

Büyük Veri Üzerine Eleştirel Sorular: Kültürel, Teknolojik ve Bilimsel Bir Olgu Hakkında Eleştirel Sorgulamalar¹

danah boyd ve Kate Crawford²

Çeviren: Bilgesu Savcı³

Teknoloji ne iyi, ne kötü ne de nötrdür... teknolojinin toplumsal evrenle öyle bir ilişkisi vardır ki, teknik gelişmeler kendi araç ve etkinliklerinin varoluş amaçlarının çok daha ötesine geçen çevresel, toplumsal ve insani sonuçlara neden olur (Kranzberg 1986, s.545).

Şu anda var olmayan bir söyleme ihtiyacımız vardır: Hem en esnek tasarıma hem de olası çoksesseliliğe ve çok süremliliğe izin veren bir bakış açısından, veri tabanlarımızda görünür kılabileceğimiz, çeşitlilik gösteren zaman, uzam ve maddeselliklerle ilgili bir söylem. İşlenmiş veri anlayışı yalnızca bir tezat değil, aynı zamanda kötü bir fikirdir; onun yerine, veri özenle işlenmelidir (Bowker 2005, ss. 183- 184).

Büyük Veri çağı yoldadır. Bilgisayar uzmanları, fizikçiler, ekonomistler, matematikçiler, siyaset bilimcileri, biyo-bilişimciler, sosyologlar ve diğer bilginler, yalnızca insanlar, şeyler ve bunların etkileşimleri tarafından ve haklarında üretilen devasa miktardaki bilgilere ulaşmak için çırpınıp durmaktadırlar. Çeşitli gruplar türeyimsel zincirleri, sosyal medya etkileşimlerini, sağlık kayıtlarını, telefon günlüklerini, devlet kayıtlarını ve insanların diğer tüm dijital izlerini incelemenin muhtemel faydaları ve getirileri

¹ Critical questions for Big Data”adlı bu yazı daha önce *Information, Communication & Society* Sayı. 15, No. 5, Haziran 2012’de ss. 662 –679’de yayınlanmıştır. Yayınevinin izni ile çevirisi yayınlanmıştır.

² **danah boyd**, Uzm. Araştırmacı, Microsoft Research; Yrd. Doç., New York Üniversitesi; Üye, Harvard’s Berkman Center for Internet & Society. Çalışmalarında özellikle gençlerin sosyo-teknik pratiklerini konu alarak, insanların sosyal medyayı günlük yaşamlarına nasıl dâhil ettikleri üzerine odaklanmaktadır. **Kate Crawford**, Doç. New South Wales Üniversitesi; Yönetici Araştırmacı, Microsoft Research New England. Cep telefonu ve sosyal medya kullanımı hakkında geniş yelpazede çalışmalar yapmakta ve sosyal medyanın kültürel ve siyasal bağlamları konularında yayımlar üretmektedir.

³ Hacettepe Üniversitesi SBE. İletişim Bilimleri Bölümü yüksek lisans öğrencisidir.

hakkında tartışmalar yürütmektedirler. Böylece ortaya muazzam sorular çıkar: Büyük ölçekli araştırma verileri daha iyi araçlar, servisler ve kamu yararı oluşturmamıza yardımcı olacak mıdır, yoksa yeni bir gizlilik ihlali ve istilacı pazarlama dalgasının önünü mü açılacaktır? Veri analizi çevrimiçi toplulukları ve siyasal hareketleri anlamamıza yardım mı edecektir, yoksa protestocu izlerini takip etmek ve sesleri susturmak için mi kullanılacaktır? İnsan iletişimi ve kültürü üzerine çalışma disiplinlerimizi mi dönüştürecek, yoksa araştırma seçeneklerimizi daraltıp “araştırma” sözcüğünden ne anladığımızı mı değiştirecek? Büyük Veri pek çok açıdan yetersiz bir terimdir. Manovich’in (2011) gözlediği üzere, Büyük Veri, bilimsel olarak süper bilgisayarları gerekli kılacak ölçüde büyük boyutlardaki veri kümelerini temsil etmede kullanılmaktadır, fakat artık böyle makinelere başvurmadan standart yazılım programlarıyla masa üstü bilgisayarlarda da incelemeler yapılabilir. Neredeyse kesin olarak şu an ulaşılabilir veri boyutu oldukça geniştir. Ne var ki, bu durum sözü edilen yeni veri evreninin tanımlayıcı bir özelliği değildir. Aslında Büyük Verinin içerdiği verilerin birazı (örn. belirli bir başlık altında toplanan tüm Twitter mesajları), daha önce Büyük Veri olarak düşünülmemiş veri kümeleriyle (örn. nüfus sayımı verileri), öyle veya böyle aynı genişlikte değildir. Büyük Veri dediğimiz şey, verinin kendisi ile ilgili olmaktan çok, büyük veri kümelerini arama, birleştirme ve göndermenin kapasitesi ile ilgilidir.

Bizler Büyük Veriyi¹ aşağıda yer alan unsurların, karşılıklı etkileşim temeline dayanan kültürel, teknolojik ve bilimsel bir olgu olarak tanımlamaktayız:

- (1) *Teknoloji*: geniş veri kümelerini toplama, analiz etme, link verme ve karşılaştırma için hesaplama gücü ve algoritmik kesinliği en üst seviyeye çıkarmak.
- (2) *Analiz*: ekonomik, toplumsal, teknik ve yasal iddialarda bulunabilme için kalıpları tanımlamaya yönelik geniş ölçekli veri kümelerini kullanmak.
- (3) *Mitoloji*: geniş veri kümelerinin gerçeklik, nesnellik ve kesinlik açısından birlikte, daha önceleri mümkün olmayan kavrayışlar üretebilecek bilgi ve anlayışı yüksek formlar içinde sunduğuna dair yaygın görüşün varlığı.

Diğer sosyo-tekniksel olgular gibi Büyük Veri de hem ütöpik hem de distopik retoriği tetiklemektedir. Büyük Veri, bir taraftan kanser araştırmalarından terörizm ve iklim değişikliğine kadar çok çeşitli alanlarda yeni kavrayış olasılıkları sunarak türlü toplumsal sorunlara gönderme yapan güçlü bir araç olarak görülmektedir. Diğer taraftan ise, mahremiyetin istilası ve azalan sivil özgürlüğü ile artan devlet ve kurumsal kontrolü olanaklı kılan Büyük Birader’in sorunlu bir alameti olarak algılanmaktadır. Bu açıdan, tüm sosyo-tekniksel olgularda görüldüğü gibi umut ve korku akımları, henüz yolda olan inçelikli ve derin bir bakış gerektiren değişimleri karartmaktadır.

Bilgisayarlı veri tabanları yeni değildir. Amerikan Nüfus Sayımı İdaresi 1890’da dünyanın ilk otomatik işlem donanımını – delikli kart makinesi – geliştirmiş (Anderson 1988) ve 1960’larda ilişkisel veri tabanları ortaya çıkmıştır (Fry & Sibley 1974). Kişisel bilgi işlem ve İnternet, -- bilim insanlarını, pazarlamacıları, devlet ve eğitim kurumlarını ve harekete geçirilmiş bireyleri de içeren -- daha fazla insana veri üretme, paylaşma,

etkileşime geçme ve düzenleme olanağı sunar haldedir. Bu ise Savage ve Burrows'un (2007) ampirik sosyolojide kriz diye tanımladıkları şeye neden olmaktadır. Bir zamanlar belirsiz ve yönetmesi güç olan – ve böylece yalnızca sosyal bilimcilerin ilgilendikleri -- veri kümeleri şimdi bir araya toplanıp, öğrenim alanları ne olursa olsun merak eden herkes için kolayca ulaşılabilir hale getirilmektedir.

Büyük Veri çağının doğuşunu nasıl ele aldığımız son derece önemlidir. Her ne kadar bu olgu bir belirsizlik ve hızlı değişim evreninde yer alıyor olsa da, güncel kararlar geleceği şekillendirecektir. Veri toplama ve analiz etmedeki artan otomasyonla birlikte – tıpkı geniş ölçekli insan davranışı kalıplarını alarak resmedebilen algoritmalar gibi—bu etkinliklerin kimler tarafından yürütülüp düzenlendiğini sormak gerekir. Lessig (1999) toplumsal sistemlerin dört güç tarafından düzenlendiğini belirtir: Pazar, yasa, toplumsal normlar ve mimari – veya teknoloji konusunda, kod. Büyük Veriye geldiğimizde ise bu durumun karşıtlıklar içerdiğini görürüz. Pazar için Büyük Veri el değmemiş bir fırsattır: Pazarlamacılar reklam amacıyla, sigorta uzmanları hizmetlerini en iyi duruma getirmek için, Wall Street bankacıları ise piyasayı takip edebilmekte kullanırlar Büyük Veriyi. Çoktandır yasama, genellikle mahremiyetle ilgili konularda, verinin biriktirilmesi ve saklanması kontrolünü tekelsel bir biçimde kendisi düzenlenmektedir (örn., ABD'nin Do Not Track Online Act hareketi). Şahsileştirme gibi özellikler daha fazla ilgili bilgiye hızlı bir ulaşımına izin verirken, aynı zamanda zor etik sorular oluşturmakta ve sorunlu yollarla toplumu ayırıştırılmaktadır (Pariser 2011).

