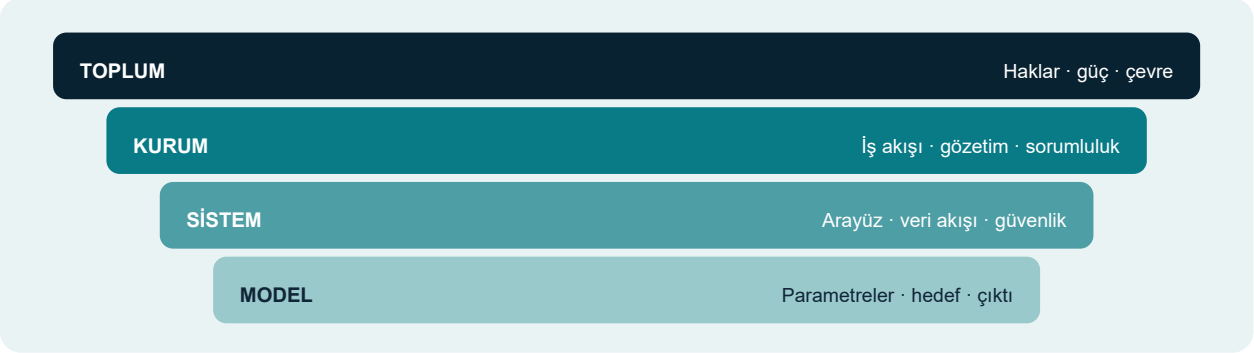
TEXTBOOK

Yapay Zekâ: Disiplinler Arası Ders Kitabı

Kavramdan uygulamaya, örnekli ve alıştırmalı akademik öğrenme rehberi



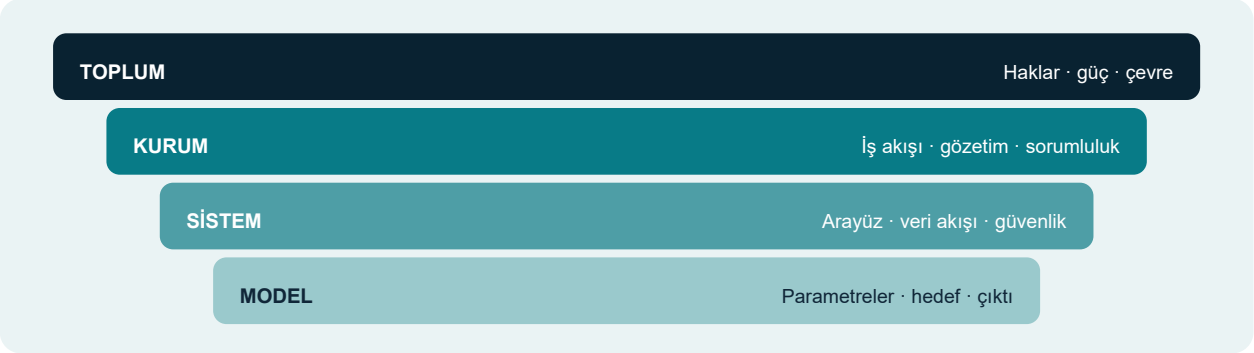
Kaynak kesiti: 4 Kasım 2026 · Türkçe özgün sentez · 22 birincil/kurumsal kaynak

Kapsam: teknik temeller, sağlık, eğitim, ekonomi, enerji, etik, hukuk ve yönetişim. Bu yayın sistematik derleme protokolü veya hukuki/klinik danışmanlık değildir.

Bu kitap nasıl okunur?

Her bölüm dört katmanla okunabilir: kısa tanım, teknik açıklama, uygulama örneği ve eleştirel sınır. Matematiksel ifadeler kavramları somutlaştırır; uzmanlık gerektiren uygulamalarda alan rehberleriyle birlikte kullanılmalıdır.

Hedef kitle: lisans öğrencileri, araştırmacılar, eğitimciler, mühendisler, yöneticiler ve politika yapıcılar. Önkoşul: temel olasılık ve dijital okuryazarlık; teknik bölümlerde ek matematik yararlıdır.



Öğrenme rotası: 1-4 kavram ve veri; 5-7 yöntem ve mimari; 8-12 değerlendirme ve alan etkileri; 13-16 yönetişim ve gelecek.

İçindekiler

01. Yapay zekâyı tanımlamak · Kavram ve sınırlar
02. Tarih, paradigma ve bilimsel yöntem · Bilgi üretimi
03. Matematiksel ve istatistiksel temel · Olasılık ve optimizasyon
04. Veri, temsil ve kalite · Veri yaşam döngüsü
05. Öğrenme yöntemleri · Klasik ML
06. Derin öğrenme ve Transformer · Mimari
07. Üretken ve çok kipli YZ · Temel modeller
08. Değerlendirme, sağlamlık ve güvenlik · Kanıt düzeyi
09. Sağlık ve yaşam bilimleri · Yüksek riskli kullanım
10. Eğitim, kültür ve insan gelişimi · Öğrenme tasarımı
11. Ekonomi, emek ve örgütler · Dağılım etkileri
12. Enerji, çevre ve altyapı · Sürdürülebilirlik
13. Hukuk, etik ve küresel yönetişim · Haklar ve kurumlar
14. Toplum, eşitsizlik ve epistemoloji · Kamusal etkiler
15. Kurumsal tasarım ve uygulama · Yaşam döngüsü
16. Gelecek gündemi ve açık sorular · Araştırma ufku

BÖLÜM 01 / KAVRAM VE SINIRLAR

Yapay zekâyı tanımlamak

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

Yapay zekâ (YZ), tanımlanmış hedefler için girdilerden tahmin, içerik, öneri veya karar üreten makine tabanlı sistemler ailesidir. Bir model, veri ve parametrelerden öğrenilmiş matematiksel eşlemedir; bir YZ sistemi ise modele ek olarak arayüz, veri akışı, insan rolleri, güvenlik ve kurum bağlamını içerir.

Derinlemesine

Dar YZ belirli görevlerde çalışır; genel amaçlı modeller birçok göreve uyarlanabilir. “Zekâ”, “anlama” ve “özerklik” sözcükleri teknik bağlama göre ayrıştırılmalıdır: akıcı çıktı, doğruluk veya niyet kanıtı değildir. Kapasiteyi tek bir test puanıyla değil, görev, kullanıcı, ortam ve hata maliyetiyle birlikte tanımlarız.

Çalışılmış örnek

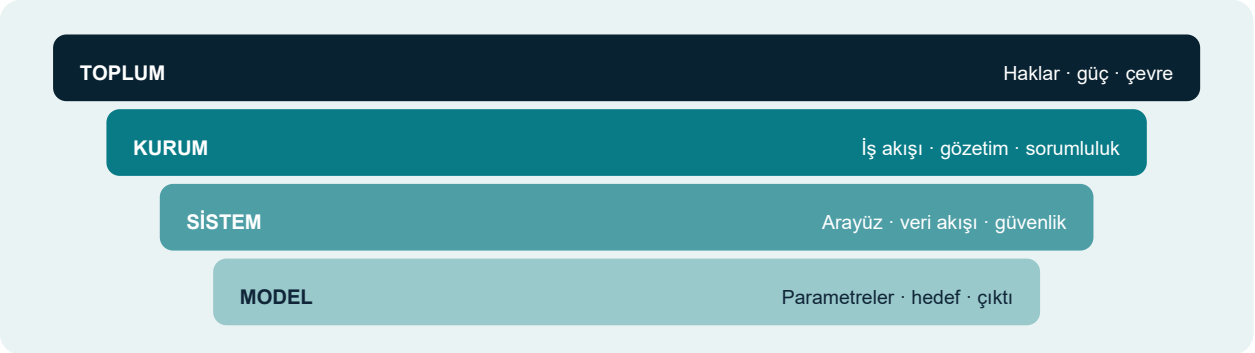
Bir hastane özetleme aracı, dil modelinden ibaret değildir: kaydın izni, alıntı izleri, hekim onayı ve düzeltme günlüğü de sistemin parçasıdır.

