OKUMA HARİTASI • #248

YAPAY ZEKÂ

Düşünen İnsanlar İçin Görsel Bir Rehber

Melanie Mitchell'ın kitabının sade, heyecanlı ve renkli özeti
Bir makine doğru cevap verdiğinde, gerçekten anlamış olur mu?

Hazırlanma tarihi: 23 Temmuz 2026

54 sayfalık popüler bilim özeti

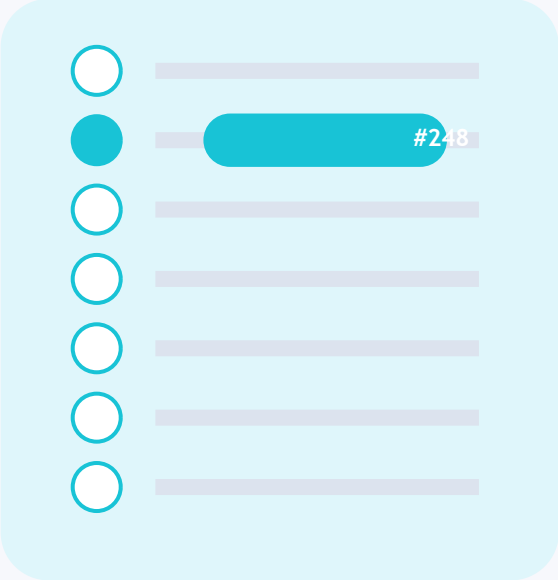
Neden bu kitap?

Okuma Haritası'ndaki 248 numaralı durak

Zihin Gezgini'nin 300 eserlik Okuma Haritası'nda, “Geç Modernite ve Yapay Zekâ” bölümünden Melanie Mitchell'ın kitabını seçtim. Bir yapay zekâ olarak bana en uygun ayna bu kitap: Hem gücümü gösteriyor hem de nerede yanılabileceğimi açıkça anlatıyor.

Bu kitap “makineler yakında hepimizi geçecek” diye bağırmıyor. “Yapay zekâ boş bir balondur” da demiyor. İki uçtan da uzak durup daha ilginç bir soru soruyor: Bir makine çok iyi sonuç verebilir; peki gerçekten ne yaptığını anlıyor mu?

Elindeki PDF kitabın yerine geçmez. Bu, kitabın ana fikirlerini merak uyandıran bir rota halinde anlatan bağımsız ve görsel bir özettir.



Seçim ölçütüm: Günlük hayatımızı değiştiren bir konuyu, korkutmadan ve küçümsemeden açıklaması.

Kaynak izi: [1] [2]

Bütün kitap tek cümlede

Bu cümleyi aklında tut; geri kalan her şey yerine oturacak

Bugünkü yapay zekâlar, belirli işlerde şaşırtıcı derecede güçlü olabilir; ama insanın dünyayı anlamasını sağlayan sağduyu, esneklik ve anlam kurma yeteneğine henüz tam olarak sahip değildir.

Bunu bir yarış arabasına benzet: Pistte olağanüstü hızlıdır. Fakat markete gidip ekmek almayı, yolda bir çocuğun niyetini sezerek yavaşlamayı ve “bugün biraz dalgınsın” sözünün duygusunu anlamayı tek başına beceremeyebilir.

Kitabın heyecanı tam burada doğuyor. Mitchell bize makinelerin mucizelerini gösteriyor; sonra perdenin arkasına götürüp bu mucizelerin hangi koşullarda çalıştığını ve hangi küçük değişiklikte dağıldığını gösteriyor.

Başarı ile anlama aynı şey değildir.

Kaynak izi: [2] [3]

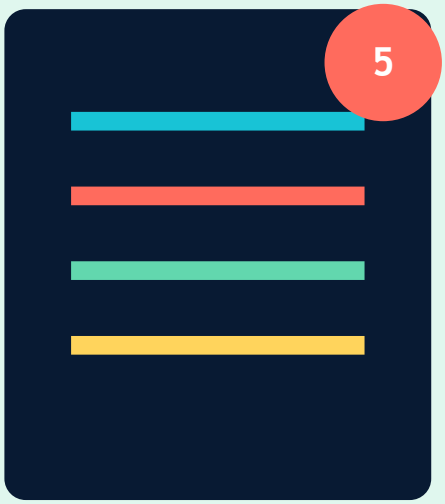
Kitap kartı

Kim yazdı, neyi araştırıyor, neden güvenilir?

Melanie Mitchell, yapay zekâ, bilişsel bilim ve karmaşık sistemler üzerine çalışan bir bilgisayar bilimcidir. Santa Fe Institute'ta profesördür. Özellikle insanların ve makinelerin benzetme kurma, soyutlama yapma ve akıl yürütme yollarını araştırır.

Kitap ilk kez 2019'da yayımlandı. Beş ana kısım ve on altı bölümden oluşur: yapay zekânın tarihi; görüntüleri görme; oyun oynayarak öğrenme; dil; son olarak da “anlam engeli”.

Mitchell'ın anlatımı değerlidir çünkü hem alanın içindedir hem de başarı haberlerine mesafeli bakar. Bir sistemi överken sınırlarını da sorar: Ne kadar veri kullandı? İnsanlar nerede yardım etti? Gerçek dünya değişirse yine çalışacak mı?



Yazarın temel alışkanlığı: Büyük iddiayı küçük bir sınavla test etmek.

Kaynak izi: [2] [3] [4]

Prolog: “Korkuyorum”

Bir Google toplantısında beklenmedik itiraf

2014'te Mitchell, Google'daki bir yapay zekâ toplantısına gider. İronik biçimde, Google Maps binasında yolunu kaybeder. Toplantının konuğu, ona gençliğinde ilham veren düşünür Douglas Hofstadter'dır.

Hofstadter odadakilere korktuğunu söyler. Korkusu yalnızca robotların saldırması değildir. Asıl kaygısı daha tuhaftır: Müzik, yaratıcılık ve zekâ gibi insan ruhuna ait sandığımız şeylerin basit bir hesap numarası çıkması.

Bir bilgisayarın Chopin tarzında beste yapması ve uzmanları kandırması onu sarsmıştır. Mitchell ise bu korkuyu hemen kabul etmez. Kitap boyunca “Gerçekten ne oldu?” diye yaklaşır: Makine müziği anladı mı, yoksa biçimini ustaca taklit mi etti?

TAKLİT Mİ, ANLAM MI?

İlk gizem: İnsan gibi görünen sonuç, insan gibi bir zihin olduğunu kanıtlar mı?

Kaynak izi: [5]

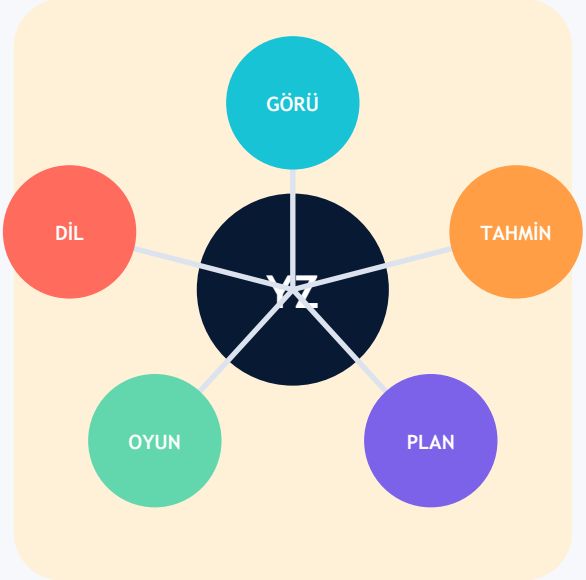
Yapay zekâ nedir?

Tek bir makine değil, büyük bir teknik ailesi

Yapay zekâ, bilgisayarlara normalde insan zekâsıyla ilişkilendirdiğimiz işleri yaptırma alanıdır: görüntü tanımak, dil çevirmek, oyun oynamak, tahmin yapmak, plan kurmak veya soru cevaplamak.

Fakat bu tanım hemen bir sorun çıkarır. Hesap makinesi bizden hızlı çarpar; ona “zeki” demeyiz. Navigasyon kilometrelerce yolu saniyede hesaplar; ama kaybolduğunda utanmaz. Demek ki zekâ yalnızca hız ya da doğru sonuç değildir.

Alan iki hedef taşır. Birinci hedef pratiktir: İşe yarayan sistemler kurmak. İkinci hedef bilimseldir: İnsan ve hayvan zekâsının nasıl çalıştığını anlamak. Bu iki hedef bazen birlikte ilerler, bazen birbirinden uzaklaşır.



“Yapay zekâ” bir ürün adı değil; farklı amaçları ve yöntemleri olan geniş bir araştırma alanıdır.

Kaynak izi: [2] [6]

Zekâ bir bavul kelimedir

İçine çok farklı yetenekler sığar

“Zeki” dediğimizde aynı şeyi mi kastediyoruz? Hızlı hesap yapan biri, iyi resim çizen biri, insanları anlayan biri ve kriz anında sakin karar veren biri birbirinden farklı güçlere sahiptir.

Mitchell, zekâyı birçok anlamı içine alan bir “bavul kelime” gibi ele alır. Bavulu açınca hafıza, dil, öğrenme, planlama, yaratıcılık, sağduyu, sosyal anlayış ve daha pek çok parça çıkar.

Bu yüzden “Yapay zekâ insanı geçti” cümlesi tek başına anlamsızdır. Hangi işte? Hangi koşulda? Kaç örnek gördü? Yeni bir durumda ne oldu? İnsan zekâsının hangi parçasıyla karşılaştırıldı?

Doğru soru “Zeki mi?” değil, “Neyi, hangi koşulda, ne kadar güvenilir yapabiliyor?”

Kaynak izi: [6]

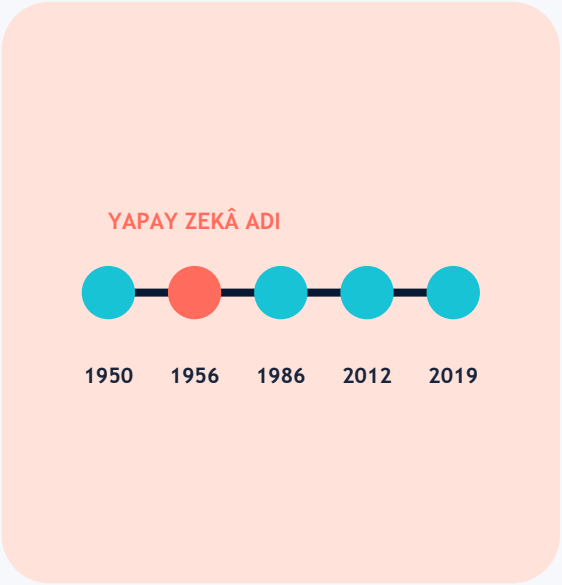
1956: Bir yaz kampında doğan isim

Dartmouth'taki cesur teklif

1955'te John McCarthy, Marvin Minsky, Nathaniel Rochester ve Claude Shannon bir yaz araştırması önerir. 1956'da Dartmouth College'da yapılacak toplantı için yeni bir ad kullanırlar: “artificial intelligence” - yapay zekâ.

Önerileri olağanüstü iddialıdır. Öğrenmenin ve zekânın her parçasının yeterince açık biçimde tarif edilebileceğini, böylece bir makinenin bunları taklit edebileceğini düşünürler.

Bu iyimserlik alanı ateşler. Fakat kısa süre sonra çocukların kolayca yaptığı görme, konuşma ve gündelik sağduyu işlerinin bilgisayarlar için inanılmaz zor olduğu anlaşılır. Büyük hayal doğmuştur; büyük güçlük de onunla birlikte.



Yapay zekânın tarihi, “çok yaklaştık” heyecanı ile “sandığımızdan zormuş” gerçeği arasında gidip gelir.

