Ekonomi ve Dış Politika Araştırmalar Merkezi

edam

SİBER POLİTİKALAR & DİJİTAL DEMOKRASİ PROGRAMI

EYLÜL 2023

DOĞRULAMA KURULUŞLARI, İLERİ TEKNOLOJİLER VE ALGORİTMİK ENFORMASYON MANİPÜLASYONUYLA MÜCADELE

Akın Ünver, Doç Dr Özyeğin Üniversitesi

GİRİŞ

Dijital platformların yükselişi ve çevrimiçi bilginin katlanarak artmasıyla birlikte, doğruluk kontrolü büyük ölçüde teknolojik gelişmelere bağlı olarak önemli bir dönüşüm geçirmiştir. Yapay Zeka (AI) ve Doğal Dil İşleme (NLP), büyük miktarda metinsel verinin gerçek zamanlı olarak analiz edilmesini sağlayarak doğruluk kontrol süreçlerinde devrim yaratmıştır. Giderek daha karmaşık hale gelen algoritmalar yanlış bilgi kalıplarını belirleyebilmekte, yanıltıcı iddiaları tespit edebilmekte ve doğruluk kontrol çalışmalarını kolaylaştırarak doğruluk kontrol uzmanlarının internette dolaşan büyük bilgi hacmine ayak uydurabilmesini sağlamaktadır. Doğruluk kontrolü yapanlar arasında giderek daha popüler hale gelen otomatik doğruluk kontrol araçları, genellikle dezenformasyon ve bilgi manipülasyonunun yaratılmasına ve yayılmasına yardımcı olan teknolojilere dayanmaktadır. Bu araçlar, iddiaların doğruluğunu değerlendirmek için makine öğrenimi algoritmalarını kullanmakta ve doğruluk kontrolörlerine neredeyse anlık sonuçlar sağlamaktadır, ancak çoğu yöntemin doğruluğu uzun zamandır kusurludur ve güven aralıkları açısından tanımlanmıştır. Bu araçlar, doğruluk kontrol sürecini otomatikleştirerek daha fazla verimlilik ve ölçeklenebilirlik vaat etmekte ve yanlışlara daha etkili bir şekilde karşı koymak için doğru bilginin yayılmasını kolaylaştırmaktadır.

Blok zinciri teknolojisi, bilgi doğrulamada güven ve şeffaflığı artırmanın bir yolunu sunarak doğruluk kontrol arenasına da girmiştir. Blok zinciri, doğrulanmış bilgileri zaman damgasıyla damgalayarak ve güvence altına alarak, iddiaların ve doğruluklarının değişmez bir kaydını sağlar. Blok zincirinin merkezi olmayan yapısı, bilgi manipülasyonuna karşı koyma ve güvenilir bir hakikat kaynağı yaratma fırsatı sunmaktadır. Deepfake tespit teknolojileri, başlangıçta deepfake içeriğin yükselişiyle mücadele etmek için geliştirilmiş olsa da, doğruluk kontrolünde de uygunluk bulmuştur. Doğruluk kontrolcüleri, kitleleri aldatmak için manipüle edilmiş ses ve video içeriğinin kullanımı konusunda giderek daha fazla endişe duymakta ve bu da deepfake tespitini araç setlerinin önemli bir bileşeni haline getirmektedir.

Doğruluk kontrolünde yeni teknolojilerin potansiyel faydalarına rağmen, etkinliği, zorlukları ve potansiyel önyargıları ile ilgili tartışmalar çoktur. Önemli tartışmalardan biri, doğruluk kontrolörlerinin dijital çağda yanlış bilginin hızla yayılmasına ayak uydurma becerisi etrafında dönmektedir. Her gün üretilen içeriğin büyük hacmi, doğruluk kontrolörleri için önemli zorluklar yaratmaktadır, çünkü bazı yanlış iddialar tamamen çürütülmeden önce ilgi çekebilmektedir. Bir başka tartışmalı konu da doğruluk kontrolörlerinin ve doğruluk kontrol kuruluşlarının algılanan önyargılarıyla ilgilidir. Bazı eleştirmenler, doğruluk kontrolörlerinin kendi ideolojik eğilimleri olabileceğini ve bunun da iddialara ilişkin değerlendirmelerini etkileme potansiyeli taşıdığını savunmaktadır. Doğruluk kontrolü yapılacak iddiaların seçimine ilişkin tartışmalar da ortaya çıkmaktadır, zira önceliklendirme yanlışlıkla doğruluk kontrolü yapan kişinin kendi önyargılarını veya çıkarlarını yansıtabilir.

Doğruluk kontrolü ve ifade özgürlüğü arasındaki gerilim, bilgi doğrulama çabalarının merkezinde yer alan sürekli bir tartışmadır. Doğruluk kontrolü yanlış bilgiye karşı koymayı amaçlarken, bazıları aşırı agresif doğruluk kontrolünün sansür suçlamalarına ve ifade özgürlüğünün ihlaline yol açabileceğini savunmaktadır. Dezenformasyonla mücadele ile demokratik değerlerin korunması arasında bir denge kurmak, doğruluk kontrolü yapanların karşılaştığı karmaşık bir zorluktur. Dahası, yanlış bilgi ve dezenformasyon kampanyalarının küresel doğası, doğruluk kontrolörleri arasında sınır ötesi iş birliğine ilişkin soruları gündeme getirmektedir. Bilgi manipülasyonu genellikle birden fazla ülkeye yayılan devlet destekli aktörlerden kaynaklandığından, dünya çapındaki doğruluk kontrol kuruluşları arasındaki iş birliği çabaları, yabancı bilgi müdahalesiyle etkili bir şekilde mücadele etmek için hayati önem taşımaktadır.

YENİ DOĞRULUK KONTROL TEKNOLOJİLERİ İLERİ VERİ ANALİZİ VE YAPAY ZEKA

Birkaç öncü kuruluş, ileri teknolojilerin entegrasyonu yoluyla doğruluk kontrolü ortamını yeniden şekillendiriyor. İngiltere merkezli Full Fact, habercilikte doğruyu yanlıştan ayırmak için doğal dil işleme ve makine öğrenimi yeteneklerini harmanlayarak öncü konumdadır. Doğruluk kontrol sürecinin bazı kısımlarını otomatikleştirmeye yönelik devam eden çabaları, erişimlerini ve etkinliklerini genişletme konusundaki kararlılıklarının altını çiziyor. Doğruluk kontrolü alanındaki öncülerden biri olan Snopes da titiz doğrulama sürecini makine öğreniminin sunduğu içgörülerle destekliyor. Bu teknoloji, sosyal medya gürültüsünü eleme ve uydurmaları ortaya çıkarma kapasitelerini artırıyor. Aynı zamanda, Annenberg Kamu Politikaları Merkezi'nin bir projesi olan FactCheck.org, yanlış bilgilerin sosyal medya ağlarında yayılmasını izlemek için sofistike veri analitiği kullanıyor. Çalışmaları, yanlış bilginin yayılma dinamikleri hakkında değerli bir anlayış sağlamaktadır.

Agence France-Presse, dijital adli tıptan yararlanan AFP Fact Check adlı kendi doğruluk kontrol birimini harekete geçirdi. Tersine görüntü arama ve coğrafi konum belirleme gibi araçları kullanarak, dolaşımda olan görüntü ve videoların gerçekliğini incelemektedirler. Associated Press de sosyal medya içeriğini ve görsel medyayı doğrulamak için dijital araçları devreye sokuyor. Sosyal medya kalıplarını incelemek için yapay zeka güdümlü analitik kullanarak, daha yakından incelenmesi gereken potansiyel yanlış iddialar için bir radar sağlıyorlar. Arjantin'de Chequeado, canlı konuşmalara ayak uydurmak için makine öğrenimi ve otomasyondan yararlanıyor ve ifadeleri önceden var olan doğruluk kontrollü veritabanıyla çapraz referanslandırıyor. Bu gerçek zamanlı doğrulama süreci, teknolojinin yanlışlıklara anında meydan okuma potansiyelini örnekliyor. Bu kuruluşlar doğruluk kontrolü çabalarında teknolojinin gücünden yararlanırken, insani muhakemenin özü de ayrılmaz bir parça olmaya devam ediyor. Teknolojinin ne kadar güçlü olsa da, insan muhakemesinin yerini almaktan ziyade, gerçeğin amansız arayışında nasıl tamamlayıcı bir araç olarak hizmet ettiğini göstermektedirler.

İşte küresel doğruluk kontrol ekosisteminde en sık kullanılan yeni teknolojilerden bazıları:

1. Otomatik Doğruluk Kontrol Araçları:

Yapay zeka destekli otomatik doğruluk kontrol araçları, doğruluk kontrol ekosisteminde giderek yaygınlaşmaktadır. Bu araçlar, metin içeriğini analiz etmek, doğrulama gerektiren iddiaları belirlemek ve bunları saygın kaynaklar ve veri tabanları ile çapraz referanslamak için doğal dil işleme (NLP) algoritmalarını kullanır. Sonuç olarak, doğruluk kontrolörleri iş akışlarını kolaylaştırabilir, potansiyel olarak yanıltıcı iddiaları daha verimli bir şekilde belirleyebilir ve çabalarını en kritik bilgilere

odaklayabilir. Otomatik doğruluk kontrol araçlarının geliştirilmesi ve kullanılması, doğruluk kontrol süreçlerinin hızını ve doğruluğunu önemli ölçüde artırabilir. Doğal Dil İşleme (NLP) algoritmaları, doğruluk kontrolörlerinin metin içeriğini analiz etmesine ve gerçeklere dayalı yanlışlıkları veya yanlış iddiaları tespit etmesine yardımcı olabilir. NLP, bilgisayarların insan dilini anlamasını, yorumlamasını ve üretmesini sağlamaya odaklanan yapay zekanın bir alt alanıdır. Doğruluk kontrolörleri, NLP teknolojilerinden yararlanarak büyük hacimli metin verilerini işleyebilir, doğrulama gerektiren iddiaları belirleyebilir ve bilgilerin doğruluğunu daha yüksek verimlilik ve doğrulukla değerlendirebilir. Bu genişletilmiş analiz, NLP'nin teknik özelliklerini ve doğruluk kontrolü ve bilgi doğrulama üzerindeki etkisini incelemektedir.

Otomatik doğruluk kontrol hattının temel aşamaları olan iddia tespiti ve çıkarımı, yapay zeka ve doğal dil işleme (NLP) kullanımı ile derinden iç içe geçmiştir. Bunlar, her biri metni ayrıştırmada ve önemli unsurları belirlemede benzersiz bir rol oynayan çok sayıda tekniğin bir kombinasyonudur. Bu tekniklerden biri, yapılandırılmamış metin verilerinden yapılandırılmış bilgileri çeken Bilgi Çıkarımıdır. Bu süreç iki önemli bileşen içerir. İlk olarak, metin içindeki kişiler, kuruluşlar ve tarihler gibi farklı varlıkları tanımlayan Adlandırılmış Varlık Tanıma (NER) vardır. İkinci olarak, tanımlanan varlıklar arasındaki ilişkileri belirlemek için İlişki Çıkarımı (RE) uygulanır. [1] Bu süreçler birlikte, iddiaların çıkarılması için zengin bir temel sağlar.

Öte yandan, makine öğrenimi, metin sınıflandırmasında kullanım alanı bulur; burada cümleleri veya metin bölümlerini doğrulama gerektiren gerçek iddialar olarak tarar ve etiketler. Bu görev genellikle, halihazırda olgusal iddiaları tanımlayan manuel olarak eklenmiş veriler üzerinde eğitilen denetimli öğrenme modellerini içerir ve daha sonra bu modeller yeni veri kümelerindeki benzer iddiaları tespit etmek için kullanılır. NLP alanı, çıkarmayı amaçladığımız iddialar ve kanıtlar gibi metin içindeki tartışmacı yapıları tanımlamaya odaklanan özel bir alt alan olan argüman madenciliği sunar. Bu alt alandan türetilen teknikler, bu iddiaların bir metin bütününden tespit edilmesi ve çıkarılmasında çok önemlidir. Derin öğrenme modelleri, özellikle BERT, GPT ve RoBERTa gibi dönüştürücüler, NLP görevlerinde umut verici sonuçlar göstermiştir ve iddia tespiti ve çıkarımı için etkili bir şekilde kullanılabilir.[2]

Bu sofistike teknikleri tamamlayan, iddiaları belirlemek için dilbilimsel kuralları veya sezgisel yöntemleri takip etmek üzere manuel olarak programlanan kural tabanlı sistemler gibi daha basit yaklaşımlar vardır. Bu sistemler, gerçeklere dayalı iddiaların varlığına işaret eden belirli anahtar kelimeleri tarar. NLP'nin karmaşık yapısı, bir cümledeki kelimelerin anlamsal rollerini ve ilişkilerini tanımlayan ve iddia yapılarının tanınmasına yardımcı olan Anlamsal Rol Etiketleme (SRL) gibi görevleri de içerir.[3] Araştırmacılar, kontrol edilmeye değer gerçek iddiaları tespit etmek için denetimli makine öğrenimini kullanan benzersiz bir algoritma olan ClaimBuster'ın

1 Bose, Priyankar, Sriram Srinivasan, William C. Sleeman IV, Jatinder Palta, Rishabh Kapoor ve Preetam Ghosh. "Klinik metinlerde yeni adlandırılmış varlık tanıma ve ilişki çıkarma teknikleri üzerine bir araştırma." Uygulamalı Bilimler 11, no. 18 (2021): 8319.