Büyük Veriyi içeren oldukça değerli ve kavrayıcı nitelikte güncel çalışmalar bulunmaktadır. Fakat hala tüm bu verilerin ne anlamının olduğu, kimlerin hangi verilere ulaştığı ve veri analizinin nasıl gelişip ne işe yaradığı ile ilgili eleştirel soruların sorulması gereklidir. Bu makalede, Büyük Veri ile ilgili konular hakkındaki tartışmaları tetiklemek için altı farklı itici güç öneriyoruz. Bizler bilgisayar ve bilişim uzmanları ile düzenli bir iletişim halinde olan ve sosyal bilimlerle ilgili medya çalışmaları içinde bulunan bilim insanlarıyız. Sosyal bilimcilere aşikâr gelebilirken, farklı disiplinlerden kimseler için genellikle şaşırtıcı olabilecek çeşitli tehlikeleri tanımlıyor olmamıza rağmen, sorduğumuz soruların cevaplarını bulmak kolay değildir. İlgimiz ve tecrübelerimiz sosyal medya alanında olduğu için buradaki temel meselemiz de çoğunlukla sosyal medya bağlamında Büyük Veridir. Bununla beraber, sormakta olduğumuz sorular diğer disiplinlerde olanlar için de önemlidir. Aynı zamanda sorduğumuz sorular yalnızca bir başlangıç olduğunun da farkındayız ve umuyoruz ki bu makale Büyük Verinin içinde gömülü olan varsayımları sorgulamak için diğerlerini de tetikleyecektir. Büyük Veri kültürü içinde, tüm alanlardaki araştırmacıların – buna bilgisayar bilimi, işletme ve tıp da dâhil – özellikle de onun çeşitli alanlardaki genişleyen etki ve olasılığı açısından bir payları vardır. İşte bu olguyu, ilgili varsayımları ve önyargıları eleştirel olarak sorgulamanın tam zamanı olduğuna inanıyoruz.

1. Büyük Veri bilginin tanımını değiştirir

Yirminci yüzyılın ilk yıllarında, Henry Ford özel düzenekler ve standardize edilmiş ürünler kullanarak bir seri üretim imalatı sistemi tasarlamıştı. Bu sistem kısa zamanda

teknolojik gelişimin geçerli bir yöntemi haline geldi. “Fordizm”, otomasyon ve montaj üretim hattı demekti; bu durum, devam eden süreçlerde, yani nitelikli zanaatkarlar ile yavaş işin ötelendiği yeni makine-işi çağında, imalatın ortodoksisi haline geldi (Baca 2004). Fakat söz konusu olan yalnızca birkaç aletin değişmesi değildi. Fordizimin, yirminci yüzyıla vurduğu damga hücrenel bir seviyedeydi: Çalışma hayatında yeni bir işgücü ve insan ilişkileri anlayışı açısından da geniş perspektifte yeni bir toplum anlayışı üretmişti.

Büyük Veri yalnızca geniş veri kümeleri, araçlar ve bu veri kümelerinin kullanılış ile inceleniş süreçleri değil, aynı zamanda düşünce ve araştırmada berimsel (bilgisayımusal) eğilim demektir (Burkholder 1992). Nasıl ki Ford araba yapmanın yollarını değiştirdiyse -- ardından da iş’in kendini dönüşüme uğrattıysa -- Büyük Veri de hem bilginin nesnelere hâlihazırda değiştirmekte olan bir bilgi sistemi oluşturdu, hem de insan ve toplum ağlarını anlayış biçimlerimizin bilgisini sunma gücüne ulaşıyor oldu. ‘Araçları değiştirin, böylece bu araçlarla yürüyen tüm sosyal teoriyi de değiştirmiş olacaksınız’ diye hatırlatır Latour (2009: s.9).

Büyük Veri, araştırmadan ne anladığımız konusunda radikal bir dönüşüm yarattı. Berimsel sosyal bilimlere yorumlayan Lazer ve diğerleri (2009) Büyük Verinin ‘eşi görülmemiş uzaklıktaki alanda, derinlikte ve ölçekteki veriyi toplama ve analiz etme’ (s.722) imkânı sunduğunu belirtirler. Söz konusu yalnızca ölçek olmadığı gibi onu yakınlığın sınırları içinde düşünmek de yeterli olmayacaktır. Bu, Moretti’nin (2007) metinlerin uzak ve yakından analizi dediği şey de değildir. Aksine, epistemoloji ve etik düzeyinde temel bir değişimdir. Büyük Veri bilginin inşası, araştırma süreci ve enformasyonla, gerçekliğin doğası ve sınıflandırmasıyla nasıl ilişkilendiğimiz hakkında kilit soruları yeniden çerçevlendirmektedir. Tıpkı Du Gay ve Pryke’nin (2002) belirttiği gibi ‘araçların hesabını tutma... basitçe ekonomik faaliyetin ölçülmesine yardımcı olmakla kalmaz, ölçtüğü gerçekliği biçimlendirir de’ (ss. 12-13). Bu nedenle de Büyük Veri nesnelere, bilme biçimlerinin ve toplumsal yaşam tanımlarının topraklarını da yeniden belirler.

Wired’ın baş editörü Anderson, “Perabayt Çağı” olarak terimselleştirdiği noktayı düşünerek şöyle der:

Bu dünya devasa boyutlardaki verinin ve uygulamalı matematiğin, uygulamaya konulabilecek her bir diğer aracın yerini aldığı bir mecradır: Dilbilimden sosyolojiye insan davranışı teorilerinin hepsini dışında bir yerde. Taksonomiye, ontolojiye ve psikolojiye unuttun. İnsanların bir şeyi neden yaptıklarını kim bilir? Mesele bunu yapmış olmalarıdır ve bizler bu eşi benzeri görülmemiş, sınırsız alanın izini sürebilir ve onu ölçebiliriz. Yeteri kadar veriyle, numaralar kendi hikâyelerini anlatırlar (2008).

Numaralar kendi hikâyelerini mi anlatırlar? Cevabın ‘hayır’ olduğuna inanıyoruz. İlginç bir şekilde Anderson’ın yeterince fark gözetmeden tüm teori ve disiplinleri ötelmesi yalnızca bir etkidir: Diğer analiz biçimlerinin oldukça kolayca tali bir hale getirildiği pek çok Büyük Veri tartışmasındaki gizli kibirli eğilimleri su yüzüne çıkarır. İnsanların neden bir şeyler yaptığını veya yazdığını sorgulayan diğer yöntemler, çok

sayıdaki numaranın arasında kaybolup gitmektedir. Bu alan, entelektüel becerinin daha önceki biçimlerine yer bırakmaz. Berry'nin (2011, s.8) belirttiği gibi Büyük Veri, “felsefenin düzenleyici gücünü bünyesine barındırmayan, karmaşık miktarlarda bilgi ve enformasyon” sağlar. Kant'ın tüm kurumların akılcı temeli olarak gördüğü felsefe yerine, artık ‘berimselliğin bir tür ontoteoloji, yeni bir tarihsel bildirişimin kümelenişi olması açısından yeni bir ontolojik “çığır”ın yaratımı olduğu anlaşılabilir’ (Berry 2011, s.12).

Yeni ortodoksluklar billurlaşmadan Büyük Verinin bildirişim modelleriyle ilgili zor sorular sormamız gerekmektedir. Ford'a dönecek olursak, onun yeniliği bileşik ve bütünsel görevleri parçalayarak, basit, otomize ve mekanik görevler haline getirmek için seri üretim hattını kullanmasıydı. Bunu ise tamamıyla önceden belirlenmiş ve işçi faaliyetini sınırlayan özel araçlar tasarlayarak yapmıştı. Benzer şekilde Büyük Verinin araçlarını da kendi sınır ve kısıtlamalarının kendileri tarafından sağlanacağı yönde düzenledi. Örneğin Twitter ve Facebook, oldukça düşük seviyede arşivleme ve arama fonksiyonları gösteren Büyük Veri kaynaklarına birer örnektirler. Bunun bir sonucu olarak ise eski verilere ulaşmanın zorluğu veya imkânsızlığından dolayı arama yapan kişiler şu ana veya yakın geçmişe – bir seçimin topladığı tepkileri ve TV finalini veya doğal bir felakete— odaklanmaya yatkınlık göstermektedirler.

