BÜYÜK VERİ ANALİTİĞİ GÜVENLİĞİ VE MAHREMİYETİ ÇALIŞTAYI

6 Ekim 2022
BİLKENT OTEL VE KONFERANS MERKEZİ, ANKARA

Düzenleyen:

Sponsor:

FAMECRYPT

EKİM 2022

ÇALIŞTAY SONUÇ BİLDİRGESİ

Gazi Üniversitesi Gazi AI R&D Center tarafından düzenlenen ve FAME CRYPT tarafından desteklenen bu etkinliğin amacı; büyük veri analitiği, güvenliği ve mahremiyeti konusunda yapılan güncel çalışmaları paylaşmak, sektör-kurum-üniversite temsilcilerini bir araya getirerek bu alanda yapılan çalışmaları ve ilerlemeleri paylaşmak, büyük veri analitiği, güvenliği ve mahremiyeti konularındaki gelişmeleri, araştırma ve uygulamaları değerlendirmek ve en önemlisi ise bu konuda ülkemizde yapılan çalışmaları değerlendirmek ve yapılacak olan yeni çalışmalara ışık tutmak ve yön vermek amacıyla fikirler, projeler, çalışmalar ve raporlar üretmektir.

Yapay Zeka ve Büyük Veri çalışmalarına farklı bilim disiplinlerinden birçok bilim insanı ve araştırmacı katkıda bulunmuşlardır. Başta bilgisayar bilimleri ve mühendisliği, istatistik ve matematik olmak üzere nörologlar, biyologlar, fizyologlar, psikologlar ve fizikçiler bazen tek çoğu zaman iş birliği içerisinde çalışarak Yapay Zeka ve Büyük Verinin bugünkü seviyeye gelmesine büyük katkıda bulunmuşlardır. Verilerin niteliği, çeşitliliği, boyutunun hızla artığı bir dönemde büyük veri analitiği çalışmalarına daha çok ihtiyaç vardır. Bu yaklaşımları destekleyen; SwissAILab, IDSIA, Nnaisense, Vicarious, Maluuba, OpenCog Foundation, Adaptive AI, LIDA, Numenta, Redwood Neuroscience Institute, Machine Intelligence Research Institute, OpenAI, MIT Laboratuvarı; Google, IBM, Apple, Amazon, Huawei gibi şirketler; Human Brain Project, DeepMind, OpenAI gibi 45 aktif araştırma projesi bulunmakta ve bu projelerin katkısı ise hem YZ çalışmalarının Genel Yapay Zeka çalışmalarına dönüşmesi hem de büyük veri teknolojilerinin de geliştirilmesine katkılar sağlamaktadır.

Doğal dil işleme, makine öğrenmesi, yapay sinir ağları ve derin öğrenme gibi yaklaşımlar, gerek veri ve verilere bağlı analizlerin modellenmesi gerekse daha hızlı çözüm sağlayan donanımların ve altyapıların geliştirilmesi ile son yıllarda tekrar ön plana çıkmıştır. Özellikle, yapay sinir ağları gizli katman ve düğüm sayılarının büyümesine karşın donanımsal gelişmelerin yetersiz kalması bu çalışmaları biraz sekteye uğratsa da GPU ve diğer donanımsal gelişmeler sayesinde yüksek hesaplama maliyetlerinin düşürülmesiyle bu yaklaşımlar tekrar ilgi odağı olmuştur. Kaggle'ın **açık kaynak platformu;** tüm dünyanın ilgisini çeken ve açık kaynak yapay zekâ çalışmalarının en güzel örneklerinden birisidir. Bu platform sayesinde, kullanıcılar hem kendilerini geliştirmekte hem de bilimin ve toplumların gelişimine doğrudan katkılar sağlamaktadırlar.