2 Casillas, Ramón, Helena Gómez-Adorno, Victor Lomas-Barrie ve Orlando Ramos-Flores. "COVID-19 İddiaları Üzerinde Yorumlanabilir Bert Tabanlı Bir Mimari Kullanarak Otomatik Doğruluk Kontrolü." Uygulamalı Bilimler 12, no. 20 (2022): 10644.

3 Màrquez, Lluís, Xavier Carreras, Kenneth C. Litkowski ve Suzanne Stevenson. "Anlamsal rol etiketleme: özel sayı için bir giriş." Computational linguistics 34, no. 2 (2008): 145-159.

geliştirilmesiyle örneklendiği gibi, bu alanda adımlar atmışlardır.[4] Bu algoritma, konuşma parçası etiketleri, sayıların varlığı ve adlandırılmış varlıklar gibi özellikleri incelemektedir. Bu yöntemlerdeki ilerlemelere rağmen, özellikle model eğitimi, denetimi ve doğrulaması için insan müdahalesi vazgeçilmezdir. Bu gereklilik, insan dilinin karmaşıklığı ve değişkenliğinden kaynaklanmaktadır. Tam otomatik ve hatasız bir iddia tespit sistemi geliştirme hedefi halen devam eden bir çalışmadır ve araştırma çabalarının sürdürülmesini gerektirmektedir.

Comparison	BERT October 11, 2018	RoBERTa July 26, 2019	DistilBERT October 2, 2019	ALBERT September 26, 2019
Parameters	**Base:** 110M **Large:** 340M	**Base:** 125 **Large:** 355	**Base:** 66	**Base:** 12M **Large:** 18M
Layers / Hidden Dimensions / Self-Attention Heads	**Base:** 12 / 768 / 12 **Large:** 24 / 1024 / 16	**Base:** 12 / 768 / 12 **Large:** 24 / 1024 / 16	**Base:** 6 / 768 / 12	**Base:** 12 / 768 / 12 **Large:** 24 / 1024 / 16
Training Time	**Base:** 8 x V100 x 12d **Large:** 280 x V100 x 1d	1024 x V100 x 1 day (4-5x more than BERT)	**Base:** 8 x V100 x 3.5d (4 times less than BERT)	[not given] **Large:** 1.7x faster
Performance	Outperforming SOTA in Oct 2018	88.5 on GLUE	97% of BERT-base's performance on GLUE	89.4 on GLUE
Pre-Training Data	BooksCorpus + English Wikipedia = 16 GB	BERT + CCNews + OpenWebText + Stories = 160 GB	BooksCorpus + English Wikipedia = 16 GB	BooksCorpus + English Wikipedia = 16 GB
Method	Bidirectional Transformer, MLM & NSP	BERT without NSP, Using Dynamic Masking	BERT Distillation	BERT with reduced parameters & SOP (not NSP)

Ana metin dönüştürücü modellerinin karşılaştırılması. Kaynak: Siddharth Godbole, Karolina Grubinska & Olivia Kelnreiter "Transformatörler Kullanılarak Ekonomik Belirsizliğin Tanımlanması - Mevcut Yöntemlerin Geliştirilmesi" (https://humboldt-wi.github.io/blog/research/information_systems_1920/uncertainty_identification_transformers/)

2. Duygu Analizi ve Duruş Tespiti:

Duygu analizi (sentiment analysis) ve duruş tespiti (stance detection), metin içeriğinde ifade edilen tutum, duygu ve bakış açılarını anlamaya yardımcı olarak otomatik doğruluk kontrolünde önemli bir rol oynar. İşte bu alanlardaki bazı gelişmiş yöntem ve tekniklere genel bir bakış:

Duyguları analiz etmek ve metinsel bir bağlamda duruşları tespit etmek, derin öğrenme modellerinden topluluk yöntemlerine kadar çeşitli sofistike tekniklere dayanır. Bu yenilikçi prosedürler, özellikle yüksek kaliteli, çeşitli eğitim verileri ve alana özgü bilgi ile eşleştirildiğinde, bu görevlerin potansiyelini önemli ölçüde geliştirmiştir. Tekrarlayan Sinir Ağları (RNN'ler), Evrişimli Sinir Ağları (CNN'ler) ve özellikle BERT gibi Transformatör tabanlı modeller dahil olmak üzere Derin Öğrenme Modelleri, duygu analizindeki etkinliklerini kanıtlamıştır.[5] Bu modeller, duyarlılık etiketli veri kümeleri kullanılarak özelleştirilebilir ve böylece herhangi bir metnin duyarlılığını doğru bir şekilde tahmin etmelerini sağlar.

4 Hassan, Naeemul, Gensheng Zhang, Fatma Arslan, Josue Caraballo, Damian Jimenez, Siddhant Gawsane, Shohedul Hasan ve diğerleri. "Claimbuster: İlk uçtan uca doğruluk kontrol sistemi." Proceedings of the VLDB Endowment 10, no. 12 (2017): 1945-1948.

5 Balakrishnan, Vimala, Zhongliang Shi, Chuan Liang Law, Regine Lim, Lee Leng Teh, Yue Fan ve Jeyarani Periasamy. "Twitter Felaket Tespitinde Transformatör-Derin Sinir Ağı Modellerinin Kapsamlı Bir Analizi." Matematik 10, no. 24 (2022): 4664.

Bağlamsal Gömüler de duygu analizinde önemli bir rol oynamaktadır. Word2Vec ve GloVe gibi önceden eğitilmiş kelime katıştırmaları yaygın olarak kullanılırken, ELMo, GPT ve BERT gibi modellerden elde edilen bağlamsal katıştırmaların, kelimelerin bağlamsal anlamını yakalama yetenekleri nedeniyle daha da etkili olduğu kanıtlanmıştır.[6] Unsur Tabanlı Duyarlılık Analizi (ABSA), duyarlılığı yalnızca genel belge düzeyinde değil, unsur veya varlık düzeyinde hedefleyen daha özel bir yaklaşımdır. ABSA'nın, bir ürün veya hizmetin belirli yönleri hakkında görüş bildiren geri bildirim veya incelemeleri analiz ederken özellikle yararlı olduğu kanıtlanmıştır.[7]

Çok Modlu Duygu Analizi, yalnızca metni değil aynı zamanda görüntüler, videolar veya ses gibi diğer modaliteleri de dikkate alarak kapsamlı bir duygu anlayışına olanak tanır.[8] Dikkat Mekanizmaları ise modellerin duyarlılığı tahmin ederken bir metnin en ilgili kısımlarına odaklanmasını sağlayarak özellikle uzun belgeler için doğruluğu artırır. Bir sınıflandırma görevi olarak yapılandırılabilecek bir görev olan duruş tespiti, Destek Vektör Makineleri, Rastgele Ormanlar veya derin öğrenme modelleri gibi makine öğrenimi modellerini kullanır. Bu modeller, duruşları tahmin etmek için etiketli veri kümeleri üzerinde eğitilir.[9]

Duygu analizine benzer bir teknik olan transfer öğrenme, BERT gibi önceden eğitilmiş dil modelleri kullanılarak duruş algılama görevlerine de uygulanabilir. Bu modellerin duruş etiketli veriler üzerinde ince ayarı genellikle daha iyi performansla sonuçlanır. Benzer şekilde, dikkat mekanizmaları ve tekrarlayan sinir ağları gibi sinir ağı mimarileri, kelimeler arasındaki bağlam ve ilişkileri daha derinlemesine anlamak için duruş tespitinde kullanılır. Doğruluk kontrolü söz konusu olduğunda, bir iddianın duruşunu belirlemek çok önemlidir. İddiaların duruşunu sınıflandırmak için doğruluk kontrolü veri kümelerinden ve alana özgü bilgilerden yararlanarak gelişmiş teknikler kullanılır. Örneğin, 2023 Hindistan Genel Seçimlerinde bu teknikler çeşitli siyasi iddiaların doğruluğunu kontrol etmek için kullanılmıştır.[10]

Son olarak, duruş tespitinin genel performansını ve sağlamlığını artırmak için topluluk yöntemleri kullanılır. Oylama topluluğu veya ağırlıklı ortalama gibi teknikler kullanılarak birden fazla duruş algılama modelinin tahminleri birleştirilerek sonuçlar daha da optimize edilir.[11] Bu yöntemler umut verici sonuçlar ortaya koymuş olsa da, otomatik duygu analizi ve duruş tespitinin dil, bağlam ve öznelliğin karmaşıklığı nedeniyle zorlu görevler olmaya devam ettiğini unutmamak çok önemlidir. Performanslarının çoğu, eğitim verilerinin kalitesine ve çeşitliliğine ve doğruluk kontrol görevinin alan özelliğine bağlıdır. Herhangi bir yapay zeka sistemi gibi, bu teknikler de güvenilir sonuçlar elde etmek için doğrulama ve insan gözetimi gerektirir.

6 Dharma, Eddy Muntina, F. Lumban Gaol, H. Leslie, H. S. Warnars ve B. Soewito. "Konvolüsyon sinir ağı (cnn) metin sınıflandırmasına yönelik word2vec, glove ve fasttext arasındaki doğruluk karşılaştırması." J Theor Appl Inf Technol 100, no. 2 (2022): 349-359.

7 Wang, Jie, Bingxin Xu, ve Yujie Zu. "Yön tabanlı duygu analizi için derin öğrenme." 2021 Uluslararası Makine Öğrenimi ve Akıllı Sistem Mühendisliği Konferansı (MLISE) içinde, s. 267-271. IEEE, 2021.

8 Zadeh, Amir, Minghai Chen, Soujanya Poria, Erik Cambria ve Louis-Philippe Morency. "Multimodal duygu analizi için tensör füzyon ağı." arXiv ön baskı arXiv:1707.07250 (2017).

9 Al Amrani, Yassine, Mohamed Lazaar, ve Kamal Eddine El Kadiri. "Duygu analizi için rastgele orman ve destek vektör makinesi tabanlı hibrit yaklaşım." Procedia Bilgisayar Bilimleri 127 (2018): 511-520.

10 Santos, Fátima C. Carrilho. 2023. "Dezenformasyonun Otomatik Tespitinde Yapay Zeka: Tematik Bir Analiz" Gazetecilik ve Medya 4, no. 2: 679-687. https://doi.org/10.3390/journalmedia4020043

11 Siddiqua, Umme Aymun, Abu Nowshed Chy, ve Masaki Aono. "Dikkat tabanlı bir sinirsel topluluk modeli kullanarak tweet duruşu algılama." Hesaplamalı dilbilim derneğinin kuzey Amerika bölümünün 2019 konferansı bildirilerinde: İnsan dili teknolojileri, cilt 1 (uzun ve kısa bildiriler), s. 1868-1873. 2019.