Arama fonksiyonlarının belirli türlerindeki otomasyonu gözlüyorsak, makine araçlarının kendilerine monte edilmiş haldeki noksanlıklarını da düşünmeliyiz. Yalnızca Anderson'ın dediği gibi ‘bilim Google'dan ne öğrenebilir?’ diye sormak yeterli olmamakla birlikte sormamız gereken Büyük Verinin çarklarının öğrenmenin anlamını nasıl değiştirdiği ve bu bilme sistemleriyle ne tür yeni olasılıkların ve sınırların gelebileceğidir.

2. Nesnellik ve doğruluk iddiaları yanlış yönlendiricidir

Latour (2009), ‘numaralar, numaralar, numaralar’ der. ‘Sosyoloji nicel bir bilim olma amacına takıntılıdır. Latour'a göre toplumsal alan bilgisinde ölçülebilir olanla ölçülemez olan arasına çektiği çizgi nedeniyle de bunu hiçbir zaman başaramamıştır. Büyük Veri, beşeri bilimlere nicel bilimin ve nesnel yöntemin konumu teorisi için yeni bir yol sunmaktadır: Daha fazla toplumsal alanı ölçülebilir hale getirmiştir. Gerçekte, Büyük Veri ile çalışmak hala öznel ve sunduğu ölçüm –özellikle de sosyal medya sitelerinden gelen mesajları düşündüğümüzde, her zaman nesnel gerçeklik hakkında bir iddia sahibi olmayabilir. Fakat nitel araştırmacıların işi hikâye yorumlamak iken, nicel araştırmacıların işinin olgusal gerçekler üretmek olduğuna dair yanlış bir inanç hala sürmektedir. Böylece, Büyük Veri bilimsel yöntemle ilgili uzun zamandır süregelen tartışmalarda yer alan yerleşik ayrımların ve sosyal bilimler ile beşeri incelemelerin meşruiyetinin, yeniden içselleştirilmesini tehlikeye atar.

Nesnellik sorusu, hem bilim felsefesinde hem de bilimsel yöntemin ilk dönem tartışmalarında oldukça merkezi bir soru olagelmıştır (Durkheim 1895). Nesnellik iddiaları nesnel dünyasına, kendi içinde kendisi için var olan şeylere bağlılık önerisi sunmaktadır. Öte yandan öznel, kuşku ile yaklaşılıp, pek çok farklı bireysel ve toplumsal koşulla çeşitlilik gösteren renklerle bezeliymiş gibi algılanır. Bilimsel yöntem kendini öznel alandan hipotezlerin önerilip test edildiği ve nihai olarak bilgide ilerleme sağlayan

tarafsız bir sürecin uygulandığı alana doğru kaydırır. Bununla birlikte, nesnellik iddialarında bulunanlar da öznelardir ve bu iddialar öznel gözlemlere ve seçimlere dayanır.

Tüm arařtırmacılar veri yorumcularıdır. Gitelman'ın (2011) gözlemlendiđi gibi, ilk olarak verinin veri olarak hayal edilmesi gerekir ve verinin bu tasarım süreci yorumamacı bir safhayı içinde barındırır: 'Her disiplin ve disiplinvari kurumun veri tasarımında kendine özgü standartları ve normları vardır'. Bilgisayar bilimciler, sosyal bilimlerin eylemleri içinde yer almaya başlamış olsalar da yaptıkları işin yorumlama değil, gerçek olgular yaratmak olduğunu iddia etmeye dair bir eğilim vardır. Bir model matematiksel olabilir, bir deney geçerli görülebilir fakat arařtırmacı bunun ne anlama geldiđini sormaya başladığı anda yorumlama süreci de başlamış demektir. Bu, tüm yorumların aynı şekilde oluşturulduklarını söylemek olmadığı gibi tüm sayıların nötr olduğunu söylemek de değildir.

Neyin ölçülmesi gerektiđini tayin edecek olan tasarım seçenekleri de yine yorumlamaktan gelmektedir. Örneđin, sosyal medya verileri alanında, bir 'veri temizleme' süreci bulunur ki bu hangi kalıpların ve deđişkenlerin sayılacağı ve hangilerinin dikkate alınmayacağına karar vermektir. Bu süreç de içkin olarak öznelardir. Bollier'in açıkladığı gibi,

Büyük bir yığın halindeki işlenmemiş enformasyon, Büyük Veri, kendi kendini açıklamaz. Fakat veriyi anlamlı kılmanın özel yöntemleri, her tür felsefi tartışmaya açıktır. Veri 'nesnel bir gerçeđin' temsili olabilir mi? Diğer taraftan her yorum ya da verilerin 'temizlenmesi' işlemi illa ki öznel filtrelemeyle yanlı bir hale mi gelecektir? (2010, s. 13)

Bu sorulara ek olarak, bir de veri hataları konusu vardır. İnternet kaynaklarından gelen geniş veri kümeleri genellikle hem güvenilmez hem de kesinti ve kayıplara yatkındır. Bu tarz hata ve boşluklar, bir de çoklu veri kümeleri bir arada kullanıldığı anda daha da artmaktadır. Sosyal bilimciler verinin toplanması ve verilerindeki sapmaları hesaplamaya çalışmakla ilgili uzunca bir eleştirel sorgulama tarihine sahiptirler (Cain & Finch 1981; Clifford & Marcus 1986). Bu durum, veri kümelerinin boyutuna bakmaksızın özelliklerini ve sınırlarını anlamayı gerektirir. Bir veri kümesi milyonlarca başka veri parçası içerebilir, fakat bu onun rasgele veya temsili olduğu anlamına gelmez. Bir veri kümesiyle ilgili istatistiksel iddialarda bulunmak için o verinin nereden geldiđini bilmemiz gerekir; bu verinin içindeki zayıflıkları bilmek ve hesaplamak da benzer şekilde önemlidir. Dahası, arařtırmacıların da veriyi yorumlama aşamalarında görülen, kendi baştan belirlenimlerini açıklayabilmeleri gerekir. Bunu yapmak ise bir kişinin kimliği ve bakış açısının yine o kişinin arařtırması hakkında bilgi verdiđini fark etmeyi gerektirir (Behar & Gordon).

Büyük Veri oldukça sıklıkla gerçekte orada olmayan, fakat sırf devasa miktardaki verinin her yöne yayılabilecek bağlantılar sunabilmesi nedeniyle kalıpların görülmesi anlamında apofeni eylemini mümkün kılar. Leinweber (2007), verdiđi deđerli bir örnekle, veri madenciliđi tekniklerinin S&P 500 hisse senetlerindeki deđişimle Bangladeş'teki yağ üretimi arasındaki güçlü fakat gayri meşru alakanın gösterilebileceđini kanıtlamıştır.

Yorumlama, veri analizinin merkezinde yer alır. Büyüklüğünün önemi olmaksızın bir veri, sınırlamaya ve yanlılığa tabidir. Bu yanlılık ve sınırlamalar anlaşılıp ortaya çıkarılmadığında sonuç yanlış yorumlama olacaktır. En etkili veri analizi, araştırmacılar verinin çözümlenmesinin altında yatan karmaşık yöntem bilimsel süreci hesaba kattıklarında ortaya çıkmaktadır.

3. Daha büyük her zaman daha iyi demek değildir

Sosyal bilimciler uzun zamandır yaptıkları işteki ihtimamın, veri toplama ve analizindeki sistematik yaklaşımlarında temel bulduğunu dile getirmektedirler (McCloskey 1985). Etnograflar, yorumlarındaki taraflılığı açıklarken düşünümsellik üstünde dururlar. Deneyselciler, deney tasarımlarını kontrol ve standardize ederler. Anket araştırmacıları, örneklem mekanizması ve soruların etkilerinin detaylarından bir sonuca varırlar. Nicel araştırmacılar istatistiksel değerleri ölçüp biçerler. Bunlar sosyal bilimcilerin birbirlerinin çalışmalarının geçerliliğini değerlendirmede başvurdukları yolların tümü değilse de bir kaçıdır. Büyük Veri de sırf bizlere geniş miktarda veri sağladığı için bu yöntem bilimsel konulardan azade olmuş değildir. Mesela bir örnekleme anlamak, şu an her zaman olduğundan daha da önemlidir.

