

Orijinal Kaynak: Zittrain, Jonathan. (2019, July 23). “[The Hidden Costs of Automated Thinking](#)”, *The New Yorker*.

Atıf Şekli: Zittrain, Jonathan. (2021, Şubat 13). “Otomatik Düşünmenin Gizli Bedelleri”, Çev. Ezgim Özgül, *Sosyal Bilimler*, Link: <https://www.sosyalbilimler.org/otomatik-dusunme-gizli-bedel>

Otomatik Düşünmenin Gizli Bedelleri

Jonathan Zittrain
Türkçesi: Ezgim Özgül

Birçok ilaç gibi, Provigil adı altında satışa sunulan uyanık kalma ilacı modafinil de küçük ve sıkıca katlanmış bir kâğıt broşürle birlikte gelir. Broşürün, ilacın kullanım şekli, dikkat edilmesi gerekenler ve ilacın moleküler yapısının bir diyagramına sahip içeriği çoğunlukla zararsız ve sıkıcıdır. Bununla birlikte, “Çalışma Prensibi” başlıklı kısım, kendi başına uykusuzluğa neden olabilecek **bir cümle içerir**: Modafinilin uyanıklığa yol açan mekanizma veya mekanizmaları bilinmemektedir.”

Provigil’in bu gizemli hikâyesi sadece ona özgü değildir. Pek çok ilaç, kimse henüz tam olarak nasıl çalıştığını bilmezken yasal onay alır ve geniş kitlelere reçete edilir. Bu gizem, genellikle deneme yanılma yoluyla ilerleyen ilaç geliştirme sürecinin bir parçasıdır. Her yıl, hücre kültürlerinde veya hayvanlarda pek çok sayıda yeni ilaç test edilir; bunların en iyi ve en güvenilir olanları insanlarda denenir. Bazı durumlarda bir ilacın başarısı, etki mekanizmalarının nasıl işlediğini açıklamayı hedefleyen yeni araştırmalara ilham verir ama her zaman böyle olmaz. Örneğin, Aspirin 1897’de keşfedildi ve **1995’e kadar** kimse nasıl çalıştığını tam olarak açıklamayı başaramadı. Benzer durumlar tıbbın başka alanlarında da görülür. **Derin beyin stimülasyonu**, parkinson veya benzeri hastalıkların yol açtığı spesifik hareket bozukluklarından mustarip kişilerin beyinlerine yerleştirilen elektrotlarla onları tedavi etmeyi amaçlar. Yirmi yıldan fazla bir süredir yaygın olarak kullanılan ve bazı bilim insanlarının, bilişsel becerileri güçlendirme dahil olmak üzere başka amaçlar için de kullanılabileceğini söyledikleri bu tedavi yönteminin nasıl çalıştığını kimse tam olarak bilmiyor.

Açıklamaya sahip olan çözümler ile karakterize olan bu keşif yaklaşımı benim entelektüel borç (*intellectual debt*) dediğim şeyi ortaya çıkartıyor. Bir şeyin işe yaradığını onun tam olarak neden işe yaradığını anlamadan bilmek mümkündür. Bu işe yarayışı hemen kullanıma sunmak ve işe yarayış mekanizmalarının daha sonra ortaya çıkacağını varsaymak mümkündür. Bazı durumlarda, bu entelektüel borcu hızlı bir şekilde öderiz.

Ancak diğler durumlarda, onlarca yıl tam olarak açıklanamayan bilgilere güvenerek hareket ederiz.