3. Gerçek Zamanlı Doğrulama ve Çürütme:

Çevrimiçi bilgi yaymanın hızlı ve dinamik yapısı, otomatik doğruluk kontrolünün temel bileşenleri olarak gerçek zamanlı doğrulama ve yanlışlama ihtiyacını gerekli kılmaktadır. Bu önemli çalışmaya yardımcı olmak için çok çeşitli gelişmiş yöntemler ve teknikler geliştirilmiştir. Otomatik sistemler, haber web sitelerini, sosyal medya platformlarını ve diğer çevrimiçi kaynakları sürekli olarak tarayıp izleyerek yeni iddiaları ve gerçek zamanlı olarak paylaşılan bilgileri belirleme yeteneğine sahiptir. Çeşitli kaynaklardan verimli bir şekilde veri toplamak için API'ler ve web kazıma araçları kullanılmaktadır. Örneğin, 2022 Avustralya orman yangınları sırasında Avustralya Associated Press, gerçek zamanlı olarak dolaşıma sokulan çok sayıda bilgi ve yanlış bilgiyi elemek için böyle bir API tabanlı sistem kullanmıştır.[12]

Doğruluk kontrolü, doğrulama için iddialara öncelik verilmesini de içerir. Gelişmiş algoritmalar buna viralite, potansiyel etki ve kaynağın güvenilirliği gibi faktörlere göre karar verir. Makine öğrenimi modelleri, hangi iddiaların yanlış bilgi olma ihtimalinin daha yüksek olduğunu veya acil doğruluk kontrolü gerektirdiğini belirlemek için geçmiş verileri kullanır. Bu süreç, kaynağın itibarının, geçmişteki doğruluğunun ve önyargısının analiz edilmesini içerir. Kaynaklara güvenilirlik puanları atanır ve bu da doğruluk kontrolü kararlarına önemli ölçüde yardımcı olabilir. Örneğin Japonya'da, 2023 Tokyo depremi sırasında olay hakkında haber yapan çeşitli çevrimiçi kaynakların güvenilirliğini değerlendirmek için yapay zeka destekli bir doğruluk kontrol aracı kullanıldı.[13] 2023'teki yıkıcı Tokyo depreminin ardından, çeşitli kaynakların olayla ilgili haber yapmaya başlamasıyla çevrimiçi bilgi dalgası yaşandı. Bu haber yağmuru, doğru güncellemelerden yaygın spekülasyonlara, yanlış bilgilere ve hatta kötü niyetli dezenformasyona kadar uzanıyordu. Sonuç olarak, hem yerel hem de küresel ölçekte bireylerin gerçeği kurgudan ayırt etmesi zorlaştı. Bu çıkmaza yanıt olarak Japon teknoloji uzmanları, olay hakkında haber yapan çok sayıda çevrimiçi kaynağın güvenilirliğini değerlendirmek üzere tasarlanmış yapay zeka destekli bir doğruluk kontrol aracını devreye soktu.[14] Bu araç Tokyo Üniversitesi'ndeki bir grup araştırmacı tarafından geliştirilmişti ve deprem olduğunda henüz test aşamasındaydı. Araç, çevrimiçi kaynaklardan gelen bilgileri, doğrulanmış haber kaynakları, resmi hükümet açıklamaları ve sismik araştırma kurumları da dahil olmak üzere çok sayıda güvenilir veri tabanıyla çapraz referanslayan bir derin öğrenme algoritması kullanıyor. Bu sayede bir bilginin doğru olma olasılığını analiz edebiliyor.

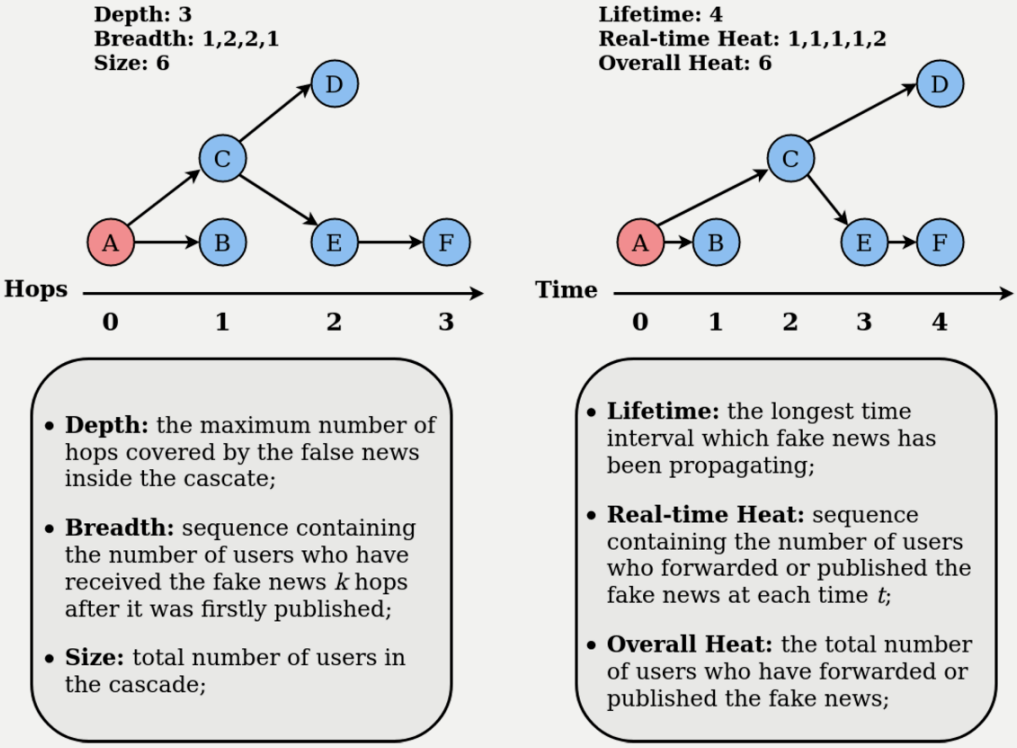
2023 yılında, Hong Kong'daki bir dizi protesto sırasında dünya, multimedya içeriğinin doğruluğunun kontrol edilmesinde sofistike yapay zeka tekniklerinin uygulanmasına tanık oldu. Bu teknolojinin merkezinde, doğrulanmış bilgilerin geniş veri tabanları

12 Brookes, Stephanie ve Lisa Waller. "Doğruluk kontrolünün üretiminde ve kaynak sağlanmasında uygulama toplulukları." Journalism (2022): 14648849221078465.

13 Japonya'nın kuzeydoğusunda meydana gelen son depremin ardından yanlış söylentiler yeniden yayıldı. The Mainichi. 15 Şubat 2023. https://mainichi.jp/english/articles/20210215/p2a/00m/0na/016000c

14 Bazarkina, Darya, Yury Kolotaev, Evgeny Pashentsev ve Daria Matyashova. "Psikolojik Alanda Yapay Zeka Tehditlerinin Mevcut ve Potansiyel Kötü Amaçlı Kullanımı: Japonya Örneği." The Palgrave Handbook of Malicious Use of AI and Psychological Security içinde, s. 419-451. Cham: Springer International Publishing, 2023.

olan bilgi grafikleri vardı. Bu veri tabanları sürekli güncellenerek çok çeşitli konuları, olayları ve daha önce değerlendirilmiş iddiaları kapsıyordu. Doğruluk kontrol yapay zeka sistemleri, yeni iddiaları hızlı bir şekilde çapraz referanslandırmak için bu bilgi grafiklerinden yararlandı. Geçmişte zaten doğrulanmış veya çürütülmüş olan



Atlama tabanlı (sol) ve zaman tabanlı (sağ) perspektifte sahte haber basamakları. Her iki perspektifte de kök düğüm A, sahte haberi yayınlayan veya oluşturan ilk kullanıcıyı temsil ederken, geri kalan düğümler sahte içeriği aktif olarak ileten veya paylaşan kullanıcıları temsil etmektedir. Kaynak: de Oliveira, Nicollas R., Pedro S. Pisa, Martin Andreoni Lopez, Dianne Scherly V. de Medeiros ve Diogo M. F. Mattos. 2021. "Doğal Dil İşlemeye Dayalı Sosyal Ağlarda Sahte Haberlerin Belirlenmesi: Trendler ve Zorluklar" Information 12, no. 1: 38. https://doi.org/10.3390/info12010038

içeriği hızla işaretleyebildiler ve böylece verimliliği artırdılar. Buna paralel olarak yapay zeka sistemleri, metinsel iddiaları gerçek zamanlı olarak doğrulamak için NLP ve makine öğrenimi modellerini kullandı. Bu sadece kelime kelime bir analiz değildi. Modeller, iddiaların yapıldığı bağlamı yorumlayarak metni bütünüyle değerlendirdi. Yapay zeka iğneleme, kültürel referanslar ve yerelleştirilmiş jargon gibi nüansları anlayabildi ve bu da bilginin doğruluğunun incelikli ve kapsamlı bir şekilde değerlendirilmesine yol açtı. Doğruluk kontrol süreci metinsel içerikle sınırlı kalmadı. Çevrimiçi olarak paylaşılan çok sayıda görüntü ve videonun gerçekliğinin yoğun bir inceleme altında olduğu Hong Kong protestoları bağlamında, yapay zeka çok önemli bir rol üstlendi. Yapay zeka, gelişmiş tersine görsel arama tekniklerini kullanarak bir görselin orijinal kaynağını, önceki kullanım durumlarını ve yapılan değişiklikleri izleyebildi. Benzer şekilde, videoların gerçekliğini kontrol etmek için video analiz araçları kullanıldı ve manipüle edilmiş kareler veya derin sahtecilikler gibi unsurlar tespit edildi. Yapay zeka bir iddianın yanlış veya yanıltıcı olduğunu tespit ettiğinde, iddiayı çürüten ayrıntılı raporlar oluşturmak için otomatik sistemler başlatıldı. Bu sistemler sadece yanlış bilgiyi işaretlemekle kalmadı; aynı zamanda

yanlışlamayı desteklemek için açıklamalar ve kanıtlar da oluşturdu. Bu şeffaflık, kullanıcıların anlamasını sağlamada çok önemliydi. Halkın belirli bir iddianın neden yanlış olduğunu görmesine yardımcı oldu, altta yatan gerçeklerin daha iyi anlaşılmasını kolaylaştırdı ve yapay zekanın doğruluk kontrol yetenekleri etrafında bir güven ortamı oluşturdu.

Kitle kaynak kullanımı çabaları, bir kullanıcı topluluğunun bilgileri kolektif olarak doğrulamasına olanak tanıyan işbirlikçi platformlarla gerçek zamanlı doğruluk kontrolünü daha da geliştirmektedir. Kullanıcılar, 2023'te Avrupa çapında doğruluk kontrol platformu "FactCheckEU"da görüldüğü gibi, doğrulama için iddialar sunabilir ve iddiaları desteklemek veya çürütmek için kanıt sağlayabilir. Otomatik uyarılar ve bildirimler gerçek zamanlı doğrulama sistemlerinde kritik bir rol oynamaktadır. Bu sistemler, potansiyel olarak yanlış veya yanıltıcı bir iddia tespit edildiğinde kullanıcıları uyarabilir ve böylece yanlış bilginin hızla yayılmasını önlemeye yardımcı olabilir. Otomasyon gerçek zamanlı doğruluk kontrolünde hayati öneme sahip olsa da, insan doğruluk kontrolörleri, özellikle karmaşık veya bağlama bağlı iddialar için hala temel uzmanlık ve muhakeme sağlar. Yapay zeka sistemleri, görevleri önceliklendirerek, bilgileri özetleyerek ve bilgi grafiklerinden kanıtlar sunarak insan doğruluk kontrolörlerine yardımcı olur. Örneğin, Brezilya'daki COVID-19 salgını sırasında, ortak bir insan-YZ çabası, virüs ve aşılarla ilgili yanlış bilgilerle mücadelede etkili oldu.[15] Sürekli gelişen yanlış bilgi ortamına ve yapay zeka teknolojilerinin devam eden gelişimine ayak uydurmak için, bu yöntem ve teknikleri sürekli olarak güncellemek ve iyileştirmek çok önemlidir. Otomatik doğruluk kontrol sistemlerinin güvenilirliğini ve doğruluğunu sağlamak, insan gözetimi ve doğrulamasının kritik bir rol oynamasıyla birlikte çok önemli olmaya devam etmektedir.

4. Diller Arası Doğruluk Kontrolü

Birden fazla dilde iddiaları ve bilgileri doğrulama uygulaması olan mütercim doğruluk kontrolü, diller arasındaki dilsel ve kültürel farklılıklar nedeniyle benzersiz bir dizi zorluk sunar. Bununla birlikte, bu zorlukların üstesinden gelmek için gelişmiş yöntemler ve teknikler geliştirilmiş ve bu da diller arası doğruluk kontrolü için sağlam bir çerçeve oluşturulmasına yol açmıştır.

Çok dilli BERT (mBERT) ve XLM-R gibi dönüştürücü tabanlı dil modelleri gibi çok dilli önceden eğitilmiş dil modellerinin kullanımı, bu çerçevenin bir ayağı olarak durmaktadır. Geniş çok dilli derlemler üzerinde eğitilen bu modeller, birden fazla dili ele alma konusunda yetkin olup, doğruluk kontrolü de dahil olmak üzere çeşitli NLP görevlerinde etkili olmaktadır.[16] Örneğin Hindistan'da mBERT, Hintçe, Bengalce ve Tamilce dahil olmak üzere çeşitli dillerdeki bilgileri doğrulamak için kullanılmış

15 Abonizio, Hugo Queiroz, Ana Paula Ayub da Costa Barbon, Renne Rodrigues, Mayara Santos, Vicente Martínez-Vizcaíno, Arthur Eumann Mesas ve Sylvio Barbon Junior. "Brezilya'daki COVID-19'da dezenformasyon ve sahte haberlere karşı insanların bir chatbot ile etkileşimi: CoronaAI vakası." Uluslararası Tıp Bilişimi Dergisi (2023): 105134.