Twitter istatistiki çözümlene bağlamında bir örnektir. Twitter, verilerini elde etmek –veya kazımak – kolay olduğundan, araştırmacılar tarafından farklı pek çok kalıbı (örn. ruhsal durum ritimleri (Golder & Marcy 2011)), medya etkinliklerine katılımı (Shamma ve diğ. 2010), siyasal başkaldırıcıyı (Lotan ve diğ. 2011) ve konuşmalı etkileşimleri (Wu ve diğ. 2011) incelemek için kullanılmaktadır. Çoğu araştırmacı yayınlarında Twitter verilerinin sınırlılıklarını tartışırken özenli olsa da, bu tarz araştırmalar etrafında dolaşımda olan yaygın söylem, ulaşılabilir tweetlerin kaba sayısına odaklanma eğilimindedir. Hatta bilimsel haber yayınları için bile odak noktası kaç milyon insanın ‘çalışıldığı’dır (Wang 2011).

Twitter, ‘tüm insanlar’ı temsil etmemekle birlikte, ‘insanlar’ ve ‘Twitter kullanıcıları’ nı aynı kefeye koymak da bir hatadır: Kullanıcılar çok özel bir alt kümedir. Öyle ki ne Twitter kullanıcılarının sayısı ile dünya nüfusunun sayısı eşittir ne de hesap ve kullanıcı sayısı. Bazı kullanıcılar birden fazla hesaba sahipken, bazı hesaplar da birden fazla kişi tarafından kullanılmaktadır. Twitter’a ağ üzerinden bağlanan kimi insanların hesapları dahi yoktur. Bir yanda ise doğrudan belirli bir kullanıcı olmayıp otomatik içerikler üreten insansız ‘boot’lar bulunmaktadır. Dahası, ‘aktif’ hesap kavramı bile problemlidir. Kullanıcılardan bazıları Twitter aracılığıyla sık sık içerik gönderiminde bulunurken, bazıları bu ortama yalnızca ‘dinleyici’ olarak katılmaktadırlar (Crawford 2009, s.532). Twitter Şti. aktif kullanıcıların %40’ının yalnızca dinlemek için giriş yaptıklarını ortaya çıkarmıştır (Twitter 2011). ‘Kullanıcı’, ‘katılımcı’ ve ‘aktif’ kavramlarının gerçekten ne anlama geldiğinin eleştirel bir biçimde ele alınması gerekmektedir.

Büyük Veri ile toplam veri de aynı anlama gelmemektedir. Bir veri kümesinin örneklemini göz önünde bulundurmadan, eldeki veri kümesi boyunun bir anlamı yoktur. Örneğin, bir araştırmacı tweetlerin belirli bir gündem etrafında ne sıklıkla dolaşıma sokulduğunu anlamaya çalışabilir fakat eğer Twitter – tıpkı pornografiye veya istenmeyen iletişimlere gönderme yapan- problemlili sözcük ve içeriklere sahip tweetlerin tamamını akıştan

kaldırırsa, bir gündemin dolaşıma sokulma sıklığı tamamlanamayacaktır. Tweetlerin sayısına bakmaksızın veri baştan çarpıklaştığı için de temsili bir veriye ulaşılamayacaktır.

Kaynak belirsiz olduğunda da örnekleme anlamak güçleşmektedir. Twitter Şti, API²’ları aracılığıyla malzemesinin bir kısmını halka açık hale getirmekte. ‘Firehouse’ teorik olarak gönderilen tüm tweetleri içermekte ve kullanıcının özel veya ‘korunmuş’ olarak seçtiği tüm tweetleri de bariz bir biçimde dışarda bırakmaktadır. Fakat bazı kamuya ulaşılabilir tweetler firehouse’dan kaçabilir. Birkaç şirket firehouse’a ulaşabilirken, bu tür bir erişim olanağından yararlanabilen araştırmacıların sayısı oldukça azdır. Çoğu araştırmacı, ya ‘gardenhouse’ (kamuya açık tweetlerin kabaca %10’u) veya ‘spritzer’a (kamuya açık tweetlerin kabaca %1’lik kısmı) ulaşabilmekte, ya da kamusal akıştan³ gelen içeriklerin farklı alt kümelerine erişim sağlamak için API’larını kullanabilecekleri ‘white-listed’ hesapları kullanmaktadırlar. Hangi tweetlerin bu veri akışı içinde yer aldığı veya temsili olarak örneklendiği net değildir. Belki API tweetlerden rasgele bir örnek derleme yapıyor veya her saat başı ilk birkaç bin tweet’i topluyor ya da yalnızca ağ çizelgesinin belirli bir bölümünden tweetler çekiyordur. Araştırmacıların, inceledikleri verinin nitelikleri hakkında bunları bilmeden iddialarda bulunmaları zordur. Veri, tüm tweetleri temsil etmekte midir? Hayır, çünkü korunmuş hesapların⁴ tweetlerini dışarda bırakmaktadır. Yine eldeki veriler de, kamuya açık tweetleri eksiksiz temsil edebilir mi? Muhtemelen, bazen edemeyecektir.

Twitter, Büyük Veri madenciliği için popüler bir kaynak olmaya başlamıştır fakat Twitter verisiyle çalışmanın, çok nadiren bu işi yapanlar için dikkat çekici olmasının yanı sıra, ciddi yöntem bilimsel zorlukları bulunmaktadır. Araştırmacıların, bir veri kümesini ele aldıklarında, yalnızca veri kümesinin sınırlılıkları değil aynı zamanda bu veri kümesine hangi soruyu sorabileceklerini ve mümkün olan yorumların sınırlılıklarını da anlamaları – ve insanlara açıklayabilmeleri- gerekir.

Bu durum, özellikle araştırmacılar çok sayıda geniş veri kümelerini birleştirdiklerinde geçerlilik kazanır. Elbette buradan verileri birleştirmenin değerli kavrayışlar getirmeyeceği sonucu çıkmaz. Acquisti ve Gross’un (2009) yaptıkları gibi çalışmalar, erişime açık veri tabanlarının tıpkı bir bireyin Sosyal Güvenlik numarasını açığa çıkarmak gibi ciddi gizlilik ihlalleri üretmek üzere bir araya getirildiğini gözler önüne sermesi açısından büyük etkilere sahiptir. Fakat açık finansal veri deposu FreeRisk’in kurucularından biri olan Jesper Anderson’ın belirttiği gibi; çoklu kaynaklardan gelen verileri birleştirmek eş benzeri görülmemiş sorunlar yaratmaktadır. ‘Bu kaynakların her biri hata eğilimlidir ... Bence biz [farklı veri kümelerini birleştirdiğimizde] sorunu iyice büyütmekten başka bir şey yapmıyoruz’ (Bollier 2010, s.13).

Son olarak, bu berimsel devir boyunca ‘küçük veri’nin değerini anlamak gittikçe daha da önemlidir. Araştırmanın sağlayacağı kavrayışlar en küçük veriler de dâhil her boyutta görülebilir. Bazı durumlarda, yalnızca bir bireye odaklanmak son derece değerli olabilir. Örneğin, mavi yakalı işçilerin enformasyon pratiklerini anlamak için yalnızca hidro-elektrik kamu kuruluşunda depo denetçisi bir işçinin verilerini çalışmış olan Veinot’un (2007) çalışmasına bakalım. Bu sıra dışı çalışmayı yaparken Veinot, odağını bilindik yaklaşımın beyaz yakalı işçilerinden, ‘enformasyon pratikleri’ni yeniden tanımladığı ofislerin ve kentlerin bağlamının dışına doğru kaydırmıştı. Onun bu çalışması,

milyonlarca Facebook veya Twitter hesabının taranmasıyla keşfedilemeyecek bir hikaye anlatmış ve mümkün olan en küçük katılımcı sayısına rağmen araştırma alanına müthiş bir katkı sağlamıştır. Verinin boyutu söz konusu araştırma sorusuna tam olarak uymalıdır; bazı durumlarda küçük en iyisidir.

4. Bağlamından koparılmış Büyük Veri, anlamını yitirir

Geniş veri kümeleri modellenebilir olduğu için, veriler sıklıkla matematiksel bir modele sığacak kadar küçültülür. Fakat bağlamının dışına çıkarılmış veri anlam ve değerini kaybeder. Sosyal ağ sitelerinin doğuşu ‘toplumsal çizelge’ de endüstri çıkışlı bir takıntıya sebep olmuştur. Binlerce araştırmacı, iletiler ve hesaplar arasındaki bağlantıyı incelemek ve sosyal ağlar hakkında iddialarda bulunabilmek için Facebook ve Twitter’a veya daha başka sosyal medyalara akın etmekte. Fakat sosyal medya aracılığıyla görünürleşen ilişkiler, sosyologların ve antropologların 1930’lardan beri inceledikleri sosyogramlarla ve akrabalık ağlarıyla her zaman eşdeğer değildir (Radcliffe-Brown 1940; Freeman 2006). İnsanlar arasındaki ilişkiyi bir çizelge olarak gösterebiliyor olmak, bu ilişkilerin eşit değerinde bilgi taşıdığı anlamına gelmemektedir.