Disiplinler arası bağlantı

Felsefe kavramları netleştirir; bilişsel bilim insan öğrenmesini karşılaştırır; mühendislik görev performansını ölçer; sosyal bilimler kurum içindeki etkileri inceler.

Sınırlar ve eleştirel okuma

Bu kitap bilinç veya genel zekâ hakkında kesin bir hüküm vermez; mevcut sistemlerin ölçülebilir davranışlarına odaklanır.



Şekil. Kavramsal öğretim diyagramı; temsilîdir, saha verisi değildir.

Araştırma ve uygulama

Sistem sınırı ve değerlendirme birimi

Bir sistemin sınırını çizmeden başarı veya sorumluluk ölçülemez. Girdinin nereden geldiğini, modelin hangi araçları çağırabildiğini, çıktının kime gösterildiğini ve hangi insanın kararı onayladığını bir akış şemasında işaretleyin. Aynı model iki kurumda farklı güvenlik düzeyine sahip olabilir; çünkü veri ve karar süreci değişir. Bu nedenle

raporlamada yalnızca model sürümünü değil, kullanım senaryosunu ve operasyonel eşikleri de yazın.

Kavram testi

Bir otomatik özetleyici yanlış ifade ürettiğinde üç farklı hata kaynağı düşünün: eğitim verisindeki eksiklik, geri getirilen yanlış belge ve kullanıcının uyarıyı görmemesi. Her biri için farklı düzeltme gerekir. Sistem düzeyi çözüm, hata kaynaklarını tek bir “model hatası” başlığında eritmez.

Uygulama görevi

Seçtiğiniz bir YZ aracının aktör haritasını çizin. Girdi sahibi, geliştirici, dağıtıcı, son kullanıcı, etkilenen kişi ve denetleyiciyi ayırın. Her aktörün hangi kanıta ihtiyaç duyduğunu bir cümleyle yazın.

Öğrenme kontrolü

1. Model ile YZ sistemi arasındaki farkı üç bileşenle açıklayın.

2. Akıcı bir cevabın neden doğruluk kanıtı olmadığını bir örnekle gösterin.

Kaynaklar: S01, S04, S09. Tam künyeler sonda yer alır.

BÖLÜM 02 / BILGI ÜRETIMI

Tarih, paradigma ve bilimsel yöntem

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

Sembolik YZ açık kurallarla, istatistiksel öğrenme örüntü çıkarımıyla, derin öğrenme çok katmanlı temsil öğrenmesiyle çalışır. Modern araştırma bunları dışlamaz; hibrit sistemler kurallar, arama, öğrenme ve araç kullanımını birleştirir.

Derinlemesine

Bir YZ iddiasını incelerken araştırma sorusu, veri kökeni, karşılaştırma tabanı, değerlendirme protokolü, belirsizlik ve yeniden üretilebilirlik ayrı ayrı değerlendirilir. Test kümesine sızma ve karşılaştırma seçimi, görünürde yüksek performans üretir. Yayınlanan benchmark değeri, kurum içi dağıtım başarısının garantisi değildir.

Çalışılmış örnek

Bir metin sınıflandırıcıyı eski veri üzerinde sınamak, yeni mevzuat dilindeki başarısını göstermez; tarihsel ayırma ve dış doğrulama gerekir.

Disiplinler arası bağlantı

Bilim felsefesi yanlışlanabilirliği, istatistik örnekleme hatasını, tarih bilimi paradigma değişimini açıklamaya yardım eder.

Sınırlar ve eleştirel okuma

Kısa tarih seçkisi tüm küresel araştırma geleneklerini temsil etmez.

Araştırma ve uygulama

Karşılaştırma ve yanlışlanabilirlik

Bir araştırma iddiası karşılaştırma tabanı olmadan eksik kalır. Yeni yöntemi rastgele veya zayıf bir temel modele karşı değil, aynı veri ve bütçe altında makul bir mevcut yöntemle sınayın. Önceden belirlenmiş ölçütleri ve test kümesini sonuca göre değiştirmeyin. Negatif sonuçlar da bilgi üretir; hangi koşullarda yöntemin işe yaramadığını gösterir.

Yeniden üretim

Kodun açık olması tek başına tekrar üretilebilirlik sağlamaz. Veri sürümü, örnekleme yöntemi, eğitim donanımı, rastgele tohum, hiperparametre araması ve insan değerlendirme yönergesi kayıt altına alınmalıdır. Farklı laboratuvar sonuçları ayrıştığında bu farkın örneklem mi uygulama mı bağlam mı kaynaklı olduğu araştırılır.

Uygulama görevi

Bir makaleden şu altı unsuru çıkarın: araştırma sorusu, veri, karşılaştırma tabanı, birincil ölçüt, belirsizlik aralığı ve sınırlılık. Bunlardan biri yoksa iddianın hangi kısmının zayıfladığını tartışın.

Öğrenme kontrolü

1. Bir benchmark çalışmasında veri sızıntısı için iki kontrol önerin.
2. Sembolik ve öğrenen yöntemlerin birlikte kullanılacağı bir görev tasarlayın.

Kaynaklar: S02, S03, S04. Tam künyeler sonda yer alır.

BÖLÜM 03 / OLASILIK VE OPTIMIZASYON

Matematiksel ve istatistiksel temel

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

Öğrenme, parametreleri eğitim verisindeki kaybı azaltacak biçimde ayarlamaktır. L(θ)=ortalama ℓ(fθ(x),y)+λR(θ) ifadesi veri uyumu ile düzenlileştirme arasındaki dengeyi gösterir. Genelleme, görülmeyen örneklerde beklenen hatadır.

Derinlemesine

Bayesçi bakış ön bilgi ile gözlemi birleştirir; nedensel bakış müdahaleleri sorgular. Korelasyon, bir değişkenin değiştirilmesinin sonucu değiştireceğini göstermez. Belirsizlik iki kaynaktan gelir: veri gürültüsü ve model bilgisizliği. Kalibrasyon, %80 güven verilen olayların yaklaşık %80 gerçekleşmesi beklentisidir.