16 Xu, Haoran, Benjamin Van Durme ve Kenton Murray. "Bert, mbert, or bibert? a study on contextualized embeddings for neural machine translation." arXiv preprint arXiv:2109.04588 (2021).

ve yerel lehçeleri dikkate alan otomatik bilgi manipülasyonu çabalarına karşı önemli bir siper görevi görmüştür.[17] Bu, çok dilli bilgi grafiklerinin uygulanmasıyla desteklenmiştir. Bunlar, çeşitli dillerden gelen verileri hizalayabilen ve böylece doğruluk kontrolörlerinin doğrulama için mevcut verilerden yararlanmasına olanak tanıyan diller arası bilgi havuzlarıdır. Bunun en iyi örneği, 2022 Avrupa Birliği seçimleri sırasında doğruluk kontrolü için çok dilli bir bilgi grafiği olan Wikidata'nın kullanılmasıdır.[18] Paralel doğruluk kontrol veri kümeleri de diller arası doğruluk kontrolünde ayrılmaz bir rol oynamaktadır. Paralel bir doğruluk kontrolü veri kümesi, belirli bir olay (seçim veya doğal afet gibi) hakkında İngilizce olarak doğruluğu kontrol edilmiş iddialardan oluşan bir veri kümesi ve aynı olay hakkında İspanyolca veya Mandarin gibi başka bir dilde doğruluğu kontrol edilmiş iddialardan oluşan başka bir veri kümesi içerebilir. Bu paralel veri kümeleri, çok dilli doğruluk kontrolü için makine öğrenimi algoritmalarının eğitiminde son derece değerli olabilir ve bu paralel veri kümelerinden öğrenerek, bir algoritma potansiyel olarak yanlış bilgi veya gerçeklerin farklı dillerde veya kültürel bağlamlarda nasıl farklı sunulabileceğini anlayabilir ve böylece iddiaları diller veya bağlamlar arasında doğrulama yeteneğini geliştirebilir.

Birden fazla dilde doğruluğu kontrol edilmiş iddiaları içeren paralel veri kümeleri oluşturarak, bu veri kümeleri farklı dillerde iddia doğrulama için denetimli öğrenme yaklaşımlarını mümkün kılar. Bunun başarılı bir uygulaması, İngilizce ve Fransızca olmak üzere iki resmi dile sahip bir ülke olan Kanada'da görülmüştür. Araştırmacılar, 2021 Kanada Federal Seçimleri sırasında hem İngilizce hem de Fransızca dillerinde çeşitli kaynaklardan gelen iddiaları içeren paralel doğruluk kontrol veri kümeleri oluşturdu.[19] Bu veri kümeleri yalnızca dilsel farklılıkları değil, aynı zamanda her iki dildeki yanlış bilgilerin kültürel ve bölgesel nüanslarını da yakaladı. Daha sonra, seçim kampanyası sırasında her iki dilde de ortaya atılan yanlış iddiaları tespit edip çürütmede başarılı olan bir yapay zeka modelini eğitmek için kullanıldı. Paralel doğruluk kontrolü veri kümelerinin bu şekilde kullanılması, seçim sürecinin bütünlüğünün sağlanmasında etkili olmuş ve çok dilli ve çok kültürlü bağlamlardaki potansiyellerini ortaya koymuştur.

İddianın dilinin mevcut doğruluk kontrol veri kümelerinin diliyle eşleşmediği senaryolarda, makine çevirisi hayati bir araç haline gelir. İddialar bir dilden diğerine çevrilir ve bu çevrilmiş iddiaları doğrulamak için mevcut modellerin kullanılmasına izin verilir. Diller arası transfer öğrenimi de, bir dilde önceden eğitilmiş modellere başka bir dilden veri kullanarak ince ayar yaparak özel görev uyarlamasına olanak tanır. Birden fazla dilde konuşanların yer aldığı kitle kaynak platformları da diller arası doğruluk kontrolünü kolaylaştırır. Dünya çapındaki bireylerin dil uzmanlığından yararlanarak, iddialar çok sayıda dilde doğrulanabilir. Bu yöntem, bir dizi kitle kaynaklı çabanın ve küresel gönüllülerin çok sayıda dilde yanlış bilgileri doğruladığı COVID-19 salgını sırasında özellikle kullanılmıştır.

17 Kar, Debanjana, Mohit Bhardwaj, Suranjana Samanta ve Amar Prakash Azad. "Söylenti yok lütfen! covid sahte tweet tespiti için çok indisli-dilli bir yaklaşım." 2021 Grace Hopper Celebration India (GHCI) içinde, s. 1-5. IEEE, 2021.

18 Rudnik, Charlotte, Thibault Ehrhart, Olivier Ferret, Denis Teyssou, Raphaël Troncy, ve Xavier Tannier. "Vikiveri tarafından kullanılan bir olay bilgi grafiği kullanarak haber makalelerini arama." İçinde 2019 world wide web konferansının tamamlayıcı bildirileri, s. 1232-1239. 2019.

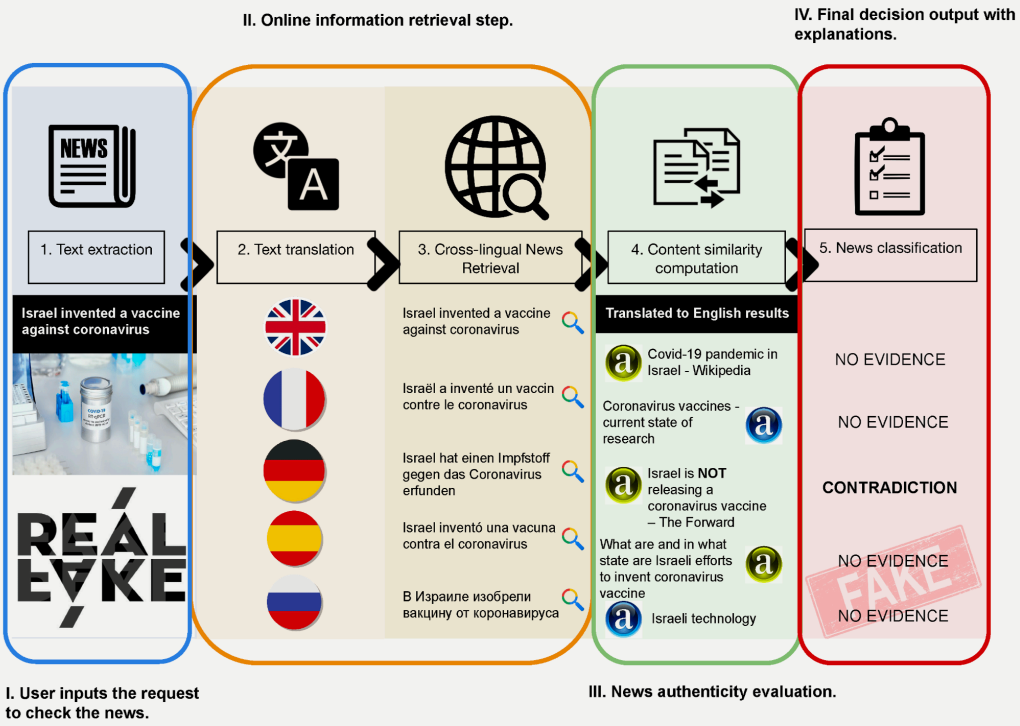
19 Al-Rawi, Ahmed ve Abdelrahman Fakida. "The methodological challenges of studying "fake news"." Journalism Practice 17, no. 6 (2023): 1178-1197.

İtalya'da Bruno Kessler Vakfı COVID-19 İnfodemik Gözlemevi'ni kurdu.[20] Bu çaba, sosyal medya platformlarında COVID-19 ile ilişkili bilgi bolluğunu -ya da "infodemik"- izlemek için makine öğrenimini kullandı. Platform sadece trend olan iddiaları tespit etmekle kalmadı, aynı zamanda bunları web sitesinde yayınlayarak küresel toplumu doğruluk kontrol sürecine katılmaya ve bilinçli tartışmalar yapmaya davet etti. Bu arada, partizan olmayan saygın bir doğruluk kontrol platformu olan Factcheck.org, web sitesinin özel bir bölümünü pandemi hakkındaki bilgilerin doğruluğunu kontrol etmeye ayırdı. Bu bölüm, dünya çapındaki kullanıcılar için bir merkez görevi gördü ve karşılaştıkları iddiaları platformun profesyonel doğruluk kontrol ekibinin doğrulaması için göndermelerine olanak tanıdı. Eş zamanlı olarak, Poynter Enstitüsü'ndeki Uluslararası Doğruluk Kontrol Ağı (IFCN), benzeri görülmemiş bir küresel doğruluk kontrol operasyonu olan CoronaVirusFacts/DatosCoronaVirus Alliance'ı koordine etti.[21] Bu ittifak, 70'ten fazla ülkeden 100'den fazla doğruluk kontrol kuruluşunu, pandemi hakkındaki yanlış iddiaları ortaklaşa çürütmek için birbirine bağladı. Özel mesajlaşma alanında WhatsApp, bazı ülkelerde kullanıcıların mesajları doğruluk kontrol kuruluşlarına iletmelerini sağlayan yeni bir özellik başlattı.[22] Yanlış bilginin bu tür uygulamalar aracılığıyla hızla yayıldığı göz önüne alındığında, bu adım yanlış bilginin yayılmasını kontrol altına almak için çok önemli bir önlemdi. Ayrıca, çevrimiçi doğruluk kontrolünün öncülerinden biri olan Snopes, pandemiyle ilgili yanlış bilgi yaydığı bilinen sahte haber sitelerini ayırt etmek için kapsamlı bir rehber sağladı. Bu rehber, internet kullanıcılarına bilgi kaynaklarının güvenilirliğini değerlendirmede yardımcı olmak açısından çok değerli olduğunu kanıtladı. Türkiye merkezli önde gelen bir doğruluk kontrol kuruluşu olan Teyit, kullanıcıların COVID-19 salgınıyla ilgili iddiaları doğrulatmak üzere gönderebilecekleri interaktif bir platform sundu. Teyit, bu süreci kamuya açarak daha yaygın ve demokratik bir doğruluk kontrolü sürecine olanak sağlamıştır.

Cümleler arasındaki anlamsal benzerliği ölçen teknikler, diller arası doğruluk kontrolü için uyarlanmıştır. Anlamsal benzerlik ölçüsü, farklı dillerdeki eşdeğer iddiaları bulmak ve bilgileri doğrulamak için kullanılabilir.[23] Ayrıca, diller arası bilgi tabanlarının entegrasyonu doğrulama için değerli bağlam ve kanıtlar sağlar. Görüntüler, videolar veya ses gibi diğer modaliteler dahil edildiğinde doğruluk kontrolü daha da güçlü hale gelir, bu teknik çok modlu diller arası doğruluk kontrolü olarak adlandırılır. Örneğin, Tokyo'daki 2022 Olimpiyatları sırasında, paylaşılan bilgilerin doğruluğunu sağlamak için multimedya içeriği diller arasında çapraz kontrole tabi tutulmuştur.[24] Dilden bağımsız temsillerin öğrenilmesine yönelik araştırmalar, doğruluk kontrol modellerinin etiketli verilerin az olabileceği düşük kaynaklı dillerdeki iddiaları ele almasını sağlayabilir. Bu teknik, düşük kaynaklı dillerin yaygın olduğu 2022'deki Sudan siyasi krizi sırasında doğruluk kontrolü çabalarında kritik bir rol oynamıştır.

20 https://covid19obs.fbk.eu/
21 https://www.poynter.org/coronavirusfactsalliance/
22 Saurwein, Florian, ve Charlotte Spencer-Smith. "Sosyal medyada dezenformasyonla mücadele: Avrupa'da çok düzeyli yönetişim ve dağıtılmış hesap verebilirlik." Digital journalism 8, no. 6 (2020): 820-841.
23 Sravanthi, Pantulkar, ve B. Srinivasu. "Cümleler arasındaki anlamsal benzerlik." Uluslararası Mühendislik ve Teknoloji Araştırma Dergisi (IRJET) 4, no. 1 (2017): 156-161.
24 Japon doğruluk kontrol kurumu internetteki yanlış bilgilere karşı koymaya hazırlanıyor. The Yomiuri Shimbun. 29 Eylül 2022. https://japannews.yomiuri.co.jp/society/general-news/20220929-61235/

Diller arası doğruluk kontrolü hala zorlu ve gelişmekte olan bir araştırma alanıdır. Dil farklılıkları, yazı stillerindeki çeşitlilikler ve farklı diller için kaynakların mevcudiyeti, göreve karmaşıklık katmanları ekler. Ayrıca, yüksek kaynaklı diller için iyi performans gösteren modeller, veri eksikliği ve dilsel çeşitlilik nedeniyle düşük kaynaklı dillere iyi genelleme yapamayabilir. Bu nedenle, diller arasında doğruluk ve kapsam sağlamak ve doğru doğrulama sağlamak için genellikle otomatik teknikler, insan uzmanlığı ve iş birliğine dayalı çabaların bir kombinasyonu gereklidir.