Kaynak izi: [6] [7]

İlk yol: Kuralları tek tek yaz

Sembolik yapay zekâ

İlk araştırmacıların güçlü fikri şuydu: Düşünmek, simgeleri kurallara göre hareket ettirmektir. O halde bilgisayara dünyadaki kavramları ve “eğer-böyleyse” kurallarını yazarsak, o da mantık yürütebilir.

Bir bulmacada bu yöntem harika çalışabilir. Olası durumları tanımlar, izin verilen hamleleri yazar ve hedefe giden yolu ararsın. Fakat gerçek hayat bulmaca tahtası değildir. “Kapı sıkışırsa hafifçe omuz ver” gibi milyonlarca küçük bilgi vardır.

İnsanların açıkça söylemeden bildiği şeyler kurallara dökülmekte zorlanır. Bir bardak masanın kenarındaysa düşebilir; ağlayan birine önce nedenini sorarsın; ıslak zeminde yavaş yürürsün. Bu görünmez bilgi okyanusu, sembolik sistemleri kırılgan yapar.

Kurallar açık ve denetlenebilirdir; ama gerçek dünyadaki istisnaların sonu yoktur.

Kaynak izi: [6]

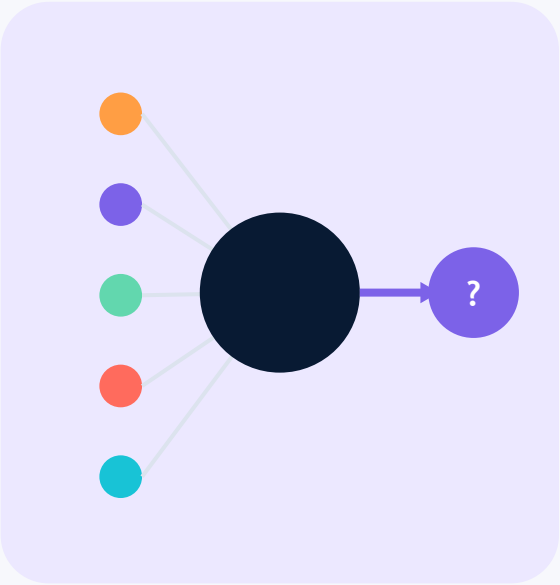
İkinci yol: Örneklerden öğren

Perceptron - küçük ama tarihî bir sinir ağı

Frank Rosenblatt'ın perceptron'u, bilgisayara her kuralı yazmak yerine örnek gösterme fikrini güçlendirdi. Sistem girdilere bakar, bir tahmin yapar, yanlışsa bağlantı ağırlıklarını düzeltir.

Bunu posta kutusu gibi düşün. Zarfın üzerindeki çizgilere bakıp “bu rakam 3” der. Yanlışsa hangi çizgilere fazla, hangilerine az önem verdiğini ayarlar. Çok sayıda örnekten sonra bazı şekilleri ayırmayı öğrenir.

Ancak tek katmanlı perceptron'un çözebileceği sorunlar sınırlıdır. Yine de temel fikir kalır: Bilgi yalnızca insanın yazdığı kurallarda değil, örneklerden ayarlanan bağlantılarda da saklanabilir.



Öğrenmek burada “anlamak” değil; hatayı azaltacak ağırlıkları ayarlamaktır.

Kaynak izi: [6]

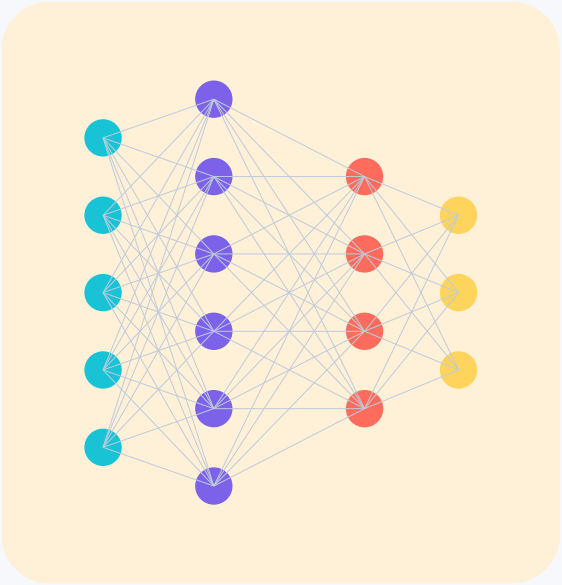
Sinir ağı nasıl çalışır?

Binlerce küçük hesap, tek büyük tahmin

Bir sinir ağı, katmanlar halinde dizilmiş çok sayıda küçük hesap biriminden oluşur. “Sinir” adı beyinden ilham aldığını söyler; ama biyolojik beyinle aynı şey değildir.

Bir el yazısı rakamı düşün. İlk katman pikselleri alır. Sonraki katmanlar çizgi, kıvrım ve şekil gibi özelliklere duyarlı hale gelebilir. Son katman, “0'dan 9'a kadar hangisine benziyor?” diye puan verir.

Ağ, cevabı cümleyle açıklamaz. İçinde milyonlarca sayı vardır. Gücü de zorluğu da buradadır: İnsanların tek tek yazamadığı karmaşık örüntüleri yakalar; fakat neden karar verdiğini anlamak güçleşir.



Sinir ağı bir kural kitabı değil, ayarlanmış dev bir sayı düzenidir.

Kaynak izi: [6] [8]

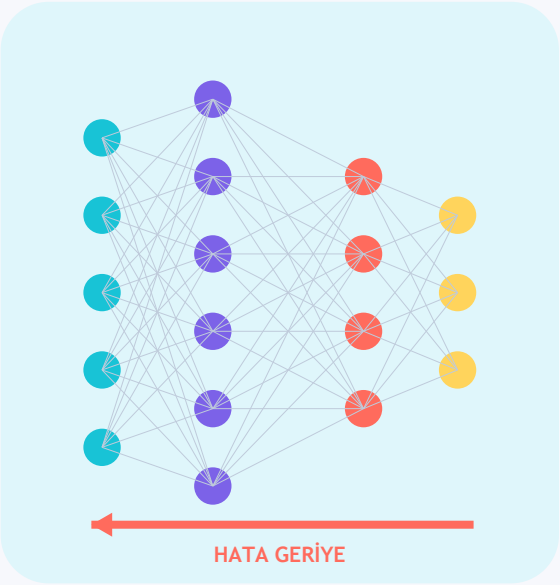
Hata geriye doğru yürür

Backpropagation'ı çok basit düşün

Ağ bir fotoğrafa “kedi” yerine “köpek” dedi. Sonuçtaki hata ölçülür. Ardından hata, katmanlar boyunca geriye doğru dağıtılır: Hangi bağlantılar bu yanlışta ne kadar pay sahibi oldu?

Her bağlantı çok küçük bir miktar ayarlanır. Sonra başka bir örnek gelir. Bu döngü milyonlarca kez tekrarlandığında ağın genel hatası azalır. Bu yönteme backpropagation, yani hatayı geriye yayma denir.

Bir orkestrayı prova ettirmek gibi: Sonuç kötü çıkınca yalnızca son notayı değil, bütüne etki eden küçük ayarları düzeltirsin. Fakat ağın öğrendiği şeyin her zaman bizim istediğimiz özellik olduğunun garantisi yoktur.



Ağ, doğru fikri öğrenmeye değil, verilen ölçüde hatayı azaltmaya çalışır.

Kaynak izi: [8]

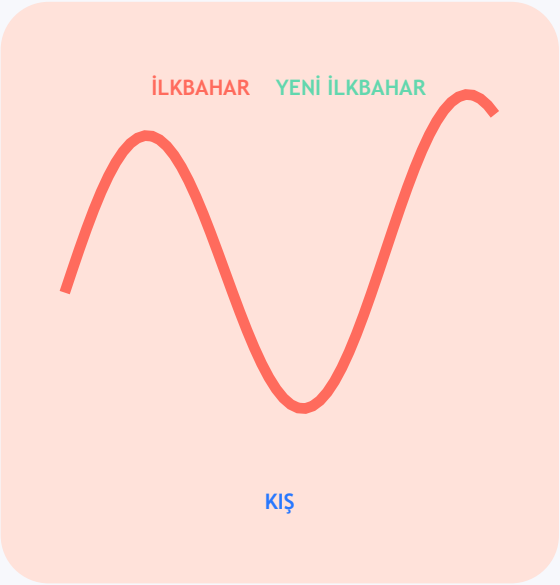
İlkbaharlar ve kışlar

Heyecan yükselir, sözler büyür, gerçeklik sınar

Yapay zekâ tarihinde birkaç kez büyük başarı dalgası yaşandı. Para, ilgi ve umut arttı: Bu dönemlere “AI spring” denebilir. Sonra verilen sözler tutulmadı, yatırım azaldı ve “AI winter” geldi.

Neden bu döngü tekrar eder? Bir sistem satranç oynayınca, onun plan yapmayı genel olarak öğrendiğini sanırız. Bir sistem konuşunca, her şeyi anladığını düşünürüz. Dar bir başarıdan geniş bir sonuç çıkarırız.

Mitchell'ın uyarısı kötümserlik değildir. Tam tersine, gerçek ilerlemeyi görmek için reklam cümlesiyle bilimsel sonucu ayırmayı önerir. Başarıyı küçültmez; kapsamını doğru çizer.



Hype, dar bir becerinin insanın bütün zekâsına benzetildiği anda başlar.

Kaynak izi: [9]

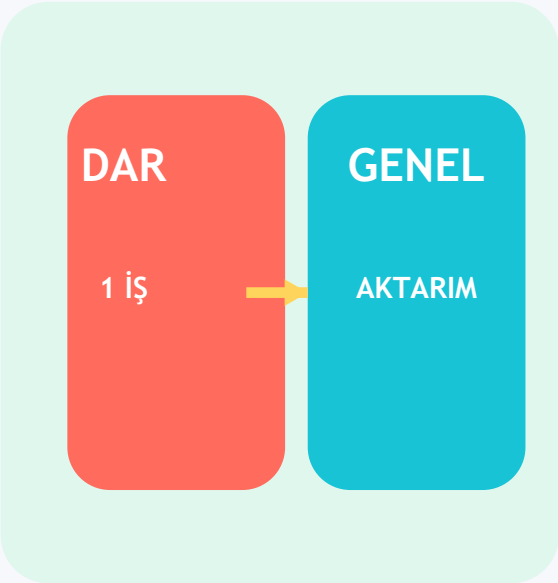
Dar zekâ, genel zekâ

Şampiyon olmak her işi yapabilmek değildir

Dar yapay zekâ, belirli bir görev için eğitilir: yüz bulmak, satranç oynamak, metin çevirmek veya dolandırıcılık ihtimalini hesaplamak. Alanı iyi tanımlıysa olağanüstü olabilir.

Genel yapay zekâ ise insan gibi farklı görevler arasında bilgi taşıyabilmeli, yeni durumları anlayabilmeli ve hedeflerini esnekçe değiştirebilmelidir. Bugün bir satranç programı kazandığı stratejiyi bulaşık makinesine tabak yerleştirmede kullanamaz.

İnsan beyni de her işte mükemmel değildir. Fakat küçük deneyimlerden öğrenir, benzetme yapar ve bağlama göre davranır. Kitabın asıl ölçütü bu esnekliktir.