Gazi Üniversitesi Gazi AI Center tarafından düzenlenen "**Büyük Veri Analitiği, Güvenliği ve Mahremiyeti Ulusal Çalıştayı**", 6 Ekim 2022, Perşembe günü Bilkent Otel ve Konferans Merkezinde başarıyla tamamlanmıştır. Amacı; günümüzde çok farklı veri kaynağından elde edilen ve farklı alanlarda kullanılan büyük verinin analitiği, güvenliği ve mahremiyeti konusunda yapılan güncel çalışmaları paylaşmak, sektör-kurum-üniversite temsilcilerini bir araya getirerek bu alanda yapılan çalışmaları ve ilerlemeleri paylaşmak, büyük veri analitiği, güvenliği ve mahremiyeti konularındaki gelişmeleri, araştırma ve uygulamaları değerlendirmek ve en önemlisi ise bu konuda ülkemizde yapılan çalışmaları değerlendirmek ve yapılacak olan yeni çalışmalara ışık tutmak ve yön vermek amacıyla fikirler, projeler, çalışmalar ve raporlar üretmek olarak belirlenen çalıştayda farklı konular sunulmuş ve tartışılmıştır.

Çalıştay, üç oturum ile gerçekleştirilmiş olup alanında uzman birbirinden değerli 8 akademisyen ve uzman davetli konuşmacı ülkemizde yapılan büyük veri çalışmaları, uygulamalar, bu alanda yapılan akademik çalışmalar ile kurulan merkezlerin faaliyetleri, dünyada büyük veri analitiği, güvenliği ve mahremiyeti konusunda yapılan çalışmalara genel bakış, ülkemizde öğretim üyelerimizin bu alana katkıları ile bu alanda henüz yapılamayanlar, büyük veri alanında yaşanan sorunlar ve çözüm önerileri konularını aktarmıştır. Çalıştayda sunum yapan davetli konuşmacılar ve sunulan konular aşağıda maddeler halinde verilmiştir:

- Çalıştay Başkanlığını Gazi Üniversitesi Gazi AI R&D Center Müdürü Prof. Dr. Şeref SAĞIROĞLU'nun yaptığı etkinlikte; Sağıroğlu üniversitelerimizde, merkezlerimizde, enstitülerimizde öğretim elemanlarının yaptığı çalışmaları, ülkemize ve dünya bilimine katkılarını değerlendirmiş ve Gazi Üniversitesinde Büyük Veri ve Büyük Veri Uygulamaları konusunda yaptıkları çalışmalardan bazılarını sunmuştur. Ayrıca dünyada büyük veri konusunda yapılan akademik çalışmaları ülkeler ve üniversiteler bazında değerlendirmiş ve bu konuda nasıl ilerlemeler sağlanabileceği hakkındaki konuyu tartışmaya açarak önerilerde bulunmuştur. Gazi Üniversitesi olarak ülke büyük veri biliminin gelişimine katkı sağlayan önemli üniversitelerden birisi olduklarını; bu alanın gelişimi için ülkemizde ilk büyük veri analitiği merkez laboratuvarını kuran ilk kamu üniversitesi olduklarını, "Büyük Veri Analitiği, Güvenliği ve Mahremiyeti" konusunda lisansüstü il İngilizce programını açtığını, bilgisayar mühendisliği bölümünde "veri bilimi" dersini zorunlu hale getiren ilk bölüm