Örnek bir çok dilli otomatik doğruluk kontrolü akış şeması. Kaynak: Dementieva, D.; Kuimov, M.; Panchenko, A. Multiverse: Sahte Haber Tespiti için Çok Dilli Kanıt. J. Imaging 2023, 9, 77. https://doi.org/10.3390/jimaging9040077

5. Deepfake ve Medya Doğrulama:

Deepfake ve medya doğrulama, manipüle edilmiş içerik ve yanlış bilginin yayılmasına karşı koymak için çok önemli olan otomatik doğruluk kontrolü alanında çok önemlidir. Bu alanın gelişimini destekleyen ve etkinliğini artıran gelişmiş metodolojiler ve teknikler geliştirilmiştir. En dönüştürücü ilerlemeler arasında deepfake tespit algoritmalarının uygulanması yer almaktadır. Bilgisayarla görme ve derin öğrenme tekniklerinden yararlanan bu algoritmalar, manipüle edilmiş görüntü ve videoları tanımlayabilmektedir.

Evrişimsel Sinir Ağları (CNN'ler) ve Tekrarlayan Sinir Ağları (RNN'ler) gibi gelişmiş makine öğrenimi tekniklerinin son ABD seçimlerinde kullanılması, manipüle edilmiş medya içeriğiyle mücadeledeki potansiyellerinin altını çizmektedir. Görüntü sınıflandırma yetenekleriyle tanınan CNN'ler, görüntü desenlerini piksel piksel analiz

etmek ve manipülasyona işaret edebilecek her türlü düzensizliği tespit etmek için kullanıldı. Buna karşılık, dizi tahmin özellikleriyle bilinen RNN'ler, video karelerinin zamansallığını incelemek ve doğal olmayan geçişleri veya değişiklikleri ortaya çıkarmak için kullanıldı. Bu yapay zeka teknikleri yalnızca doğruluk kontrol sürecini ölçeklendirmekle kalmadı, aynı zamanda insan gözlemcileri atlatabilecek karmaşık manipülasyonları tespit ederek doğruluğunu da artırdı.

Kişisel kimlik doğrulama alanında, yüz ve ses tanıma yöntemlerinde önemli ilerlemeler kaydedildi. Bu teknolojiler 2023 G7 Hiroşima zirvesi sırasında deepfake taklitlere karşı korunmada hayati bir rol oynamıştır. Küresel etkileri olan zirve, yanlış bilgilendirme kampanyaları için birincil hedefti. Yetkililer buna karşı koymak için, ince yüz özelliklerini ve ifadelerini tanımlayabilen son teknoloji yüz tanıma teknolojisini ve değiştirilmiş olsa bile bireysel ses özelliklerini tanıyabilen ses biyometrik sistemlerini uygulamaya koydu. Sonuç olarak, zirveye katılan dünya liderlerine atfedilen her görüntü ve video klibin gerçekliğini doğrulayabildiler. Ayrıca zirve sırasında gerçek zamanlı deepfake tespit algoritmaları da devreye sokuldu. Örneğin, zirvede üretilen dijital içeriği sürekli olarak izlemek için bir yapay zeka sistemi kullanıldı. Bu sistem, deepfake teknolojisinin göstergesi olan anormallikleri aramak için videolardaki dünya liderlerinin yüz hareketlerini ve ses kalıplarını analiz etti.

Bunun önemli bir örneği 2022'de Avustralya'daki orman yangınları sırasında yaşanmıştır. Bu tür felaketler sırasında yanlış bilgilendirme durumu daha da kötüleştirebilir ve bununla mücadele etmek için doğruluk kontrolörleri tersine görsel arama tekniklerini kapsamlı bir şekilde kullandı. Örneğin, yangınlardan kurtarıldığı iddia edilen bir grup koalayı tasvir eden bir görsel viral hale geldiğinde, doğruluk kontrolörleri tersine görsel aramaları kullandı ve görselin aslında yangınlardan yıllar önce çekildiğini tespit etti. Fact-checker'lar bu tür yanıltıcı içerikleri çürüterek kamuoyunun dikkatini orman yangınlarıyla ilgili gerçek ve acil sorunlara odaklayabilmişlerdir. Bu tekniklerin önemini vurgulayan bir başka örnek olay da Suriye İç Savaşı sırasında yaşanmıştır. Yanlış bilgi yaygındı ve şiddet ve yıkım olaylarının çeşitli görüntüleri ve videoları geniş çapta paylaşıldı. Fact-checker'lar ve Bellingcat gibi OSINT girişimleri bu videoların kaynağının izini sürmek için ters video arama tekniklerini kullandılar.[25] Birçok örnekte, videoların üzerinde oynanmış ya da tamamen farklı çatışma bölgelerinden çekilmiş olduğunu tespit ettiler, böylece yanlış iddiaları çürüttüler ve savaşın parçaladığı bölgede meydana gelen gerçek zulümlere dikkat çektiler.

Görüntülerin ve videoların meta verilerinin analiz edilmesi bir başka etkili doğrulama tekniği olarak ortaya çıkmıştır. Meta veriler, bir medya parçasının kaynağı ve olası değişiklikler hakkında değerli bilgiler sağlar. Kullanılan cihaz, konum ve düzenleme geçmişiyle ilgili ayrıntıları ortaya çıkarabilir; bu da özellikle Instagram gibi sosyal medya platformlarında paylaşılan değiştirilmiş görüntülerin çürütülmesinde yararlı olmuştur. Adli tıp uzmanlarının rolü, yapay zekanın egemen olduğu bu çağda azalmadı. Aksine, medya içeriğinde tahrifat ve tutarsızlık gibi işaretler olup

25 Ikonen, Pasi, Jere Hokkanen, Turo Uskali, Ville JE Manninen ve Heidi Kuusniemi. "Gazetecilikte Uydu Görüntüleri ve Jeo-uzamsal Teknolojinin Ağ Bağlantılı Kullanımı." Gazeteciliğin Geleceği içinde: İzleyici-Haber Medyası İlişkisinde Teknoloji Uyarımlı Evrim, s. 245-260. Cham: Springer International Publishing, 2022.

olmadığını incelemek için gelişmiş araçlar kullanma konusundaki uzmanlıkları paha biçilmezdir. Bu kişilerin katılımı özellikle Doğu Ukrayna'daki çatışma sırasında önem kazanmış, görüntü ve videoların gerçekliğinin doğrulanmasına yardımcı olmuşlardır.[26]

Medya doğrulamada kullanılan bir diğer yöntem de içerik tabanlı hash'tir. Bu teknikler medya içeriği için benzersiz 'parmak izleri' oluşturarak benzerlik analizi için hızlı ve verimli karşılaştırma yapılmasını kolaylaştırır. Öne çıkan bir diğer yaklaşım olan multimedya adli bilimi, ekleme, rötuşlama ve üst üste bindirme gibi görüntü ve video manipülasyonlarını tespit etmek için algoritmalar ve yöntemler geliştirmeye adanmıştır. Generative Adversarial Networks'ün (GAN'lar) ortaya çıkışı, deepfake'lerde bir artışa yol açmış ve araştırmacıları, manipüle edilmiş içerikte GAN'lar tarafından bırakılan izleri belirlemeye odaklanan GAN tespit yöntemleri geliştirmeye itmiştir. Benzer bir şekilde, metin, görüntü ve video analizinin birleşimi olan multimodal analiz, medya içeriğinin bağlamı ve potansiyel manipülasyonu hakkında kapsamlı bir anlayış sunmaktadır. Ayrıca, kuruluşlar etkili deepfake tespit yöntemlerinin geliştirilmesini teşvik etmek için zorluklara ve yarışmalara ev sahipliği yaparak sentetik medya tespitinde yeniliği teşvik etmektedir. Deepfake tespit modellerinin sağlamlığı, veri artırımı ve düşmanca eğitim teknikleriyle daha da geliştirilmiştir. Bunlar farklı manipülasyon senaryolarını ve düşman saldırılarını simüle ederek modelin dayanıklılığını artırır. Örneğin, Facebook'un Deepfake Tespit Yarışması[27] modelleri eğitmek ve etkinliklerini artırmak için bu teknikleri kullanmıştır.

Açıklanabilir yapay zeka, doğruluk kontrolörlerinin yeni teknolojileri benimseme biçiminde şeffaflığı ve kullanıcı güvenini artırmada çok önemli bir rol oynamaktadır. Özellikle canlı etkinlikler sırasında anında medya doğrulaması için gerçek zamanlı deepfake tespiti için hızlı ve verimli algoritmaların geliştirilmesi önemli bir kilometre taşıdır. Ancak, deepfake ve medya doğrulama yöntemleri geliştikçe, tespitten kaçmayı amaçlayan kötü niyetli aktörler tarafından kullanılan teknikler de gelişmektedir. Bu nedenle, otomatik doğruluk kontrol sistemleri en son gelişmelerle güncel kalmalı ve bu zorlukların üstesinden gelmek için metodolojilerini sürekli olarak geliştirmelidir. Medya doğrulama süreçlerinin doğruluğunu ve güvenilirliğini sağladıkları için insan gözetiminin ve saha uzmanlarıyla iş birliğinin değeri küçümsenmemelidir.

26 Chen, Emily, ve Emilio Ferrara. "Çatışma zamanlarında atılan tweetler: Ukrayna ve Rusya arasındaki savaşla ilgili twitter söylemini izleyen halka açık bir veri kümesi." Uluslararası AAAI Web ve Sosyal Medya Konferansı Bildirileri, cilt 17, s. 1006-1013. 2023.

27 Schroepfer, Mike. "Think like the bad guys: An interview with Cristian Canton Ferrer, AI Red team lead." Meta, 9 Eylül 2020, https://tech.facebook.com/ideas/2020/9/dfdc/

KİTLE KAYNAKLI ve EŞGÜDÜMLÜ DOĞRULUK KONTROLÜ İÇİN PLATFORMLAR ARASI İŞBİRLİĞİ

Doğruluk kontrol kuruluşları ve sosyal medya platformları, yanlış bilgi ve dezenformasyonun yayılmasıyla mücadele etmek amacıyla dünya çapında giderek daha güçlü ittifaklar kurmaktadır. Bunun en önemli örneklerinden biri Facebook'un 2016 yılında başlattığı Üçüncü Taraf Doğruluk Kontrol Programıdır.[28] Bu program, Facebook'un dünya çapında partizan olmayan Uluslararası Doğruluk Kontrol Ağı'ndan (IFCN) sertifika almış 80'den fazla doğruluk kontrol kuruluşuyla güçlerini birleştirmesini içermektedir. Bu girişim sayesinde Facebook, yanlış olarak işaretlenmiş hikayelerin dağıtımını önemli ölçüde azaltmakta ve ek bağlam sağlamak için bunları doğruluk kontrolörlerinin ilgili makaleleriyle desteklemektedir.

Twitter da 2021 yılında Birdwatch programını başlatarak benzer bir girişimde bulunmuştur. Bu pilot program başlangıçta ABD'ye özel olsa ve esas olarak kullanıcıların yanıltıcı bilgileri tespit edip açıklamalarına dayansa da Twitter, Birdwatch programına yerleşik doğruluk kontrol kuruluşlarını entegre etme ve böylece güvenilirliğini ve etkinliğini artırma niyetini dile getirdi. Buna paralel olarak YouTube da doğruluk kontrol panellerini platformuna entegre etmek için adımlar atmıştır. Bu, özellikle son dakika haberleri ve sıklıkla yanlış bilgilendirmeye maruz kalan konularla ilgili arama sorguları için vurgulanmıştır.

Örneğin Hindistan'da YouTube, BOOM FactCheck ve Fact Crescendo gibi doğrulama sürecine yardımcı olan ve kullanıcılara kritik bilgiler sağlayan çeşitli kuruluşlarla ortaklıklar kurmuştur. Bilgi ve içgörülerin paylaşılması, dezenformasyon kampanyalarının ve koordineli manipülasyon çabalarının daha kapsamlı bir şekilde tespit edilmesini sağlayabilir. İşbirlikçi Platformlar Arası İzleme, çeşitli çevrimiçi platformlarda yanlış bilgilerle toplu olarak mücadele etmek ve bilgileri doğrulamak için doğruluk kontrol kuruluşlarını, sosyal medya platformlarını, teknoloji şirketlerini ve araştırmacıları bir araya getiren güçlü bir yaklaşımdır. Bu genişletilmiş analiz, İşbirlikçi Çapraz Platform İzleme ile ilişkili yeni teknolojilerin ve tekniklerin teknik özelliklerine odaklanarak, doğruluk kontrolü ve bilgi doğrulama üzerindeki etkisini araştırmaktadır.

1. Veri Paylaşımı ve Eşgüdüm:

Veri paylaşımı ve birlikte çalışabilirlik, kitle kaynak kullanımı ve iş birliğine dayalı doğruluk kontrolünün temel direkleridir. Bu önemli hususlar bilgi akışını kolaylaştırır, doğruluk kontrol kuruluşları ve bireyler arasında iş birliğini teşvik eder ve doğruluk kontrol sürecinin verimliliğini artırır. FullFact, Factmata ve MediaWise gibi birçok doğruluk kontrol kuruluşu açık API'ler sunmakta ve ortak veri standartlarını

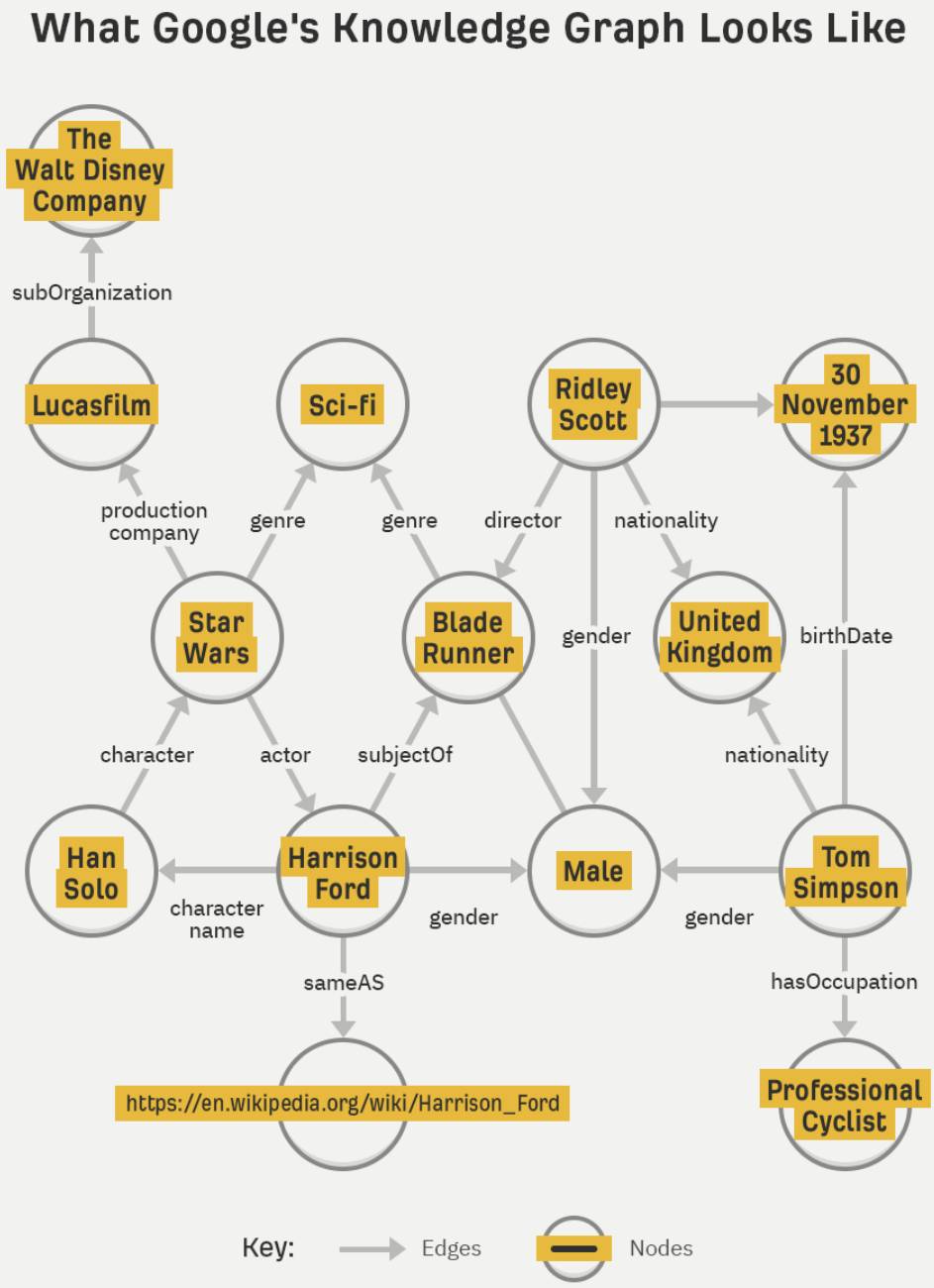
28 Meta's Third-Party Fact-Checking Program, https://www.facebook.com/formedia/mjp/programs/third-party-fact-checking

benimsemektedir. Bu yaklaşım, diğer platformların ve araçların doğruluk kontrol verilerine standartlaştırılmış bir formatta erişmesine olanak tanıyarak birlikte çalışabilirlik ve çeşitli doğruluk kontrol sistemleri arasında sorunsuz entegrasyon sağlar. Örneğin, küresel haber ajansı Reuters, doğruluk kontrol verilerini paylaşmak için böyle bir açık API kullanmakta ve dünya çapında doğruluk kontrol süreçlerinin uyumlaştırılmasını teşvik etmektedir.

Bu doğrultuda, doğruluk kontrol verilerini merkezi olmayan ve standartlaştırılmış bir şekilde yayınlamak ve birbirine bağlamak için Bağlantılı Veri ilkelerinin ve Kaynak Tanımlama Çerçevesinin (RDF) kullanımı giderek yaygınlaşmaktadır.[29] Bu yaklaşım, farklı kaynaklardan gelen verilerin birbirine bağlanmasını sağlayarak iddiaların doğrulanmasını ve çapraz referanslandırılmasını basitleştirmektedir. Örneğin BBC, doğruluk kontrol çabalarını kolaylaştırmak için bu yöntemleri kullanmaktadır.[30] Ayrıca, veri takas merkezleri olarak hareket eden çevrimiçi platformlar, doğruluk kontrol kuruluşları arasında veri paylaşımını ve iş birliğini teşvik etmede çok önemli olmuştur. Bu platformlar, Avrupa Birliği'nin doğruluk kontrolü için dijital platformu tarafından örneklenen bir yaklaşımla, doğruluk kontrolü yapanların verilere erişmesi ve katkıda bulunması için veri kümeleri, API'ler ve araçlar barındırmaktadır.

Doğruluk kontrol verileri ayrıca bir bilgi grafiğine entegre edilebilir, böylece diğer ilgili bilgilerle birlikte bağlanabilir ve kullanılabilir. Bilgi grafikleri gelişmiş sorgulama ve akıl yürütmeyi kolaylaştırarak iş birliğine dayalı doğruluk kontrol çabalarının etkinliğini artırır. Google'ın Bilgi Grafiği bunun en iyi örneğidir ve farklı bilgi parçalarını birbirine bağlamak için etkili bir yol sunar.[31] Birleştirilmiş öğrenmenin ortaya çıkmasıyla birlikte, birden fazla taraf ham verileri paylaşmadan makine öğrenimi modellerini iş birliği içinde eğitebilir. Doğruluk kontrolü alanında uygulanan bu yöntem, DuckDuckGo gibi gizlilik odaklı kuruluşlar tarafından etkin bir şekilde kullanılan bir yaklaşım olan veri gizliliğinden ödün vermeden model doğruluğunu artırır. Değişmez ve şeffaf kayıt tutma özellikleriyle tanınan Blockchain teknolojisi, doğruluk kontrolünde de uygulama alanı bulmuştur. Verilerin kaynağını garanti eder ve veri paylaşımını ve katkılarını izleyerek katılımcı kuruluşlar arasında güven oluşturur. Örneğin, New York Times'ın News Provenance Project'i tam da bu amaçla blok zincirini kullanmaktadır.[32] Verilerin hassasiyeti göz önünde bulundurulduğunda, veri paylaşımına izin verirken gizli bilgilerin korunmasını sağlamak için diferansiyel gizlilik gibi gizliliği koruyan teknikler kullanılmıştır.

29 LinkedData. https://www.w3.org/wiki/LinkedData
30 BBC Ontolojileri. https://www.bbc.com/ontologiesTchechmedjiev, Andon, Pavlos Fafalios, Katarina Boland, Malo Gasquet, Matthäus Zloch, Benjamin
31 Zapilko, Stefan Dietze ve Konstantin Todorov. "ClaimsKG: Doğruluğu kontrol edilmiş iddiaların bilgi grafiği." The Semantic Web-ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26-30, 2019, Proceedings, Part II 18, pp. 309-324. Springer Uluslararası Yayıncılık, 2019.
32 The News Provenance Project. https://www.newsprovenanceproject.com/



Google Bilgi Grafiklerinin nasıl çalıştığına dair örnek bir plan. Kaynak: Pecanek, Michal. 'Google'ın Bilgi Grafiği Açıklandı: SEO'yu Nasıl Etkiliyor'. Eylül 2020. https://ahrefs.com/blog/google-knowledge-graph/

İşbirliğine dayalı doğruluk kontrolü, gönüllüler, gazeteciler ve uzmanlar arasında kolektif çabaları teşvik eden özel çevrimiçi platformlara da büyük ölçüde dayanmaktadır. Wikipedia gibi bu platformlar veri paylaşımı, iddia doğrulama ve işbirlikçi düzenleme için araçlar sunmakta ve böylece büyük ölçekli doğruluk kontrol projelerini kolaylaştırmaktadır. Hızlı haber ve bilgi dünyasında, gerçek zamanlı veri paylaşım mekanizmaları çok önemli hale gelmiştir. Doğruluk kontrolcülerinin en yeni bilgilere erişmesini ve canlı doğruluk kontrol çabalarına katkıda bulunmasını sağlarlar; bu, başkanlık tartışmaları gibi etkinliklerde veya son dakika haberleri

sırasında hayati önem taşıdığı kanıtlanmış bir yaklaşımdır. Şeffaflığı sağlamak için, iş birliğine dayalı doğruluk kontrol platformları genellikle sürüm kontrolü ve revizyon izleme mekanizmaları uygular. Bu, GitHub gibi yazılım platformları tarafından kullanılan bir yöntem olan paylaşılan verilerdeki değişikliklerin izlenmesine yardımcı olur.

Ayrıca, paylaşılan verileri doğrulamak ve onaylamak için kullanılan kürasyon mekanizmaları, bilimsel veri havuzlarında sıklıkla kullanılan bir strateji olan veri kalitesini ve güvenilirliğini korumaya yardımcı olur. Sonuç olarak, kitle kaynak kullanımı ve iş birliğine dayalı doğruluk kontrolünde etkili veri paylaşımı ve birlikte çalışabilirlik, tüm katılımcılar arasında güven, şeffaflık ve iş birliğine bağlıdır. Gelişmiş veri paylaşım teknikleri kullanılarak, daha kapsamlı ve doğru doğruluk kontrolü sonuçları elde edilebilir, birden fazla kuruluşun ve bireyin yanlış bilgilerle ortaklaşa mücadele etmesine ve güvenilir bilgi paylaşımını teşvik etmesine olanak sağlanabilir.

2. API Entegrasyonu ve Gerçek Zamanlı Veri Toplama:

API entegrasyonu ve gerçek zamanlı veri toplama, kitle kaynak kullanımı ve iş birliğine dayalı doğruluk kontrol girişimlerinde kilit kolaylaştırıcılar olarak hizmet eder. Bunların en önemli işlevi, doğruluk kontrol platformları, veri kaynakları ve katkıda bulunanlar arasında sorunsuz bir iletişim sağlayarak verimli ve güncel bilgi alışverişinin önünü açmaktır. Hızla değişen bilgi ortamına ayak uydurmak için, haber web siteleri, sosyal medya platformları ve resmi açıklamalar gibi çeşitli çevrimiçi kaynaklardan veri toplamak için sofistike web kazıma ve tarama teknikleri kullanılmaktadır. Örneğin, Brandwatch gibi firmalar, doğruluk kontrolörlerinin ortaya çıktıkça en taze bilgilere erişebilmelerini sağlamak için gerçek zamanlı izleme kullanır.

Doğruluk kontrol platformları API'lerini harici uygulamalara ve katılımcılara açarak verilerine, araçlarına ve doğruluk kontrol işlevlerine erişim sağlayabilir. Bunun bir örneği, farklı doğruluk kontrol girişimleri arasında iş birliğini teşvik eden ve gerçek zamanlı verilerin entegrasyonunu destekleyen Google Fact Check Tools API'si olabilir. Sunucular arasında gerçek zamanlı iletişim, web kancaları ve WebSockets ile mümkün kılınarak yeni veriler veya katkılar mevcut olduğunda anında güncellemeler ve bildirimler sağlar. Örneğin, popüler bir iletişim platformu olan Slack, gerçek zamanlı bildirimler göndermek için web kancalarını kullanır. Doğruluk kontrol platformları ayrıca API'lerini birleştirerek çeşitli paydaşlar arasında gerçek zamanlı veri alışverişi için birleşik ve birlikte çalışabilir bir ekosistem oluşturabilir.

Apache Kafka veya RabbitMQ gibi veri akışı teknikleri, gerçek zamanlı veri akışlarını yönetmek için kullanılabilir ve ilgili taraflara verimli veri iletimi sağlar. Sosyal medya alanında, API'leri ile entegrasyon, doğruluk kontrolörlerinin Twitter, Facebook ve YouTube gibi platformlardan gerçek zamanlı olarak veri izlemelerine

ve toplamalarına olanak tanır. Bu, Facebook'un üçüncü taraf doğruluk kontrol kuruluşlarıyla yaptığı ortaklıkta görüldüğü gibi, viral yanlış bilgi kampanyalarının belirlenmesi ve çürütülmesinde özellikle yararlıdır. Coğrafi konum tabanlı API'ler ve resimler ve videolar gibi kullanıcı tarafından oluşturulan içeriği analiz edenler de masaya büyük değer katıyor. Bunlar, doğruluk kontrolü yapanların belirli konumlara veya bölgelere göre bilgileri filtreleyip toplamasına ve derin sahtekarlıkları ve manipüle edilmiş medyayı tespit etmesine olanak tanır. Örneğin, Google Cloud Vision API, kullanıcı tarafından oluşturulan içeriğin analiz edilmesine olanak tanır.