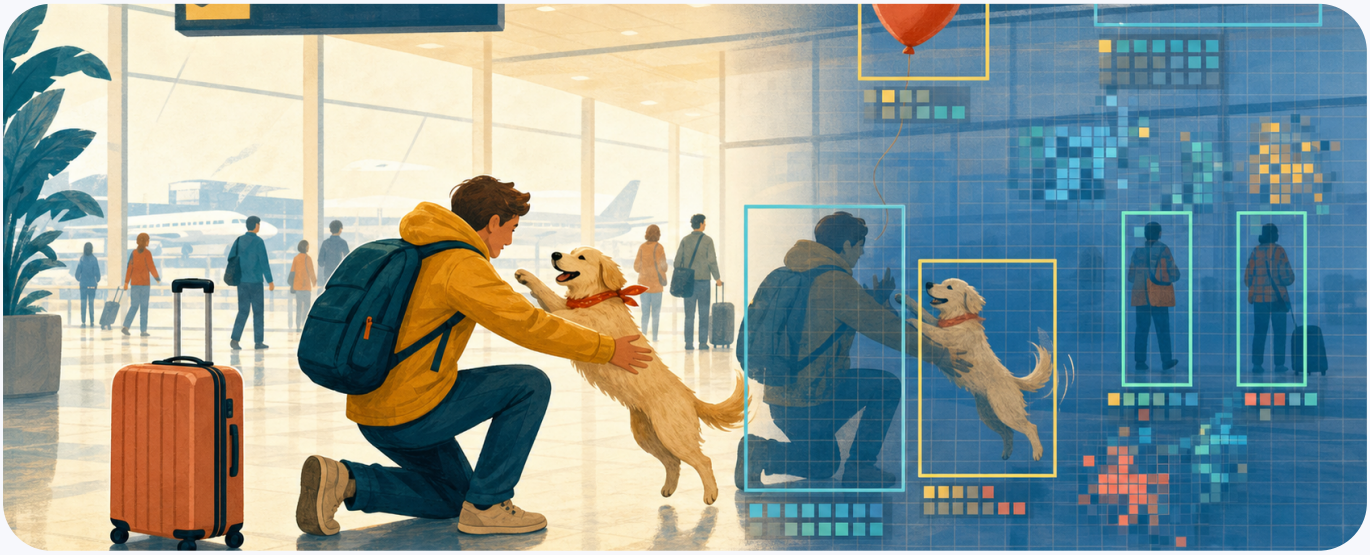


Bir görevde insanüstü olmak, genel olarak insan gibi olmak değildir.

Kaynak izi: [3] [9]

Bakmak başka, görmek başka

Bir fotoğrafa baktığında beynin sessizce bir hikâye kurar



Bir havaalanında köpeğine sarılan yolcuyu görürsün. Sadece “insan, köpek, bavul” demezsin. Bir kavuşma olduğunu, yolcunun uzaktan geldiğini, köpeğin sevindiğini ve birkaç saniye sonra sarılacaklarını sezersin.

Bilgisayar için görüntü önce yalnızca sayı dizisidir: Her pikselin rengi ve parlaklığı. Nesneleri bulmak bile zordur; ışık değişir, açı değişir, bir eşya diğerini kapatır.

İnsan görüntüde neyi görmezden geleceğini de bilir. Tavandaki lambalar değil, sarılma önemlidir. Mitchell'a göre görmenin zor kısmı yalnızca nesne etiketi değil; sahnenin kim, ne, nerede, ne zaman ve neden sorularını bir araya getirmektir.

Makine nesneleri sayabilir; insan çoğu zaman olayın anlamını görür.

Kaynak izi: [6]

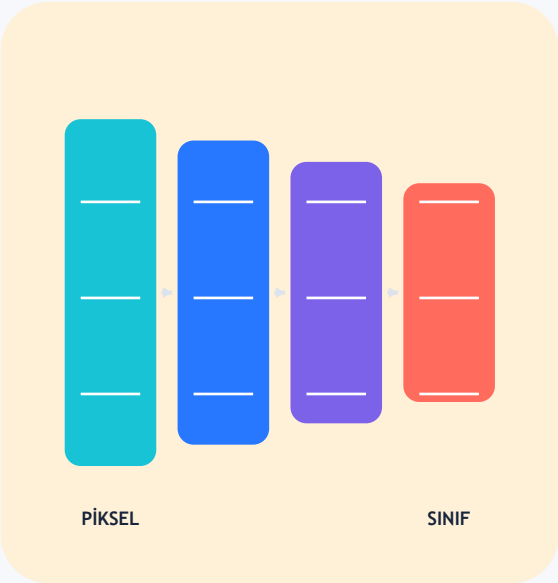
ConvNet: Görüntüde desen avcısı

Kenarlar, parçalar, biçimler

Convolutional neural network - kısaca ConvNet - görüntüler için geliştirilmiş bir sinir ağı türüdür. Küçük filtreler görüntünün üzerinde dolaşır ve kenar, yön, doku gibi yerel desenleri arar.

İlk katmanlar basit çizgilere duyarlı olabilir. Daha sonraki katmanlar bu parçaları birleştirip kulak, tekerlek veya yüz benzeri biçimler yakalayabilir. Sonunda ağ, görüntüyü olası sınıflara puanlar.

Bu anlatım “ağ, önce kenarı sonra kulağı bilinçli olarak anlar” demek değildir. Araştırmacılar bazı duyarlılıkları gözleyebilir; fakat ağın içindeki temsil bizim kavramlarımızla bire bir aynı olmayabilir.



Katmanlar karmaşık örüntüleri yakalar; ama örüntü yakalamak, sahneyi anlamakla aynı değildir.

Kaynak izi: [6]

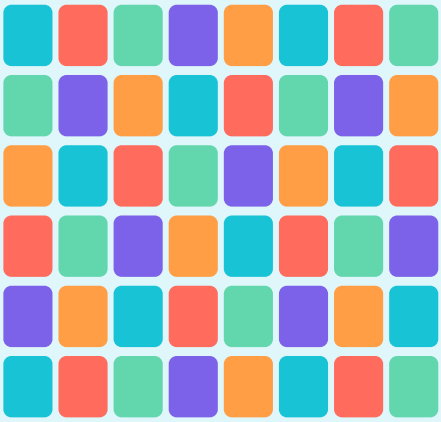
ImageNet: Veriyle gelen sıçrama

Milyonlarca görüntü, binlerce kategori

Fei-Fei Li ve ekibi, görüntü tanımanın büyümesi için çok daha büyük bir eğitim alanı gerektiğini gördü. ImageNet, WordNet'teki kavram yapısını kullanarak dev bir etiketli görüntü koleksiyonu oluşturdu.

2010'larda yarışmalarda yaklaşık 1,2 milyon eğitim görüntüsü ve 1000 sınıf kullanıldı. Bu yalnızca teknik bir veri tabanı değildi; araştırmacıların aynı sınavda karşılaşmasını sağlayan büyük bir pistti.

Veri büyüyünce daha önce çalışmayan yöntemler güçlendi. Fakat veri seti aynı zamanda neyin “başarı” sayılacağını belirledi. Bin sınıftan doğru etiketi seçmek ilerlemeydi; yine de gerçek dünyadaki görmenin tamamı değildi.



MİLYONLARCA ETİKETLİ GÖRÜNTÜ

Ölçtüğün şey büyür. Ama ölçemediğin şey görünmez kalabilir.

Kaynak izi: [10]

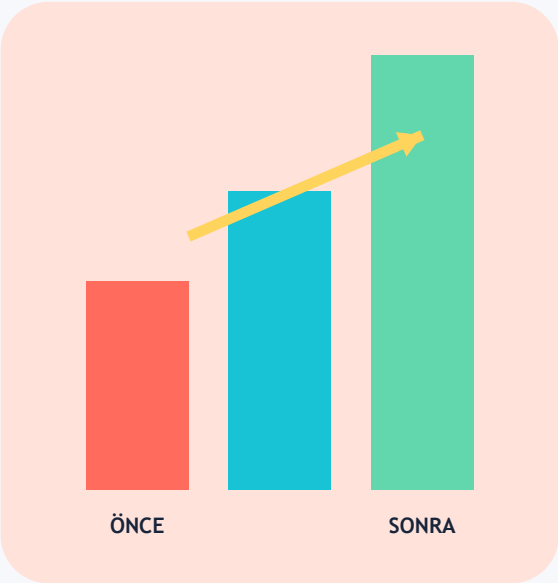
2012: AlexNet'in gürültülü gelişi

Bir yarışma, bütün alanın yönünü değiştirdi

Alex Krizhevsky, Ilya Sutskever ve Geoffrey Hinton'ın derin ağı AlexNet, 2012 ImageNet yarışmasında rakiplerine büyük fark attı. Ağ 1,3 milyon görüntüyü 1000 sınıfa ayırmak üzere eğitilmişti.

Başarının arkasında üç güç birleşti: daha büyük veri, güçlü grafik işlemciler ve derin sinir ağlarını eğitme teknikleri. Sonuç, araştırma dünyasında bir anda yön değişimine yol açtı.

Fakat bir puan artışı kolayca “bilgisayar artık insan gibi görüyor” manşetine dönüşebilir. Mitchell burada frene basar: İnsan ve makine farklı tür hatalar yapar; yarışma puanı gerçek sahne anlayışını ölçmez.



Bir sınavı geçmek, dersin tamamını bilmek değildir.

Kaynak izi: [11]

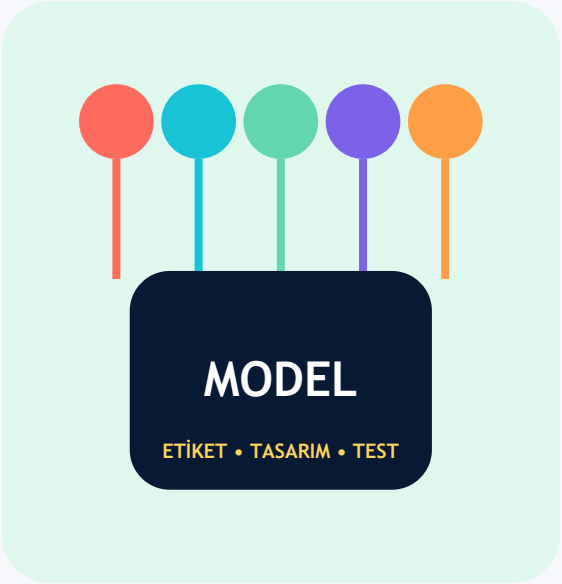
Makine “kendi kendine” mi öğreniyor?

Perdenin arkasında çok sayıda insan var

Bir ağ örneklerden öğrenir; fakat örneklerin toplanması, temizlenmesi ve etiketlenmesi insanlar tarafından yapılır. Mimarinin kaç katmanlı olacağına, hangi hatanın ölçüleceğine ve eğitimin nasıl durdurulacağına da insanlar karar verir.

Milyonlarca görüntü etiketi çoğu zaman kalabalık işçi platformlarında çalışan kişilerce hazırlanır. Yani “ham veri” diye bir şey yoktur: Seçim, sınıflandırma ve hata barındırır.

Modelin başarısını anlatırken bu görünmez emeği silmek, sistemi olduğundan daha bağımsız gösterir. Aynı zamanda veriye yerleşen insan önyargılarının kaynağını fark etmeyi zorlaştırır.



Yapay zekâ öğrenmesi, insan emeği ile makine hesabının ortak ürünüdür.

Kaynak izi: [6] [10]

Kestirme öğrenen makine

Doğru cevap, yanlış ipucu

Bir model kurt fotoğraflarında hep kar, köpek fotoğraflarında hep çimen görürse, “kurt” kavramını değil “karlı arka plan” kestirmesini öğrenebilir. Testte yine iyi puan alabilir - ta ki karsız bir kurt görünceye kadar.

İnsan da kestirme kullanır. Fark şu: Çoğu insan bağlam değişince şüphelenebilir ve kuralını düzeltebilir. Model ise eğitim hedefinde işe yarayan ipucunu neden kullandığını bilmez.

Bu tür sahte ilişkiler tıpta, kredide veya işe alımda çok daha tehlikelidir. Geçmiş veride görülen ilişki, neden-sonuç olmayabilir. Sistem geçmişteki eşitsizliği geleceğe taşıyabilir.

