

VERİ BİLİMİ DERSİ

# AI Etik ve Capstone Projesi

*Hafta 15 · Modül 15 (Final)*

*Sorumlu AI, Gelecek Vizyonu ve Büyük Final Projesi*

**Dr. Murat Altun**

Veri Bilimi ve Yapay Zekâ Eğitimi · 2026

6

Saat

2

Notebook

5

Proje Seçeneği

# İçindekiler

01

## AI Etik

Bias · Fairness · Transparency · EU AI Act · Sorumlu AI İlkeleri

Slayt 3-9

02

## AI'nın Geleceđi

AGI · Multimodal · Otonom Ajanlar · Kariyer Yolları · 15 Hafta Özeti

Slayt 10-12

03

## Capstone Projesi

5 Proje Seçeneđi · Deđerlendirme · Notebook'lar · Ödev

Slayt 13-19

## Tanım

AI etiđi, yapay zekâ sistemlerinin tasarlanması, geliştirilmesi ve kullanılması sırasında ortaya çıkan ahlaki soruları, toplumsal etkileri ve sorumlulukları inceleyen disiplinler arası bir alandır.

## Neden Önemli?

- AI kararları milyarlarca insanı etkiliyor
- Önyargılı modeller eşitsizliđi derinleştirebilir
- Şeffaflık olmadan güven inşa edilemez

## Gerçek Dünya Etkileri

- Ceza adalet sistemlerinde ırkçı tahminler
- Kredi skorlamasında cinsiyet ayrımcılıđı
- Deepfake ile dezenformasyon yayılımı

78%

AI Liderlerinin  
Etik Endişeleri

40+

Ülkede AI  
Düzenleme Çalışması

\$4.7T

2030 AI Pazar  
Büyüküğü

## 1 Veri Yanlılığı (Data Bias)

- Eğitim verisi toplumu temsil etmiyorsa
- Tarihi önyargılar veriye yansır
- Örneğin: Yüz tanımada koyu ten rengi için düşük başarı oranı

## 2 Algoritma Yanlılığı

- Model mimarisinin kendisi yanlı olabilir
- Optimizasyon hedefi etik değil
- Proxy değişkenler üzerinden dolaylı ayrımcılık

## ⚠ VAKA ÇALIŞMASI: Amazon İşe Alım AI'ı

### Problem

Amazon'ın işe alım AI'ı kadın adayları sistematik olarak düşük puanladı

### Neden

10 yıllık çoğunlukla erkek çalışan verisiyle eğitildi — model 'erkek = başarılı' öğrendi

### Sonuç

Sistem 2018'de kapatıldı. AI bias'ın en bilinen örneklerinden biri oldu

Adalet Türü	Tanım	Örnek
Demografik Eşitlik	Her gruba eşit oranda olumlu karar	Kredi onayı her etnik grup için %50
Fırsat Eşitliği	Gerçek pozitifler için eşit doğru pozitif oranı	Hastalığı olan herkes eşit tanı şansı
Bireysel Adalet	Benzer kişilere benzer kararlar	Aynı niteliklere sahip kişiler aynı sonuç
Kalibre Edilmiş	Tahmin olasılıkları gerçek oranlarla örtüşsün	%80 güven = gerçekte %80 doğru

## Adalet Paradoksu

Tüm adalet tanımlarını aynı anda sağlamak matematiksel olarak imkânsızdır (Chouldechova, 2017). Bu nedenle hangi adalet tanımının uygulanacağı bağlama (domain) göre karar verilmelidir. Sağlık alanında fırsat eşitliği, kredi skorlamasında kalibre edilmiş adalet tercih edilir.



## Explainable AI (XAI)

Model kararlarının neden verildiğini açıklayabilme yeteneği. Kullanıcı güveni ve regülasyon uyumu için kritik.



## Black Box Problemi

Derin öğrenme modelleri milyarlarca parametreyle karar verir. İç mantık insan için anlaşılmaz. Hayati kararlarda kabul edilemez.



## LIME & SHAP

LIME: Yerel açıklamalar üretir (her tahmin için).  
SHAP: Shapley değerleriyle global özellik önemi hesaplar.

Veri Girişi



Model Kararı



XAI Açıklaması



Kullanıcı Güveni

## YASAK

### Kabul Edilemez Risk

Sosyal skorlama, biyometrik k tle g zetimi, manipulatif AI

## YÜKSEK

### Yüksek Risk

Saęlık AI, iŐe alım, eęitim, kredi skorlama, adalet sistemi

## SINIRLI

### Sınırlı Risk

Chatbotlar, deepfake i erik — Őeffaflık y ks ml l ę 

## MİNİMAL

### Minimal Risk

Spam filtre, oyun AI,  neri sistemleri — d zenleme yok

Y r rl k: 2024 onay → 2025 yasak uygulamaları → 2026 y ksek risk y ks ml l kleri → 2027 tam uygulama

1

## Güvenlik

AI sistemleri güvenli ve sağlam olmalı. Adversarial saldırılara dayanıklı, hata durumlarında güvenli mod.