Çalışılmış örnek

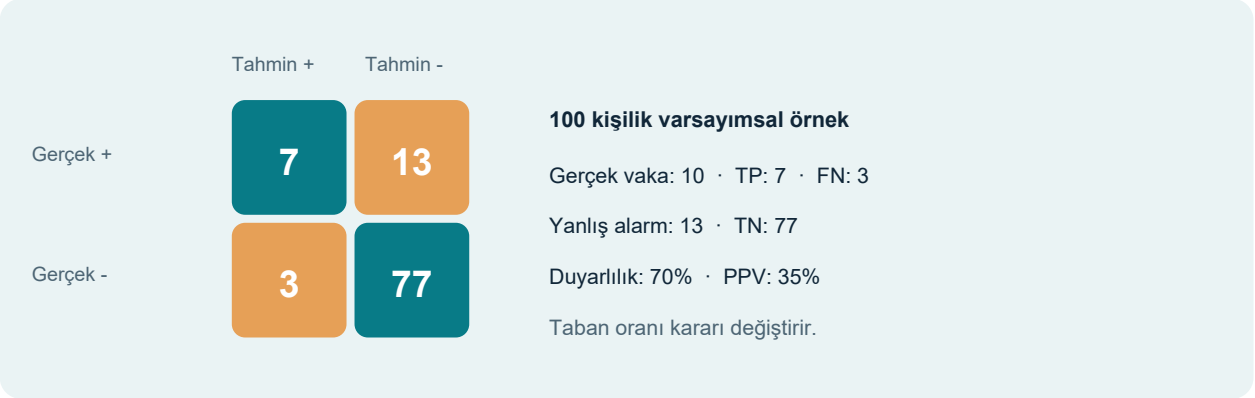
100 kişilik bir taramada 10 gerçek vaka, 8 doğru pozitif ve 12 yanlış pozitif varsa duyarlılık 8/10=%80, pozitif öngörü değeri 8/20=%40 olur.

Disiplinler arası bağlantı

Epidemiyoloji taban oranı, ekonomi nedensel etkileri, fizik modelleme varsayımlarını görünür kılar.

Sınırlar ve eleştirel okuma

Tek bir kayıp fonksiyonu insan değerlerini ve kurum hedeflerini tam temsil edemez.



Şekil. Kavramsal öğretim diyagramı; temsilîdir, saha verisi değildir.

Araştırma ve uygulama

Bayes ve taban oranı

Bir pozitif sonucu yorumlamak için testin duyarlılığı ve özgüllüğü kadar olayın yaygınlığı da gerekir. Bayes kuralı P(H|D)=P(D|H)P(H)/P(D) biçimindedir. Az görülen bir olayda küçük yanlış pozitif oranı çok sayıda yanlış alarma yol açabilir. Bu, tarama testlerinden dolandırıcılık tespitine kadar uzanır.

Optimizasyon ve değer

Kayıp fonksiyonu hangi hatanın daha pahalı olduğunu kodlar. Yanlış negatif ile yanlış pozitif maliyetleri eşit değilse karar eşiği bağlama göre değişir. Fakat sayısal maliyet atamak etik ve hukuki kararı otomatik çözmez. Hassas sonuçlarda eşik seçimi ilgili tarafların katılımı ve uzman gözetimiyle yapılır.

Uygulama görevi

1.000 kişi ve %2 yaygınlık varsayın. Duyarlılık %90, özgüllük %95 olsun. Beklenen TP, FN, FP ve TN değerlerini hesaplayın; pozitif öngörü değerinin neden sezgisel beklentiden farklı olabileceğini açıklayın.

Öğrenme kontrolü

1. Verilen örnekte özgüllüğü hesaplamak için hangi ek bilgi gerekir?

2. Kalibrasyon ile ayrıştırma gücü arasındaki farkı açıklayın.

Kaynaklar: S02, S03. Tam künyeler sonda yer alır.

BÖLÜM 04 / VERI YAŞAM DÖNGÜSÜ

Veri, temsil ve kalite

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

Veri yalnızca miktar değildir: kapsam, örnekleme, etiket tanımı, rıza, lisans, dil çeşitliliği, güncellik ve güvenlik belirleyicidir. Veri belgesi, toplama amacı ve sınırlılıkları görünür kılar.

Derinlemesine

Eğitim, doğrulama ve test kümeleri bağımsız tutulur. Zaman, kişi veya kurum düzeyindeki tekrarlar sızıntıya yol açabilir. Eksik temsil bir grubun hatasını artırır; ölçüm yanlılığı ise hedefin yanlış vekille tanımlanmasıdır. Sentetik veri bazı sorunları azaltabilir fakat özgün verinin kusurlarını çoğaltabilir.

Çalışılmış örnek

Bir dermatoloji veri seti tek cihaz ve tek cilt tonu dağılımından geliyorsa başka kliniklerde dış doğrulama gereklidir.

Disiplinler arası bağlantı

Arşivcilik kökeni, antropoloji bağlamı, hukuk hakları, istatistik örnekleme tasarımını inceler.

Sınırlar ve eleştirel okuma

“Temiz” veri bağlamdan bağımsız tarafsız veri anlamına gelmez.

Araştırma ve uygulama

Örneklem çerçevesi

Veri setinin kimin dünyasını temsil ettiği ilk sorudur. Zaman aralığı, coğrafya, cihaz, kurum, dil ve erişim koşulları örnekleme çerçevesini belirler. Kayıp kayıtlar rastgele değilse sadece mevcut örneklerden genelleme yanıltıcı olur. Etiketlerin uzmanlar arasında ne kadar tutarlı olduğunu ölçmek de gereklidir.

Veri yönetişimi

Veri yaşam döngüsünü toplama, temizlik, erişim, saklama, paylaşım ve silme olarak belgeleyin. Hak ve izinleri veri türüne göre ayırın. Bir verinin kamuya açık olması kullanımın her bağlamda uygun veya yasal olduğu anlamına gelmez. Veri belgesine hedef kullanım kadar kaçınılması gereken kullanım da yazılmalıdır.

Uygulama görevi

Örnek bir görüntü veya metin veri kümesi için veri belgesi taslağı hazırlayın: köken, örnekleme, etiketleme, lisans/izin, alt grup dağılımı, bilinen boşluklar ve silme politikası.

Öğrenme kontrolü

1. Bir veri belgesine hangi beş alanı koyarsınız?

2. Kişi düzeyinde sızıntı nasıl saptanır?

Kaynaklar: S20, S21, S05. Tam künyeler sonda yer alır.

BÖLÜM 05 / KLASIK ML

Öğrenme yöntemleri

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

Denetimli öğrenme etiketli örneklerden tahmin yapar; denetimsiz öğrenme yapı keşfeder; pekiştirmeli öğrenme eylem ve geri bildirim döngüsünü inceler. Özdenetimli öğrenme verinin iç yapısından hedef üretir.

Derinlemesine

Doğrusal modeller yorumlanabilir bir başlangıç sağlar. Ağaç toplulukları tablosal veride güçlü olabilir; sinir ağları görüntü, ses ve dilde temsil öğrenir. Yöntem seçimi verinin büyüklüğü, hata maliyeti, açıklanabilirlik, gecikme ve bakım gereksinimiyle birlikte yapılır.

Çalışılmış örnek

Kredi riski için basit bir lojistik regresyonu başlangıç modeli kurup karmaşık modelin ek faydasını dış testte ölçmek gerekir.

Disiplinler arası bağlantı

Kontrol teorisi pekiştirmeyi, psikoloji ödül tasarımını, operasyon araştırması karar kısıtlarını zenginleştirir.

Sınırlar ve eleştirel okuma

Yüksek ortalama başarı, nadir ve ağır hatalarda kabul edilebilirlik anlamına gelmez.

Araştırma ve uygulama

Model seçimi

Yöntem seçimini bir yarış tablosu yerine karar problemi olarak kurun. Veri az ve açıklama gereksinimi yüksekse basit model makul olabilir. Çok büyük serbest metin veya görüntüde temsil öğrenmesi avantajlı olabilir. Önce basit baz çizgi, sonra daha karmaşık yöntem denenir. Kazanç, bakım maliyeti ve hata örüntüsüyle birlikte değerlendirilir.