Bilgi grafiği API'leri, yapılandırılmış bilgilere erişim sağlayarak ve mevcut gerçeklere karşı iddia doğrulamayı kolaylaştırarak doğruluk kontrolcülerinin ufkunu daha da genişletir. Bunun da ötesinde, veri doğrulama ve kalite kontrol kontrolleri gerçekleştiren API'ler ve makine öğrenimi API'leri, birden fazla kaynaktan toplanan gerçek zamanlı verilerin doğruluğunu, güvenilirliğini ve alaka düzeyini sağlamada önemli rol oynamaktadır. Blockchain teknolojisi, doğası gereği, gerçek zamanlı olarak toplanan verilerin kaynağını ve geçmişini izlemeye olanak tanır, böylece şeffaflık ve güven sağlar. Örneğin, IBM'in Blockchain Transparent Supply çözümü, verilerin ve geçmişinin net bir şekilde kanıtlanmasını sağlar.

API'lerdeki hız sınırlama ve azaltma mekanizmaları, veri akışını kontrol etmeye ve doğruluk kontrol sunucularına aşırı yüklenmeyi önlemeye yardımcı olarak verimli ve kesintisiz operasyonlara olanak tanır. Bu araç ve teknolojilerden yararlanmak, kitle kaynak kullanımı ve iş birliğine dayalı doğruluk kontrol girişimlerinin bilgiye verimli bir şekilde erişmesine ve analiz etmesine olanak tanıyarak, çabalarını yanlış bilgilerle mücadelede ve doğru raporlamayı teşvik etmede daha etkili hale getirir. Ancak, bu gelişmelerin yanı sıra, veri kalitesi, gizlilik ve güvenliğin sağlanması bu işbirlikçi çabalarda büyük önem taşımaya devam etmektedir. Bu tür hedeflere bu gelişmiş tekniklerin kullanılması yoluyla ulaşılabilir.

3. Gerçek Zamanlı Doğrulama ve Erken Uyarı Sistemleri:

Gerçek zamanlı uyarılar ve erken uyarı sistemleri, kitle kaynak kullanımı ve iş birliğine dayalı doğruluk kontrol girişimlerinde önemli bir rol oynamaktadır. Vekiller Kullanarak Erken Model Tabanlı Olay Tanıma veya EMBERS, Virginia Tech'teki Discovery Analytics Center tarafından geliştirilen yenilikçi bir araçtır.[33] Bu platform, tweetlerden restoran rezervasyonlarına ve hatta döviz kurlarına kadar açık kaynaklı verilerden yararlanarak protestolar veya ayaklanmalar gibi önemli toplumsal olayları meydana gelmeden önce tahmin ediyor. Aslında EMBERS, 2013 yılında Brezilya'daki protestoları başarılı bir şekilde tahmin ederek öngörü potansiyelini ortaya koydu. Küresel ölçekte, Küresel Olaylar, Dil ve Ton Veritabanı veya GDELT'e sahibiz. Bu sistem, dünyanın dört bir yanındaki kaynaklardan gelen medyayı çok

33 Saraf, Parang, ve Naren Ramakrishnan. "EMBERS autogsr: Sivil huzursuzluk olaylarının otomatik kodlanması." 22Nd ACM SIGKDD Uluslararası Bilgi Keşfi ve Veri Madenciliği Konferansı Bildirileri, s. 599-608. 2016.

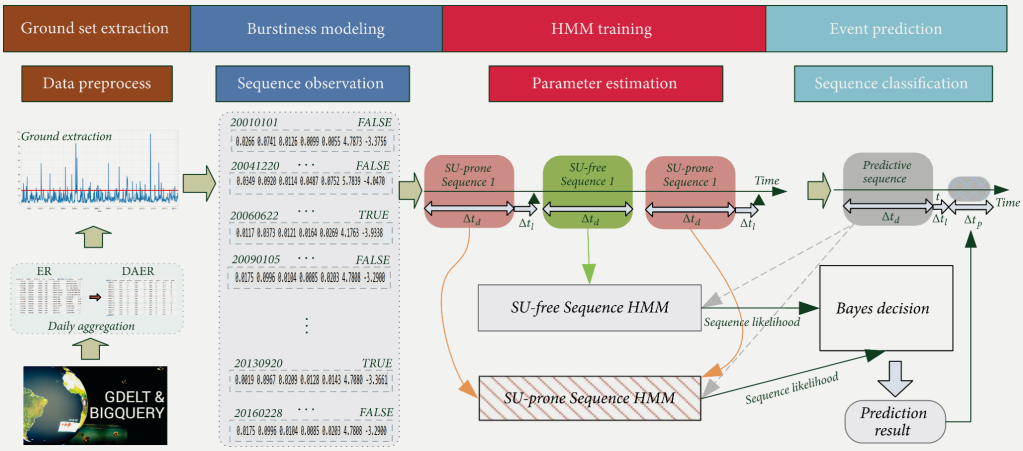
sayıda dilde özenle izliyor. Doğal dil işlemeyi kullanan GDELT, olaylar hakkında veri üretir, konumlarını belirler ve ilgili aktörleri tanımlar. Bu kapsamlı analiz, toplumsal huzursuzluğun kaynayan birikimini etkili bir şekilde tespit edebilir. Ayrıca, ICEWS olarak bilinen Entegre Kriz Erken Uyarı Sistemine sahibiz. Savunma İleri Araştırma Projeleri Ajansı (DARPA) tarafından başlatılan bu platform, protestolar veya ayaklanmalar gibi siyasi krizleri tahmin etmek için uluslararası haberler de dahil olmak üzere çok çeşitli verilerden yararlanıyor.[34] Son olarak, kolluk kuvvetlerinde kullanılan algoritma tabanlı bir yazılım olan PredPol uygulamasında farklı bir yaklaşım görülmektedir. PredPol, geçmiş verileri analiz ederek sivil huzursuzluk da dahil olmak üzere potansiyel suç faaliyeti alanlarını tahmin etmektedir.

Temel amaçları, potansiyel olarak yanıltıcı veya yanlış bilgileri başlangıçta hızlı bir şekilde tespit etmek ve işaretlemektir. Bu şekilde, doğruluk kontrolörleri derhal yanıt verebilir ve yanlış bilginin kök salmadan önce hızla yayılmasını durdurabilir. Bu bağlamda kullanılan etkili yöntemler arasında gerçek zamanlı veri toplama ve izleme yer almaktadır. Gelişmiş web kazıma ve veri tarama teknikleri, yeni bilgi ve iddialar bulmak için haber sitelerinden sosyal medya platformlarına ve resmi açıklamalara kadar çeşitli çevrimiçi kaynakları sürekli olarak tarar. Bu gerçek zamanlı veri toplama yöntemi, Import.io gibi web veri entegrasyon platformlarının kullandığı yaklaşımda olduğu gibi, doğruluk kontrolörlerinin en taze bilgilere anında erişmesini garanti eder.

Duygu analizi ve konu modelleme tekniklerinin bu çabada çok değerli olduğu kanıtlanmıştır. Trend konuların belirlenmesine ve belirli konuları çevreleyen duyarlılığın ölçülmesine yardımcı olurlar. Örneğin, IBM'in Watson Tone Analyzer'ı, olumsuz duygulardaki ani artışları veya belirli konular etrafındaki tartışmaları tespit edebilir; bu da potansiyel yanlış bilginin doğuşuna işaret edebilir. Twitter veya Facebook gibi platformlardaki paylaşımların viralliğini ve etkileşimini analiz etmek, hızla yayılan bilgilerin belirlenmesine yardımcı olur. Belirli bir içerikle yüksek düzeyde etkileşim genellikle acil doğruluk kontrolü yapılmasını gerektirir. Benzer şekilde, sosyal medya analiz araçları da bilginin farklı platformlardaki erişimini ve etkisini izleyerek olası yanlış bilgilerin etkili kaynaklarını ve yayıcılarını belirler.

Bu çabaları destekleyen teknoloji altyapısında, olay odaklı mimarilerin uygulanması, gerçek zamanlı işlemeye ve gelen verilere ve uyarılara anında yanıt verilmesine olanak tanır. Apache Kafka gibi sistemlerde kullanılan bu mimari, ortaya çıkan bilgilere dayalı olarak derhal harekete geçilmesini sağlar. Makine öğrenimi de özellikle anomali tespiti için bu çabalara yardımcı olur. Algoritmalar, veri modellerindeki anormallikleri tespit etmek üzere eğitilebilir, böylece potansiyel olarak yanlış bilgilerin yayılmasındaki ani artışlar belirlenebilir. Bu uygulamaya iyi bir örnek Twitter'ın gerçek zamanlı anomali tespit sistemidir. Doğal Dil İşleme (NLP), anahtar kelime çıkarma ve adlandırılmış varlık tanıma gibi tekniklerle, gerçek zamanlı veri akışlarındaki gerçek iddiaları ve potansiyel yanlış bilgileri tespit etmek için uygulanabilir.

34 Wang, Wei, Ryan Kennedy, David Lazer ve Naren Ramakrishnan. "Toplumsal olayların küresel olarak izlenmesi için büyüyen sancılar." Science 353, no. 6307 (2016): 1502-1503.



Parametre tahmini ve dizi sınıflandırması yoluyla tahminler üretmek için olay veri kümelerini entegre eden örnek bir iş akışı. Kaynak: Qiao, Fengcai, Pei Li, Xin Zhang, Zhaoyun Ding, Jiajun Cheng ve Hui Wang. "GDELT kullanarak gizli Markov modelleri ile sosyal huzursuzluk olaylarını tahmin etmek." Discrete Dynamics in Nature and Society 2017 (2017).

Bu teknikler, kontrol edilmeye değer iddiaların belirlenmesini kolaylaştırır. Benzer şekilde, anahtar kelime tabanlı uyarılar oluşturmak, yanlış bilgiyle ilişkili belirli terimler veya ifadeler tespit edildiğinde doğruluk kontrolörlerinin bildirim almasına yardımcı olur. Fact-checker'lar aynı anda birden fazla platformu izleyerek gözlemlerini daha da genişletebilirler. Bu, yanlış bilgiyi farklı kanallarda yaymaya yönelik koordineli çabaları ortaya çıkarabilir ve müdahale önlemlerinin etkinliğini artırabilir. Gerçek zamanlı doğrulama API'leri, doğruluk kontrol iş akışına entegre edilerek iddiaların doğruluğunun hızlı bir şekilde değerlendirilmesini ve zamanında yanıt verilmesini sağlayabilir. Örneğin, Google Fact Check Tools API gerçek zamanlı doğrulama yetenekleri sağlar. İşbirliğine dayalı uyarı sistemleri, erken uyarı ve çürütme konusunda kolektif bir yaklaşımı teşvik eder.

Kullanıcıların ve doğruluk kontrolcülerinin potansiyel yanlış bilgileri göndermelerine ve gerçek zamanlı uyarılar almalarına olanak tanıyan platformlar, yanlış bilgi saldırısına karşı iş birliğine dayalı bir duvar inşa edilmesine yardımcı olur. Deeptrace Labs tarafından kullanılanlar gibi deepfake'lerin ve manipüle edilmiş medyanın erken tespiti için gelişmiş algoritmalar, erken teşhis ve çürütmeye yardımcı olarak ek bir savunma hattı sağlar. Gerçek zamanlı uyarıların ve erken uyarı sistemlerinin etkinliği büyük ölçüde veri toplama kalitesine ve hızına, tespit algoritmalarının hassasiyetine ve doğruluk kontrolü yapan kuruluşlar ve bireyler arasındaki zamanında iş birliğine bağlıdır. Bu sistemler, gelişmiş teknolojilerden ve iş birliğine dayalı çabalardan yararlanarak yanlış bilginin etkisini azaltmada önemli bir rol oynar ve böylece gerçek zamanlı olarak paylaşılan bilgilerin güvenilirliğini artırır.