2

## Adalet

Tüm gruplar için adil sonuçlar. Bias tespiti ve azaltma süreçleri zorunlu.

3

## Gizlilik

Kişisel verilerin korunması. Differential privacy, federe öğrenme, veri minimizasyonu.

4

## Şeffaflık

Model kararları açıklanabilir olmalı. Dokümantasyon, model kartları, veri sayfaları.

5

## Hesap Verebilirlik

AI kararlarından sorumlu kişi/kurum belirlenmeli. Denetim ve log mekanizmaları.

6

## Kapsayıcılık

Engelli bireyler dahil herkes için erişilebilir. Çok dilli, kültürel farkındalık.

```
# Fairness metrikleri hesapla
import pandas as pd
from sklearn.metrics import confusion_matrix

def demographic_parity(y_pred, group):
    """Demografik eşitlik: P(Y=1|G=a) = P(Y=1|G=b)"""
    rates = {}
    for g in group.unique():
        mask = (group == g)
        rates[g] = y_pred[mask].mean()
    return rates

def equalized_odds(y_true, y_pred, group):
    """Fırsat eşitliği: TPR ve FPR gruplar arası eşit"""
    results = {}
    for g in group.unique():
        mask = (group == g)
        tn, fp, fn, tp = confusion_matrix(
            y_true[mask], y_pred[mask]).ravel()
        tpr = tp / (tp + fn) if (tp + fn) > 0 else 0
        fpr = fp / (fp + tn) if (fp + tn) > 0 else 0
        results[g] = {"TPR": tpr, "FPR": fpr}
    return results
```



## AGI (Genel Yapay Zekâ)

İnsan seviyesinde genel zekâ. Henüz teorik aşamada ama ilerlemeler hızlanıyor. Tahmin: 2030–2050 aralığı.



## Multimodal Modeller

Metin + görüntü + ses + video birleşik işleme. GPT-4V, Gemini, Claude örnekleri. Gerçek dünya algısı.



## Otonom Ajanlar

AI'nın kendi başına karar verip eylem yapması. Tool-use, planlama, multi-agent işbirliği.



## AI-Human Collaboration

İnsan + AI birlikte çalışma. Augmented intelligence yaklaşımı. Yaratıcılık ve karar desteđi.

Rol	Temel Beceriler	Ortalama Maş (ABD)	Talep
Data Scientist	Python, ML, İstatistik, SQL	\$130K+	★★★★★
ML Engineer	MLOps, Docker, Cloud, CI/CD	\$145K+	★★★★★
AI Researcher	Matematik, Paper yazma, PyTorch	\$160K+	★★★★★
Prompt Engineer	LLM'ler, Prompt Design, RAG	\$120K+	★★★★★
AI Ethics Officer	Hukuk, Sosyoloji, AI bilgisi	\$110K+	★★★
Computer Vision Eng.	OpenCV, YOLO, CNN, 3D Vision	\$140K+	★★★★★
NLP Engineer	Transformers, Tokenization, LLM	\$135K+	★★★★★

H1-3	Python, NumPy, Pandas, Görselleştirme
H4-5	İstatistik, A/B Testi, Olasılık
H6-7	ML Temelleri, Regresyon, Sınıflandırma
H8-9	Ensemble, XGBoost, Hiperparametre
H10-11	NLP, Transformers, Sentiment Analizi
H12-13	Deep Learning, CNN, Transfer Learning
H14	GenAI, RAG, Prompt Engineering
H15	AI Etik, Gelecek, Capstone (BU HAFTA!)

## OpenCV + MediaPipe ile Hareket Sayma ve Form Analizi

Kamera görüntüsünden vücut pozisyonlarını algılayarak egzersiz tekrarlarını otomatik sayan ve yanlış formu uyarıyan bir gerçek zamanlı AI asistanı.

1

### Pose Estimation

MediaPipe Pose ile 33 landmark tespit

Python

2

### Açı Hesaplama

Diz/dirsek açıları ile hareket fazları

OpenCV

3

### Tekrar Sayacı

Durum makinesi: yukarı → aşağı → tekrar++

MediaPipe

4

### Form Kontrolü

Yanlış açılarda uyarı gösterimi

Streamlit

## Collaborative Filtering + GenAI ile Ruh Haline Göre Öneri

Kullanıcı tercihleri ve ruh haline dayalı kişisel film önerileri sunan, hem geleneksel ML hem de GenAI kullanan hibrit öneri motoru.

### Collaborative Filtering

- User-Item matris faktörizasyonu
- Surprise / Implicit kütüphaneleri
- MovieLens 100K veri seti
- Cosine similarity + KNN
- Cold-start problemi çözümleri

### GenAI Öneri Katmanı

- Gemini API ile ruh hali analizi
- “Bugün üzgünüm” → komedi/drama önerisi
- Film özetinden embedding çıkarma
- Prompt: “Bu filme benzer 5 öner”
- Gradio ile etkileşimli arayüz

## RAG + Gemini API ile Doküman QA Sistemi

PDF, Word, CSV dokümanlarını yükleyin, sorular sorun, kaynaklı cevaplar alın. Kurumsal bilgi yönetimi için RAG pipeline.



## Kullanılan Teknolojiler

LangChain

RAG orkestrasyon

ChromaDB

Vektör veritabanı

Gemini API

LLM cevaplayıcı

Streamlit

Web arayüz

## Transfer Learning + Gradio Arayüz ile Görüntü Tanıma

Önceden eğitilmiş CNN modellerini (ResNet50, EfficientNet) fine-tune ederek özel görüntü sınıflandırma ve Gradio ile kullanıcı dostu arayüz.

### Model Mimarisi

ResNet50 (ImageNet ağırlıkları)

- ↓ Freeze: İlk 140 katman dondur
- ↓ Fine-tune: Son katmanları eğit
- ↓ Global Average Pooling
- ↓ Dense(256, ReLU) + Dropout(0.5)
- ↓ Dense(num\_classes, Softmax)

### Özellikler

- Data Augmentation (döndürme, zoom, çevirme)
- Learning Rate Scheduler (ReduceLRonPlateau)
- Early Stopping ile overfitting önleme
- Gradio arayüz: sürükle-bırak görüntü yükle
- Gerçek zamanlı confidence gösterimi
- Hugging Face Spaces'e deploy

**%40**

Teknik  
Uygulama

**%20**

Sunum ve  
Dokümantasyon

**%20**

Yenilikçilik  
ve Yaratıcılık

**%20**

Deploy ve  
Demo

Kriter	Açıklama	Puan
Kod Kalitesi	Temiz, yorumlu, modüler kod. PEP 8 uyumu.	15
Model Başarısı	Uygun metriklerle ölçülmüş performans.	15
Veri İşleme	EDA, temizleme, feature engineering kalitesi.	10
Sunum	Slayt kalitesi, akıcı anlatım, soru cevaplama.	10
Dokümantasyon	README, notebook açıklamaları, kullanım kılavuzu.	10
Yenilikçilik	Farklı yaklaşım, özgün fikir, ek özellikler.	20
Deploy	Çalışan demo, Streamlit/Gradio/HF Spaces.	20

## 1 capstone\_sablon.ipynb

- Capstone proje şablonu
- Proje yapısı: veri, model, değerlendirme
- README şablonu ve dokümantasyon
- requirements.txt örneği
- Git repo yapılandırması
- Streamlit/Gradio demo şablonu
- Model kaydetme ve yükleme
- Deployment kontrol listesi

## 2 ai\_etik.ipynb

- Bias analizi: COMPAS veri seti
- Fairness metrikleri hesaplama
- Demographic parity implementasyonu
- Equalized odds kontrolü
- SHAP ile model açıklanabilirlik
- LIME yerel açıklama örnekleri
- Model Card şablonu oluşturma
- Etik kontrol listesi çalışması

## Capstone Proje Teslimi

- 5 projeden birini seçin
- GitHub repo oluşturun
- Veri toplama ve EDA
- Model geliştirme ve değerlendirme
- Streamlit/Gradio demo
- Sunum hazırlığı (10 dk)

## Kaynaklar ve Okumalar

### EU AI Act (Resmi Metin)

[artificialintelligenceact.eu](https://artificialintelligenceact.eu)

### Google Responsible AI

[ai.google/responsibility](https://ai.google/responsibility)

### Microsoft RAI Toolkit

[github.com/microsoft/responsible-ai](https://github.com/microsoft/responsible-ai)

### Fairlearn Kütüphanesi

[fairlearn.org](https://fairlearn.org)

### SHAP Dokümantasyonu

[shap.readthedocs.io](https://shap.readthedocs.io)

### Model Cards (Google)

[modelcards.withgoogle.com](https://modelcards.withgoogle.com)

## TESLİM TARİHİ VE FORMAT

Format: GitHub repo linki + Jupyter Notebook + çalışan demo linki + 10 dakikalık sunum  
Son teslim: Eğitim bitiminden 2 hafta sonra. Geç teslim her gün için %5 puan düşümü.

# Büyük Final — 15 Haftalık Yolculuğun Özeti

1 AI etiği sadece teknik değil, toplumsal bir sorumluluktur.

2 Bias farkındalığı olmadan adil AI sistemleri inşa edilemez.

3 Sorumlu AI ilkeleri her projede tasarım aşamasından itibaren uygulanmalıdır.

4 15 haftada Python'dan GenAI'ye, istatistikten deep learning'e uzanan bir yolculuk tamamlandı.

5 Capstone projesi öğrendiklerinizi gerçek dünyada uygulamanın en güçlü yoludur.

*“Yapay zekâ gücü, onu sorumlu kullananların elindedir. Şimdi sıra sizde!”*