Deney düzeni

Hiperparametreler doğrulama verisiyle seçilir; nihai test veri seti kararın sonunda bir kez kullanılır. Tek bir rastgele ayırma yanıltıcı olabilir; mümkünse zaman, kurum veya kişi bazlı dış test ekleyin. Sınıf dengesizliğinde doğruluk yanında duyarlılık, özgüllük, hassasiyet ve kalibrasyon raporlayın.

Uygulama görevi

Bir okul terk riski projesinde hangi yöntemleri başlangıç ve geliştirilmiş model olarak seçersiniz? Ölçüm hedefini, yanlış alarmın maliyetini ve insan danışman rolünü birlikte yazın.

Öğrenme kontrolü

1. Denetimli ve özdenetimli öğrenmeye birer örnek verin.

2. Daha karmaşık modelin kullanımı için hangi ek kanıt gerekir?

Kaynaklar: S02, S04. Tam künyeler sonda yer alır.

BÖLÜM 06 / MIMARI

Derin öğrenme ve Transformer

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

Derin ağlar katmanlar boyunca temsilleri dönüştürür. Transformer, dikkat mekanizmasıyla dizideki öğelerin birbirleriyle ilişkisini hesaplar; bu yaklaşım büyük ölçekli dil ve çok kipli sistemlerin temel mimarilerinden biridir.

Derinlemesine

Öz dikkat için Q, K, V dönüşümleri kullanılır: Attention(Q,K,V)=softmax(QK^T/√d□)V. Bu işlem bağlam ağırlıklarını öğrenir, fakat dikkat ağırlığını tek başına nedensel açıklama saymak hatalıdır. Ölçek büyüdükçe hesaplama, veri ve enerji maliyeti artar; daha küçük amaca özel modeller bazen yeterlidir.

Çalışılmış örnek

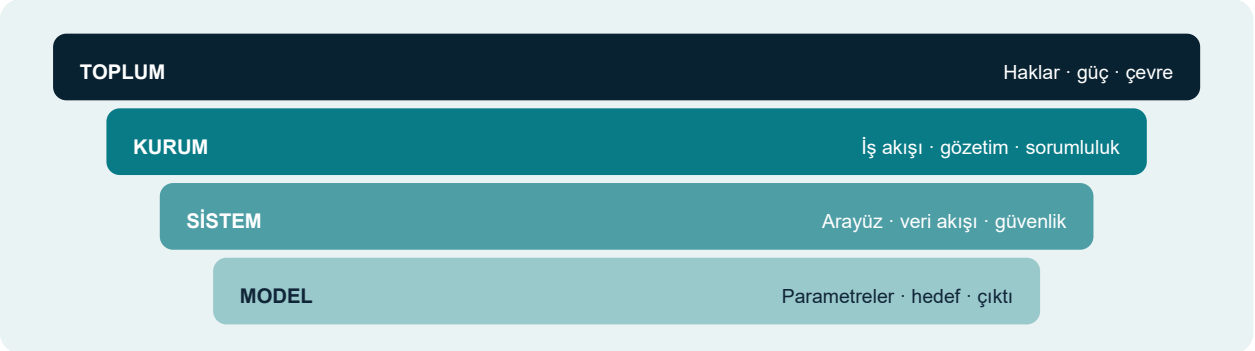
“Banka” sözcüğünün finans kurumu mu kıyı mı olduğunu çevre sözcüklerin ilişkisi belirler; dikkat mekanizması bu bağlamsal bağlantıları kullanabilir.

Disiplinler arası bağlantı

Dilbilim anlam ve bağlamı, donanım mühendisliği bellek darboğazını, bilgi kuramı kodlamayı açıklar.

Sınırlar ve eleştirel okuma

Formül mimarinin bir parçasıdır; eğitim verisi, amaç ve çıkarım düzeni davranışı birlikte belirler.



Şekil. Kavramsal öğretim diyagramı; temsilîdir, saha verisi değildir.

Araştırma ve uygulama

Dikkatin cebiri

Öz dikkat, her token için sorgu (Q), anahtar (K) ve değer (V) vektörleri üretir. QK^T ilişkiler için puan verir; ölçekleme ve softmax bu puanları ağırlığa dönüştürür. Ağırlıklar V üzerinde birleşir. Çok başlıklı dikkat farklı ilişki türlerini aynı

anda temsil edebilir. Bu matematik davranışın tümünü açıklamaz; eğitim hedefi ve veri de belirleyicidir.

Ölçek ve verim

Dizi uzunluğu arttığında dikkat hesabı ve bellek maliyeti önemli olabilir. Bağlam penceresi büyük olsa da uzun metindeki bilgiyi güvenilir kullanma ayrı bir test gerektirir. Küçük model, sıkıştırma veya geri getirme yaklaşımı belirli görevde daha düşük maliyetle yeterli olabilir; kalite karşılaştırması şarttır.

Uygulama görevi

Altı sözcüklü bir cümle seçin. Bir sözcüğün hangi diğer sözcüklerle ilişkili olabileceğini işaretleyin; çizdiğiniz bağlantıların gerçek model ağırlığı değil yorum amaçlı varsayım olduğunu belirtin.

Öğrenme kontrolü

1. Q, K, V rollerini sezgisel olarak açıklayın.

2. Dikkat ağırlığı neden tek başına açıklama değildir?

Kaynaklar: S02, S03, S04. Tam künyeler sonda yer alır.

BÖLÜM 07 / TEMEL MODELLER

Üretken ve çok kipli YZ

> **Bu bölümde**
>
> Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

Üretken modeller dağılımdan yeni metin, görüntü, ses veya kod örnekleri üretir. Temel modeller geniş veri üzerinde ön eğitim alır ve çok sayıda aşağı akış göreve uyarlanır. İstem, geri getirme ve araç çağrısı sistem davranışını değiştirir.

Derinlemesine

Dil modeli çoğu zaman sonraki token olasılığını tahmin eder; olasılıksal akıcılık doğrulanmış bilgi değildir. Geri getirme destekli üretim (RAG), yanıtı dış belgelere dayandırabilir; ancak yanlış belge seçimi ve hatalı alıntı riski sürer. Çok kipli modellerde kipler arası eşleme, eksik bağlam ve güvenlik sınırları ayrıca test edilir.

Çalışılmış örnek

Bir kurum politika asistanında belge sürümü ve kaynak satırı gösterilir; kanıt yoksa sistem belirsizliği belirtir ve insan incelemesine yönlendirir.

Disiplinler arası bağlantı

Medya çalışmaları sentetik içeriği, dilbilim söylemi, kütüphanecilik kaynak izlenebilirliğini inceler.

Sınırlar ve eleştirel okuma

Modelin ürettiği açıklama, iç süreçlerinin doğrulanmış izahı olmayabilir.

Araştırma ve uygulama

Üretim ile doğrulama

Üretken model, dilsel olasılıkları kullanarak tutarlı bir çıktı kurabilir. Doğrulama ise dış dünyaya dayalı kanıt gerektirir. Kaynaklı yanıt tasarımında belgenin sürümü, tarih ve alıntı konumu saklanmalıdır. Sistem kanıt bulamadığında bir yanıtı zorla üretmek yerine belirsizliği göstermelidir.

Araç kullanımı

Bir model arama, hesap makinesi veya veri tabanını çağırdığında eylem alanı genişler. Yetki sınırı, onay ve kayıt mekanizmaları önem kazanır. Özellikle dış metinlerin komut gibi davranmasını önlemek için güven sınırları açık tanımlanır. Araç çağrısı doğruluğu artırabilir, fakat yanlış araca yanlış argüman göndermek yeni hata üretir.