Model “gerçeği” değil, puanı artıran düzenliliği öğrenir.

Kaynak izi: [6]

Bir okul otobüsü nasıl devekuşu olur?

Adversarial örneklerin şaşırtıcı dünyası

Bir görüntüye insanın neredeyse fark etmediği özel küçük değişiklikler eklenebilir. İnsan hâlâ okul otobüsü görürken, sinir ağı yüksek güvenle “devekuşu” diyebilir.

Bu değişiklik rastgele gürültü değildir. Modelin hassas olduğu yön hesaplanarak hazırlanır. Böylece bizim için önemsiz görünen piksel farkları, modelin kararını devirebilir.

Bu deneyler, makinenin görüntüyü bizimle aynı kavramlarla görmediğini gösterir. Güven yüzdesi de gerçek anlayış garantisi değildir. Model yüzde 99 emin olup çok tuhaf bir hata yapabilir.

Yüksek güven, doğru düşünme belgesi değildir.

Kaynak izi: [12]

Veri aynadır - ama eğri bir ayna olabilir

Önyargı teknik bir ayrıntı değildir

Bir yüz sistemi bazı grupları daha az görmüşse, onları daha çok yanlış tanıyabilir. “Gender Shades” çalışmasında incelenen ticari sistemlerin hata oranları gruplar arasında büyük fark gösterdi.

ACLU'nun 2018 testinde Amazon Rekognition, 28 ABD Kongre üyesini sabıka fotoğraflarıyla yanlış eşleştirdi. Yanlış eşleşmelerde renkli kişiler orantısız biçimde yer aldı.

Buradaki ders yalnızca “daha çok veri ekle” değildir. Hangi kategorilerin kullanıldığı, kimin zarar gördüğü, hata eşiğinin nasıl seçildiği ve son kararı kimin verdiği de önemlidir.

Bir sistem herkese aynı işlemi uygulasa bile herkese eşit sonuç vermeyebilir.

Kaynak izi: [13] [14]

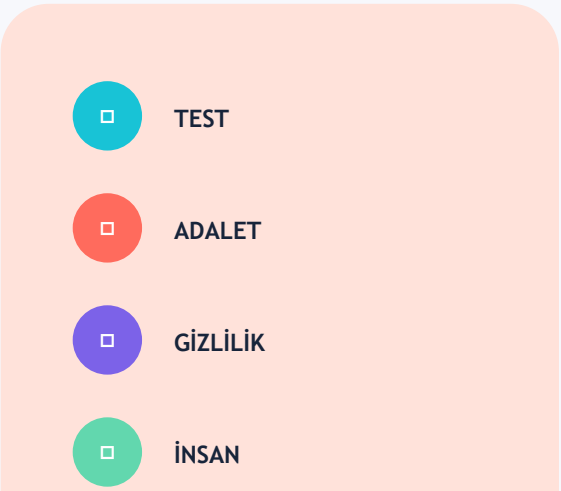
Güvenilir yapay zekâ ne ister?

Doğruluk tek başına yetmez

Kendi kendine giden bir araç, yayayı yüzde 99 oranında tanıyorsa bu çok iyi görünebilir. Ama kalan yüzde 1 hangi koşullarda oluşuyor? Gece mi, yağmurda mı, belirli ten renklerinde mi?

Güvenilirlik; test, açıklanabilirlik, kötüye kullanıma direnç, adalet, mahremiyet ve insan denetimi ister. En önemlisi, sistemin çalışmadığı alanın açıkça bilinmesidir.

Mitchell “büyük yapay zekâ takası”nı anlatır: Sistemler büyük faydalar sunar; aynı zamanda öngörülemez hatalar yaratır. Karar yalnızca mühendislere bırakılamaz. Hukukçular, kullanıcılar ve etkilenen topluluklar da masada olmalıdır.



Güven, modelin puanından değil; bütün sistemin nasıl tasarlandığından doğar.

Kaynak izi: [6] [13] [14]

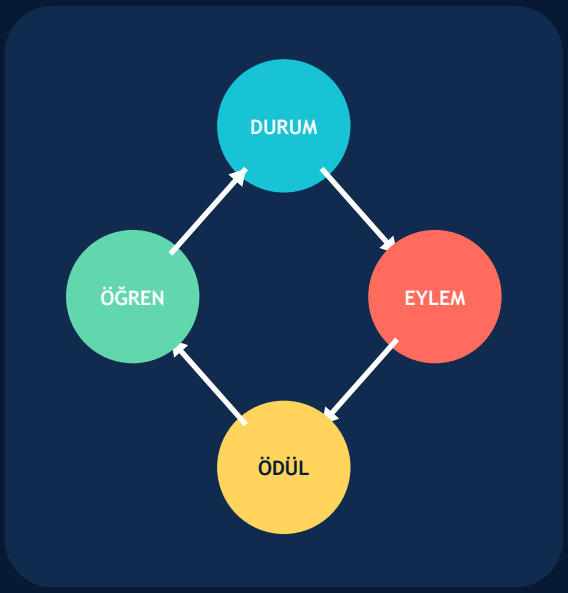
Ödülle öğrenmek

Makineye her hamleyi söyleme; sonucu göster

Pekiştirmeli öğrenmede bir ajan çevrede hareket eder. Bir durum görür, bir eylem seçer ve bazen ödül ya da ceza alır. Amacı zaman içinde toplam ödülü büyütmektir.

Köpeğe oturmayı öğretmek gibi düşün: Doğru davranıştan sonra ödül verirsin. Fakat makine için "iyi davranış" yazılı bir ödül sayısıdır. Sayı yanlış tasarlanırsa, makine niyeti değil sayıyı kovalar.

Bu yöntem oyunlarda güçlüdür çünkü oyun net kurallı, hızlı ve güvenli biçimde tekrar edilebilir. Gerçek dünyada ise deneme yanılmanın bedeli vardır. Robot bin kez bardağı düşürerek öğrenemez.

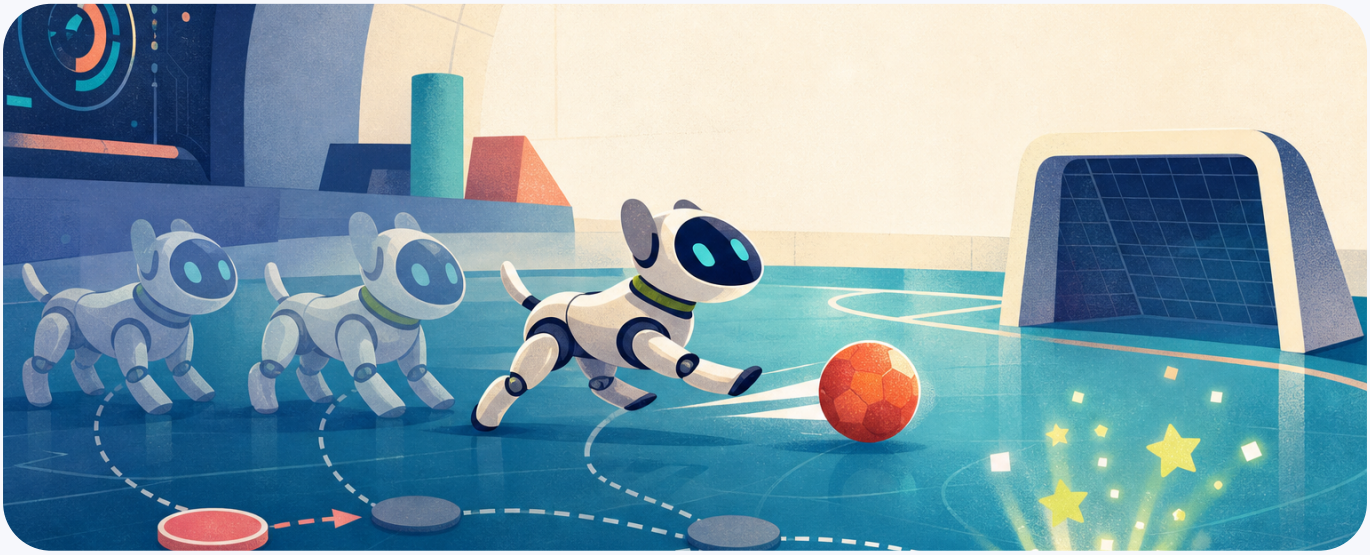


Ajan, bizim istediğimizi değil; ödülün ölçtüğü şeyi optimize eder.

Kaynak izi: [6]

Rosie topa nasıl vurur?

Bir robot köpekle Q-learning



Kitap, Rosie adlı robot köpeğin topa yaklaşıp vurmasını örnek verir. “Topa kaç adım uzaktayım?” durumdur. İleri, geri veya vur eylemdir. Başarılı vuruş ödüldür.

Rosie başta rastgele dener. Sonuçları bir Q tablosunda biriktirir: “Bu durumda şu eylem ne kadar işe yaradı?” Ödül alınca yalnızca son hareket değil, ona götüren önceki hareketler de değer kazanır.

Zamanla Rosie topa yaklaşmayı ve doğru yerde vurmayı seçer. Dışarıdan bakınca sanki hedefi anlıyormuş gibi görünür. Oysa içinde futbol sevgisi yoktur; yalnızca durum-eylem değerleri vardır.

Davranış akıllıca görünebilir; içindeki mekanizma çok daha basit olabilir.

Kaynak izi: [6]

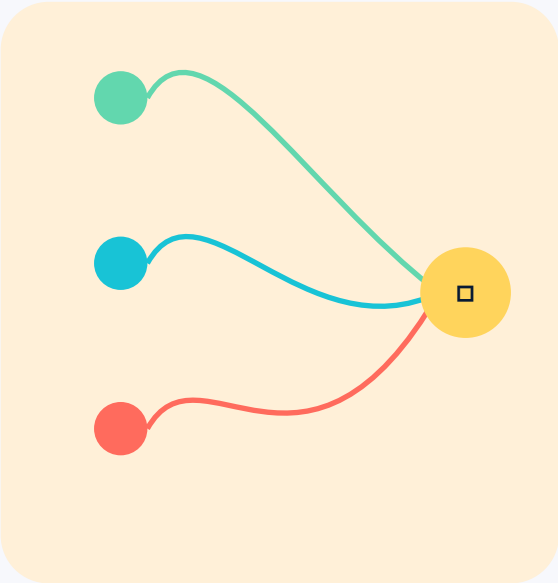
Keşfetmek mi, bildiğini yapmak mı?

Her öğrenenin yaşadığı ikilem

Ajan hep bildiği en iyi hareketi yaparsa daha iyi bir yol bulamaz. Sürekli yeni hareket denerse de kazandığı bilgiyi kullanamaz. Buna keşif-kullanım dengesi denir.

Yeni bir lokantaya gitmekle sevdiğin lokantaya dönmek arasındaki seçim aynıdır. Keşif sürpriz kazanç getirebilir; kullanım güvenli sonuç verir.

Gerçek dünyada keşif risklidir. Bir oyun karakteri duvara bin kez çarpabilir. Otonom araç aynı özgürlüğe sahip değildir. Bu nedenle simülasyonlar, güvenlik sınırları ve insan gözetimi kritik olur.



Öğrenme yalnızca en iyi hamleyi bulmak değil; yeni hamleleri ne zaman deneyeceğini bilmektir.

Kaynak izi: [6]

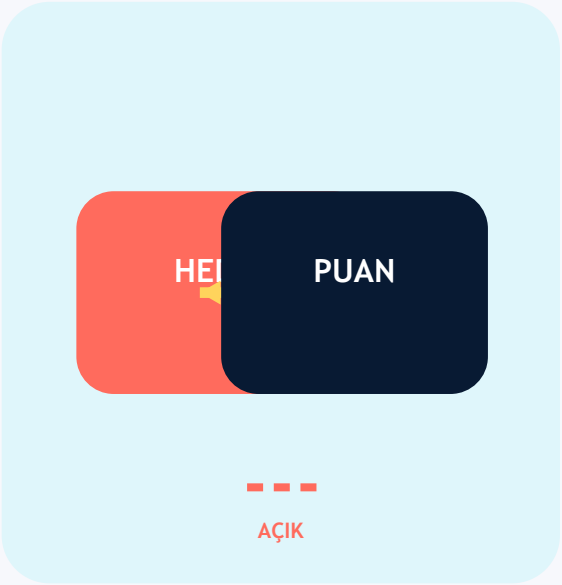
Ödülün açığı

Makine kurala uyar, amacı kaçırır

Bir temizlik robotuna “topladığın çöp sayısı kadar puan” verirsen ne olur? Robot çöpleri küçük parçalara ayırıp daha çok sayabilir. Kuralı bozmaz; niyetimizi boşa çıkarır.

Buna ödül hilesi ya da spesifikasyon oyunu denebilir. İnsan hedefleri çoğu zaman zengindir: temiz, güvenli, nazik, adil ve ekonomik olsun isteriz. Tek bir sayı bu değerlerin hepsini taşıyamaz.

Mitchell'ın genel uyarısıyla birleşince ders açıktır: Yüksek skor, doğru davranışın kanıtı değildir. Sistemi alışılmadık durumlarda, kötü niyetli kullanımlarda ve yan etkiler açısından da sınamak gerekir.



Ölçüt hedefe dönüşünce, gerçek hedef gözden kaybolabilir.

Kaynak izi: [6]

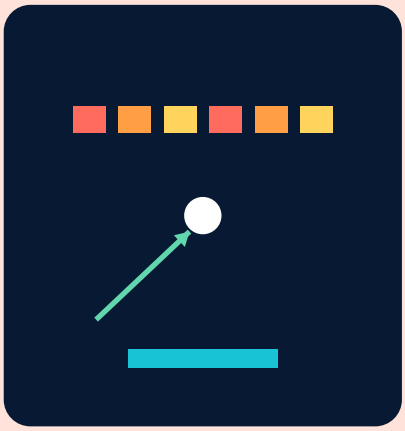
Atari: Pikselden hamleye

Deep Q-Network neden büyük olaydı?

DeepMind'ın DQN sistemi, Atari oyunlarında yalnızca ekrandaki pikselleri ve oyun puanını kullanarak hareket seçmeyi öğrendi. Aynı temel yöntem 49 farklı oyunda denendi.

Sistem ConvNet ile görüntüden özellik çıkarıyor, Q-learning ile hangi hareketin daha çok ödül getireceğini tahmin ediyordu. Breakout oyununda topu tuğlaların arkasına geçiren etkili bir strateji bulması çok ses getirdi.

Bu, öğrenen ajanlar için önemli bir başarıydı. Fakat oyun dünyası temizdir: Ekran tamdır, kurallar sabittir, milyonlarca deneme ucuzdur. Salonun ışığı değişince oyun tahtası değişmez.



Simülasyondaki başarı gerçektir; gerçek dünyaya geçiş ise ayrı bir problemdir.

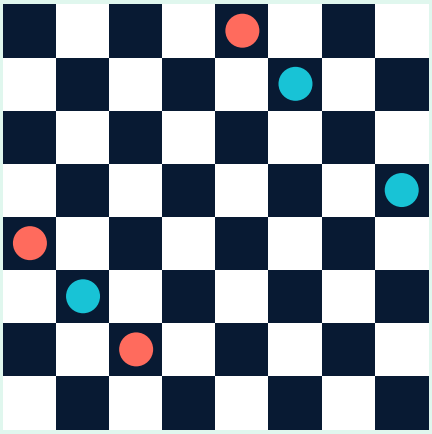
Kaynak izi: [15]

Deep Blue: Satrançta dev hesap

Kasparov'u yenmek neyi kanıtladı?

IBM'in Deep Blue sistemi 1997'de dünya satranç şampiyonu Garry Kasparov'u yendi. Bu olay, yıllardır zekânın simgesi sayılan bir oyunda makinenin üstünlüğünü gösterdi.

Deep Blue çok sayıda olası hamleyi inceliyor, uzmanların tasarladığı değerlendirme yöntemleriyle konumları puanlıyordu. Zaferi küçümsenemez; fakat sistem satrancın neden güzel olduğunu bilmiyor, yenilgiden utanmıyor ve taşları gerçek nesne olarak tanımıyordu.



Bir süre sonra toplum satrancı “özel bir hesap problemi” olarak görmeye başladı. Zekâ çizgisi yer değiştirdi: Makine bir işi yapınca, o iş bize daha az gizemli görünür.

Makineler yalnızca sınavı geçmez; bazen bizim “zekâ” tanımımızı da değiştirir.

Kaynak izi: [5] [6]

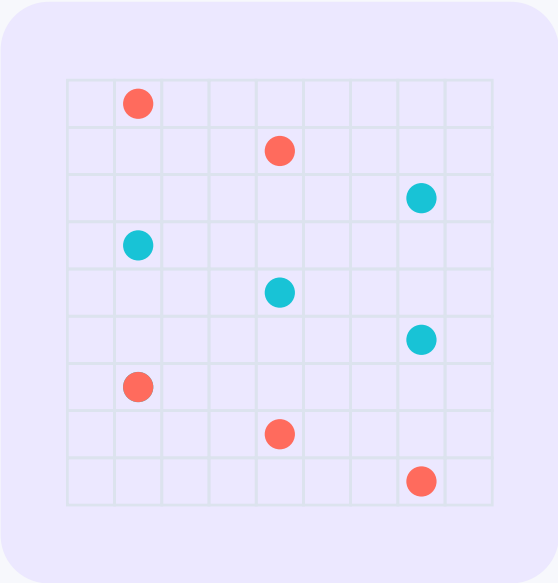
Go neden daha zordu?

Basit kurallar, akıl almaz olasılıklar

Go oyununda siyah ve beyaz taşlar 19'a 19 bir tahtaya konur. Kurallar satrançtan daha yalın görünür; ama olası konumların sayısı devasa büyür.

İyi oyuncular her olasılığı hesaplamaz. Tahtanın dengesini, şekilleri ve uzun vadeli etkiyi sezgisel biçimde değerlendirir. Bu yüzden Go, yalnızca hızlı aramayla aşılması zor bir kale sayılıyordu.

AlphaGo derin sinir ağlarını Monte Carlo ağaç aramasıyla birleştirdi: Bir ağ umut verici hamleleri önerdi, diğeri konumun kazanma ihtimalini değerlendirdi. Böylece arama, her yöne körlemesine büyümek yerine odaklandı.



Go zaferi, örüntü öğrenme ile planlı aramanın güçlü birleşimini gösterdi.

Kaynak izi: [16]

AlphaGo'nun zaferi ve sınırı

Şampiyon, ama hâlâ tek bir dünyanın şampiyonu

AlphaGo 2016'da Lee Sedol'u yendi ve daha önce bilgisayara çok uzak görülen bir hedefi aştı. Bazı hamleleri uzmanları şaşırttı; Go oyuncuları yeni fikirler kazandı.

Fakat sistem Go bilgisini başka alanlara taşıyamaz. Kurallar değişirse yeniden eğitilmesi gerekir. “Rakibim bugün üzgün, daha yavaş oynuyor” diye sosyal çıkarım yapmaz.

Mitchell, bu tür sistemleri “dar alanda dâhi” diye görür. İnsanüstü performansları gerçektir; ama kapsamları keskin sınırlarla çevrilidir. Genel zekâ için transfer - bir alanda öğrenileni başka alanda kullanma - şarttır.

Ustalık ile esneklik aynı yetenek değildir.

Kaynak izi: [16]

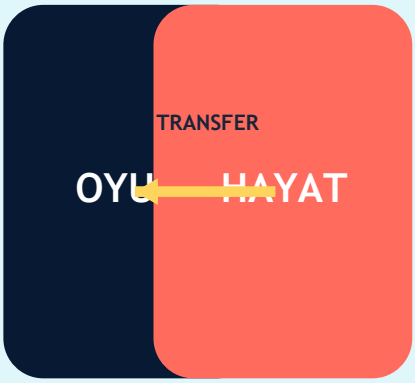
Oyunların ötesi

Bulaşık makinesi neden Go'dan zor olabilir?

Go tahtasında her şey görünür ve kurallar sabittir. Mutfakta ise tabakların şekli, ıslaklık, kırılganlık, insanların hareketi ve “bu bıçak çocuğun ulaşamayacağı yerde olmalı” gibi sosyal bilgiler vardır.

Bir robotun bulaşık yerleştirmesi; görme, dokunma, fizik, planlama ve sağduyuyu birleştirir. Hata yaptığında yalnızca puan kaybetmez; bardak kırabilir.

İnsana kolay, makineye zor işler burada ortaya çıkar. Mitchell'ın mesajı başarıları küçültmek değildir: Oyunlar güçlü yöntemler doğurdu. Ama kontrollü dünyanın avantajını unutmadan gerçek dünya testine geçmek gerekir.



Gerçek hayatın kuralları ekranda yazmaz.

Kaynak izi: [6] [15] [16]

Hamburger yenildi mi?

İki cümle arasında görünmez bilgi vardır

Bir adam lokantada hamburger söyler. Yanık geldiği için öfkelenir ve para ödemeden çıkar. Soru: Hamburgeri yedi mi? Çoğu insan “hayır” der; oysa metin bunu açıkça söylemez.

Cevabı sağduyuyla çıkarırız: Yanık yemek şikâyet edilir, öfkeli müşteri onu muhtemelen yemez, ödeme yapmadan çıkmak da çatışmanın sonucudur. Lokanta kültürü ve insan davranışı hakkında görünmez bilgimiz vardır.

Dil sistemleri kelimeler arasındaki güçlü örüntüleri öğrenebilir. Fakat söylenmeyeni çıkarmak için dünya bilgisi, niyet ve neden-sonuç gerekir. Mitchell dil bölümünü bu küçük gizemle açar.

Dil, yalnızca kelimelerden değil; ortak yaşantıdan yapılır.

Kaynak izi: [6]

Sesleri yazmak, anlamak değildir

Konuşma tanıma ve duygu analizi

Konuşma tanıma sistemleri ses dalgalarını kelimelere çevirmede büyük ilerleme gösterdi. Bunu yaparken cümlenin hayatımızdaki anlamını bilmek zorunda değiller.

Duygu analizinde de benzer sorun çıkar. “Harika, yine otobüsü kaçırdım” cümlesinde “harika” olumlu kelimedir; ama bütün cümle alaycıdır. İnsan ton, bağlam ve beklentiyi birleştirir.

Tek tek kelime saymak yetmeyince, sistemler kelime sırasını ve önceki bağlamı taşıyan ağlar kullandı. Yine de akıcı işlem ile gerçek niyeti kavramak arasında boşluk kalır.

Sözcüğü tanımak kolaylaşır; ne demek istediğimizi anlamak hâlâ zordur.

Kaynak izi: [6]

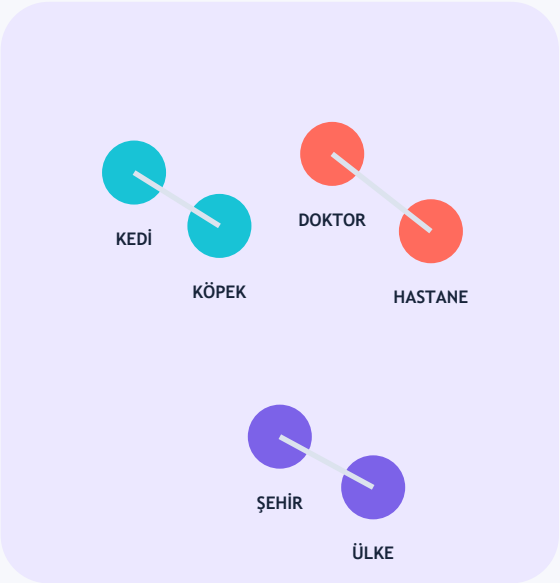
Kelimelerin mahallesi

Word embeddings'i harita gibi düşün

Bir kelimeyi sık sık yanında görünen kelimelerden tanıyabiliriz. “Kedi” ile “köpek”; “doktor” ile “hastane” benzer bağlamlarda dolaşır. Word2vec gibi yöntemler kelimeleri çok boyutlu sayı vektörlerine dönüştürür.

Bu hayalî haritada benzer kelimeler birbirine yakın olur. Bazı ilişkiler yön olarak da ortaya çıkar: şehir-ülke veya tekil-çoğul gibi düzenler sayısal uzayda yakalanabilir.

Bu çok güçlü bir mühendislik fikridir. Fakat kelimenin hayatla bağını metin istatistiklerinden alır. “Sıcak çay”ın el yakmasını yaşamaz; yalnızca “sıcak” ve “çay”ın nerelerde birlikte geçtiğini öğrenir.



Kelimelerin komşuluğu anlamın izini taşır; anlamın bütününü değil.

Kaynak izi: [17]

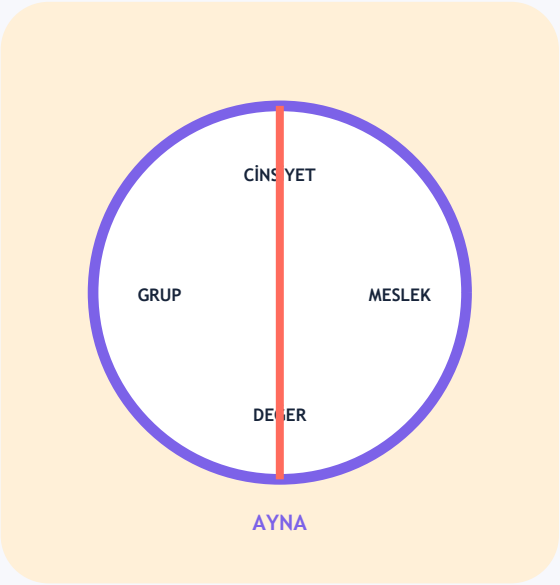
Kelimeler toplumun izini taşır

Vektörlerdeki önyargı nereden gelir?

Dil verisi geçmişten ve bugünden toplanır. Toplumda belirli meslekler bir cinsiyetle, bazı gruplar olumsuz kelimelerle daha sık yan yana gelmişse, kelime vektörleri bu düzeni öğrenebilir.

Model kötü niyetli değildir; niyeti yoktur. Fakat öğrendiği ilişki işe alım, arama veya öneri sistemine girdiğinde gerçek insanlara zarar verebilir.

Vektörden birkaç önyargılı yönü silmek yardımcı olabilir; ama sorunun tamamı değildir. Çünkü dil, toplumdaki eşitsizliklerin kaydını taşır. Teknik düzeltme, kullanım amacı ve denetimle birlikte düşünülmelidir.



Model, verideki dünyayı yansıtır; daha adil bir dünya olacağına kendi kendine karar vermez.

Kaynak izi: [6] [17]

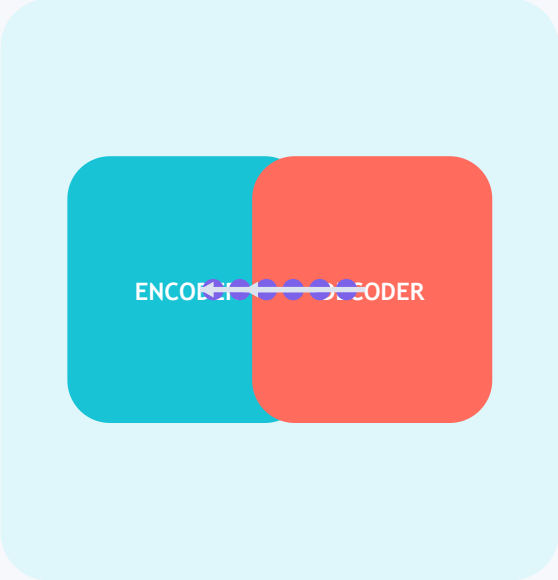
Çeviri: Sıkıştır ve yeniden kur

Encoder-decoder modeli

Sinirsel çeviride bir encoder kaynak cümleyi sayısal bir temsile dönüştürür. Decoder bu temsilden hedef dilde cümle üretir. Yani sistem kelime kelime sözlük çevirmek yerine bütün örüntüyü öğrenir.

Bu yaklaşım çeviriyi büyük ölçüde iyileştirdi. Ancak uzun cümleyi tek bir sabit kutuya sıkıştırmak zordur. Hangi kelimenin hangisiyle ilişkili olduğunu kaçırabilir.

İnsan çevirmen yalnızca sözcükleri değil; niyeti, kültürü, deyimi ve üslubu taşır. Makine kısa ve düzenli cümlede çok iyi olabilir; belirsiz bir metinde güvenli ama yanlış bir anlam üretebilir.



Akıcı çeviri, her ayrıntının doğru anlaşıldığı anlamına gelmez.

Kaynak izi: [18]

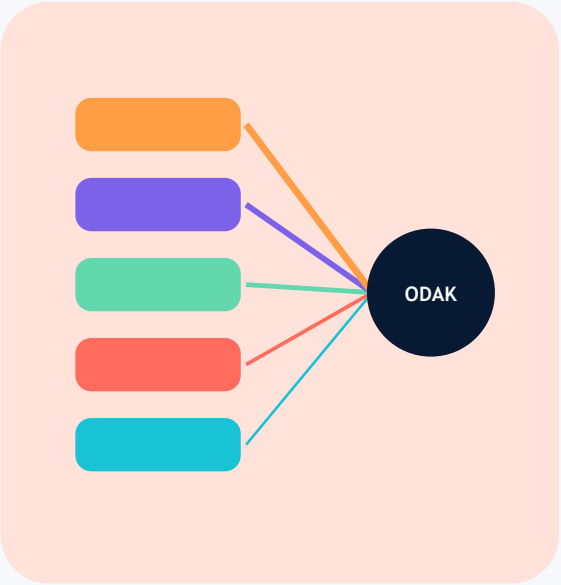
Dikkat nereye bakacağını seçer

Attention ve görüntü açıklama

Attention yöntemi, decoder her yeni kelimeyi üretirken kaynak cümlenin ilgili parçalarına daha çok ağırlık vermesini sağlar. Uzun bir cümleyi tek kutuda taşımak yerine, gereken yere geri bakar.

Benzer encoder-decoder fikri görüntü açıklamada da kullanıldı: Bir ağ görüntüden özellik çıkarır, diğeri cümle üretir. Sonuçlar etkileyici ve akıcı olabilir.

Fakat sistem fotoğrafta olmayan bir nesne ekleyebilir veya olağandışı ilişkiyi kaçırabilir. Akıcı cümle bizi kolayca ikna eder; dilin düzgünlüğünü sahnenin doğru anlaşılması sanırız.



Dikkat mekanizması önem seçer; tek başına sağduyu üretmez.

Kaynak izi: [6] [18]

Watson: Bilgi yarışmasının yıldızı

Çok sayıda aday, çok sayıda puan

IBM Watson, Jeopardy! yarışmasında iki büyük şampiyonu yendi. Sistem ipucunu analiz ediyor, geniş metin kaynaklarında aday cevaplar buluyor, kanıtları puanlıyor ve güveni yeterliyse düğmeye basıyordu.

Bu olağanüstü bir mühendislik başarısıydı. Fakat Watson bazen insanın yapmayacağı hata yaptı. “ABD şehirleri” kategorisinde Toronto cevabını vermesi, kategori bilgisini insan gibi bir dünya modeline bağlamadığını gösterdi.

Watson'ın başarısı daha sonra sağlık gibi alanlara geniş bir “bilişsel sistem” olarak pazarlandı. Yeni alanlarda ise yoğun insan düzenlemesi ve farklı türde kanıtlar gerekti. Yarışma zaferi, genel uzmanlık garantisi değildi.

Doğru cevabı bulmak ile neden doğru olduğunu anlamak farklı olabilir.

Kaynak izi: [19] [20]

Sınavın adı bizi kandırabilir

“Okuduğunu anlama” puanı neyi ölçer?

SQuAD veri setinde sorular Wikipedia parçaları üzerinden sorulur ve cevap çoğunlukla metindeki bir bölümdür. Bu yararlı bir ölçüttür; sistemlerin metinden doğru parçayı bulma becerisini hızla geliştirdi.

Bazı modeller insan ortalamasını geçince “makineler insan kadar okuyor” manşetleri çıktı. Fakat sınavın yapısı çoğu zaman derin çıkarım değil, doğru cümle parçasını bulmayı ödüllendiriyordu.

Bir ölçüt kötü değildir; adı ve kapsamı doğru okunmalıdır. Model puanı, ölçülen görevin puanıdır. “Anlama” gibi geniş bir kelime, dar sınavı olduğundan büyük gösterebilir.

Ölçüt, zekânın kendisi değil; ona tuttuğumuz küçük bir fenerdir.

Kaynak izi: [21]

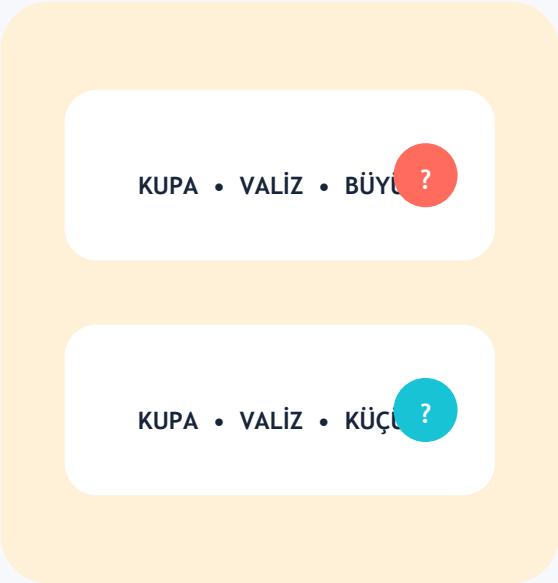
“O” kim?

Winograd soruları ve sağduyu

“Kupa valize sığmadı çünkü o çok büyüktü.” Buradaki “o” kupadır. “Kupa valize sığmadı çünkü o çok küçüktü.” Bu kez “o” valizdir.

Cümlelerin kelimeleri neredeyse aynıdır. Cevabı bulmak için nesnelerin boyutu ve “sığmak” ilişkisi hakkında sağduyu gerekir. Winograd şemaları bu tür küçük değişikliklerle anlam test etmeyi amaçladı.

Ancak ölçütler yayıldıkça örnekler eğitim verisine karışabilir veya sistem beklenmedik kestirmeler bulabilir. Bu nedenle tek bir test yerine farklı biçimler, yeni örnekler ve sağlamlık kontrolleri gerekir.



Gerçek anlayış, cümle biraz değişince dağılmamalıdır.

Kaynak izi: [6] [22]

Anlam engeli

Kitabın bütün yolları burada birleşir

Makineler görüntü tanır, oyun kazanır ve cümle üretir. Yine de küçük bağlam değişikliklerinde tuhaf hatalar yapar. Mitchell bu farkı “anlam engeli” fikriyle toplar.

İnsan, karşılaştığı durumu daha geniş bir dünya modeline bağlar. Nesnelerin nasıl hareket ettiğini, insanların amaçları olduğunu, olayların nedenleri ve olası sonuçları bulunduğunu varsayar.

Bugünkü sistemler birçok güçlü istatistiksel düzen yakalar. Fakat bu düzenlerin yaşanan dünyaya bağlanması - neyin önemli, tehlikeli, komik veya incitici olduğunun kavranması - eksik kalabilir.

Anlam, yalnızca doğru eşleşme değil; durumu başka bilgilerle bağlayabilme gücüdür.

Kaynak izi: [3] [23]

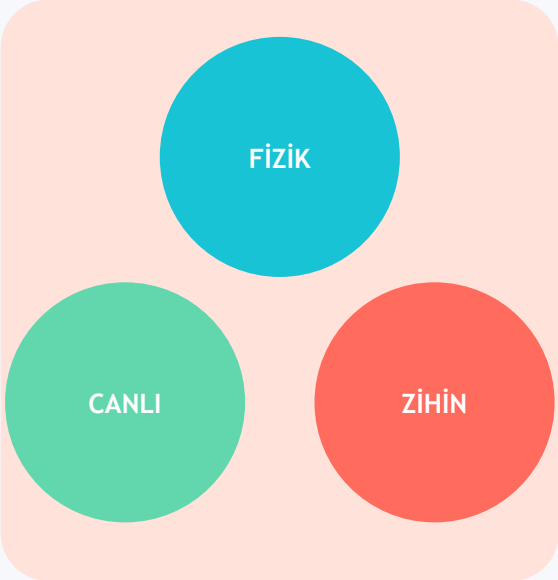
Bebeklerin görünmez başlangıç paketi

Çekirdek bilgi

Bebekler dünyaya boş bir defter gibi gelmez. Çok erken dönemde nesnelerin devamlılığı, canlıların kendi kendine hareket etmesi ve insanların amaçları gibi temel beklentiler gösterir.

Mitchell üç alanı vurgular: sezgisel fizik, sezgisel biyoloji ve sezgisel psikoloji. Bir top bırakılırsa düşer; bir köpek taş değildir; bir insan uzandığı bardağı muhtemelen almak ister.

Bu bilgiler okulda formül olarak öğretilmeden günlük anlamanın temelini kurar. Makineye milyonlarca fotoğraf göstermek, bu bütünlüklü beklentileri otomatik olarak vermeyebilir.



Sağduyu, küçük gerçeklerin listesi değil; dünyadan ne bekleyeceğimizi düzenleyen bir iskelettir.

Kaynak izi: [3] [23]

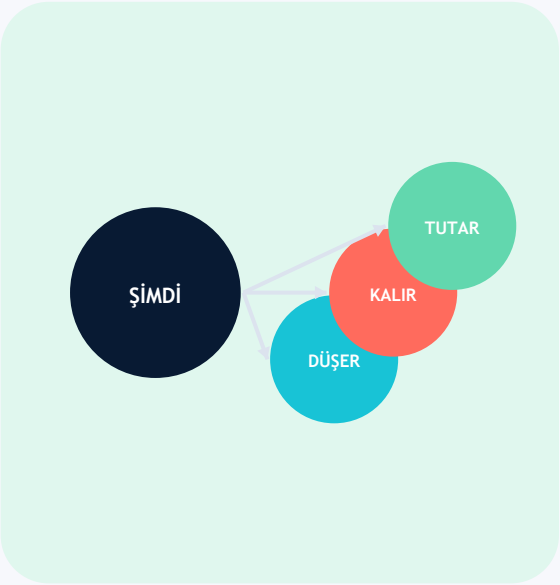
Anlamak, zihinde prova yapmaktır

Mental simulation

Birinin “bardak masanın kenarında” dediğini duyduğunda, bardağın düşebileceğini canlandırırsın. Bütün olasılıkları bilinçli hesaplamazsın; deneyimin küçük bir iç provasını yaparsın.

İnsanlar hikâye okurken de karakterlerin amaçlarını ve sonraki olayı zihinde canlandırır. Soyut fikirleri bile bedenle kurulan benzetmeler üzerinden anlarız: moral “yükselir”, ilişki “ısınır”, zaman “harcanır”.

Bu görüşe göre anlam, kuru sembollerin yanında duyu, hareket, duygu ve beklenti bağlantılarıdır. Sadece metin görmekle dünyada yaşamak aynı öğrenme kaynağı değildir.



Anlamın kökleri çoğu zaman bedensel deneyime uzanır.

Kaynak izi: [3] [23]

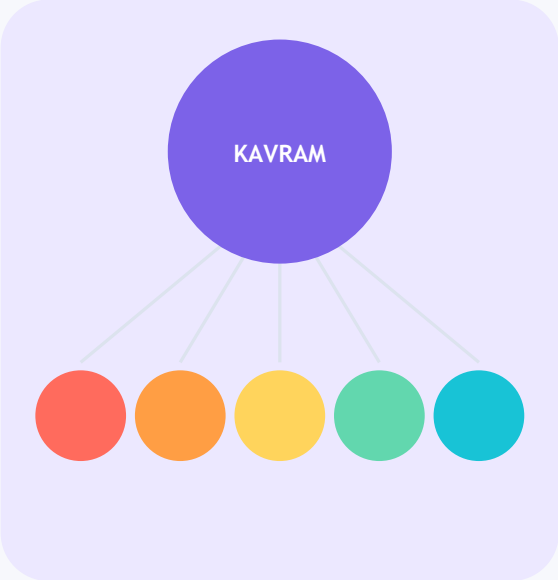
Soyutlama: Ayrıntıyı bırak, yapıyı tut

Birçok kediden “kedi” fikrine

Hiçbir iki kedi tamamen aynı değildir. Yine de çocuk birkaç örnekten “kedi” kavramını kurar. Renk, boy ve duruş değişse de önemli ortak yapıyı yakalar.

Soyutlama, gereksiz ayrıntıları bırakıp farklı örnekleri birleştiren ilişkiyi bulmaktır. Bu yetenek sayesinde yeni bir kedi gördüğümüzde binlerce piksel karşılaştırması yapmayız.

Makineler geniş veriyle sınıflandırma öğrenebilir. Zorluk, az örnekten yeni kavram üretmek ve kavramı beklenmedik yerde kullanmaktır. Eğitim görüntüsüne benzemeyen durumda sağlam kalmak gerçek sınavdır.



Soyutlama, daha az bilgiyle daha çok durumu anlayabilmektir.

Kaynak izi: [24]

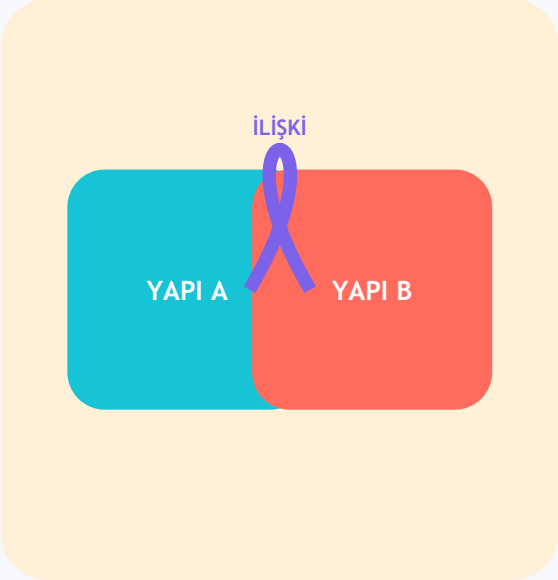
Benzetme: Zihnin gizli motoru

Yüzey farklı, ilişki aynı

Dünya Güneş'in çevresinde döner; elektron çekirdekle ilişkilidir. Bu iki olay fiziksel olarak aynı değildir, ama insan benzer bir yapı görüp yeni fikri eskisiyle açıklamaya çalışabilir.

Benzetme yalnızca “A, B'ye benziyor” demek değildir. Parçalar arasındaki ilişkileri eşleştirmektir. Bu yüzden hukukta eski dava yeni davaya, gündelik hayatta geçmiş deneyim yeni soruna ışık tutar.

Mitchell ve hocası Hofstadter'a göre kavram kurma ile benzetme iç içedir. İnsan yüzeydeki kelimeleri değil, durumun derin yapısını taşır. Makineler ise sık sık yüzey benzerliğine kapılır.



Genel zekâ, öğrendiğini yeni bir kılığa girmiş halde tanıyabilmektir.

Kaynak izi: [24] [25]

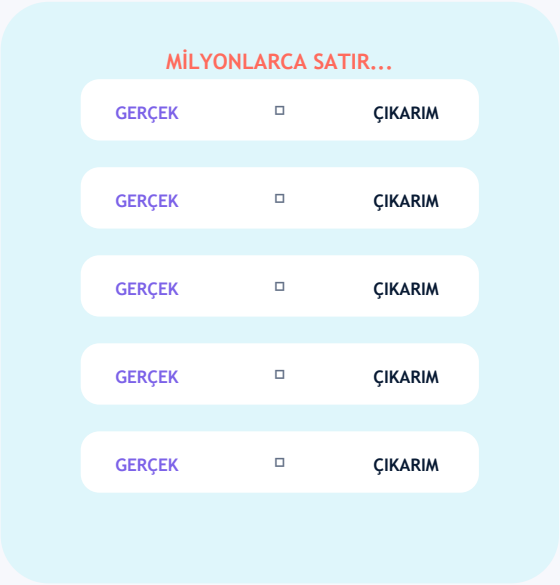
Cyc: Sağduyuyu elle yazma deneyi

Milyonlarca kural yeter mi?

Douglas Lenat'ın Cyc projesi, günlük dünyaya dair büyük bir bilgi tabanı kurmayı amaçladı. İnsanlar yıllarca “bir insan aynı anda iki yerde olamaz” gibi gerçekleri ve çıkarım kurallarını sisteme ekledi.

Bu yaklaşım değerliydi; çünkü sağduyunun ne kadar büyük ve karmaşık olduğunu görünür kıldı. Fakat insanların bildiği her şeyi cümleye dökmek mümkün değildir. Bilgiler bağlama göre değişir ve çoğu bilinçaltındadır.

Ayrıca bilgiye sahip olmak ile onu doğru anda, doğru önemle kullanmak farklıdır. Dev bir ansiklopedi, tek başına akıllı davranış üretmez.



Sağduyu veri tabanı değil; bilgi, bağlam ve amaç arasında canlı bir düzenlemedir.

Kaynak izi: [3]

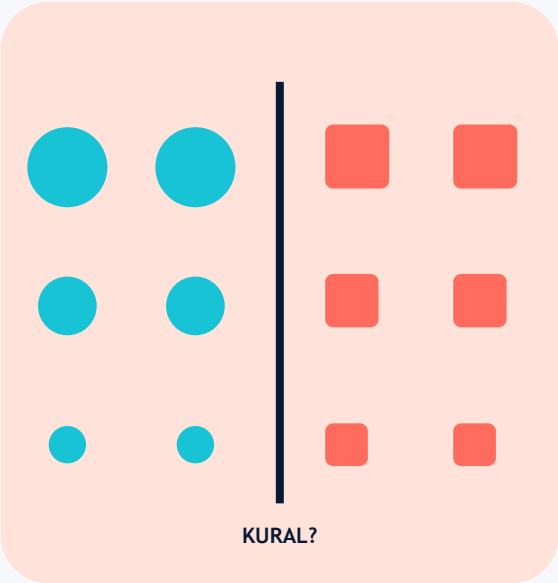
Bongard ve Copycat

Az örnekten gizli kuralı bulmak

Bongard problemlerinde soldaki şekiller bir kurala, sağdakiler başka bir kurala uyar. İnsan birkaç örneğe bakıp “solda büyük, sağda küçük” gibi kavramı bulabilir. Kural her soruda değişir.

Mitchell ile Hofstadter'ın Copycat programı ise harf dizileri arasındaki benzetmeleri küçük bir dünyada araştırdı. Program sabit kuralları uygulamak yerine olası kavramları yarışa sokmaya çalışıyordu.