Tarihsel olarak sosyologlar ve antropologlar insanların ilişkileri hakkındaki verileri anketler, görüşmeler, gözlemler ve deneyler yoluyla bir araya getirmişlerdir. Bu verileri kullanarak ise insanların ‘kişisel ağları’ nı, bireylerin geliştirdiği ve yönettiği ilişkiler kümesini tanımlamaya odaklanmışlardır (Fisher 1982). Söz konusu bağlantılar, zamanla kişisel bağları tanımak için geliştirilmiş bir dizi ölçünün temel alınmasıyla dönüşüme uğramıştır. Büyük Veri, veri izlerinden çıkarılmış iki popüler sosyal ağ türünü incelemiştir; ‘eklemler’ ve ‘davranışsal ağlar’.

Eklemler, insanların e-posta veya cep telefonu adres rehberleri, çabuk mesajlaşmadaki simge listeleri, sosyal ağ sitelerindeki ‘Arkadaş’ listeleri ve diğer sosyal medya türlerindeki ‘Takipçi’ listeleri gibi teknik mekanizmalar yoluyla bireyselleştirdikleri bağlantılardır. İnsanların bu listelerin her birine, diğer kişileri eklemelerindeki motivasyon geniş bir çeşitlilik göstermektedir. Fakat sonuç itibarıyla bu listelerin arkadaşları, iş arkadaşlarını, tanıdıkları, ünlüleri, arkadaşların arkadaşlarını, kamuya mal olmuş kişileri ve ilginç bulunan yabancıları içerdiği görülmektedir.

Davranışsal ağlar, iletişim kalıplarından, hücre yapılanmalarından ve sosyal medya etkileşimlerinden doğmaktadır (Onella ve diğ. 2007; Meiss ve diğ. 2008). Bu ağlar birbirlerine cep telefonu iletileri gönderen, Facebook fotoğraflarında beraberce etiketlenen, bir başkasına e-posta yollayan ve en azından cep telefonlarına göre fiziksel olarak aynı mekânda bulunan insanları içermektedir.

Hem davranışsal hem de eklemler, araştırmacılar için büyük bir öneme sahiptir fakat yine de kişisel ağlarla eşit değerinde değildir. Örneğin, tartışmaya açık olmakla birlikte ‘bağ kuvveti’ kavramının bireysel ilişkilerin önemine işaret ettiği düşünülmektedir (Granovetter 1973). Cep telefonları verilerinin, çalışanların eşlerinden daha çok iş arkadaşlarıyla vakit geçirdiğini gösteriyor olması, bu insanlar için iş arkadaşlarının her zaman eşlerinden daha değerli olduğu anlamına gelmemektedir. Sıklık veya kamusal eklemleme yoluyla bağ kuvvetini ölçmek yaygın bir hatadır: Bağ kuvveti – ve bu kavram

etrafında şekillenen teorilerin pek çoğu – insanların diğerleriyle olan ilişkilerinden ne anladıkları ve bu ilişkilere nasıl değerler yüklediklerinin örtük bir hesabıdır. Her bağlantı bir diğeriyle aynı olmadığı gibi, ilişkinin sıklığı da onun gücünün bir göstergesi olmayabilir. Dahası, bir bağlantının yokluğu da mutlaka eksik bir ilişkinin oluşturulması gerektiğine işaret etmeyebilir.

Veri, genelleme demek değildir. Veri soyutlamalarını incelemenin bir değeri vardır fakat bağlamın alıkonulması özellikle de bazı inceleme alanları için hala önemini korumaktadır. Bağlamın bir ölçekte değerlendirilmesi zordur ve veriler bir modele sığdırılmaya çalışıldığında bu durum daha da zorlaşır. Bağlamı, Büyük Verinin ışığında kontrol etmek süregelen bir görevdir.

5. Erişilebilir demek, etik demek değildir

2006 yılında Harvard temelli bir araştırma grubu, ilgi alanlarında ve arkadaşlıklarında zamanla nasıl bir değişim olduğunu incelemek için 1,700 üniversitelinin Facebook profillerini toplamaya başladı (Lewis ve diğ. 2008). Anonim olduğu farz edilen bu veriler diğer araştırmacıların da incelemesine izin verecek şekilde tüm dünyaya yayıldı. Diğer araştırmacılar, veri kümesinin parçalarını anonimleştirmenin imkânsız olduğunu hemen keşfettiler: Yapılan şey mahremiyete gölge düşürmekti çünkü öğrencilerin hiç biri verilerinin toplandığından haberdar değildi (Zimmer 2008).

Konu, manşetlere taşındı ve araştırmacıları zor duruma soktu: Sosyal medya sitelerindeki bu ‘kamusal’ veri denen şeyin konumu neydi ki? İzin istenmeden de kolayca kullanılabilir miydi? Araştırmacıların en ahlaki şekilde davranmaları için neler gerekliydi? Güvenlik şirketleri bu durumları hâlihazırda daha iyi gizlilik duvarlarına ihtiyaç duyulan kilit bir savaş alanı olarak görmekte-dirler. İşin zor tarafı da gizlilik ihlallerini özelleştirmenin kolay olmayışıdır – o anda bir ihlal söz konusu mudur? Peki ya 20 yıl sonra? ‘İnsan deneklerle ilgili her veri gizlilik konusunu ortaya çıkarır ve bu verilerin istismar edilme riskini hesaplamak zordur’ (Nature, Berry 2011 içinde).

Etik kurullar (Institutional Review Boards-IRB) -ve diğer araştırma etiği komitele-ri– 1970’lerde insan konulu araştırmaları denetlemek üzere kurulmuştur. Uygulamada doğal olarak sorunlar bulunmasına rağmen (Schrag 2010), etik kurulların amacı belirli bir araştırma alanına ait etik değerler geliştirmek için sınırlar çizmek ve kuvvetler ayrılığının özneleri korumak için uygulamaya konulduğundan emin olmaktır. ‘Onanmış rıza’ ve insan deneklerin gizlilik haklarının korunması gibi faaliyetler medikal ve sosyal bilim alanlarında daha önce karşılaşılmış olan istismarlar göz önünde bulundurularak, katılımcıları güçlendirmeyi amaçlamaktadır (Blass 2004; Reverby 2009). Her ne kadar etik kurullar belirli bir çalışmanın yol açabileceği zararları her zaman tahmin edemeseler de – ki araştırmacıları etik olmayan bir zeminde çalışma yürütmekten oldukça sıklıkla alıkoymaktadırlar-araştırmacıları kendi çalışmalarındaki etik ilkeleri eleştirel biçimde düşünmeye telkin etmeleri açısından bir değere sahiptirler.

Büyük Veri olgusunu destekleyen etik çıkarımlar hakkında çok az şey bilinmektedir. Herhangi bir kişiyi, geniş ölçekli veriler içinde değerlendirmek gerekir mi? Birinin ‘kamuya açık’ blog iletileri, bağlamından koparıp yazarın hiç tahmin etmeyeceği yönde

incelenirse ne olur? Böyle habersizce incelenmek veya mercek altına alınmak kişiler için ne anlam ifade eder? Bireylerin veya toplumların araştırma sürecinden etkilenmediklerinden şüphe götürmez biçimde emin olmak kimin sorumluluğudur? ‘Bilgilendirilmiş rıza’ nasıl bir şeydir?

Araştırmacıların her bir tweet göndericisinden rıza almalarını istemek mantıksız olabilir fakat onların da sırf veriler erişime açık olduğu için eylemlerinin etiğe uygun olduğunu düşünmeleri sorunlu bir yaklaşımdır. Bir içeriğin kamusal erişime açık olması, bu içeriğin her isteyen tarafından tüketilebileceği anlamına gelmemektedir. Çevrimiçi veri toplama ve veri analizi etiği içerisinde pek çok problem bulunmaktadır (Ess 2002). Yalnızca veriler görünüşte kamu erişimine açık diye araştırma etiği geliştirme süreci savsaklanamaz. Araştırmacılar hem kendilerine hem de meslektaşlarına veri toplama, analiz etme ve yayınlama konusunda etik sorular sormaya devam etmelidirler.

Araştırmacıların etik davranabilmeleri için, hem araştırma alanına hem de araştırmanın deneklerine hesap verebilirliğin önemini kavramış olmaları mühimdir. Hesap verebilirlik burada mahremiyetten daha geniş bir kavram olarak kullanılmaktadır. Öyle ki, Troshynski ve diğerlerinin (2008) belirttiği gibi bu kavram geleneksel mahremiyet beklentilerinin sorgulanmadığı yerlerde bile uygulanabilir. Hesap verebilirlik çok yönlü bir ilişkidir: Üstlere, iş arkadaşlarına, katılımcılara ve kamuya karşı bir hesap verebilirlik durumu mevcut olabilir (Dourish & Bell 2011). Akademik bilim insanları, insan deneklerle çalışırken bu insanların denek haklarını ve refahlarını korumak için belirli profesyonel standartlara tabiidirler. Fakat pek çok etik kurulu, tek başına Büyük Veri madenciliği ve anonimleştirme sürecinin, verilerin kişisel olarak tanınabilir hale gelmesi hatasına yol açabileceğini anlamamaktadırlar. Hesap verebilirlik, etik kurulların insanlara kesinlikle koruma sağladığını farz etmekten ziyade, Büyük Verinin bölümlenmesi konusunda ciddi bir düşünömselliği gerektirmektedir.

Büyük Veri çalışmalarında aynı zamanda gerçeklik, kontrol ve iktidarla ilgili dikkate değer sorular bulunmaktadır: Araştırmacılar araç ve erişim olanaklarına sahipken, sosyal medya kullanıcıları bir bütün olarak bu olanaklardan yoksundurlar. Kullanılan veriler yüksek ölçüde içeriğe duyarlı alanlarda yaratılırken, eğer kullanıcılara sorulacak olsa bazıları mutlaka ürettikleri verinin başka yerlerde kullanılmasına izin vermeyecekleridir. Pek çoğu ise verilerini gelecekte kullanmak üzere toplayıp depolayan birim ve algoritmaların nasıl çeşitlilik gösterdiğinden bihaberdir. Araştırmacıların, kullanıcıların hayal ettiği seyirci konumunda bulunmaları nadirendir. Kullanıcılar gönderdikleri enformasyonun sağladığı pek çok farklı kullanımın, kar ve diğer kazançların her zaman farkında olmayabilirler. Veri erişime açık veya yarı açık olabilir fakat bu yine de her türden kullanım için verilen izinle aynı anlamda değildir. Araştırmacılar, kamuda olmakla (örn. bir parkta oturmak) kamulaşmak (yani aktif olarak ilgiye tabii olmak) arasında büyük bir fark olduğunu nadiren kabul etmektedirler (boyd & Marwick 2011).

6. Büyük Veriye sınırlı erişim, yeni dijital ayrımlar yaratır

Büyük Veri üzerine bir makalede, Golder (2010) sosyolog Homans’dan (1974) bir alıntı yapar: ‘Sosyal bilimlerin yöntemi zaman ve para açısından değerlidir ve her geçen

gün biraz daha değeli hale gelmektedir'. Tarihsel olarak baktığımızda veri toplamanın zor, zaman alıcı ve yoğun kaynak gerektirici olduğunu görürüz. Büyük Veriyi saran heveslerin çoğu, onun devasa boyutta veri kümesi sunuyor olmasından kaynaklanmaktadır.

Fakat kimlerin erişim izni vardır? Ne amaçla? Ne bağlamda ve hangi koşullarda? Sosyal medya kaynaklarından edinilen veri kümelerinin kullanıldığı araştırmalar alanındaki patlama, erişimin doğrudanlığına vurgu yapar gibi olsa da, durum hiç de öyle değildir. Manovich'in (2011) belirttiği gibi 'büyük ölçekli toplumsal verilere yalnızca sosyal medya şirketleri erişebilmektedir – özellikle de işlem verilerine. Facebook için çalışan bir antropolog veya Google için çalışan bir sosyolog, geri kalan bilim insanları topluluğunun erişemeyeceği veriye erişebilecektir'. Bazı şirketler verilerine erişimi tamamiyle kısıtlamaktadırlar; bazıları da erişim hakkını belirli bir ücret karşılığında satmaktadır; kimileri ise küçük veri kümelerini üniversite temelli araştırmacılara açık hale getirmektedirler. Bu ise sistem içinde gözle görülür bir eşitsizlik yaratmaktadır: Paralı – veya şirket dâhilinde olanlar- dışarıda kalanlardan farklı türde çalışmalar üretebilirler. Erişim olanağı olmayanlar, ayrıcalıklı erişime sahip kişilerin metodolojik iddialarını ne dönüştürebilmekte ne de bu iddiaları geliştirebilmektedirler.

Şunu akılda tutmak da son derece önemlidir; Büyük Veri zengini sınıf, üniversite sistemi ile güçlenmektedir: Üst düzey, sağlam kaynaklı üniversiteler veri erişimini satın alabilecekken, büyük sosyal medya şirketleri de en çok, zirvedeki üniversitelerin öğrencileri ile çalışmak isteyeceklerdir. Dışarıda kalanların bu tür teklifler almaları ve becerilerini geliştirmeleri bu koşullarda daha az mümkündür. Sonuç ise, bilim insanları arasında derin ayrışmaların yaşanması yönünde olacaktır.

Erişim sorularına ek olarak, beceriyle ilgili sorular da bulunmaktadır. Etik kurulların kavgası, veri kazıma ve büyük şeritler halindeki veri analizi gibi beceriler, genellikle berimsel arka planı olanlarla atfedilmektedir. Berimsel beceriler en değerli beceriler olarak konumlandırıldığında, kimlerin bu bağlamda avantajlı iken kimlerin ise avantaj sahibi olmadığı konusunda sorular doğmaktadır. Bu durum kendi içinde, bilgisayar bilimcileri ve sosyal bilimcilerin aynı anda değerli bakış açıları üretebileceğini akla getirmek yerine, 'numaraları kimin okuyabileceği' ekseninde yeni hiyerarşiler yaratmaktadır. Bu önemli oranda, aynı zamanda cinsiyetler arası bir ayrımdır. Şimdilik berimsel becerilere sahip olan çoğu araştırmacı erkektir ve feminist tarihçilerin ve bilim felsefecilerinin işaret ettikleri gibi soruları kimin sorduğu, nelerin sorulacağını da tayin eder (Harding 2010; Forsythe 2001). Gelecekte hangi araştırma becerilerinin değerli görüleceği ve bu becerilerin nasıl kazandırılacağı konusunda karmaşık sorular bulunmaktadır. Öğrencileri nasıl bir öğretime tabi tutmak gerekir ki aynı zamanda algoritmalarla ve veri analiziyle olduğu kadar sosyal teori ve inceleme konusunda da rahatlıkla çalışsınlar?

Sonuç olarak, Büyük Veriye erişimin zorluğu ve pahası, araştırma bulguları açısından sınırlanmış bir kültür üretmektedir. Büyük Veri şirketleri hem verilerini erişilebilir kılmaktan sorumlu değildir hem de bu verileri kimlerin görebileceği konusunda bütünsel bir kontrole sahiptirler. Sahipli veri kümelerine erişim olanağı olan araştırmacılar, sonucun kendileri için erişimin kesilmesine yol açacağını düşündüklerinde, bir sosyal medya şirketiyle çatışacak sorular sorma konusunda daha az cüretkâr davranmaktadırlar.

Sorulabilecek -özel veya kamusal- araştırma soruları üzerindeki soğutucu etkiler, Büyük Verinin geleceğini değerlendirirken hepimizin göz önünde bulunması gereken şeylerdir.

Büyük Veri etrafındaki güncel ekosistem, yeni bir tür dijital ayırım yaratmaktadır: Büyük Veri zengini ve Büyük Veri fakiri. Bazı şirket araştırmacıları akademisyenlerin sosyal medya veri setlerini çalışmakla uğraşmamalarını belirtecek kadar ileri bile gitmektedirler. Twitter’da, araştırma yapma izninde olan profesör Jimmy Lin, akademisyenlerin, endüstrinin “daha iyisini yapacağı” araştırmalara katılmamaları gerektiğini belirtmiştir (Conover 2011). Araştırmayı ‘içerden’ veya ‘dışardan’ diye açıkça ayırma çabaları – bu durum yeni de değildir– araştırma cemiyetinin kuyusunu kazmaktadır. Derrida (1996), ‘etkili demokrasi anlayışını ölçmek için, katılım ve arşivlere erişim, inşa ve yorumlama süreçleri gibi elzem kriterleri her zaman kullanabiliriz’ (s.4) diye belirtir.

Eşitsizlikler ne zaman bir sistem içinde açık şekilde yer alsalar, orada sınıf temelli yapılar oluşur. Manovich (2011), Büyük Veride payda sahibi olan üç sınıftan bahseder: ‘veriyi yaratanlar (bilinçli veya yalnızca dijital izler bırakarak), veriyi toplamak için gerekli araçlara sahip olanlar ve veriyi analiz etmede uzmanlaşmış olanlar’. Son grubun en küçük ve en ayrıcalıklı grup olduğunu biliyoruz: Bunların aynı zamanda Büyük Verinin kullanım kurallarını tayin eden ve kimlerin katılımcı olacağını belirleyen kişiler olduklarını da bilmekteyiz. Her ne kadar kurumsal eşitsizlikler akademi dünyasında boş verilmiş bir sonuç olsa da, her daim sorgulanıp incelenmelidir. Çünkü tüm bunlar verilerde ve araştırılacak konular bağlamında önceden belirlenimlere neden olmaktadır.

Büyük Veri olgusunun bazı geniş tarihsel ve felsefi değişimlerle iç içe olduğunu ifade etmek, onun tümüyle açıklanabilir olduğu söylemek değildir; akademi, bu berimsel devrin ardındaki tek itici güç değildir. Verilerden en üst düzeyde fayda sağlama ve biriktirmeye yönelik ve onu daha fazla hedeflenmiş reklam, ürün tasarımı, trafik planlaması veya ceza siyasetine liderlik eden enformasyon haline getirmek isteyen derin bir hükümet ve endüstri faktörü bulunmaktadır. Fakat Büyük Verinin sayılabilir hale getirilmesinin ve gelecekteki araştırma günlüklerinde nasıl bir yer işgal edeceğinin, ciddi ve geniş ölçekli belirtileri olduğunu düşünüyoruz. Suchman’ın (2011) Levi Strauss aracılığıyla gözlemlediği gibi ‘kendi kendimizin araçlarıyız’. Araçların, bizim kullanımımızla birlikte dünyanın değişimine nasıl katıldıkları hakkında düşünmemiz gerekir. Büyük Veri çağı, daha yeni başlamıştır fakat bizim bu yeni araştırma dalgasının varsayımları, değerleri ve önyargılarıyla ilgili hali hazırda sorular sormakta oluşumuz da son derece önemlidir. Bilgi üretimi için yatırım yapılmış bilim insanları olarak, bu tür sorgular yaptığımız için önemli parçalarıdır.

Teşekkürler

Heather Casteel’e, bu makalenin hazırlanmasına yardım ettiği için teşekkür etmek istiyoruz. Eytan Adar, Tarleton Gillespie, Bernie Hocan, Mor Naaman, Jussi Parikka, Christian Sandvig ve ufuk açıcı diyalogları, önerileri ve geri dönütleri için tüm Microsoft Sosyal Medya Araştırmaları Topluluğuna (Microsoft Research Social Media Collective) büyük ölçüde müteşekkirimiz. Oxford Internet Enstitüsünün (Oxford Internet Institute) 10.yıl dönümünde geri bildirimlerde bulunan herkese borçluyuz. Sonuç olarak, anonim eleştirmenlerin yardımcı yorumları için teşekkür ediyoruz.

Notlar

- 1 Tartışmakta olduğumuz kavramın kendisini açıkça belirtebilmek için makale boyunca ‘Büyük Veri’ terimini büyük harflerle yazmayı tercih ettik.
- 2 API, uygulamalı program ara yüzünü (application programming interface) belirtmek için kullanılmaktadır; bu ise geliştiricilerin yapılaşmış verilere erişmek için kullandıkları bir dizi araca işaret etmektedir.
- 3 Twitter’ın sunduklarının ayrıntıları <https://dev.Twitter.com/docs/streaming-api/methods> adresinde bulunabilir. Beyaz olarak listelenmiş hesaplar, araştırmacılar tarafından genellikle kullanılan hesaplardır fakat artık ulaşılabilir değildir.
- 4 Her ne kadar korumalı hesapları tanımlama çabaları korunan hesapların %10’undan daha az olduğuna işaret etse de, korumalı hesapların sayısı bilinmemektedir.

Referanslar

- Acquisti, A. ve Gross, R. (2009). Predicting social security numbers from publicdata, *Proceedings of the National Academy of Science*, sayı. 106, no. 27, ss. 10975–10980.
- Anderson, C. (2008). The end of theory, will the data deluge makes the scientificmethod obsolete?, *Edge*, [Online] Erişim http://www.edge.org/3rd_culture/anderson08/anderson08_index.html (25 Temmuz 2011).
- Anderson, M. (1988). *The American Census: A Social History*, Yale University Press, New Haven, CT.
- Baca, G. (2004). Legends of Fordism: between myth, history, and foregone conclusions, *Social Analysis*, sayı. 48, no. 3, ss. 169–178.
- Behar, R. ve Gordon, D. A. (Der.) (1996). *Women Writing Culture*, University of California Press, Berkeley, CA.
- Berry, D. (2011). The computational turn: thinking about the digital humanities, *Culture Machine*, sayı. 12, [Online] Erişim: <http://www.culturemachine.net/index.php/cm/article/view/440/470> (11 Temmuz 2011).
- Blass, T. (2004). *The Man Who Shocked the World: The Life and Legacy of Stanley Milgram*, Basic Books, New York.
- Bollier, D. (2010). The promise and peril of big data, [Online] Erişim: http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf (11 Temmuz 2011).
- Bowker, G. C. (2005). *Memory Practices in the Sciences*, MIT Press, Cambridge, MA.
- Boyd, D. ve Marwick, A. (2011). Social privacy in networked publics: teens’ attitudes, practices, and strategies, paper given at Oxford Internet Institute, [Online] Erişim: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1925128 (28 Eylül 2011).
- Burkholder, L. (Der.) (1992). *Philosophy and the Computer*, Westview Press, Boulder, San Francisco ve Oxford.
- Cain, M. ve Finch, J. (1981). Towards a rehabilitation of data, içinde P. Abrams, R. Deem, J. Finch ve P. Rock (Der.) *Practice and Progress: British Sociology 1950–1980*, ss.105-119. George Allen and Unwin, London.
- Clifford, J. ve Marcus, G. E. (Der.) (1986). *Writing Culture: The Poetics and Politics of Ethnography*, University of California Press, Berkeley, CA.

Conover, M. (2011). Jimmy Lin, *Complexity and Social Networks Blog*, [Online] Erişim: http://www.iq.harvard.edu/blog/netgov/2011/07/the_international_conference_o.html (9 Aralık 2011).

Crawford, K. (2009). Following you: disciplines of listening in social media', *Continuum: Journal of Media & Cultural Studies*, sayı. 23, no. 4, ss. 532–533.

Derrida, J. (1996). *Archive Fever: A Freudian Impression* Çev. Eric Prenowitz, University of Chicago Press, Chicago.

Dourish, P. ve Bell, G. (2011). *Divining a Digital Future: Mess and Mythology in Ubiquitous Computing*, MIT Press, Cambridge, MA.

Du Gay, P. ve Pryke, M. (2002). *Cultural Economy: Cultural Analysis and Commercial Life*, Sage, London.

Durkheim, E. (1895/1982). *Rules of Sociological Method*, The Free Press, New York, NY.

Ess, C. (2002). Ethical decision-making and Internet research: recommendations from the aoir ethics working committee, Association of Internet Researchers, [Online] Erişim: <http://aoir.org/reports/ethics.pdf> (12 Eylül 2011).

Fischer, C. (1982). *To Dwell Among Friends: Personal Networks in Town and City*, University of Chicago, Chicago.

Forsythe, D. (2001). *Studying Those Who Study Us: An Anthropologist in the World of Artificial Intelligence*, Stanford University Press, Stanford.

Freeman, L. (2006). *The Development of Social Network Analysis*, Empirical Press, Vancouver.

Fry, J. P. ve Sibley, E. H. (1996). [1974] Evolution of database management systems, *Computing Surveys*, sayı. 8, no. 1.1, ss. 7–42. yeniden basıldı: (1996) *Great Papers in Computer Science*, (Der.) L. Laplante, IEEE Press, New York.

Gitelman, L. (2011). Notes for the Upcoming Collection 'Raw Data' is an Oxymoron, [Online] Erişim: <https://files.nyu.edu/lg91/public/> (23 Temmuz 2011).

Golder, S. (2010). Scaling social science with hadoop, *Cloudera Blog*, [Online] Erişim: <http://www.cloudera.com/blog/2010/04/scaling-social-science-with-hadoop/> (18 Haziran 2011).

Golder, S. ve Macy, M.W. (2011). Diurnal and seasonal mood vary with work, sleep and Daylength across diverse cultures, *Science*, sayı. 333, no. 6051, ss. 1878–1881, [Online] Erişim: <http://www.sciencemag.org/content/333/6051/1878>.

Granovetter, M. S. (1973). The strength of weak ties, *American Journal of Sociology*, sayı. 78, no. 6, ss. 1360–1380.

Harding, S. (2010). 'Feminism, science and the anti-Enlightenment critiques', içinde (Der.) A. Garry ve M. Pearsall, *Women, Knowledge and Reality: Explorations in Feminist Philosophy*, Unwin Hyman, Boston, MA, ss. 298–320.

Homans, G. C. (1974). *Social Behavior: Its Elementary Forms*, Harvard University Press, Cambridge, MA.

Kranzberg, M. (1986). Technology and history: kranzberg's laws, *Technology and Culture*, sayı. 27, no. 3, ss. 544–560.

Latour, B. (2009). Tarde's idea of quantification, içinde (Der.) M. Candea *The Social after Gabriel Tarde: Debates and Assessments*, ss. 145-162. Routledge, London. [Online] Available at: <http://www.bruno-latour.fr/articles/article/116-TARDE-CANDEA.pdf> (19 Haziran 2011).

Lazer, D., Pentland, A., Adamic, L., Aral, S., Baraba'si, A., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. ve Van Alstyne, M. (2009). Computational social science, *Science*, sayı.323, no. 5915, ss. 721–723.

Leinweber, D. (2007). Stupid data miner tricks: overfitting the S&P 500, *The Journal of Investing*, sayı. 16, no. 1, ss. 15–22.

Lessig, L. (1999). *Code: and Other Laws of Cyberspace*, Basic Books, New York, NY.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A. ve Christakis, N. (2008). Tastes, ties, and time: a new social network dataset using Facebook.com, *Social Networks*, sayı. 30, no. 4, ss. 330–342.

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I. ve boyd, D. (2011). The revolutions were tweeted: information flows during the 2011 Tunisian and Egyptian revolutions, *International Journal of Communications*, sayı. 5, ss. 1375–1405, [Online] Erişim: <http://ijoc.org/ojs/index.php/ijoc/article/view/1246>.

Manovich, L. (2011). Trending: the promises and the challenges of big social data, İçinde (Der.) M.K. Gold, *Debates in the Digital Humanities*, The University of Minnesota Press, Minneapolis, MN, [Online] Erişim: http://www.manovich.net/DOCS/Manovich_trending_paper.pdf (15 Temmuz 2011).

McCloskey, D. N. (Der.) (1985). From methodology to rhetoric, *The Rhetoric of Economics*, University of Wisconsin Press, Madison, ss. 20–35.

Meeder, B., Tam, J., Gage Kelley, P. ve Faith Cranor, L. (2010). RT @IWantPrivacy: widespread violation of privacy settings in the Twitter social network, paper presented at Web 2.0 Security and Privacy, W2SP 2011, Oakland, CA.

Meiss, M. R., Menczer, F. ve Vespignani, A. (2008). Structural analysis of behavioral networks from the Internet, *Journal of Physics A: Mathematical and Theoretical*, sayı. 41, no. 22, ss. 220–224.

Moretti, F. (2007). *Graphs, Maps, Trees: Abstract Models for a Literary History*, Verso, London.

Onnela, J. P., Sarama'ki, J., Hyvo"nen, J., Szabo', G., Lazer, D., Kaski, K., Kerte'sz, J. ve Baraba'si, A. L. (2007). Structure and tie strengths in mobile communication networks, *Proceedings from the National Academy of Sciences*, sayı. 104, no. 18, ss. 7332–7336.

Pariser, E. (2011). *The Filter Bubble: What the Internet is Hiding from You*, Penguin Press, New York.

Radcliffe-Brown, A. R. (1940). On social structure, *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*, sayı. 70, no. 1, ss. 1–12.

Reverby, S. M. (2009). *Examining Tuskegee: The Infamous Syphilis Study and Its Legacy*, University of North Carolina Press, Chapel Hill, NC.

Savage, M. ve Burrows, R. (2007). The coming crisis of empirical sociology, *Sociology*, sayı. 41, no. 5, ss. 885–899.

Schrag, Z. M. (2010). *Ethical Imperialism: Institutional Review Boards and the Social Sciences, 1965–2009*, Johns Hopkins University Press, Baltimore, MD.

Shamma, D. A., Kennedy, L., ve Churchill, E. F. (2010). Tweetgeist: Can the Twitter Timeline Reveal the Structure of Broadcast Events?. *Paper presented at the Computer-Supported Cooperative Work-2010*, Association for Computing Machinery, February 6–10, Savannah, Georgia USA. Erişim: <http://research.yahoo.com/pub/3041>.

Suchman, L. (2011). Consuming anthropology, içinde (Der.) A. Barry ve G. Born *Interdisciplinarity: Reconfigurations of the Social and Natural Sciences*, eds, Routledge,

London,[Online] Erişim: http://www.lancs.ac.uk/fass/doc_library/sociology/Suchman_consuming_anthroplology.pdf.

Troshynski, E., Lee, C. ve Dourish, P. (2008). Accountabilities of presence: reframing location-based systems, *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, Nisan 5–10, Florence, Italy.

Twitter (2011). One hundred million voices, *Twitter Blog*, [Online] Erişim:<http://blog.twitter.com/2011/09/one-hundred-million-voices.html> (12 Eylül 2011).

Veinot, T. (2007). The eyes of the power company: workplace information practices of a vault inspector, *The Library Quarterly*, sayı. 77, no. 2, ss. 157–180.

Wang, X. (2011). Twitter posts show workers worldwide are stressed out on the job', *Bloomberg Businessweek*, [Online] Erişim: <http://www.businessweek.com/news/2011-09-29/Twitter-posts-show-workers-worldwide-arestressed-out-on-the-job.html> (12 Mart 2012).

Wu, S., Hofman, J. M., Mason, W. A. ve Watts, D. J. (2011). Who says what to whom on Twitter', *Proceedings of the International World Wide Web Conference (WWW 2011)*, Mart 28-Nisan 1, Hyderabad, India, ss. 705 –714.

Zimmer, M. (2008). More on the “Anonymity” of the Facebook dataset – it’s Harvard College, *MichaelZimmer.org Blog*, [Online] Erişim: <http://www.michaelzimmer.org/2008/01/03/more-on-the-anonymity-of-the-facebook-dataset-its-harvard-college/> (20 Haziran 2011).

Büyük Veri çağı başlamıştır. Bilgisayar uzmanları, fizikçiler, ekonomistler, matematikçiler, siyaset bilimcileri, biyo-bilişimciler, sosyologlar ve diğer bilginler, yalnızca insanlar, şeyler ve bunların etkileşimleri tarafından ve haklarında üretilen devasa miktardaki bilgilere ulaşmak için çırpınıp durmaktadırlar. Çeşitli gruplar türeyimsel zincirleri, sosyal medya etkileşimlerini, sağlık kayıtlarını, telefon günlüklerini, devlet kayıtlarını ve insanların diğer tüm dijital izlerini incelemenin muhtemel faydaları ve getirileri hakkında tartışmalar yürütmektedirler. Böylece ortaya muazzam sorular çıkar: Büyük ölçekli araştırma verileri daha iyi araçlar, servisler ve kamu yararı oluşturmamıza yardımcı olacak mıdır, yoksa yeni bir gizlilik ihlali ve istilacı pazarlama dalgasının önünü mü açılacaktır? Veri analizi çevrimiçi toplulukları ve siyasal hareketleri anlamamıza yardım mı edecektir, yoksa protestocu izlerini takip etmek ve sesleri susturmak için mi kullanılacaktır? İnsan iletişimi ve kültürü üzerine çalışma disiplinlerimizi mi dönüştürecek, yoksa araştırma seçeneklerimizi daraltıp “araştırma” sözcüğünden ne anladığımızı mı değiştirecek? Bu açılardan, Büyük Veriyi sosyo-tekniksel bir olgu olarak ele aldığımızda, varsayım ve önyargıların eleştirel bir biçimde eklemlenmesinin gerekli olduğunu iddia etmekteyiz. Bu makalede ütopyacı ve distopyacı retoriği harekete geçirmekte olup, teknoloji, analiz ve mitoloji birlikteliğine dayanan kültürel, teknolojik ve bilimsel bir olgu olarak Büyük Veri etrafında toplanan konular hakkındaki görüşleri keskinleştirmek için altı farklı tartışma konusu öneriyoruz:

Anahtar sözcükler: Büyük Veri; mantıksal analiz; sosyal medya; iletişim çalışmaları; sosyal ağ siteleri; bilim felsefesi; epistemoloji; etik; Twitter