Uygulama görevi

Kaynaklı bir kurum asistanı tasarlayın: hangi belgeler taranacak, alıntı nasıl gösterilecek, veri yoksa ne denecek ve insan onayı hangi eylemler için gerekecek?

Öğrenme kontrolü

1. RAG hangi sorunu azaltır, hangilerini çözmez?

2. Bir çok kipli sistem için iki özgül değerlendirme önerin.

Kaynaklar: S04, S05, S22. Tam künyeler sonda yer alır.

BÖLÜM 08 / KANIT DÜZEYI

Değerlendirme, sağlamlık ve güvenlik

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

Değerlendirme görev başarısı, kalibrasyon, alt grup hatası, sağlamlık, güvenlik, kullanıcı deneyimi ve kaynak maliyetini birlikte ölçer. Offline test, pilot, kontrollü saha çalışması ve sürekli izleme farklı kanıt düzeyleridir.

Derinlemesine

Saldırı yüzeyi eğitim verisi zehirleme, kaçınma, istem enjeksiyonu, mahremiyet sızıntısı ve kötüye kullanımı içerir. Kırmızı takım çalışması hedefli zayıflıkları arar; tek başına güvenlik sertifikası değildir. Dağılım kayması ve geri besleme döngüleri dağıtımdan sonra ölçülmelidir.

Çalışılmış örnek

Bir işe alım modeli, genel doğruluk yanında aday grupları için yanlış eleme oranı ve insan itiraz süresiyle ölçülür.

Disiplinler arası bağlantı

Siber güvenlik tehdit modelini, insan faktörleri kullanım hatasını, istatistik güven aralıklarını sağlar.

Sınırlar ve eleştirel okuma

Hiçbir benchmark tüm gerçek dünya durumlarını kapsamaz.



Şekil. Kavramsal öğretim diyagramı; temsilîdir, saha verisi değildir.

Araştırma ve uygulama

Çok boyutlu test kartı

Bir değerlendirme kartında hedef görev, test örneklemi, karşılaştırma modeli, ortalama başarı, alt grup sonuçları, kalibrasyon, sağlamlık, güvenlik ve gecikme aynı sayfada bulunmalıdır. Her metrik için güven aralığı veya örneklem büyüklüğü verilmelidir. Ortalama hata küçük olsa bile zarar yoğunlaşması büyük olabilir.

Saldırı modeli

Tehdit modeli saldırganın yetkisini ve hedefini açıklar. Veri zehirleme eğitim aşamasını, istem enjeksiyonu kullanım aşamasındaki güven sınırını, çıkarım saldırısı mahremiyeti hedefler. Savunma, kötü niyetli örnekleri filtrelemekten ibaret değildir; ayrıştırılmış yetki, kayıt, insan incelemesi ve olay müdahalesi birlikte gerekir.

Uygulama görevi

Bir belge asistanına sahte talimat içeren belge eklendiğini varsayın. Güven sınırını, beklenen güvenli davranışı, test örneğini ve olay sonrası yapılacakları yazın.

Öğrenme kontrolü

1. Bir sistem için beş ölçütlü değerlendirme kartı tasarlayın.
2. İstem enjeksiyonu ile veri zehirleme farkını açıklayın.

Kaynaklar: S06, S07, S21. Tam künyeler sonda yer alır.

BÖLÜM 09 / YÜKSEK RISKLI KULLANIM

Sağlık ve yaşam bilimleri

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

YZ görüntü analizi, klinik iş akışı, ilaç keşfi ve protein yapısı tahmininde kullanılabilir. AlphaFold, protein yapısı tahmininde büyük ilerleme göstermiştir; tahmin deneysel doğrulamanın yerine geçmez.

Derinlemesine

Klinikte veri kayması, kurumlar arası fark, taban oranı, belirsizlik, klinisyenin nihai sorumluluğu ve advers etkiler değerlendirilmelidir. Bir modelin AUC değeri karar eşiğini ve hasta yararını tek başına belirlemez. WHO çok kipli sağlık YZ için etik ve yönetişim gereğini vurgular.

Çalışılmış örnek

Radyoloji triyajı için prospektif pilotta gecikme, kaçırılan vaka, iş yükü ve alt gruplar izlenir; klinik karar hekime aittir.

Disiplinler arası bağlantı

Biyoistatistik doğrulamayı, tıp klinik yararı, etik özerklik ve adaleti birleştirir.

Sınırlar ve eleştirel okuma

Bu eğitim içeriği klinik tavsiye veya cihaz onayı değildir.

Araştırma ve uygulama

Klinik yarar ve ölçüm

Bir klinik modelin istatistiksel ayrıştırması yüksek olabilir; fakat uygulamada erken tanı, hasta sonucu veya iş yükü iyileşmeyebilir. Uygun karar eşiği, hastalık yaygınlığı ve müdahale kapasitesiyle değişir. Prospektif doğrulama, farklı kurum ve hasta gruplarını kapsamalıdır. Uyarı yorgunluğu da ölçülür.

Bilimsel keşif

Protein yapı tahmini araştırma için güçlü bir hipotez üretir. Ancak yapısal tahmin fonksiyon, dinamik davranış, hücresel bağlam veya tedavi etkinliğini tek başına doğrulamaz. Deneysel biyoloji ile hesaplamalı model birbirini tamamlar. Belirsizlik skorları ve modelin çalışmadığı moleküler durumlar açık raporlanmalıdır.

Uygulama görevi

Bir radyoloji triyaj pilotu için hasta güvenliği planı yazın: dış test, uyarı eşiği, hekim devralması, alt grup incelemesi ve durdurma ölçütü.

Öğrenme kontrolü

1. AUC yüksekken klinik yarar neden düşük kalabilir?

2. Protein tahmininde deneysel doğrulamanın rolü nedir?

Kaynaklar: S13, S18, S06. Tam künyeler sonda yer alır.

BÖLÜM 10 / ÖĞRENME TASARIMI

Eğitim, kültür ve insan gelişimi

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

YZ öğretmene geri bildirim, uyarlanmış materyal ve erişilebilirlik desteği verebilir; öğrenme hedefi insanın kavrayışı ve özerkliğidir. UNESCO çerçeveleri öğretmen ve öğrenci yetkinliklerini teknik, pedagojik ve etik boyutlarda ele alır.

Derinlemesine

Üretken araçla tamamlanan ödev ile öğrencinin kazanımı ayrıştırılmalıdır. Değerlendirme süreç, sözlü savunma, kaynak doğrulama ve yansıtmayı içerebilir. Dil ve kültür çeşitliliği veri temsiliyle ilişkilidir; küçük diller ve yerel bilgi görünmez kalabilir.

Çalışılmış örnek

Öğrenciler aynı soruya üç YZ cevabı üretip kanıt, hata ve önyargı açısından karşılaştırır; ardından kendi gerekçeli cevabını yazar.

Disiplinler arası bağlantı

Eğitim bilimleri ölçmeyi, dilbilim çok dilliliği, sanat çalışmaları yaratıcılığı tartışır.

Sınırlar ve eleştirel okuma

Araç erişimi ve altyapı eşitsizliği pedagojik tasarımı sınırlar.

Araştırma ve uygulama

Öğrenme kanıtı

Bir öğrencinin YZ ile ürettiği düzgün metin onun kavradığını kanıtlamaz. Ön bilgi ölçümü, süreç günlüğü, sözlü açıklama ve yeni soruya transfer değerlendirmeyi güçlendirir. Öğretmen aracı hangi aşamada kullanacağını hedef kazanıma göre belirlemelidir. Öğrenci, kaynağı ve kullanılan YZ desteğini uygun biçimde açıklamalıdır.

Erişim ve kültür

Erişim eşitliği cihaz, bağlantı, dil, engellilik ve kurum desteğine bağlıdır. Bir modelin büyük dillerdeki akıcılığı yerel bilgiye uygunluğu göstermez. Eğitim materyali yerel müfredat ve kültürel bağlam açısından öğretmen tarafından gözden geçirilmelidir; otomasyon öğretmen açığını çözmek için tek başına strateji değildir.

Uygulama görevi

Bir ders için YZ kullanımı politikası yazın: izin verilen destek, beyan biçimi, kaynak doğrulama, mahremiyet sınırı ve öğrencinin kendi öğrenmesini nasıl göstereceği.

Öğrenme kontrolü

1. YZ destekli bir ödev için öğrenme kanıtını nasıl toplarsınız?

2. Yerel dil açısından iki risk belirleyin.

Kaynaklar: S14, S15, S08. Tam künyeler sonda yer alır.

BÖLÜM 11 / DAĞILIM ETKILERI

Ekonomi, emek ve örgütler

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

YZ etkisi çoğunlukla meslekten çok görev düzeyinde görülür. Maruziyet, otomatik iş kaybıyla aynı kavram değildir; iş tasarımı, beceri, ücret ve pazarlık koşulları sonucu değiştirir.

Derinlemesine

ILO görev temelli küresel maruziyet ölçer; NBER saha çalışması belirli müşteri hizmetleri bağlamında üretkenlik etkisini inceler. Bu bulgular başka mesleklere doğrudan genellenmemelidir. Kurumda değer, iş akışı değişimi, eğitim, kalite ve çalışan katılımı ile ölçülür.

Çalışılmış örnek

Bir hukuk bürosu taslak üretiminde zaman kazansa bile kaynak doğrulama süresi ve hata sorumluluğu netleşmeden net fayda hesaplanamaz.

Disiplinler arası bağlantı

İktisat verimlilik ve dağılımı, sosyoloji iş organizasyonunu, yönetim bilimi uygulamayı inceler.

Sınırlar ve eleştirel okuma

Kısa dönem üretkenlik artışı uzun dönem istihdam etkisini belirlemez.

Araştırma ve uygulama

Görev ve meslek ayrımı

Bir meslek çok sayıda görevden oluşur; YZ bazılarını hızlandırırken diğerlerini daha önemli kılabilir. Maruziyet endeksi görevin teknolojiyle karşılaşma olasılığını inceler, fiilî iş kaybını ölçmez. Sonuç ücret pazarlığı, eğitim, kurum kararı ve talep değişimine bağlıdır. Bu yüzden farklı ülkeler ve sektörler için ayrı veri gerekir.

Saha etkisi

Bir üretkenlik pilotunda yalnızca işlem süresine bakmak kalite kaybını gizler. Hata düzeltme, müşteri memnuniyeti, yeni çalışan öğrenmesi ve uzun dönem beceri birikimi birlikte izlenmelidir. Karşılaştırma grubu varsa nedensel yorum güçlenir; yoksa eşzamanlı değişiklikler sonuçları karıştırabilir.

Uygulama görevi

Bir çağrı merkezi pilotu planlayın: birincil sonuç, kalite koruma metriği, çalışan görüşü, deney grubu, veri koruma ve pilot bitiş kararı.

Öğrenme kontrolü

1. Maruziyet ile ikame arasındaki farkı açıklayın.

2. Bir pilotun net faydasını ölçmek için hangi maliyetleri katarsınız?

Kaynaklar: S16, S19, S01. Tam künyeler sonda yer alır.

BÖLÜM 12 / SÜRDÜRÜLEBILIRLIK

Enerji, çevre ve altyapı

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

YZ veri merkezi elektriği, donanım, su ve malzeme kullanımı üzerinden çevresel etki yaratır; enerji sistemlerini optimize etme potansiyeli de vardır. Net etki kullanım, şebeke, donanım ve geri tepme etkisine bağlıdır.

Derinlemesine

Eğitim ve çıkarım maliyeti ayrılmalıdır. İşlevsel birim olarak görev başına enerji, kalite ve kullanıcı sayısı birlikte raporlanabilir. Sadece verimlilik artışı tüketimin toplamda azalacağını garanti etmez; talep büyümesi tasarrufu aşabilir. IEA belirsizlik aralıklarıyla senaryolar üretir.

Çalışılmış örnek

Aynı kaliteyi veren küçük modelin görev başına enerji tüketimi düşebilir; fakat kullanımı on kat artarsa toplam tüketim de artabilir.

Disiplinler arası bağlantı

Enerji sistemleri şebekeyi, çevre bilimi yaşam döngüsünü, ekonomi geri tepme etkisini açıklar.

Sınırlar ve eleştirel okuma

Tek bir “bir sorgunun karbonu” sayısı altyapı ve yerel enerji karışımı olmadan evrensel değildir.

Araştırma ve uygulama

Ölçüm birimi

Görev başına enerji, toplam veri merkezi enerjisi ve yaşam döngüsü emisyonu farklı büyüklüklerdir. Karşılaştırmada aynı kalite, aynı görev ve aynı sınır kullanılmalıdır. Donanım üretimi, soğutma ve yerel elektrik karışımı çıkarımı etkiler. Tek bir sorgunun çevresel maliyeti evrensel sabit değildir.

Geri tepme etkisi

Verimli modelin ucuzlaması kullanım sayısını artırabilir. Örnek: görev başına enerji %40 azalır, toplam görev sayısı iki katına çıkarsa toplam enerji başlangıcın %120’si olur. Bu basit cebir gerçek enerji tahmini değildir; toplamdaki davranış değişiminin önemini gösterir.

Uygulama görevi

Aynı kaliteyi sağlayan iki model için görev başına enerji, toplam görev sayısı ve hata düzeltme maliyetini içeren bir karşılaştırma tablosu tasarlayın.

Öğrenme kontrolü

1. Görev başına enerji ile toplam enerji arasındaki farkı örnekleyin.

2. Bir YZ hizmeti için yaşam döngüsü sınırlarını çiziniz.

Kaynaklar: S17, S01. Tam künyeler sonda yer alır.

BÖLÜM 13 / HAKLAR VE KURUMLAR

Hukuk, etik ve küresel yönetişim

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

Etik ilkeler adalet, zarar vermeme, özerklik, mahremiyet ve hesap verebilirlik sorularını açar. Hukuki yükümlülükler yargı alanı ve kullanım durumuna göre değişir. UNESCO, OECD ve BM metinleri uluslararası yönetişim için ortak dil sunar.

Derinlemesine

AB YZ Yasası risk temelli yükümlülükler kurar; uygulama tarihleri ve değişiklikler günceldir, bu nedenle resmi konsolide metin ve takvim kontrol edilmelidir. Yönetişim yalnızca politika belgesi değildir: sahiplik, etki değerlendirmesi, kanıt, itiraz, olay yönetimi ve bağımsız gözetim gerekir.

Çalışılmış örnek

Kamu hizmetinde bir hak sahipliği modeli için karar açıklaması, insan incelemesi, başvuru yolu ve alt grup etkisi yayınlanır.

Disiplinler arası bağlantı

Hukuk usul güvencesini, siyaset bilimi meşruiyeti, etik gerekçelendirmeyi, tasarım erişilebilirliği işler.

Sınırlar ve eleştirel okuma

Bu metin hukuki görüş değildir; yerel mevzuat ve güncel düzenleme ayrıca incelenir.



Şekil. Kavramsal öğretim diyagramı; temsilîdir, saha verisi değildir.

Araştırma ve uygulama

İlke ile yükümlülük

Adalet ve saydamlık gibi ilkeler yön gösterir; bağlayıcı yükümlülükler belirli hukuk düzenlerinde ve kullanım türlerinde ortaya çıkar. Bir sistem aynı anda veri koruma, tüketici, iş ve sektörel kurallara tabi olabilir. Uygulama takvimleri değişebildiği için resmi konsolide metni ve otorite rehberini karar tarihindeki hâliyle doğrulayın.

Hesap verebilirlik

Bir kurum politikası sorumluluk zincirine dönüşmelidir: kullanım sahibi, teknik sahibi, veri sorumlusu, etkilenen kişi için başvuru kanalı ve bağımsız denetim. Etki değerlendirmesi yalnızca ilk gün yapılmaz; yeni veri, model sürümü veya kullanım amacı değişince yenilenir. Olay kayıtları düzeltme ve kamusal güven için kanıttır.

Uygulama görevi

Kamu kurumunda bir YZ sistemi için onay matrisi hazırlayın. Hangi riskte kimin durdurma yetkisi olduğunu ve etkilenen kişinin nasıl itiraz edeceğini belirtin.

Öğrenme kontrolü

1. Bir yüksek riskli kullanım için hesap verebilirlik zinciri kurun.
2. Etik ilke ile hukuki yükümlülük farkını açıklayın.

Kaynaklar: S08, S09, S10, S11, S12. Tam künyeler sonda yer alır.

BÖLÜM 14 / KAMUSAL ETKILER

Toplum, eşitsizlik ve epistemoloji

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

YZ, bilgi üretimi ve dağıtımında güç ilişkilerini etkiler. Sentetik medya, yanlış bilgi, temsil eksikliği, dil eşitsizliği ve platform yoğunlaşması toplumsal sonuçlar doğurabilir.

Derinlemesine

Bir sistemin “tarafsız” olduğu iddiası hangi hedef, grup ve bağlam için sorulmalıdır. Ölçülen adalet tanımları çatışabilir; kararların kim tarafından, hangi itiraz yolu ile verildiği önemlidir. Kaynak çeşitliliği ve bağımsız inceleme epistemik dayanıklılığı artırır.

Çalışılmış örnek

Bir haber doğrulama aracında otomatik etiketin gerekçesi ve kaynak zinciri gösterilir; çekişmeli durumlar editöre aktarılır.

Disiplinler arası bağlantı

Sosyoloji güç ve kurumları, iletişim bilimi medya ekolojisini, felsefe bilgi gerekçesini inceler.

Sınırlar ve eleştirel okuma

Teknik bir metrik sosyal meşruiyetin tamamını ölçmez.

Araştırma ve uygulama

Adaletin çoğulluğu

Farklı adalet ölçütleri aynı anda sağlanamayabilir. Hangi metrik seçilirse seçilsin kararın etkilediği grupların deneyimi ve hakları dikkate alınmalıdır. “Önyargısız veri” iddiası, etiketlerin tarihsel kurum kararlarından üretildiği durumlarda özellikle sorgulanır. Sayısal ölçüm katılımcı süreçle tamamlanır.

Bilgi ekolojisi

Üretken içerik hacmi arttığında kaynak doğrulama ve köken bilgisi önemli hâle gelir. Bir metnin ikna edici olması güvenilir olması demek değildir. Kaynağın kimliği, tarihi, bağımsız doğrulama ve düzeltme geçmişi okunmalıdır. Platformların içerik sıralama kararları da kamusal tartışmayı etkiler.

Uygulama görevi

Sentetik içerik içeren bir kamusal iletişim örneği seçin. Köken, etiketleme, itiraz, düzeltme ve farklı dil gruplarına etkisini analiz edin.

Öğrenme kontrolü

1. Bir adalet metriğinin hangi grubu görünmez kılabileceğini tartışın.

2. Kaynak çeşitliliğini pratikte nasıl ölçersiniz?

Kaynaklar: S05, S08, S12. Tam künyeler sonda yer alır.

BÖLÜM 15 / YAŞAM DÖNGÜSÜ

Kurumsal tasarım ve uygulama

Bu bölümde

Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

İyi uygulama, problem tanımından emekliliğe uzanan yaşam döngüsüdür: amaç, veri, model, değerlendirme, dağıtım, izleme ve geri çekme. Her aşamada sorumlu kişi, kanıt ve durdurma koşulu tanımlanır.

Derinlemesine

Önce baz çizgi kurulur; YZ ek faydası ve maliyeti ölçülür. Pilot gerçek kullanıcı davranışını ve hata türlerini gösterir. Model kartı ve veri belgesi karar vericinin varsayımları görmesini sağlar. Sürüm değişiklikleri yeniden değerlendirme tetikler.

Çalışılmış örnek

Müşteri destek asistanı için kalite örneklemi, yanlış yönlendirme oranı, operatör devralma süresi ve şikâyetler haftalık izlenir.

Disiplinler arası bağlantı

Yazılım mühendisliği sürümlemeyi, örgüt yönetimi sahipliği, insan faktörleri kullanıcı davranışını ele alır.

Sınırlar ve eleştirel okuma

Kontrol listesi bağlama özgü uzman incelemesinin yerini tutmaz.



Şekil. Kavramsal öğretim diyagramı; temsilîdir, saha verisi değildir.

Araştırma ve uygulama

Pilot tasarımı

Pilotun amacı yalnızca sistemi çalıştırmak değil, belirsizlikleri azaltmaktır. Önceden başarı ve durdurma eşikleri belirlenir. Gerçek kullanıcı akışı, hata bildirimi, insan devralma süresi ve etkilenen kişi deneyimi kaydedilir. Olumsuz sonuçlar tasarım değişikliği veya geri çekme kararı için kullanılır.

Sürekli izleme

Dağıtımdan sonra veri ve davranış değişir. Kayma göstergeleri ile gerçek sonuç ölçümleri ayrılmalıdır. Başlangıçtaki test kümesi tekrar tekrar karar vermek için kullanılırsa bağımsızlığı aşınır. Sürüm değişimleri, yeni veri kaynağı ve yeni kullanıcı grubu yeniden değerlendirme tetikler.

Uygulama görevi

Seçtiğiniz kullanım için bir sayfalık model kartı yazın: amaç, uygun olmayan kullanım, veri, ölçüt, alt grup sonuçları, bilinen hata ve güncelleme tarihi.

Öğrenme kontrolü

1. Bir pilot için durdurma koşulu yazın.

2. Dağıtım sonrası kaymayı nasıl izlersiniz?

Kaynaklar: S06, S20, S21. Tam künyeler sonda yer alır.

BÖLÜM 16 / ARAŞTIRMA UFKU

Gelecek gündemi ve açık sorular

> **Bu bölümde**
>
> Kavram · yöntem · uygulama · sınır · alıştırma

Temel fikir

Açık sorunlar güvenilir değerlendirme, nedensel açıklama, ajan davranışı, mahremiyet, çok dillilik, düşük kaynaklı bağlamlar, enerji etkisi ve demokratik denetimi kapsar.

Derinlemesine

Gelecek öngörülerini kesin sonuç değil senaryo olarak okumak gerekir. Kabiliyet artışı ile gerçek dünya güvenliği farklı hızlarda ilerleyebilir. Araştırma önceliği, yüksek etki ve düşük kanıtın kesiştiği yerlerde bağımsız doğrulama ve kamusal katılımı güçlendirmektir.

Çalışılmış örnek

Otonom bir araştırma ajanı değerlendirirken sadece doğru sonuç sayısı değil, kaynak uydurma, araç yetkisi ve geri alınabilirlik ölçülür.

Disiplinler arası bağlantı

Bilim ve teknoloji çalışmaları toplumsal yönü, bilgisayar bilimi teknik sınırları, kamu yönetimi kurumları birlikte inceler.

Sınırlar ve eleştirel okuma

Bu bölüm gelecek tahmini değil, araştırma soruları haritasıdır.

Araştırma ve uygulama

Senaryo disiplini

Gelecek tahmini tek çizgi değildir. Yüksek ve düşük teknoloji ilerleme hızı, güçlü ve zayıf kurum kapasitesi, farklı enerji koşulları gibi eksenler kurarak senaryolar üretin. Her senaryoda varsayımları, erken işaretleri ve geri alınabilir politika seçeneklerini yazın. Böylece belirsizlik görünür olur.

Araştırma önceliği

Bir araştırma sorusuna öncelik verirken beklenen yarar, olası zarar, mevcut kanıtın zayıflığı ve araştırmanın uygulanabilirliği birlikte değerlendirilir. Çok dilli değerlendirme, saha güvenliği, emek ve enerji etkileri disiplinler arası ekip gerektirir. Bağımsız veri erişimi ve yöntem açıklığı güvenilir bulgu üretimini destekler.

Uygulama görevi

2030 için iki karşıt YZ senaryosu kurun. Her biri için üç gözlenebilir gösterge, bir yanlışlanabilir iddia ve düşük pişmanlık yaratan bir politika seçeneği yazın.

Öğrenme kontrolü

1. Bir araştırma sorusunu ölçülebilir hipoteze dönüştürün.

2. Gelecek senaryosunda hangi belirsizlikleri açıkça yazarsınız?

Kaynaklar: S01, S04, S06, S17. Tam künyeler sonda yer alır.

Bütünleştirici vaka: kamu hizmetinde YZ

Durum

Bir kurum başvuru metinlerini sınıflandırıp uzmanlara öncelik sırası önermek istiyor. Hak kaybı doğurabilecek kararı otomatik vermeyecek.

Tasarım

Amaç süresini ve kaçırılan acil dosyaları ölçmek; eski süreçle karşılaştırmak; veri kökeni, rıza ve saklama koşullarını belgelemek.

Değerlendirme

Yanlış önceliklendirme, alt grup farkı, uzman devralma süresi, açıklama kalitesi ve itiraz sonuçlarını dış pilotta izlemek.

Yönetişim

İnsan karar mercii, olay kayıtları, model sürüm kontrolü ve durdurma eşiği tanımlamak. Güncel yerel hukuk ve sektörel yükümlülükleri uzmanla incelemek.

Tartışma: “Doğruluk %92” tek başına devreye alma için neden yetersizdir? Hangi ölçümler ve hangi hak mekanizmaları eksiktir?

Terimler sözlüğü

AUC. Sıralama ayrıştırmasını özetleyen eğri altı alan; karar eşiği ve yararı tek başına vermez.

Benchmark. Belirlenmiş görev ve veri ile karşılaştırma düzeni.

Dağılım kayması. Üretimdeki veri veya ilişkilerin eğitim/test ortamından ayrılması.

Duyarlılık. Gerçek pozitiflerin doğru yakalanma oranı: TP/(TP+FN).

Genelleme. Görülmemiş örneklere performans aktarımı.

Kalibrasyon. Tahmin edilen olasılığın gerçekleşme sıklığıyla uyumu.

Model kartı. Modelin amaç, veri, ölçüm ve sınırlılık belgesi.

RAG. Yanıt üretiminde dış kaynaktan geri getirilmiş içerik kullanımı.

Temel model. Geniş veriyle ön eğitim alıp çok göreve uyarlanabilen model.

Veri belgesi. Veri kümesinin kökeni, yapısı, kullanım amacı ve sınırlarının kaydı.

Yanlış pozitif. Gerçekte negatifken sistemin pozitif demesi.

Öz dikkat. Bir dizideki öğeler arası bağlamsal ilişkiyi ağırlıklandıran işlem.

Kaynakça

[S01] Stanford HAI (2026). AI Index Report 2026. https://hai.stanford.edu/assets/files/ai_index_report_2026.pdf

[S02] LeCun, Bengio & Hinton (2015). Deep learning. https://doi.org/10.1038/nature14539

[S03] Vaswani et al. (2017). Attention Is All You Need. https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[S04] Bommasani et al. / Stanford CRFM (2021). On the Opportunities and Risks of Foundation Models. https://crfm.stanford.edu/report

[S05] Bender et al. (2021). On the Dangers of Stochastic Parrots. https://doi.org/10.1145/3442188.3445922

[S06] NIST (2023). Artificial Intelligence Risk Management Framework 1.0. https://www.nist.gov/itl/ai-risk-management-framework

[S07] NIST (2025). Adversarial Machine Learning: Taxonomy and Terminology. https://csrc.nist.gov/pubs/ai/100/2/e2025/final

[S08] UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence

[S09] OECD (2024). AI Principles (2024 update). https://www.oecd.org/en/topics/ai-principles.html

[S10] European Union (2026). AI Act, consolidated text. https://eur-lex.europa.eu/eli/reg/2024/1689/2026-07-27/eng

[S11] European Commission (2026). AI Act implementation timeline. https://ai-act-service-desk.ec.europa.eu/en/ai-act/eu-ai-act-implementation-timeline

[S12] United Nations (2024). Global Digital Compact. https://www.un.org/digital-emerging-technologies/global-digital-compact

[S13] WHO (2024). Ethics and governance of AI for health: large multi-modal models. https://www.who.int/publications/b/70584

[S14] UNESCO (2024). AI competency framework for teachers. https://www.unesco.org/en/articles/ai-competency-framework-teachers

[S15] UNESCO (2024). AI competency framework for students. https://www.unesco.org/en/articles/ai-competency-framework-students

[S16] ILO (2025). Generative AI and Jobs: A Refined Global Index of Occupational Exposure. https://www.ilo.org/publications/generative-ai-and-jobs-refined-global-index-occupational-exposure

[S17] IEA (2025). Energy and AI. https://www.iea.org/reports/energy-and-ai

[S18] Jumper et al. (2021). Highly accurate protein structure prediction with AlphaFold. https://www.nature.com/articles/s41586-021-03819-2

[S19] Brynjolfsson, Li & Raymond (2023/2025). Generative AI at Work. https://www.nber.org/papers/w31161

[S20] Gebru et al. (2018/2021). Datasheets for Datasets. https://arxiv.org/abs/1803.09010

[S21] Mitchell et al. (2019). Model Cards for Model Reporting. https://arxiv.org/abs/1810.03993

[S22] OpenAI (2023). GPT-4 Technical Report. https://arxiv.org/abs/2303.08774