olduklarını, ülkemizde büyük veri konusunda ilk doktora mezununu vermenin yanında, bu alanda ülkemizde en çok uzmanı yetiştiren üniversite olduklarını, Yapay Zeka ve Büyük Veri Kitap Serisi (4 Cilt), Büyük Veri ve Açık Veri gibi konularda açık kaynak kitap serilerini açık kaynak olarak tüm ülkeye açtıklarını ve herkesin bu kaynakları ücretsiz olarak indirip paylaşabileceğini; verinin huge, large, big, smart ve intelligent gibi formları olsa da verinin veri olduğu ve özelliklerine göre bunların sınıflandırılıp kullanılabileceğini, sadece işleme teknik ve teknolojilerinin değişerek geliştiğini, asıl olanın ise veriden değer üretmek olduğunu ve bunu da en iyi yapanların ise veriye sahip olan gelişmiş ülkelerin olduğunu berlitmiştir.
- İTÜ Araştırma Dekanı Prof. Dr. Altan ÇAKIR ise İTÜ'nün Büyük Veri alanında geliştirdiği ve kullandığı uygulamaları tanıtmış, dünyada büyük veri teknolojileri ve bu teknolojilerinin kullanım alanları ile bu alanın gelişimine projeksiyon tutmuştur.
- TÜBİTAK BİLGEM'den Baş Araştırmacı Ruşen HALEPMOLLASI, TÜBİTAK BİLGEM bünyesinde yapılan büyük veri çalışmaları ve projelerini tanıtmıştır.
- Ülkemizde büyük veri alanında ilk doktora çalışmasını yapan ve Huawei'de baş araştırmacı olarak çalışan Dr. Umut DEMİREZEN ise büyük veri konusunda yaptığı sektörel çalışmaları ve bunun ülkeye katkılarını somut olarak gösteren uygulamalarını tanıtmıştır.
- Bilkent Üniversitesinden Prof. Dr. Selim AKSOY "Uzaktan Algılamada Büyük Veri Analitiği" konularında yaptıkları çalışmaları, uydulardan elde edilen ve coğrafi bilgi sistemlerinde işlenen görüntü bazlı büyük veri analiz çalışmalarını sunmuştur.
- TOBB ETÜ'den Prof. Dr. Osman ABUL ise kişisel mahremiyet kapsamında ele alınan konum mahremiyeti ve konum ile elde edilebilen rotalama konularında yaptıkları çalışmaları ve literatüre yaptığı katkılarını aktarmıştır.
- ODTÜ Enformatik Enstitüsünden Prof. Dr. Tuğba TAŞKAYA TEMİZEL "Veri Kaynaklı Sorunlar ve Modellemeye Etkileri" konulu sunumunda, veri kalitesinin önemi, kullanılabilirliği, entegrasyonu, yönetimi, geçmiş veri yönetimi, sentetik veri üretimi, işleme ve denetimi, veri zehirleme saldırıları, ön işleme hataları gibi bu alanda karşılaşılan güçlükler ile DataOps, etiketlemeye yardımcı olan kılavuzlar, Veri Merkezli Yapay Zeka gibi konularda yapılan çalışmaları, kullanılan teknikleri ve çözümlerini aktarmıştır.
- Hacettepe Üniversitesinden son konuşmacımız Doç. Dr. Burkay GENÇ ise "Büyük Veride Görselleştirme: R Yaklaşımı" konularında yaptıkları çalışmaları ve büyük verinin görselleştirilmesinde yapılması gereken adımları sunmuşlardır.

Çalıştayın son oturumunda ise; sunulan çalışmalar ışığında ülkemizin büyük veri alanındaki bilgi birikimi, ülkemize olan katkıları, gelecekte daha çok katkı sağlanması için yapılması gerekenler, bu alanın geliştirilmesinin önündeki sorunlar, yapılabilecek ortak çalışmalar, büyük verinin çalışmalarının bilime ve ülke gelişimine katkıları, açık veri ve açık bilim bakış açımızın genişletilmesi için atılması gereken adımlar ve en önemlisi ise bu konuda yapılması gerekenler kapsamlı olarak değerlendirilmiş ve çözüm önerileri tartışılmıştır. Yapılan değerlendirmeler ve öneriler aşağıda verilmiştir.

GENEL DEĞERLENDİRMELER VE ÖNERİLER

Çalıştayda yapılan değerlendirmeler sonucunda sunulan öneriler ve görüşler aşağıda maddeler halinde paylaşılmıştır.

- Ülkemizde; 204 üniversite bulunduğu; 4.672 araştırma merkezi ve enstitüsü olduğu istatistiklere bakıldığında bunların bilim üretmekte yetersiz kaldıkları; 4 yıllık 4.500.000 üniversite lisans öğrencisi, 350.000 üzerinde yüksek lisans ve 110.000 civarında da doktora öğrencisi olduğu ve büyük veri, veri bilimi ve yapay zeka alanında 20'nin üzerinde lisans ve lisansüstü program olduğu görülmüştür. Ülkemizin bu kapasitesinin mutlaka veri bilimi, yapay zeka ve büyük veri analitiği alanlarına kaydırılması gereklidir.
- Web of Science'a (WoS) göre büyük veri alanında; Dünyada 218.663 civarında yayın üretildiğini, Büyük Veri Güvenliği konusunda yapılan çalışma sayısı 16654 olduğu, ve ülkemizde ise büyük veri alanında 2544 yayın yapıldığı; 1.702'sinin makale 768'inin ise bildiri olduğu ve kalanların ise kitap bölümü, rapor gibi yayınlar olduğu; 2.286 yayının İngilizce 252 yayının Türkçe olduğu; bu alana en çok katkıyı bilgisayar ile elektrik-elektronik mühendisliği bölümlerinin sağladığı; son iki yılda yıllık 350'nin üzerinde yayın yapıldığı; en çok yayının IEEE dergilerinde yapıldığı; 893 eserin açık kaynaklı dergilerde yayımlandığı; bu alana katkı sağlayan ilk beş üniversitelerin sırasıyla Hacettepe, ODTÜ, İTÜ, Yıldız Teknik ve Gazi Üniversitesi olduğu; yapılan yayınların 192'sine Tübitak'ın maddi destek verdiği ayrıca ABD, Çin ve İngiltere araştırma enstitülerinin de fonladığı bilinmektedir. Ülkemizde de bunun artırılması için gerekli önlemler alınmalı ve geliştirilmelidir.
- Büyük veri teknolojilerinin ve uygulamalarının üniversitelerde uygulanması ile gerek üretilen çıktıların detaylarının anlaşılması gerekse öğretim üyelerinin performanslarının konu bazlı, çıktı bazlı, araştırma konusuna katkı bazlı, hedeflere uygunluk temelli gibi pek çok konuda katkı sağlanabileceği hiçbir zaman unutulmamalıdır.
- Ülkemizde büyük veri ve yapay zeka konularında açılan programların sayısının giderek artığı bilinse de bu konuda faaliyet gösteren yeni programların açılmasının yerinde olacağı değerlendirilmektedir.
- Büyük veri konusunun veri ile ilgili bir konu olması sebebiyle ülkemizde değer üretilebilecek veri setlerinin üretilmesi ve bunların anonimleştirilerek çık erişime açılması bu mümkün değil ise anonimleştirilmiş verilerin kısıtlı olarak kullanıma açılması konularının netleştirilmesi gereklidir. KVKK'da verilerin anonimleştirilerek araştırmalarda kullanılabileceği konusu desteklenmesine karşın, bunun gerçek hayatta çokta uygulanmadığı veya uygulanamadığı bilinmektedir. Bu sorunun çözülmesi için somut adımlar atılmalıdır.
- Yapay zeka konularında ise çalışmaların artması ve veriden değer üretilebilmesi için mutlaka açık veri yaklaşımlarının geliştirilmesi, verilerin anonimleştirilmesi veya sentetik veriler üretilerek araştırmacılara sunulmalıdır. SSB'nin üzerinde çalıştığı ve savunma sanayinde yapılacak olan çalışmalara öncülük etmesi açısından açacağı yeni merkez gibi yapıların sayıları artırılmalıdır.
- Ülkemizde büyük veri yaklaşımlarının geliştirilmesi ve geliştirilen algoritmaların veya yaklaşımların diğer sektörlere de aktarılması için ortak platformlar kurulmalı ve üretilen çıktılar paylaşılmalıdır.
- Büyük veri konusunda yapılacak çalışmaların teşvik edilmesi önemlidir. Bu konuda çalışmaların ve güçlerin birleştirilmesi için, bu konuda uzmanlığı yüksek olan üniversitelerin koordinatörlüğünden veya sorumluluğunda "ortak araştırma merkezleri" kurulmalıdır.
- Yapılan tezlerin, projelerin veya çalışmalardan daha fazla paylaşılması için ortak platformlar kurulmalıdır. Her yıl "Büyük Veri Teknolojileri ve Uygulamaları Zirvesi" yapılmalı ve deneyimler paylaşılmalıdır.
- Ülkemizde bu konuya olan ilginin artması, öğretim üyelerinin ve öğrencilerinin dikkatini çekecek teşvik sistemleri oluşturulmalı ve desteklenmelidir. 100-2000 Projesi önemli bir adım olup bunun destek boyutu büyütülmelidir.

- Büyük verilerin kurumlara ve şirketlere farklı bakış açıları kazanmalarına katkı sağlayacağı ortadadır. Etkilerini gösterecek, büyük veri yarışmaları, hackatonları veya etkinlikleri yapılmalı ve desteklenmelidir.
- Büyük verilerin sektörlerin gelişmesi ve büyümesine etkileri de yüksek olacaktır. Bunu artıracak olan çalışmalar ATO, İTO, ASO, EMO gibi meslek odaları destekleri alınarak yapılmalıdır.
- Ülkemizde araştırma ve geliştirme faaliyetlerine katkı sağlayan 5000'e yakın enstitü ve araştırma merkezi bulunmaktadır. Bunların ülkeye katkılarının etkinleştirilmesi ve katkılarının artırılması için önlemler alınmalı, bunun içinde büyük veri teknolojilerinden de faydalanılmalıdır.
- Veri boyutu büyüdükçe, mahremiyet ve güvenlik ihlallerinde de artış görülmektedir. Bunların önlenmesine yönelik özel projeler yapılmalıdır. Mahremiyet korumalı büyük veri yayınlama modelleri geliştirilmeli ve paylaşılmalıdır.
- YÖK'ün devreye aldığı üniversitelerin uzmanlaşması ve uzmanlaştığı alanlarda üniversitelerin bu gibi konuların hem yaygınlaşması hem de daha çok değer üretilecek projeler hayata geçirmesi ve buna benzer projelerinde mutlaka desteklenmesi yerinde olacaktır.
- Veri mahremiyeti konusunda endişe edenlerin danışabileceği, hizmet alabileceği veya güvenle uygulama geliştirebilecekleri merkezlerin sayısı artırılmalıdır.
- Kaggle'ın **açık kaynak platformu;** tüm dünyanın ilgisini çeken ve açık kaynak yapay zekâ çalışmalarının en güzel örneklerinden birisidir. Bu platform sayesinde, kullanıcılar hem kendilerini geliştirmekte hem de bilimin ve toplumların gelişimine doğrudan katkılar sağlamaktadırlar. Ülkemizde de buna benzer projeler veya platformlar hayata geçirilmelidir.
- Verileri boyutunun arttığı, çeşitliliğinin fazlalaştığı, tehdit ve tehlikelerin çoğaldığı, ihlallerin ve saldırılarıların fazlalaştığı bir zamanda bulunuyoruz. Bundan sonrada bunların hızlıca artacağı değerlendirilmektedir. Bunun için ortak çözümler geliştirilmeli, bunun için adımlar atılmalıdır.
- Depolama büyük bir problem olup maliyetlidir. Bu maliyet göz önünde bulundurulmalı, verilerin kişilerin, firmaların, kurumların önemli varlıkları olduğu bilinse de bundan mutlaka değer elde edilecek çözümlere de ağırlık vermeleri gereklidir. Verilerden değer üretecek yaklaşımlar artık zorunlu hale getirilmelidir.
- YZ hızla gelişiyor. YZ temelli çözümler, ürünler, modeller, algoritmalar, uygulamalar geliştiriliyor, mevcut birikilerden faydalanılacak ortak platformlar kurulmalıdır.
- Siber güvenlik çözümlerinde büyük veri ve yapay zeka her gün daha çok kullanılıyor. Mahremiyete **saygı duyan yeni çözümler** geliştiriliyor. Differansiyel Mahremiyet Derin Öğrenme Yöntemlerinde başarılıdır. Bileşik (Federe) Öğrenme yaklaşımları umut vericidir. Şifreleme yaklaşımları hala önemlidir. Yeni çözümlerden mutaka faydalanılmalı ve artık verilerden değer üretilmeli veya ğretilecek ortamların sayısı artırılmalıdır.
- Tehdit boyutu büyüyor, saldırı çeşitliri artıyor, tehditler ve tehlikeler artıyor, ihlaller ve saldırılar artıyor. Bunları önlemenin tek yolu, verileri büyük bakış açısıyla analiz etmek, yeni bakış açıları kazanmak ve büyük veri tabanlı analizler ile çözümler geliştirmektir.

Kamuoyuna saygıyla duyurulur.