Geçmişte entelektüel borç, keşiflerinin deneme yanılma yöntemiyle yapıldığı tıp ve birkaç alanla sınırlıydı. Ancak yapay zekâ alanındaki yeni teknikler —özellikle de yapay öğrenme (*machine learning*)— geliştikçe kolektif entelektüel borcumuz hızla artışa geçebilir. Yapay öğrenme sistemlerinin çalışma prensibi, veri okyanuslarındaki kalıpları belirlemeye dayanır. Bu kalıpları kullanarak, belirsiz ve açık uçlu sorulara yanıt verirler. Yapay sinir ağlarına kedilerin ve kedi olmayan nesnelere resimlerini beslerseniz kedileri ayırt etmeyi öğrenirken, tıbbi kayıtlara erişim imkânı sağlarsanız hastanedeki yeni bir hastanın ölme olasılığını tahmin etmeye çalışabilir. Buna rağmen çoğu yapay öğrenme sistemi nedensel mekanizmaları ortaya çıkarmaz. Onlar istatistiksel korelasyon motorlarıdır. Neden bazı hastaların ölme olasılığının daha yüksek olduğunu düşündüklerini açıklayamazlar çünkü aslında bizim anladığımız biçimde “düşünmezler”, sadece cevap verirler. Yapay zekanın iç görülerini yaşamlarımıza entegre etmeye başladığımızda, toplum olarak her geçen gün daha fazla entelektüel borç biriktirmeye başlayacağız.

İlaçlarda teoriden bağımsız gerçekleşen yenilikler, bazı durumlarda entelektüel borcun vazgeçilmez olabileceğini gösteriyor. Milyonlarca hayat, temelde anlamadığımız mekanizmalar ile kurtarıldı. Çok az kişi hayat kurtaran bir ilacı veya en basitinden aspirini, kullanmayı kimse nasıl çalıştığını bilmiyor diyerek reddeder. Ancak entelektüel borcun birikmesinin dezavantajları var. Bilinmeyen etki mekanizmalarına sahip ilaçlar çoğaldıkça, istenmeyen ilaç etkileşimlerini ortaya çıkarmak için gereken test sayısı katlanarak artacaktır —ilaçların çalışma mekanizmaları anlaşıldığında, istenmeyen etkileşimler önceden tahmin edilebilirdir. Pratikte, etkileşimler yeni ilaçlar piyasaya sürüldükten sonra keşfedilir, bu da ilaçların piyasaya sürüldüğü, ardından dava edildiği ve en sonunda piyasadan çekildiği döngüye katkıda bulunur. Her bir kişi için teker teker bakıldığında, yeni bir ilaçla ilişkili entelektüel borç göz ardı edilebilir görünür. Ancak entelektüel borçlar tek başlarına kalmaz. Farklı alanlarda kullanılan ve bir teoriye sahip olmayan çözümler, birbirini tahmin edilemeyen şekillerde karmaşıktırabilir.

Yapay öğrenme nedeniyle bize yüklenen bu entelektüel borç, eski tarz deneme yanılma yoluyla gelen risklerin de ötesinde riskler içerir. Çoğu yapay öğrenme modeli sahip oldukları iç görüleri ilişkin nedenler sunamadığından, sağladıkları yanıtlar hakkında önceden bağımsız bir yargıya sahip değilsek, ne zaman yanıldıklarını anlamamızın bir yolu yoktur. İyi eğitilmiş bir sistemde yanlış sonuçlar oldukça nadir görünür. Ancak sistemi ne tür verilerle beslemesi gerektiğini bilen biri kasıtlı olarak sistemi kendi avantajları için kullanabilir.

Görüntü tanıma teknolojisini ele alalım; on yıl önce bilgisayarlar fotoğraflardaki nesnelere günümüzdeki gibi kolayca tanımlayamıyordu. Günümüzde görsel arama motorları, günlük olarak etkileşim kurduğumuz pek çok sistem gibi, olağanüstü derecede

yetenekli yapay öğrenme modelleri ile çalışmaktadır. Google'ın görsel aramaları, **Inception** adlı bir yapay sinir ağını kullanır. 2017'de, MIT **LabSix** isimli lisans ve lisansüstü öğrencilerinden oluşan bir araştırma grubu, bir kedi fotoğrafının piksellerini **değiştirerek** insan gözüne kedi gibi görünse de Inception'ın %99,99 kesinlikle guacamole¹ olduğundan emin olduğu (fotoğraf çok düşük bir ihtimalle brokoli veya harca aitti) bir fotoğrafı Inception'a besledi. Elbette Inception, fotoğrafın hangi özelliklerin onu kedinin kedi olduğunu kararına götürdüğünü açıklayamaz; bunun sonucunda, özel hazırlanmış veya bozuk veriler ile beslendiğinde sistemin ne şekilde başarısız olabileceğini tahmin etmenin kolay bir yolu yoktur. Böyle bir sistemin, akıllı ve kararlı bir saldırganın elinde güvenlik açıklığına dönüşecek olan boşluklara sahip olması muhtemeldir.

Yapay öğrenme sistemleri tarafından üretilen bilgiler kullanıma sunuldukça, sistemin doğası gereği bu tür hatalar olarak ortaya çıkabilir. Tıp için kullanılan yapay zekâlar, deri lezyonlarını iyi huylu veya kötü huylu olarak sınıflandırabilmek için **başarıyla eğitilmiştir**. Yine de —Harvard Tıp Fakültesi ve Massachusetts Teknoloji Enstitüsü'den bir araştırma ekibinin 2019'da yayımladıkları bir **makaleye göre**, sistemlerin kedileri guacamole ile karıştırmalarına benzer yolla, yanlış kararlar vermelerine sebep olacak şekilde kandırılacaklarını gösterdi —diğer tüm sorunların yanı sıra, saldırganlar bu güvenlik açıklarını **sigorta dolandırıcılığı** için kullanabilir. Bu tür sistemlerin öngörü gücünden etkilenerek insan yargılarını ikinci plana atabiliriz. Ancak güvenlik açıkları kaçınılmaz olmaya devam edecek ve üretilen yanıtları doğrulamak kolay bir süreç olamayacak.

Teorisiz bilginin nerede ve nasıl kullanıldığını takip etmek için bir sistem geliştirip, entelektüel borç için bir bilanço tablosu oluşturabilir miyiz? Muhasebemiz, tüm entelektüel borçların eşit derecede sorun çıkarmadığı gerçeğini gösterebilir. Bir yapay zekâ, **yeni pizza tarifleri** üretiyorsa, buna karşı çıkmayıp pizzanın tadını çıkarmak mantıklı olabilir. Fakat yapay zekâyı sağlıkla ilgili tahminlerde ve önerilerde bulunmak için kullanmaya başladığımızda, konuya tamamıyla hâkim olmak daha sağlıklı olabilir.

Küresel bir entelektüel borç bilanço tablosu oluşturmak, muhtemelen ticari sırlar ve diğer fikri mülkiyetlere yönelik yaklaşımımızı yeniden gözden geçirmeyi gerektirecektir. Kent kurallarına göre, bina sahipleri bina yenileme planlarını kamuya açıklamalıdır. Benzer şekilde, kütüphanelerden veya üniversitelerden, belirli bir kamusal kullanıma sahip gizli verileri ve algoritmaları kamu ile paylaşmalarını talep edebiliriz. Bu, araştırmacıların kullandığımız yapay zekâ modellerini ve temel verilerini inceleyebilmesine ve teoriler oluşturarak entelektüel borcumuzu hatalar ve güvenlik açıkları yaratılmadan ödememize yardımcı olacaktır.

¹ Guacamole, avokado ile yapılan bir Orta Amerika mezesidir. Aztek kökenli olan bu yiyecek, içeriğindeki yağ ve vitamin açısından değerli tutulmuştur. Genel olarak avokadonun ezilmesiyle ve domates, tuz eklenmesiyle yapılır. **Kaynak:** Wikipedia, —YHN.

Yapay öğrenme modellerinin yaygınlığının artması ve herkesin yapay öğrenme modeli dizayn edebileceği gerçeği, muhasebe sürecini zorlaştıracaktır. Ama bu hayati önem taşıyan bir mesele. Tek başına bakıldığında, gelecek tahmini içeren yanıtlar bize yardımcı olan sonuçlar üretebilir. Ancak bu sistemler tek başlarına kalmayacaklar: Yapay zekâlar dünyadaki verilerini topladıkça kendilerine ait yeni veriler üretecekler ve bu verilerin çoğu daha sonra başka sistemler tarafından kullanılacak. Bilinmeyen etki mekanizmalarına sahip ilaçların bazen birbirleri ile etkileşime girmesi gibi, entelektüel borç yüklü algoritmalar da birbirleriyle etkileşime girecektir.

En basit etkileşimler bile sorunlara yol açabilir. 2011 yılında Michael Eisen adlı bir biyolog, bir öğrencisinden, dikkate değer hiçbir özelliği olmayan *The Making of a Fly: The Genetics of Animal Design* isimli ikinci el bir kitabın en ucuz kopyasının Amazon'da \$1,7 milyona satıldığını öğrendi. \$1,7 milyon artı \$3,99 kargo ücreti. En ucuz ikinci kopya ise \$2,1 milyona satılıyordu. İki satıcının sayfaları da binlerce olumlu yorum ile doluydu. Eisen birkaç gün üst üste kitabın Amazon sayfasını ziyaret ettiğinde, fiyatların düzenli bir şekilde arttığını keşfetti. Satıcı A'nın kitabının fiyatı tutarlı bir şekilde Satıcı B'nin kitabının fiyatının %99,83'üydü. Satıcı B ise her gün kitabının fiyatını, Satıcı A'nın fiyatının %127,059'una çıkarıyordu. Eisen, Satıcı A'nın kitabın iki kopyasına sahip olduğunu ve ikinci en ucuz fiyatın altında kalmaya çalıştığını **tahmin etti**. Bu arada Satıcı B tek bir kopyaya sahipti ve bu yüzden kitabı daha yüksek fiyatlandırdı; eğer birisi kitabı satın alacak olursa, Satıcı B o müşterinin adına kitabı Satıcı A'dan sipariş edecekti.

Her satıcının stratejisi kendi içinde mantıklıydı. Mantıksız sonuçlar üreten, algoritmalarının etkileşimiymi. Vahşi doğada binlerce yapay öğrenme sistemlerinin birbirleri ile etkileşimi çok daha öngörülemez olacağı benziyor. Son teknoloji yapay öğrenme sistemlerinin hâlihazırda kullanıldığı finans piyasaları, bu tür problemler için çok uygun bir alanı da beraberinde getiriyor. 2010 yılında, algoritma ticaretinin neden olduğu otuz altı dakika süren bir “ani çöküş”, büyük ABD endekslerinden bir trilyon dolardan fazlasını sildi. 2018 sonbaharında J. P. Morgan analisti Marko Kolanovic, her zamankinden daha fazla ticaret otomatize edildiği için böyle bir çöküşün kolayca tekrarlanabileceğini **savundu**. Entelektüel borç, sistemlerin doğrudan birbirleriyle bağlantılı olmasa da bile birbirlerine dokunduğu boşluklarda birikebilir. Bir entelektüel borç bilançosu olmadan, herhangi bir entelektüel borcun üstlenmeye değer olup olmadığını —önceden veya geriye dönük olarak— belirlemenin bir yolu yoktur.

Entelektüel borcumuzdaki artış, düşünme şeklimizde de bir değişimi beraberinde getirebilir, temel bilimlerden uzaklaşıp uygulamalı teknolojiye yönelebiliriz. Zengin hükümetler tarafından desteklenen ve akademik araştırma kurumları tarafından yürütülen devasa sermayelere sahip, temel bilim araştırmalarının (örneğin parçacık hızlandırıcılar) aksine, yapay öğrenme araçları akademi tarafından olduğu kadar özel sektör tarafından da kolayca kullanılabilir. Aslında, yararlı tahminler üretmek için kullanılacak veriler, bir bilgisayar bilimi enstitüsü veya istatistik departmanından daha

yüksek ihtimalle Google veya Facebook'un kullanımındadır. İş adamları bu tür açıklanamayan bilgilerden bütünüyle memnun olabilirler, ancak entelektüel borç hâlâ artmaya devam ederken, bu veriler borcu ödemeye en çok ilgilenen akademik araştırmacılardan uzakta özel şirketlerin ellerinde olacaktır.

Yapay öğrenmeye dayalı bilginin finansmanın, bu teknolojinin çalışma mekanizmalarını anlamak için uzun süreli araştırmalar yapmak konusunda ısrarcı olan araştırmacılardan uzaklaştırılacağını tahmin etmek zor değildir. 2018'in aralık ayında, protein katlanması üzerine çalışan bir araştırmacı olan Mohammed AlQuraishi, kendi alanındaki yeni bir gelişmeyle ilgili bir **makale yazdı**. Makalenin konusu protein kıvrımlarını insan araştırmacılardan çok daha doğru bir şekilde tahmin edebilen bir makine öğrenimi modeliydi. AlQuraishi, makalede teorinin, veri karşısındaki güç kaybı ile ilgili sıkıntılarını dile getirirken bir taraftan da bu yeni gerçekle yüzleşmeye çalışıyordu. **Bir röportajında** "Kavramsal makaleler ya da yeni analitik iç görüler hakkındaki yazılar artık hiç prestijli değil" dedi. Makine keşifleri hızlandırdıkça, insanlar teorisyenleri zamanın gerisinde kalmış ve gereksiz kişiler olarak görmeye başlayacaklardır. Belirli bir alan hakkındaki teorik bilgiler, o konuda pratik yanıtlar üreten yapay öğrenme modellerinin oluşturulmasındaki uzmanlıktan daha az değerli olacaktır.

Finansal borç, kontrolü borçludan borç verene, gelecekte geçmişe doğru kaydırır. Entelektüel borcun artması da buna benzer biçimde kontrolün kimin elinde olduğunu değiştirebilir. Teorinin var olmadığı bir bilgi dünyası, sebep ve sonucun olmadığı bir dünyadır; bu noktada bize ne yapmamız ve ne zaman yapmamız gerektiğini söylemeleri için dijital danışmanlarımıza bağımlı hâle gelebiliriz. Örneğin, bir üniversiteye kabul komitesinin, zahmetli ve karmaşık eleme sürecini bir yapay öğrenme modeline çevirebileceğini hayal etmek zor değildir; böyle bir model adayları sadece akademik başarılarına göre değil, uyumlu ilişkiler kurabilme yetilerine ve cömert mezun bağışlarına göre de değerlendirecektir. Böyle bir dünyada var olabilmenin tek yolu kendi yapay zekâlarımızı kullanmaktan geçiyor olabilir: Biz de başvurduğumuz okulun beklentilerine göre sosyal medya profillerimizin en ince detayına kadar düzenleyecek yapay sinir ağlarından destek alırız.

Belki de tüm bu teknoloji gerçekten işe yarayacak ve işte bu, bir sorun hâline gelecek. Yapay zekâyâ yönelik haklı eleştirilerin çoğu nelerin yanlış gidebileceğine odaklanmıştır: önyargılar yaratabilir veya varolan önyargıları kopyalayabilir, hatalar yapabilir, kötü amaçlar için programlanabilir. Bunların yanı sıra her şey doğru gitse bile neler olabileceği ile ilgili endişelenmeliyiz.

Kayda Değer Akademik Metinler mottosuyla, 10 Ağustos 2015 tarihinde yayın hayatına başlayan *sosyalbilimler.org*, sosyal bilimler meselelerine yoğunlaşan, gönüllülük odaklı, açık erişim, akademik bir web sitesidir. Hakkında detaylı bilgi almak için sosyalbilimler.org/hakkinda sayfasını, ekibimizde gönüllü olarak görev almak için sosyalbilimler.org/basvuru sayfasını ziyaret edebilirsiniz.

Facebook, Twitter, Instagram ve YouTube'da **@sosbilorg** kullanıcı adıyla *Sosyal Bilimler*'i takip edebilirsiniz.

sosyalbilimler.org/abonelik sayfasından e-bülten abonesi olarak, her pazar günü, o hafta içinde *sosyalbilimler.org*'da yayımlanan çalışmaların tamamını size gönderilecek bir e-posta ile alabilirsiniz.

sosyalbilimler.org'da yayımlanan metin, video ve podcastlerin paylaşıldığı Telegram grubuna t.me/sosbilorg adresinden katılabilirsiniz.