4. Veri Mahremiyeti ve Kişisel Veri Koruma:

Kitle kaynak kullanımı ve iş birliğine dayalı doğruluk kontrol çabaları söz konusu olduğunda, çeşitli tarafların ve katkıda bulunanların dahil olduğu göz önüne alındığında, gizlilik ve veri korumanın önemi artmaktadır. Hassas bilgilerin

güvenli ve gizli kalmasını sağlamak çok önemli hale gelmektedir. Bu güvenliği sağlamak için gelişmiş yöntemler ve teknikler geliştirilmiş ve uygulanmıştır. Veri minimizasyonu, gizliliğin korunmasında kilit bir rol oynar. Veri minimizasyonu uygulamalarını benimseyen kuruluşlar, yalnızca gerekli ve ilgili verilerin toplanmasını ve saklanmasını sağlayarak hassas bilgilerin açığa çıkma riskini azaltır. Örneğin, Apple'ın kullanıcı verilerine ilişkin veri minimizasyonu yaklaşımı teknoloji sektöründe bir ölçüt olarak övgüyle karşılanmıştır.

Katkıda bulunanların ve kullanıcıların kimliklerinin daha fazla korunmasını sağlamak için, veriler genellikle anonimleştirilir veya takma ad verilir. Bu, verilerin bir bireye kadar izlenmesini son derece zorlaştırır. Diferansiyel gizlilik teknikleri, verilere gürültü ekleyerek bir adım daha ileri gider, bu da bireysel gizlilik koruması sağlar, ancak yine de doğru bir toplu analiz sağlar. Verilerin iletim ve depolama sırasında güvende tutulması, veri korumanın bir diğer hayati yönüdür. Bunu başarmak için şifreleme ve güvenli yuva katmanı (SSL) gibi güvenli protokoller uygulanır.[35] Bu, verilerin yanlış ellere geçmesini önler. Benzer şekilde, erişim kontrolü ve rol tabanlı izinler, hassas verilere erişimi sınırlandırmaya hizmet ederek yalnızca yetkili personelin bu tür bilgileri görüntüleyebilmesini ve işleyebilmesini sağlar.

Bir diğer kritik süreç de Veri Koruma Etki Değerlendirmelerinin (DPIA'lar) yapılmasıdır.[36] DPIA'lar kitle kaynak kullanımı ve doğruluk kontrolü sürecindeki potansiyel gizlilik risklerini belirleyip ele alabilir ve belirlenen riskleri azaltmak için proaktif önlemler alınmasına olanak tanır. Federe öğrenme ve güvenli çok partili hesaplama gibi gizliliği koruyan makine öğrenimi teknikleri, model eğitimini mümkün kılarken gizliliğin korunmasında kullanışlı olmaktadır. Bu yaklaşım, kişiselleştirilmiş metin tahminleri için Google'ın Gboard klavye uygulaması tarafından kullanılan bir uygulama olan ham verilerin paylaşılmasına gerek kalmadan model eğitimine olanak tanır. Katılımcı kuruluşlar arasında güvenilir veri paylaşım anlaşmaları, gizlilik düzenlemelerine uyulmasını sağlamaya ve veriye katkıda bulunanların haklarını korumaya hizmet eder. Verilerin nasıl kullanılabileceği konusunda net kurallar ve sınırlar belirlerler. Şeffaflık ve rıza, gizliliğin korunmasında iki temel unsurdur. Veri toplama ve işleme için açık kullanıcı onayı alınması esastır. Veri kullanımı hakkında açık ve şeffaf bilgiler sağlamak, kullanıcıların verileri hakkında bilinçli kararlar vermelerini sağlar.

Her doğruluk kontrol platformu, kullanıcı verilerinin nasıl işlendiğini özetleyen ve kullanıcıların gizlilik haklarını kullanmaları için mekanizmalar sağlayan kapsamlı bir gizlilik politikasına sahip olmalıdır. Bu politikalar açık olmalı ve kullanıcılar tarafından kolayca erişilebilir olmalıdır. Düzenli denetim ve uygunluk takibi, gizlilik önlemlerine sürekli olarak uyulmasını ve olası ihlallerin tespit edilip derhal ele alınmasını sağlar. Bu süreç, kullanıcının sisteme olan güvenini ve itimadını korumaya yardımcı olur. Veri silme ve saklama politikaları, verilerin gerekenden daha uzun süre saklanmaması ve artık ihtiyaç duyulmadığında uygun şekilde imha edilmesi gerektiğini belirtir. Bu politikalar verilerin atıl kalmamasını ve tehlikeye girme riski taşımamasını sağlar.

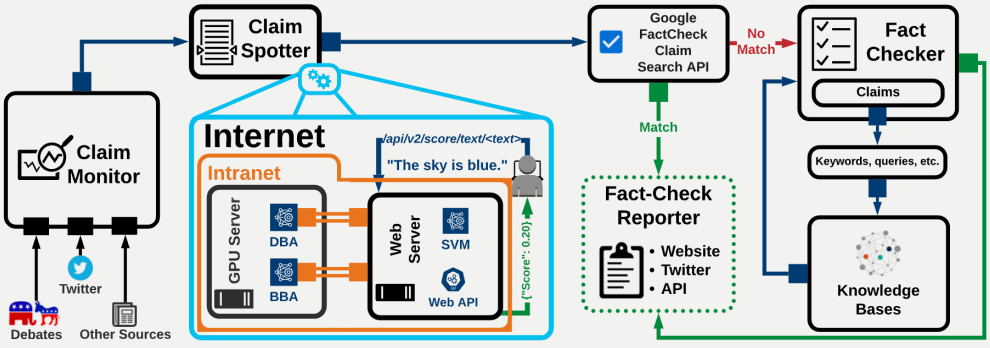
35 Dastres, Roza ve Mohsen Soori. "Ağ ve web güvenliğinde güvenli soket katmanı (SSL)." Uluslararası Bilgisayar ve Bilişim Mühendisliği Dergisi 14, no. 10 (2020): 330-333.

36 Demetzou, Katerina. "Veri Koruma Etki Değerlendirmesi: Hesap verebilirlik için bir araç ve Genel Veri Koruma Tüzüğü'nde netleştirilmemiş 'yüksek risk' kavramı." Computer Law & Security Review 35, no. 6 (2019): 105342.

Kişisel olarak tanımlanabilir bilgiler (PII) gibi hassas verilerle ilgili durumlarda, kullanıcıların açık ve bilgilendirilmiş onayı tartışılmaz bir gerekliliktir.[37] Ayrıca, uçtan uca şifrelenmiş mesajlaşma gibi güvenli iletişim kanallarının kullanılması, veri paylaşımı ve iletişim sırasında hassas bilgilerin korunmasını sağlar. Son olarak, üçüncü taraf tedarikçiler söz konusuysa, verileri korumak için sıkı güvenlik ve gizlilik standartlarına uyduklarından emin olmak çok önemlidir. Bu titizlik, koruyucu önlemleri doğrudan kuruluşun ötesine taşır. Kuruluşlar, bu gelişmiş gizlilik ve veri koruma yöntemlerini kitle kaynak kullanımı ve iş birliğine dayalı doğruluk kontrolü sürecine dahil ederek, hassas bilgilerin bütünlüğünü ve gizliliğini korurken katkıda bulunanlar ve kullanıcılar arasında güven oluşturabilir. Bu önlemlerin uygulanması sadece iyi bir uygulama olmakla kalmaz, aynı zamanda gizlilik düzenlemeleriyle de uyumludur ve doğruluk kontrolü ekosisteminde sorumlu veri işleme uygulamalarını teşvik eder.

5. Blockchain'in Veri Bütünlüğündeki Rolü:

Merkezi olmaması ve değişmezliği ile bilinen blok zinciri teknolojisi, kitle kaynak kullanımı ve iş birliğine dayalı doğruluk kontrolü alanında oyunun kurallarını değiştiren bir unsur olduğunu kanıtlamaktadır. Benzersiz özellikleri veri bütünlüğünü desteklemekte, şeffaflığı artırmakta ve katılımcılar arasında güveni teşvik etmektedir. Bu alanda blok zincirinin çeşitli yenilikçi uygulamaları ortaya çıkmıştır. Blok zinciri teknolojisi, veri katkılarının ve doğruluk kontrol sonuçlarının değişmez bir kaydını oluşturarak veri kaynağı ve bütünlüğü sağlar. Bu, bir ürünün yaşam döngüsünün her aşamasının izlenebildiği ve doğrulanabildiği tedarik zinciri yönetiminde blok zincirinin nasıl kullanıldığına benzer.[38]



ClaimBuster's blueprint for multimodal fact-checking. Sources: https://idir.uta.edu/claimbuster/

Bu bağlamda, blok zincirine yapılan her ekleme veya güncelleme kriptografik olarak bağlantılıdır ve verilerin izlenebilirliğini ve bütünlüğünü sağlar. Blockchain'in merkezi olmayan yapısı, merkezi olmayan bir doğruluk kontrol platformunun oluşturulmasına olanak tanır. Burada, çok sayıda katılımcı merkezi bir otoriteye bağlı kalmadan

37 Schwartz, Paul M., ve Daniel J. Solove. "The PII problem: Privacy and a new concept of personal identifiable information." NYUL rev. 86 (2011): 1814.
38 Edwardsson, Malin Picha ve Walid Al-Saqaf. "Küresel bir doğruluk kontrol veritabanı oluşturmak için blok zinciri teknolojisini kullanmanın itici güçleri ve engelleri." Online Journal of Communication and Media Technologies 12, no. 4 (2022): e202228.

bilgiye katkıda bulunabilir ve doğrulayabilir, bu da daha demokratik ve güvenilir bir sistem yaratır. Ayrıca veriler ve iddialar için güvenilir zaman damgası sunarak bilgilerin blok zincirine ne zaman eklendiğini gösterir. Zaman damgası doğrulaması, Bitcoin işlemlerinin blok zincirinde zaman damgalı ve doğrulanmış olmasına benzer şekilde, doğruluk kontrol süreci sırasında verilerin doğruluğunu ve güncelliğini doğrulamada çok önemli hale gelir.[39] Blok zincirindeki akıllı sözleşmeler, kitle kaynak kullanımı ve iş birliğine dayalı doğruluk kontrolünde iş birliği sürecini otomatikleştirebilir.

Bu sözleşmeler kuralları, ödülleri ve cezaları tanımlamak için kullanılabilir, böylece katılımcılar arasında şeffaflık ve adalet sağlanır. Bu, Ethereum'da aracılara ihtiyaç duymadan anlaşmaları uygulamak için akıllı sözleşmelerin nasıl kullanıldığına benzerlik göstermektedir. Blok zinciri tabanlı platformlar, katkıda bulunanları doğruluk kontrol çabaları için teşvik etmek amacıyla token veya kripto para birimleri kullanabilir. Böyle bir mekanizma, blok zinciri tabanlı içerik platformlarının yaratıcıları ödüllendirmesine benzer şekilde, daha fazla katılımı teşvik edebilir ve sonuçların doğruluğunu artırabilir. Blockchain'in Proof of Work (PoW) veya Proof of Stake (PoS) gibi mutabakat mekanizmaları, katılımcılar arasında veri doğrulama ve anlaşma için kullanılabilir.[40] Bu mekanizmalar, bilginin çoğunluk tarafından kabul edilmesini ve doğrulanmasını sağlayarak verinin güvenilirliğine katkıda bulunur. Blok zinciri ayrıca gizlilik ve veri koruması için bir araç olarak da kullanılabilir. Kullanıcıların belirli bilgileri ifşa etmeden bu bilgilere sahip olduklarını kanıtlamalarına olanak tanıyan sıfır bilgi ispatı gibi teknikler, kullanıcıların verileri üzerindeki kontrolü ellerinde tutmalarına yardımcı olurken aynı zamanda doğruluk kontrol sürecine de katkıda bulunur.[41] Teknoloji ayrıca iş birliğine dayalı bilgi grafiklerinin oluşturulmasını da kolaylaştırabilir.

Doğrulanmış gerçekler ve veriler, merkezi olmayan ve kurcalamaya karşı dayanıklı bir şekilde saklanır ve daha önce doğruluğu kontrol edilmiş bilgilere kolay erişim ve referans sağlar. Blok zinciri, katılımcılar için itibar ve güven sistemlerini teşvik ederek katkıda bulunanların güvenilirliğini artırabilir. Bu, bir varlığın itibarının izlenebildiği ve doğrulanabildiği blok zincirindeki merkezi olmayan kimliklere benzer. Platformlar arası veri alışverişi, blok zincirinin üstün olduğu bir başka alandır. Farklı doğruluk kontrol platformları ve veritabanları arasında güvenli ve standartlaştırılmış bir veri alışverişi yolu sağlayarak birlikte çalışabilirliği teşvik eder ve doğruluk kontrol kuruluşları arasında bilgi paylaşımını kolaylaştırır. Ayrıca, blok zincirinin dağıtılmış defteri, ağ genelinde veri tutarlılığını sağlar ve böylece bilgilerin çoğaltılmasını veya değiştirilmesini önler.

Son olarak, doğruluğu kontrol edilen tüm iddialar blok zincirinde değişmez kayıtlar olarak saklanabilir ve doğrulanmış bilgilerin kapsamlı bir arşivi oluşturulabilir. Blok zinciri teknolojisinden yararlanmak, kitle kaynak kullanımı ve iş birliğine dayalı

39 Dwivedi, Ashutosh Dhar, Rajani Singh, