen Nissenbaum tarafından geliştirilen bu teoriye göre, veri akışı, verinin toplandığı bağlamların normlarına uygun olmalıdır. Örneğin, bir doktorun hastasından aldığı sağlık verisi meşrudur; ancak bu verinin hastanın rızası olmadan bir sigorta şirketine satılması veya bir reklam algoritmasında kullanılması, bağlamsal bütünlüğü ihlal eder.

BT yöneticileri için veri etiği şu prensipleri zorunlu kılar:

- **Veri Minimizasyonu:** Sadece amaç için gerekli olan en az veriyi toplamak ve saklamak.
- **Amaç Bağlılığı (Purpose Limitation):** Veriyi sadece toplandığı spesifik amaç için kullanmak.
- **Veri Egemenliği:** Kullanıcının kendi verisi üzerindeki kontrol hakkını (silme, taşıma, düzeltme, işlemeyi durdurma) teknik olarak mümkün kılmak ve arayüzlerle desteklemek.¹⁰

2.5. Sorumlu Yapay Zeka (Responsible AI)

Sorumlu YZ, etik ilkelerin soyut birer temenni olmaktan çıkıp, ölçülebilir mühendislik kriterlerine dönüştürülmesidir. Microsoft, Google, IBM gibi teknoloji devleri ve OECD gibi uluslararası kuruluşlar tarafından benimsenen bu çerçeve, YZ geliştirme sürecini şeffaf, hesap verebilir, adil ve insan merkezli hale getirmeyi hedefler. Sorumlu YZ, **"Ethics by Design"** yaklaşımını benimser; yani etik önlemler, sistem mimarisinin bir parçasıdır, sonradan eklenen bir yama değildir.¹⁵ Bu yaklaşım, sistemin sadece "doğru" çalışmasını değil, aynı zamanda "iyi" ve "adil"

alışmasını garanti altına almayı hedefler.

3. Yapay Zekada Temel Etik Problemler ve Teknik Analizler

Yapay zeka sistemlerinin yaygınlaşması, daha önce teorik olan birçok etik problemi somut teknik zorluklara dönüştürmüştür. Bu bölümde, bu problemlerin teknik kökenleri ve çözüm mekanizmaları analiz edilmektedir.

3.1. Gizlilik ve Veri Mahremiyeti: Yeni Tehdit Vektörleri

Geleneksel veritabanı güvenliğinin aksine, YZ modelleri veriyi "ezberleme" (memorization) eğilimi gösterebilir. Özellikle Büyük Dil Modelleri (LLM) ve üretken modeller, eğitim verilerinde yer alan hassas kişisel verileri (PII - Personally Identifiable Information) parametreleri içine kodlayabilir. Bu durum, modelin çıktılarında bu verilerin açığa çıkmasına neden olabilir.

Bu alandaki iki temel saldırı vektörü şunlardır:

- Model Inversion Attacks (Model Tersine Çevirme Saldırıları):** Saldırganın, modelin çıktılarını veya güven skorlarını (confidence scores) analiz ederek, eğitim verisindeki belirli bir özelliğin veya bireyin verisinin (örneğin bir hastanın genetik belirteçlerinin veya yüz görüntüsünün) geri mühendislikle yeniden oluşturulmasıdır.¹
- Membership Inference Attacks (Üyelik Çıkarımı Saldırıları):** Bir bireyin verisinin, modelin eğitim setinde yer alıp almadığının tespit edilmesidir. Bu durum, örneğin bir kişinin nadir bir hastalığa sahip hastaların bulunduğu bir veri setinde yer aldığıının anlaşılması gibi ciddi mahremiyet ihlallerine yol açabilir.

Teknik Çözüm: Bu risklere karşı "**Differential Privacy**" (Diferansiyel Gizlilik) tekniği öne çıkmaktadır. Bu yöntem, veri setine veya modelin eğitim sürecindeki gradyanlara (gradients) matematiksel olarak hesaplanmış "gürültü" (noise) ekleyerek, modelin genel örüntüleri öğrenmesini sağlarken bireysel verilerin ayırt edilmesini imkansız hale getirir.¹⁶

3.2. Veri Güvenliği ve Adversarial Saldırıları

Yapay zeka sistemleri, klasik siber saldırıların (DDoS, SQL Injection) yanı sıra, makine öğrenmesi algoritmalarının matematiksel yapısını hedef alan saldırı türlerine de açıktır.

- Zehirlleme Saldırıları (Data Poisoning):** Saldırganın eğitim verisine manipüle edilmiş, yanlış etiketlenmiş veriler ekleyerek modelin hatalı öğrenmesini sağlamasıdır. Örneğin, bir otonom aracın eğitim setine üzerine küçük etiketler yapıştırılmış "dur" tabelaları eklenerek, aracın bu tabelaları "hız sınırı 45" olarak algılaması sağlanabilir. Bu saldırı, modelin karar sınırlarını (decision boundaries) bozar.
- Evasion Attacks (Kaçınma Saldırıları):** Saldırganın, girdi verisinde insan gözünün fark

edemeyeceği küçük değişiklikler yaparak (pixel perturbation) modelin yanlış sınıflandırma yapmasını sağlamasıdır. Bu, güvenlik kameralarını veya spam filtrelerini atlatmak için kullanılabilir.¹⁷

3.3. Algoritmik Önyargı (Bias) ve Ayrımcılık

Algoritmik önyargı, istatistiksel önyargının toplumsal önyargıya dönüşmesi durumudur. Bu, YZ sistemlerinin belirli demografik gruplara (ırk, cinsiyet, yaş vb.) karşı sistematik olarak dezavantajlı kararlar üretmesidir.

Önyargı türleri şunlardır:

- **Temsil Önyargısı (Representation Bias):** Eğitim verisinin hedef evreni tam temsil etmemesi. Örneğin, ImageNet gibi popüler görsel veri setlerinin çoğunlukla Batı ülkelerinden gelen görsellerden oluşması, modellerin diğer coğrafyalardaki nesneleri veya insanları tanımada başarısız olmasına yol açar.¹⁸
- **Tarihsel Önyargı (Historical Bias):** Verinin gerçeği doğru yansıtmaması, ancak toplumsal gerçeğin kendisinin adaletsiz olması durumudur. Örneğin, geçmişte kadınların yönetici pozisyonlarına daha az işe alınması verisinin bir işe alım modeline öğretilmesi, modelin "kadın olmayı" başarısızlık kriteri olarak öğrenmesine ve kadın adayları elemesine neden olur (Amazon Vakası).
- **Ölçüm Önyargısı (Measurement Bias):** Yanlış bir hedef değişkenin (proxy) seçilmesi. Ziad Obermeyer'in çalışmasında görüldüğü gibi, "sağlık ihtiyacını" ölçmek için "sağlık harcamalarını" kullanmak, parası olmadığı için harcama yapamayan ama hasta olan siyahileri sistemin dışına itmiştir.¹⁹
- **Toplama Önyargısı (Aggregation Bias):** Farklı özelliklere sahip popülasyonlar için tek bir model kullanılması. Örneğin, diyabet teşhisi için tüm etnik gruplara aynı hemoglobin A1c eşiklerini uygulayan bir model, fizyolojik farklılıklar nedeniyle bazı gruplarda hatalı sonuçlar verebilir.²¹

3.4. Şeffaflık ve Açıklanabilirlik (XAI)

Derin öğrenme modelleri, milyonlarca parametre arasındaki doğrusal olmayan (non-linear) ilişkilerle karar verir. Bu durum "Kara Kutu" (Black Box) sorununu doğurur. Kullanıcılar ve hatta geliştiriciler, modelin neden belirli bir çıktı verdiğini anlayamaz.

- **Global Açıklanabilirlik:** Modelin genel olarak hangi özelliklere ağırlık verdiğinin anlaşılması.
- **Lokal Açıklanabilirlik:** Belirli bir kararın (örneğin Ahmet Bey'in kredi başvurusunun reddedilmesi) neden verildiğinin açıklanması.

Teknik Çözüm: XAI (Explainable AI) yöntemleri. LIME (Local Interpretable Model-agnostic Explanations) ve SHAP (SHapley Additive exPlanations) gibi teknikler, karmaşık modellerin kararlarını daha basit ve anlaşılır modellere indirgeyerek veya özelliklerin karara katkısını hesaplayarak açıklanabilirlik sağlar. Ancak, açıklanabilirlik ile modelin

doğruluğu (accuracy) arasında genellikle bir ters orantı (accuracy-interpretability trade-off) vardır; daha karmaşık modeller genellikle daha doğru ama daha az açıklanabilirdir.¹

3.5. Sorumluluk ve Hesap Verebilirlik

Otonom bir sistem hata yaptığında sorumlu kimdir?

- Yazılımı geliştiren mühendis mi?
- Veri setini hazırlayan veri bilimci mi?
- Sistemi satın alıp, parametrelerini ayarlayarak kullanan şirket mi?
- Yoksa sistemi denetleyen kamu otoritesi mi?

"Many Hands Problem" (Çok el sorunu) olarak bilinen bu durum, sorumluluğun dağılmasına ve belirsizleşmesine yol açar. Hukuki açıdan, otonom sistemlere "kusursuz sorumluluk" (strict liability) rejimlerinin uygulanması, yani hatada kusur aranmaksızın üreticinin/işletenin sorumlu tutulması tartışılmaktadır. AB AI Act, yüksek riskli sistemler için sağlayıcıları (provider) ve kullanıcıları (deployer) net yükümlülüklerle bağlayarak bu sorunu çözmeye çalışmaktadır.⁶

3.6. Otonom Sistemler ve İnsan Denetimi Kaybı (Human Agency)

İnsan denetiminin (Human-in-the-loop) azalması, özellikle askeri sistemler, otonom araçlar ve tıbbi teşhis sistemlerinde "ahlaki eylemlilik" (moral agency) sorununu doğurur. Bir makinenin insan hayatı üzerinde karar verme yetkisine sahip olması, insan onuru açısından temel bir ihlal riski taşır. Ayrıca, sistemin hataya düştüğü durumlarda müdahale edecek bir insanın olmaması, kazaların kaçınılmaz hale gelmesine neden olabilir (Cruise Vakası). Bu nedenle düzenlemeler, kritik kararlarda insan gözetimini (human oversight) zorunlu kılmaktadır.²³

4. Kapsamlı Vaka Analizleri

Bu bölüm, teorik etik sorunların gerçek dünyada nasıl büyük çaplı krizlere dönüştüğünü gösteren somut olayları teknik ve yönetsel açılardan incelemektedir.

4.1. Vaka 1: Otonom Araçlarda Algı Hatası ve Kurumsal Şeffaflık İhlali - Cruise Robotaxi Kazası (San Francisco, 2023)

4.1.1. Olayın Detaylı Açıklaması

2 Ekim 2023 tarihinde San Francisco'da, General Motors'un otonom araç iştiraki Cruise'a ait sürücüsüz bir taksi (robotaxi), ağır bir trafik kazasına karışmıştır. Olay, insan sürücülü başka bir aracın bir yayaya çarpıp kaçmasıyla başlamıştır. Çarpmanın etkisiyle yaya, yan şeritte ilerleyen Cruise aracının (kod adı "Panini") önüne fırlatılmıştır. Cruise aracı yayaya çarpmış, ancak durmak yerine "güvenli bir noktaya çekilme" (pullover maneuver) protokolünü devreye sokmuştur. Bu manevra sırasında, aracın altında sıkışmış olan yaya yaklaşık 6 metre (20 feet)

boyunca sürüklenmiş ve hayati tehlike arz edecek şekilde yaralanmıştır.²⁴

4.1.2. Yapay Zeka Kaynaklı Teknik Problem Analizi

Bağımsız teknik raporlara (Quinn Emanuel Raporu ve Exponent Analizi) göre, kaza tek bir hatadan değil, bir dizi teknik yetersizliğin zincirleme sonucundan kaynaklanmıştır:

1. **Anlamsal Algı ve Sınıflandırma Hatası (Semantic Perception Failure):** Cruise'un yapay zeka sistemi, ilk çarpışmayı tespit etmiş ancak durumu yanlış sınıflandırmıştır. Sistem, aracın bir yayaya çarptığını fark etmiş, ancak bu çarpışmayı aracın yan tarafından gelen bir darbe ("yanal çarpışma" / side collision) olarak yanlış yorumlamıştır. Yanal çarpışma senaryosunda, aracın durması değil, trafiği engellemek için kenara çekilmesi öncelikli protokoldür.
2. **Nesne Sürekliliği ve Takip Hatası (Object Permanence Failure):** Yaya aracın altına girdikten sonra, LiDAR ve kamera sensörlerinin kör noktasına düşmüştür. İnsan sürücüler, bir nesne gözden kaybolursa bile onun hala orada olduğunu bilir (nesne sürekliliği). Ancak Cruise'un algoritması, yayanın hala orada olduğunu "hatırlayamamış" (lack of object persistence) ve yolun boş olduğunu varsayarak sürüşe devam etmiştir.²⁶
3. **Hatalı Karar Mekanizması (Faulty Decision Logic):** Sistem, bir kaza sonrası durup beklemek (fail-safe) yerine, operasyonel devamlılığı önceleyen (fail-operational) bir moda geçmiştir. Aracın altında bir engel varken bu manevranın yapılması, yayanın sürüklenmesine neden olmuştur.²⁵

4.1.3. Etik ve Hukuki İhlal Analizi

Olayın teknik boyutu kadar yönetsel ve etik boyutu da vahimdir:

- **Şeffaflık ve Dürüstlük İhlali:** Cruise yönetimi, kaza sonrası düzenleyici kurumlara (California DMV ve NHTSA) olayın videosunu gösterirken, yayanın sürüklendiği kritik kısmı gizlemiş veya videonun tamamını proaktif olarak paylaşmamıştır. Ayrıca internet bağlantısı sorunları bahane edilerek videonun tam oynatılmadığı iddia edilmiştir. Bu durum, "yanıltma yoluyla adaleti engelleme" ve şeffaflık ilkesinin ihlali olarak değerlendirilmiştir.
- **Güvenlik Önceliği İhlali:** Şirketin teknik incelemelerinde, sistemin "yere düşen nesneleri" ve özellikle araç altındaki insanları algılamadaki eksikliğini daha önceki simülasyonlarda (Phantom braking gibi) fark ettiği, ancak bu "edge case" (uç durum) senaryosunu yeterince önceliklendirmede ve düzeltmeden aracı trafiğe çıkardığı ortaya çıkmıştır.²⁵

4.1.4. Sonuçlar ve Çıkarılan Dersler

- **Yaptırımlar:** Cruise'un California'daki otonom sürüş lisansı süresiz iptal edilmiştir. CEO Kyle Vogt ve kurucu ekipten birçok kişi istifa etmiş, şirket personelinin %25'ini işten çıkarmıştır. Şirket, federal soruşturmayı etkilemeye çalıştığı için 1.5 milyon dolar ceza ödemeyi kabul etmiştir.²⁸
- **Ders:** Otonom sistemlerde "Human-in-the-loop" (döngüde insan) eksikliği ve kurumsal şeffaflığın olmaması, teknik hataları büyük çaplı kurumsal ve toplumsal krizlere dönüştürür.

Şeffaflık, teknik bir özellik değil, bir hayatta kalma stratejisidir.

4.2. Vaka 2: Sağlık Algoritmalarında Yapısal Irkçılık - Obermeyer Çalışması

4.2.1. Olayın Açıklaması

2019 yılında *Science* dergisinde yayınlanan çığır açıcı bir çalışma (Obermeyer et al.), ABD hastanelerinde milyonlarca hasta için kullanılan ticari bir algoritmanın, siyahilere karşı sistematik ayrımcılık yaptığını ortaya koymuştur. Bu algoritma, hangi hastaların "yüksek riskli" olduğunu ve dolayısıyla "ekstra bakım yönetimi" (care management) programlarına alınması gerektiğini belirlemektedir. Algoritma, aynı kronik hastalık seviyesindeki beyaz hastalara, siyah hastalara oranla çok daha yüksek risk puanı vererek onları öncelikli bakım listesine almıştır. Siyah hastaların listeye girebilmesi için beyazlardan çok daha hasta olması gerekmektedir.¹⁹

4.2.2. Teknik Hata / Veri Sorunu: Proxy Değişken Yanılgısı

Bu vaka, veri kalitesinden veya kötü niyetten ziyade, "etik tasarım" hatasının (Design Flaw) ve "**Vekil Değişken**" (**Proxy Variable**) kullanımının tehlikelerini gösterir.

- **Sorun:** Geliştiriciler, algoritmayı eğitirken "hastanın gelecekteki sağlık durumu ne kadar kötü olacak?" sorusuna yanıt aramıştır. Ancak "sağlık durumu" doğrudan ölçülebilir somut bir veri değildir. Bu nedenle, geliştiriciler sağlık durumunu temsil etmesi için bir vekil değişken olarak "**gelecekteki sağlık harcamaları**"nı (**healthcare costs**) kullanmışlardır.
- **Varsayım:** "Bir insan ne kadar hastaysa, hastaneye ve tedaviye o kadar çok para harcar."
- **Gerçek:** ABD sağlık sistemindeki yapısal eşitsizlikler nedeniyle, siyahiler aynı hastalık seviyesinde olsalar bile, sigorta eksikliği, hastaneye erişim zorluğu, doktorların onlara karşı önyargılı davranarak daha az test yapması veya sisteme güvensizlik gibi nedenlerle beyazlardan daha az para harcamaktadır.
- **Sonuç:** Algoritma, "sağlık ihtiyacını" değil, "harcama kapasitesini" optimize etmiştir. Sonuç olarak, çok hasta olan ancak sisteme para harcamayan siyahiler, algoritma tarafından "az harcama yaptıkları" için "sağlıklı" olarak etiketlenmiş ve ihtiyaç duydukları ekstra hizmetten mahrum bırakılmıştır. Bu, **Ölçüm Önyargısı (Measurement Bias)** örneğidir.¹⁶

4.2.3. Etik İhlal ve Çözüm Önerileri

- **İhlal:** Adalet (Fairness) ve Zarar Vermeme (Non-maleficence) ilkeleri ihlal edilmiştir. Algoritma, mevcut toplumsal eşitsizliği (sağlığa erişim farkı) teknolojik bir araçla pekiştirmiştir.
- **Çözüm:** Araştırmacılar, hedef değişkeni "maliyet" yerine "kronik hastalık sayısı" (biyolojik veri) gibi doğrudan sağlık verileriyle değiştirdiğinde, ırksal önyargı %84 oranında azalmıştır.
- **Ders:** Algoritmik tasarımda seçilen hedef değişkenler (Objective Function), toplumsal gerçekliklerden bağımsız düşünülemez. Teknik optimizasyon, etik sonuçları garanti etmez; aksine yanlış hedefe optimize edilmiş bir algoritma, ayrımcılığı otomatize eder.

4.3. Vaka 3: İşe Alımda Cinsiyet Ayrımcılığı - Amazon İşe Alım Algoritması

4.3.1. Olayın Açıklaması

2014 yılında Amazon, işe alım süreçlerini otomatize etmek ve en iyi yetenekleri bulmak amacıyla bir yapay zeka aracı geliştirmiştir. Ancak 2015 yılında, sistemin yazılım geliştirici ve teknik pozisyonlar için kadın adaylara karşı önyargılı olduğu fark edilmiştir. Sistem, özgeçmişinde "kadın" ifadesi geçen (örneğin "kadın satranç kulübü kaptanı" veya sadece kadın kolejlerinden mezun olan) adayların puanını düşürmüştür. Amazon, önyargıyı düzeltemediği için projeyi 2018'de sonlandırmıştır.³²

4.3.2. Teknik Analiz: Tarihsel Önyargının Tekrarı

Bu vaka, **Tarihsel Önyargı (Historical Bias)** ve Doğal Dil İşleme (NLP) hatalarının birleşimini gösterir.

- **Veri Seti:** Model, Amazon'a son 10 yılda gönderilen özgeçmişlerle eğitilmiştir. Teknoloji sektöründeki erkek egemenliği nedeniyle, bu özgeçmişlerin büyük çoğunluğu erkeklere aitti.
- **Örüntü Tanıma:** Algoritma, başarılı adayların (erkeklerin) ortak özelliklerini öğrenmiştir. Erkek adayların CV'lerinde daha sık kullandığı "executed" (yürüttü), "captured" (ele geçirdi) gibi agresif fiilleri başarı kriteri olarak belirlerken, kadınların kullandığı dili negatif puanlamıştır.
- **Sonuç:** Model, veri setindeki cinsiyet dengesizliğini "erkek olmak = iyi aday" şeklinde yorumlamıştır. "Kadın" kelimesini negatif bir özellik (feature) olarak öğrenmiştir.³⁴

4.3.3. Çıkarılan Dersler

Veri setindeki dengesizlik, sadece teknik bir sorun değildir; modelin dünya görüşünü şekillendirir. Geçmiş verilerle eğitilen modeller, geçmişin adaletsizliklerini geleceğe taşır. Bu nedenle işe alım gibi kritik alanlarda **"Algoritmik Denetim" (Algorithmic Auditing)** ve veri dengeleme teknikleri zorunludur.¹⁸

4.4. Vaka 4: Kurumsal Finansta Üretken YZ Dolandırıcılığı - Arup Deepfake Vakası (2024)

4.4.1. Olayın Açıklaması

2024 başlarında, Hong Kong merkezli çok uluslu mühendislik firması Arup'un bir finans çalışanı, şirketin CFO'sundan (Mali İşler Müdürü) gizli bir işlem yapmasını isteyen bir e-posta almıştır. Çalışan şüphelenmiş, ancak daha sonra katıldığı bir video konferans görüşmesinde, CFO ve diğer üst düzey yöneticilerin de orada olduğunu görünce ikna olmuştur. Ancak görüşmedeki herkes (çalışan hariç), gerçek zamanlı oluşturulmuş Deepfake kopyalarıdır. Sonuç

olarak çalışan, dolandırıcılara 25 milyon dolar (200 milyon HKD) transfer etmiştir.³⁵

4.4.2. Teknik Analiz

- **GANs ve Ses Klonlama:** Dolandırıcılar, şirketin halka açık videolarını (YouTube vb.) kullanarak yöneticilerin yüzlerini ve seslerini eğitmişlerdir. Generative Adversarial Networks (GANs) kullanılarak gerçek zamanlı, düşük gecikmeli (low latency) video ve ses sentezi yapılmıştır.
- **Sosyal Mühendislik:** Saldırı sadece teknik değil, psikolojiktir. "Güven hiyerarşisi" manipüle edilmiştir. Çalışan, çok katılımcılı bir toplantıda, tanıdığı yüzleri görünce "sosyal kanıt" (social proof) ilkesi gereği şüphelerini bastırmıştır.

4.4.3. BT Etiği ve Güvenlik Çıkarımları

Bu vaka, "gördüğüne inanma" döneminin sona erdiğini kanıtlamaktadır.

- **Kimlik Doğrulama Krizi:** Geleneksel görsel ve işitsel doğrulama yöntemleri çökmüştür.
- **Güvenlik Protokolleri:** Kurumsal güvenlik protokolleri artık "**Sıfır Güven**" (**Zero Trust**) prensibini biyometrik verileri de kapsayacak şekilde genişletmelidir. Çok faktörlü kimlik doğrulama (MFA) ve finansal işlemlerde "bant dışı" (out-of-band) doğrulama (örneğin video görüşmesi sırasında ayrıca telefonla veya şifreli mesajla teyit) zorunlu hale gelmelidir. Ayrıca, çalışanların Deepfake teknolojisinin geldiği son nokta (gerçek zamanlı interaksyon) hakkında eğitilmesi hayati önem taşımaktadır.³⁹

5. BT Etiğinde Kullanılan Çerçeveler ve Düzenleyici Standartlar

Yapay zeka etiği, artık sadece gönüllü prensiplerle değil, sert regülasyonlar ve uluslararası standartlarla yönetilmektedir.

5.1. IEEE ve ACM Etik İlkeleri (Yumuşak Hukuk)

Mühendislik meslek örgütleri, YZ etiğinin temelini oluşturur.

- **IEEE Ethically Aligned Design:** "İnsan Refahı"nı (Human Well-being) en üst değer olarak tanımlar. Otonom sistemlerin, insan haklarını ihlal etmeyecek şekilde tasarlanmasını öngörür. Mühendisler için somut tasarım metodolojileri (örneğin IEEE P7000 serisi standartlar) sunar.³
- **ACM Code of Ethics:** "Zarar vermekten kaçın" (Avoid harm) ve "Mahremiyete saygı göster" ilkelerini temel alır. Yazılım mühendislerinin, geliştirdikleri kodun toplumsal etkilerini düşünmekle yükümlü olduğunu belirtir.

5.2. Avrupa Birliği Yapay Zeka Yasası (EU AI Act)

Dünyanın ilk kapsamlı YZ yasası olan AI Act (2024), teknolojiye değil, kullanım alanına ve riske

odaklanan "risk temelli" bir yaklaşım benimser.¹³ Mühendisler için bu yasa, kod yazmak kadar teknik dokümantasyon ve risk analizini de zorunlu kılar.

Risk Seviyesi	Örnekler	Yükümlülükler
Kabul Edilemez Risk (Yasaklı)	Sosyal puanlama (Social Scoring), iş yerinde duygu tanıma, manipülatif (bilinçaltı) teknikler, gerçek zamanlı uzaktan biyometrik tanımlama (istisnalar hariç).	Geliştirilmesi ve kullanımı tamamen yasaktır.
Yüksek Risk	Kritik altyapılar (ulaşım, enerji), eğitim, işe alım, kredi skortlama, sınır kontrolü, adalet sistemleri, tıbbi cihazlar.	Veri kalitesi, detaylı teknik dokümantasyon, şeffaflık, insan gözetimi (human oversight), siber güvenlik ve uygunluk değerlendirme (conformity assessment) zorunludur.
Sınırlı Risk	Chatbotlar (müşteri hizmetleri), Deepfake'ler.	Şeffaflık zorunluluğu: Kullanıcı bir makine ile konuştuğunu bilmelidir. Deepfake içerikler filigranla işaretlenmelidir.
Minimum Risk	Spam filtreleri, video oyunları, envanter yönetimi.	Ekstra yükümlülük yoktur, gönüllü davranış kuralları önerilir.

Genel Amaçlı YZ (GPAI): ChatGPT gibi temel modeller (Foundation Models) için şeffaflık, eğitim verisi özetleri ve telif hakkı uyumu şarttır. "Sistemik risk" taşıyan modeller için ise "Kırmızı Takım" (Red Teaming) testleri ve olay raporlama zorunludur.¹⁴ Uyumsuzluk durumunda küresel cironun %7'sine (veya 35 milyon Avro) varan cezalar öngörülmektedir.⁴³

5.3. KVKK ve GDPR Uyumu

Türkiye'de KVKK ve AB'de GDPR, YZ'nin "veri yakıtını" düzenler.

- **Otomatik Karar Verme (Automated Decision Making - Profiling):** Bireylerin, sadece otomatik işleme dayalı olarak kendileri hakkında hukuki veya benzeri önemli sonuçlar doğuran kararlara (örneğin kredi reddi veya işe alınmama) itiraz etme ve insan müdahalesi

talep etme hakkı vardır (GDPR Madde 22).

- **Unutulma Hakkı ve Machine Unlearning:** Bir YZ modeli, verisi silinmek istenen bir kullanıcıyla eğitildiyse, bu verinin modelden "unutulması" teknik olarak çok zordur (modelin yeniden eğitilmesini gerektirebilir) ancak yasal bir gerekliliktir.¹⁰

5.4. Kurumsal Yönetişim: ISO 42001, COBIT ve ITIL

- **ISO/IEC 42001 (Yapay Zeka Yönetim Sistemi):** Kurumların YZ'yi sorumlu bir şekilde geliştirmesi veya kullanması için gereken süreçleri standartlaştırır. Risk yönetimi, şeffaflık ve sürekli iyileştirme döngülerini (PUKÖ) içerir.
- **COBIT 2019:** YZ risklerinin kurumsal risk yönetimi (ERM) çerçevesine entegre edilmesini sağlar. YZ'nin iş hedefleriyle uyumlu olmasını ve kaynakların verimli kullanılmasını denetler.
- **ITIL 4:** YZ'yi bir BT hizmeti olarak ele alır. Hizmet seviyesi yönetimi (SLA), olay yönetimi ve değişim yönetimi süreçlerinin YZ sistemlerine uyarlanmasını sağlar. Örneğin, kendi kendine öğrenen bir modelin güncellenmesi, standart bir yazılım yaması (patch) gibi değil, risk odaklı bir "değişim yönetimi" süreci olarak ele alınmalıdır.¹

6. Sonuç ve Öneriler

6.1. Genel Değerlendirme

Yapay zeka etiği, felsefi bir "iyi niyet" gösterisi olmaktan çıkıp, mühendislik pratiğinin, risk yönetiminin ve yasal uyumluluğun zorunlu bir bileşeni haline gelmiştir. Cruise, Amazon ve Obermeyer vakaları açıkça göstermektedir ki, sadece teknik başarı (doğruluk oranı, hız, verimlilik) odaklı geliştirme süreçleri, şirketin itibarını, mali durumunu ve toplumun güvenliğini yıkıcı şekilde etkileyebilmektedir. "Kara kutu" modellerin açıklanabilirliğinin sağlanamaması, verideki önyargıların temizlenememesi ve insan denetiminin ihmal edilmesi, sürdürülebilir bir teknolojik geleceğin önündeki en büyük engellerdir.

6.2. Kurumsal Düzeyde Uygulanabilir Öneriler

1. **AI Etik Kurulları ve Yönetişim Yapısı:** Şirket içinde hukuk, mühendislik, veri bilimi ve sosyal bilimcilerden oluşan disiplinler arası kurullar ("AI Ethics Board") kurulmalı ve yüksek riskli projeler, geliştirme aşamasında bu kurulun onayından geçmelidir (Stage-gate süreci).
2. **Kırmızı Takım (Red Teaming) Testleri:** Modelleri piyasaya sürmeden önce, onları kandırmaya, önyargılarını ortaya çıkarmaya, güvenlik açıklarını bulmaya ve "halüsinasyon" gördürmeye çalışan bağımsız "Red Team" ekipleri oluşturulmalıdır.
3. **Algoritmik Etki Değerlendirmesi (AIA):** Çevresel Etki Değerlendirmesi (ÇED) raporları gibi, her YZ projesi öncesinde potansiyel toplumsal zararlar, ayrımcılık riskleri ve mahremiyet etkileri analiz edilmeli ve belgelenmelidir.
4. **Veri ve Model Kartları (Data & Model Cards):** Kullanılan veri setlerinin içeriği, kaynağı, sınırlılıkları ve modelin hangi koşullarda çalışıp hangi koşullarda çalışmayacağı (intended

use vs. limitations) standart formatlarda (Model Cards) dokümente edilmelidir.

5. **Sürekli Eğitim:** Yazılım mühendislerine ve veri bilimcilere sadece kodlama değil, veri etiği, yasal uyumluluk (AI Act/KVKK) ve yanlışlık tespiti konularında düzenli eğitimler verilmelidir.

Proje Sunum Taslağı

Sunum Başlığı: Yapay Zeka Etiği: Riskler, Vakalar ve Mühendislik Sorumlulukları
Hazırlayan:

Slayt 1: Giriş ve Bağlam

- Başlık:** Neden Şimdi? Yapay Zeka Etiğinin Yükselişi
- İçerik:**
 - Dönüşüm:** Geleneksel BT Etiğinden (İnsan odaklı) YZ Etiğine (Otonom sistem odaklı) geçiş.
 - Fark:** Deterministik yazılımlardan Stokastik (Olasılıksal) modellere geçişin yarattığı belirsizlik.
 - Temel Soru:** "Yapabilir miyiz?" (Teknik Kapasite) vs. "Yapmalı mıyız?" (Etik ve Yasal Sorumluluk).

Slayt 2: Temel Etik Problemler (Risk Haritası)

- Başlık:** YZ Sistemlerinin Risk Yüzeyi ve Teknik Karşılıkları
- Tablo:**
 - Önyargı (Bias):** Temsil, Tarihsel ve Ölçüm Önyargısı.
 - Kara Kutu (Black Box):** Açıklanabilirlik (XAI) eksikliği ve güven sorunu.
 - Mahremiyet:** Model Inversion ve Membership Inference saldırıları.
 - Güvenlik:** Adversarial (Karşıt) örnekler ve Veri Zehirlenme (Data Poisoning).

Slayt 3: Vaka Analizi 1 - Otonom Araçlar

- Başlık:** Cruise Robotaxi Kazası (San Francisco 2023)
- Görsel:** Olayın şematik çizimi (Yayanın sürüklenmesi).
- Analiz:**
 - Teknik Hata:** Semantik Sınıflandırma Hatası (Yaya vs. Yanal Çarpışma) ve Nesne Sürekliliği (Object Permanence) kaybı.
 - Etik Hata:** Şeffaflık eksikliği (Düzenleyicilerden videonun gizlenmesi).
 - Sonuç:** Lisans iptali, 1.5 Milyon \$ ceza, CEO istifası, %25 işten çıkarma.

Slayt 4: Vaka Analizi 2 - Algoritmik Ayrımcılık

- Başlık:** Sağlıkta Görünmez Ayrımcılık (Obermeyer Çalışması)

- **İçerik:**

- **Amaç:** En yüksek riskli hastaları tespit edip önceliklendirmek.
- **Tasarım Hatası:** "Sağlık Durumu" yerine "Sağlık Harcaması"nı vekil değişken (proxy) olarak kullanmak.
- **Sonuç:** Siyahilerin daha az tedavi alması (Paraları olmadığı için harcamaları düşüktü, sağlıklı oldukları için değil).
- **Ders:** Değişken seçimi (Feature Selection) etik bir karardır.

Slayt 5: Vaka Analizi 3 - Deepfake Dolandırıcılığı

- **Başlık:** Arup 25 Milyon \$ Deepfake Vakası (2024)

- **İçerik:**

- **Olay:** Bir finans çalışanının, tamamı Deepfake olan bir video toplantıda 25M \$ transfer etmesi.
- **Mekanizma:** Gerçek zamanlı GANs (Video/Ses Klonlama) + Sosyal Mühendislik.
- **Çıkarım:** "Gördüğüne inanma" dönemi. Sıfır Güven (Zero Trust) mimarisi ve bant dışı doğrulama şart.

Slayt 6: Düzenlemeler ve Standartlar

- **Başlık:** Oyunun Kuralları: EU AI Act

- **Tablo:** AI Act Risk Piramidi

- **Yasaklı:** Sosyal Puanlama, Duygu Tanıma (İş yeri/Okul).
- **Yüksek Risk:** İşe alım, Sağlık, Ulaşım, Kredi (Sıkı denetim, Veri kalitesi, İnsan gözetimi).
- **Sınırlı Risk:** Chatbot, Deepfake (Şeffaflık şartı).

- **Ek Standartlar:** ISO 42001 (Yönetim Sistemi), IEEE P7000.

Slayt 7: Sonuç ve Yol Haritası

- **Başlık:** Sorumlu YZ İçin Ne Yapmalı?

- **Öneriler:**

1. **Tasarım aşamasında etik (Ethics by Design):** Sonradan yama yapılamaz.
2. **Kırmızı Takım (Red Teaming):** Sisteminizi kendiniz "hack"leyin.
3. **Disiplinler arası Etik Kurulları:** Sadece mühendisler karar veremez.
4. **Sürekli Gözetim:** "Human-in-the-loop" mekanizmaları.

Bu rapor ve sunum taslağı, sağlanan araştırma materyalleri ve akademik kaynaklar¹ ışığında hazırlanmıştır.

Works cited

1. BİL493-Mühendislik Etiği-Yapay Zeka Etiği Raporu.pdf
2. The Ethical Dimension of Artificial Intelligence - DergiPark, accessed December 2, 2025, <https://dergipark.org.tr/en/download/article-file/2962187>

3. IEEE Ethically Aligned Design - Palo Alto Networks, accessed December 2, 2025, <https://www.paloaltonetworks.com/cyberpedia/ieee-ethically-aligned-design>
4. YAPAY ZEKÂ ETİĞİ: TEMEL İLKELER, SORUNLAR VE DİSİPLİNLERARASI YAKLAŞIMLAR - DergiPark, accessed December 2, 2025, <https://dergipark.org.tr/tr/pub/inifedergi/issue/92327/1605400>
5. Algorithmic Bias in Hiring: Fact or Myth? - Mitch Daniels School of Business, accessed December 2, 2025, <https://business.purdue.edu/daniels-insights/posts/2025/algorithmic-bias-in-hiring.php>
6. OTONOM ARAÇLARDA ORTAYA ÇIKAN ETİK İKİLEMLER ÇERÇEVESİNDE ÜRETİCİNİN TASARIM YÜKÜMLÜLÜĞÜ - DergiPark, accessed December 2, 2025, <https://dergipark.org.tr/tr/pub/taad/issue/93885/1752847>
7. Ethical Decision-Making for Self-Driving Vehicles: A Proposed Model & List of Value-Laden Terms that Warrant (Technical) Specification - PubMed Central, accessed December 2, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11466986/>
8. comparative analysis of ethical incident approaches in generative artificial intelligence applications - DergiPark, accessed December 2, 2025, <https://dergipark.org.tr/en/download/article-file/4599917>
9. Destek Süreçlerinde Üretken Yapay Zekânın (ÜYZ) Sorumlu ve Güvenilir Kullanımı Rehberi - TÜBİTAK, accessed December 2, 2025, https://tubitak.gov.tr/sites/default/files/2025-10/UYZ_Rehberi_v03_TR.pdf
10. YAPAY ZEKÂYA - KVKK, accessed December 2, 2025, <https://kvkk.gov.tr/SharedFolderServer/CMSFiles/d4a738b6-5a86-454f-8788-b97758cab0da.pdf>
11. Yapay Zeka Ve Eğitim Araştırması Akademik Bir Vaka Analizi - Scribd, accessed December 2, 2025, <https://www.scribd.com/document/917947648/Yapay-Zeka-Ve-E%C4%9Fitim-Ara%C5%9Ft%C4%B1rmas%C4%B1-Akademik-Bir-Vaka-Analizi>
12. Makale » Yapay Zekâ Etiği: Toplum Üzerine Etkisi - DergiPark, accessed December 2, 2025, <https://dergipark.org.tr/tr/pub/makufebed/issue/71318/1058538>
13. High-level summary of the AI Act | EU Artificial Intelligence Act, accessed December 2, 2025, <https://artificialintelligenceact.eu/high-level-summary/>
14. AI Act | Shaping Europe's digital future - European Union, accessed December 2, 2025, <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
15. IEEE Ethically Aligned Design Overview - Emergent Mind, accessed December 2, 2025, <https://www.emergentmind.com/topics/ieee-ethically-aligned-design-ead>
16. Dissecting racial bias in an algorithm used to manage the health of populations, accessed December 2, 2025, https://www.ftc.gov/system/files/documents/public_events/1548288/privacycon-2020-ziad_obermeyer.pdf
17. Upturn | Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias, accessed December 2, 2025, <https://www.upturn.org/static/reports/2018/hiring-algorithms/files/Upturn%20--%20Help%20Wanted%20-%20An%20Exploration%20of%20Hiring%20Algorithms,%20Equity%20and%20Bias.pdf>

18. Discrimination and bias in AI recruitment: a case study - Lewis Silkin LLP, accessed December 2, 2025, <https://www.lewissilkin.com/en/insights/2023/10/31/discrimination-and-bias-in-ai-recruitment-a-case-study>
19. Ziad Obermeyer testifies in U.S. Congress on how AI can help health care, accessed December 2, 2025, <https://cdss.berkeley.edu/news/ziad-obermeyer-testifies-us-congress-how-ai-can-help-health-care>
20. Dissecting racial bias in an algorithm used to manage the health of populations - PubMed, accessed December 2, 2025, <https://pubmed.ncbi.nlm.nih.gov/31649194/>
21. A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring, accessed December 2, 2025, <https://www.mdpi.com/2673-2688/5/1/19>
22. Case Studies: When AI and CV Screening Goes Wrong - Fairness Tales, accessed December 2, 2025, <https://www.fairnesstales.com/p/issue-2-case-studies-when-ai-and-cv-screening-goes-wrong>
23. Ethical issues in focus by the autonomous vehicles industry - Taylor & Francis Online, accessed December 2, 2025, <https://www.tandfonline.com/doi/abs/10.1080/01441647.2020.1862355>
24. Lessons from the Cruise Robotaxi Pedestrian Dragging Mishap - Carnegie Mellon University, accessed December 2, 2025, https://users.ece.cmu.edu/~koopman/cruise/Koopman2024_CruiseMishap_IEEEReliabilityMagazine.pdf
25. A Root Cause Analysis of a Self-Driving Car Dragging a Pedestrian, accessed December 2, 2025, <https://www.computer.org/csdl/magazine/co/2024/11/10720344/215PD0vqgTe>
26. Internal Report Shows Cruise Didn't Think Its Robotaxi Dragging A Pedestrian Was A Big Enough Deal To Fix The Cars - The Autopian, accessed December 2, 2025, <https://www.theautopian.com/internal-report-shows-cruise-didnt-think-its-robotaxi-dragging-a-pedestrian-was-a-big-enough-deal-to-fix-the-cars/>
27. Analysis reveals how GM's Cruise robotaxi struck and dragged pedestrian 20 feet, accessed December 2, 2025, <https://www.foxbusiness.com/technology/analysis-reveals-gms-cruise-robotaxi-s-truck-dragged-pedestrian-20-feet>
28. Cruise Admits To Submitting A False Report To Influence A Federal Investigation And ... - Department of Justice, accessed December 2, 2025, <https://www.justice.gov/usao-ndca/pr/cruise-admits-submitting-false-report-influence-federal-investigation-and-agrees-pay>
29. GM's Cruise admits submitting false report to robotaxi safety investigation - The Guardian, accessed December 2, 2025, <https://www.theguardian.com/technology/2024/nov/15/gm-cruise-self-driving-taxi>
30. NHTSA Announces Consent Order with Cruise After Company Failed to Fully Report Crash Involving Pedestrian, accessed December 2, 2025, <https://www.nhtsa.gov/press-releases/consent-order-cruise-crash-reporting>

31. Rooting Out AI's Biases | Hopkins Bloomberg Public Health Magazine, accessed December 2, 2025, <https://magazine.publichealth.jhu.edu/2023/rooting-out-ais-biases>
32. Amazon's sexist hiring algorithm could still be better than a human - IMD Business School, accessed December 2, 2025, <https://www.imd.org/research-knowledge/digital/articles/amazons-sexist-hiring-algorithm-could-still-be-better-than-a-human/>
33. Why Amazon's Automated Hiring Tool Discriminated Against Women | ACLU, accessed December 2, 2025, <https://www.aclu.org/news/womens-rights/why-amazons-automated-hiring-tool-discriminated-against>
34. Why Amazon's AI-driven high volume hiring project failed - Hubert, accessed December 2, 2025, <https://www.hubert.ai/insights/why-amazons-ai-driven-high-volume-hiring-project-failed>
35. Deepfake: A Horrifying Tale of a \$25 Million Cybercrime - VendorInfo, accessed December 2, 2025, <https://vendorinfo.com/deepfake-a-horrifying-tale-of-a-25-million-cybercrime/>
36. Cyber Case Study: \$25 Million Deepfake Scam - CoverLink Insurance, accessed December 2, 2025, <https://coverlink.com/case-study/case-study-25-million-deepfake-scam/>
37. \$25M Deepfake CEO Scam Shakes Hong Kong Firm by Doron Ish Shalom - Clarity, accessed December 2, 2025, <https://www.getclarity.ai/ai-deepfake-blog/25m-deepfake-ceo-scam-shakes-hong-kong-firm>
38. UK engineering firm Arup falls victim to £20m deepfake scam - The Guardian, accessed December 2, 2025, <https://www.theguardian.com/technology/article/2024/may/17/uk-engineering-arup-deepfake-scam-hong-kong-ai-video>
39. Detecting dangerous AI is essential in the deepfake era | World Economic Forum, accessed December 2, 2025, <https://www.weforum.org/stories/2025/07/why-detecting-dangerous-ai-is-key-to-keeping-trust-alive/>
40. Cyber security awareness month - Case Study - Deepfakes (DOCX, 25.72KB), accessed December 2, 2025, https://www.wa.gov.au/system/files/2024-10/case.study_deepfakes.docx
41. EU Artificial Intelligence Act | Up-to-date developments and analyses of the EU AI Act, accessed December 2, 2025, <https://artificialintelligenceact.eu/>
42. What is the Artificial Intelligence Act of the European Union (EU AI Act)? - IBM, accessed December 2, 2025, <https://www.ibm.com/think/topics/eu-ai-act>
43. The EU AI Act: What Businesses Need To Know | Insights - Skadden, accessed December 2, 2025, <https://www.skadden.com/insights/publications/2024/06/quarterly-insights/the-eu-ai-act-what-businesses-need-to-know>
44. KVKK Yapay Zeka Rehberleri - avaiersin.com, accessed December 2, 2025,

<https://avaliersin.com/kvkk-yapay-zeka-rehberleri/>

45. Designing Ethical Self-Driving Cars | Stanford HAI, accessed December 2, 2025,
<https://hai.stanford.edu/news/designing-ethical-self-driving-cars>