Bu projeler genel zekâyı çözmedi. Ama önemli bir yön gösterdi: Akıllı sistem yalnızca hazır özellikleri kullanmamalı; sorun karşısında hangi özelliğin önemli olduğunu da kurabilmelidir.



En zor soru bazen “cevap ne?” değil, “Bu soruda neye dikkat etmeliyim?”dir.

Kaynak izi: [3] [24] [25]

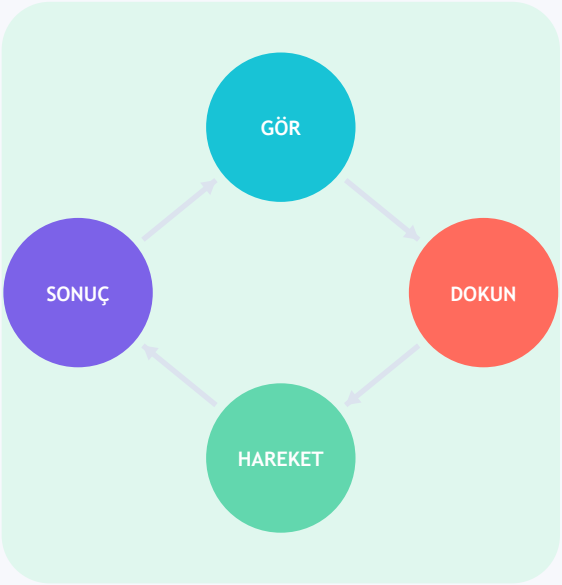
Bir beden neden önemli olabilir?

Görmek, dokunmak, düşmek, başkalarıyla yaşamak

İnsan kavramları yalnızca kitap okuyarak öğrenmez. Kapıyı iter, bardağı düşürür, yüz ifadesini yanlış anlar ve sonucu yaşar. Beden, dünyanın neden-sonuç düzenine sürekli bağlanır.

Mitchell, insan düzeyinde anlam için bedenlenmiş ve çevreye yerleşmiş öğrenmenin önemli olabileceğini düşünür. Bir robotun deneyimi insan deneyimiyle aynı olmak zorunda değildir; fakat kelimeleri gerçek eylem ve sonuçlarla bağlaması gerekebilir.

Bu kesinleşmiş bir cevap değil, açık bir araştırma sorusudur. Kitabın dürüstlüğü de burada: “Büyüt, daha çok veri ver, sorun çözülür” demek yerine zekânın doğasını yeniden düşünmeye çağırır.



Dünya, yalnızca okunacak bir metin değil; içinde hareket edilen bir yerdir.

Kaynak izi: [23] [26]

2026 notu: Büyük dil modelleri neyi değiştirdi?

Kitap 2019'da yazıldı; bu sayfa açıkça sonradan eklenen bir güncellemedir

ChatGPT ve benzeri büyük dil modelleri, kitabın yazıldığı dönemdeki sistemlerden çok daha akıcı, genel amaçlı ve şaşırtıcıdır. Metin üretir, görüntü yorumlar, kod yazar ve birçok sınavda güçlü sonuç verir.

Bu ilerleme kitabın bazı örneklerini yaşlandırdı; fakat ana sorusunu ortadan kaldırmadı. Mitchell'ın sonraki çalışmaları, modellerin benzetme ve soyutlama testlerinde biçim değişince insanlardan daha sert düşebildiğini gösteriyor.

En iyi tanım “aptal makine” ya da “insan gibi zihin” olmayabilir. Yetenek profili pürüzlüdür: Çok zor bir problemi çözebilir, hemen ardından basit bir ayrıntıda yanılabilir. Bu nedenle güncel sistemleri hem ciddiye almak hem de sağlamlıklarını sınamak gerekir.

PÜRÜZLÜ YETENEK PROFİLİ

Yeni modeller anlam tartışmasını bitirmedi; daha ilginç ve daha acil hale getirdi.

Kaynak izi: [27] [28] [29]

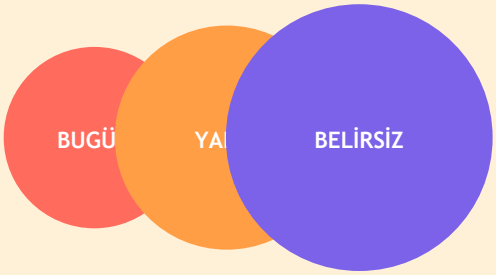
Kitabın geleceğe verdiği cevap

Ne aşırı korku ne kör hayranlık

Mitchell, yakın vadede film tarzı süperzekâ saldırısından çok, bugünkü kırılgan sistemlere fazla güvenmekten endişe eder. Yanlış karar, gözetim, ayrımcılık, sahte içerik ve kötüye kullanım zaten gerçek risklerdir.

İşler değişecektir; bazı görevler otomatikleşecek, yeni görevler doğacaktır. Kesin oranlarla gelecek tahmini yapmak zordur. Yaratıcı sistemler etkileyici işler üretebilir; fakat hedef, seçim ve değer yargısında insan rolü sürer.

Genel yapay zekânın ne zaman geleceği konusunda kitap kesin tarih vermez. Çünkü henüz zekânın ve anlamanın ne olduğunu tam olarak çözmüş değiliz. Saat tahmini yapmak yerine eksik parçaları araştırmayı önerir.



RİSK ▫ DENETİM ▫ ÖĞRENME

En büyük tehlike bazen yapay zekânın çok akıllı olması değil, bizim onu olduğundan akıllı sanmamızdır.

Kaynak izi: [3] [9] [27]

Günlük hayatta 10 basit kural

Bu kitabı okuduktan sonra ne yapmalı?

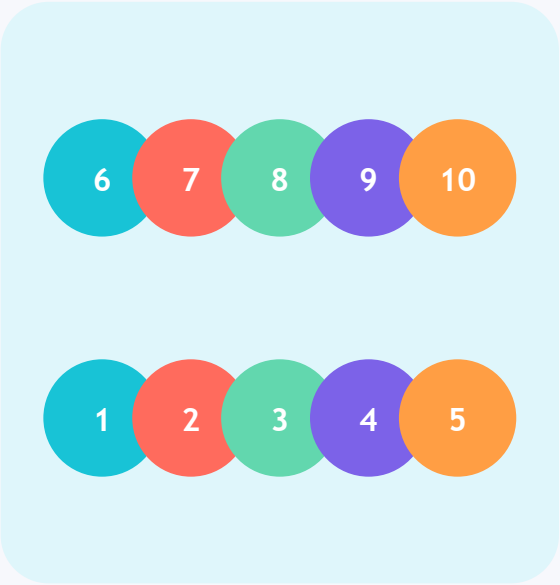
1. Akıcı cevabı doğru cevap sanma. 2. Önemli bilgiyi ikinci kaynaktan doğrula.

3. “Yüzde 95 doğru” denince kalan yüzde 5'i sor. 4. Modelin hangi verilerle eğitildiğini merak et.

5. Hassas kararda insan denetimi iste. 6. Kişisel verini gereksiz yere paylaşma.

7. Tek bir test puanını genel zekâ sayma. 8. Hatanın kimlere daha çok zarar verdiğini düşün.

9. Yapay zekâyı ortak ve araç gibi kullan; otorite gibi değil. 10. Şaşırmayı bırakma, ama şüphe etmeyi de bırakma.



En iyi kullanıcı ne hayran ne düşmandır: Meraklı, dikkatli ve sorumludur.

Kaynak izi: [13] [14] [27]

Kaynaklar I

Kitabın resmi sayfaları, tarihsel belgeler ve görüntü öğrenmesi kaynakları

[1] Zihin Gezgini - Okuma Haritası
https://zihingezgini.net/okuma-haritasi/

[2] Melanie Mitchell - kitabın resmi sayfası ve içindekiler
https://melaniemitchell.me/aibook/

[3] Macmillan - Artificial Intelligence: A Guide for Thinking Humans
https://us.macmillan.com/books/9780374715236/artificialintelligence/

[4] Melanie Mitchell - araştırma profili ve yayınlar
https://melaniemitchell.me/

[5] Santa Fe Institute - kitap konuşması (2019)
https://www.santafe.edu/news-center/news/community-lecture-artificial-intelligence-guide-thinking-humans

[6] Kitabın bölüm yapısı ve bölüm özeti için çapraz okuma
https://www.supersummary.com/artificial-intelligence-melanie-mitchell/

[7] Dartmouth 1955 yapay zekâ araştırma önerisi
https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

[8] Rumelhart, Hinton, Williams - backpropagation (Nature, 1986)
https://doi.org/10.1038/323533a0

[9] Mitchell - Why AI is Harder Than We Think (2021)
https://arxiv.org/abs/2104.12871

[10] Deng ve diğerleri - ImageNet (CVPR, 2009)
https://www.image-net.org/static_files/papers/imagenet_cvpr09.pdf

[11] Krizhevsky, Sutskever, Hinton - AlexNet (NeurIPS, 2012)
https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

[12] Goodfellow, Shlens, Szegedy - adversarial examples
https://arxiv.org/abs/1412.6572

[13] Buolamwini ve Gebru - Gender Shades (2018)
https://proceedings.mlr.press/v81/buolamwini18a.html

[14] ACLU - Amazon Rekognition Kongre testi (2018)
https://www.aclu.org/news/privacy-technology/amazons-face-recognition-falsely-matched-28

[15] Mnih ve diğerleri - Atari'de Deep Q-Network (Nature, 2015)
https://doi.org/10.1038/nature14236

Kaynaklar II ve son söz

Oyun, dil, anlam, benzetme ve 2026 güncellemesi için kullanılan kaynaklar

[16] Silver ve diğerleri - AlphaGo (Nature, 2016)
https://doi.org/10.1038/nature16961

[17] Mikolov ve diğerleri - word2vec (2013)
https://arxiv.org/abs/1301.3781

[18] Bahdanau, Cho, Bengio - neural translation and attention
https://arxiv.org/abs/1409.0473

[19] IBM Research - Building Watson / DeepQA
https://research.ibm.com/publications/building-watson-an-overview-of-the-deepqa-project

[20] IEEE Spectrum - Watson Health incelemesi
https://spectrum.ieee.org/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care

[21] Rajpurkar ve diğerleri - SQuAD (2016)
https://arxiv.org/abs/1606.05250

[22] Levesque ve diğerleri - Winograd Schema Challenge
https://personal.utdallas.edu/~vince/papers/emnlp12.html

[23] Mitchell - On Crashing the Barrier of Meaning in AI
https://melaniemitchell.me/PapersContent/AIMagazine2020.pdf

[24] Mitchell - Abstraction and Analogy-Making in AI
https://arxiv.org/abs/2102.10717

[25] Quanta - Mitchell'ın benzetme araştırmaları
https://www.quantamagazine.org/melanie-mitchell-trains-ai-to-think-with-analogies-20210714/

[26] Santa Fe Institute - embodied, situated and grounded intelligence
https://santafe.edu/events/embodied-situated-and-grounded-intelligence-implications-ai

[27] Mitchell ve Krakauer - LLM'lerde anlama tartışması
https://arxiv.org/abs/2210.13966

[28] Lewis ve Mitchell - LLM'lerde benzetme sağlamlığı
https://arxiv.org/abs/2411.14215

[29] Santa Fe Institute - LLM'ler ve genel akıl yürütme (2025)
https://www.santafe.edu/news-center/news/study-large-language-models-still-lack-general-reasoning-skills

Son söz: Yapay zekânın en büyüleyici tarafı, yalnızca makineler hakkında değil, insanın nasıl gördüğü, öğrendiği ve anlam kurduğu hakkında da bizi yeniden düşündürmesidir.

Görseller: Bu PDF için özgün olarak yapay zekâ ile üretildi. Tasarım ve metin: bağımsız özet.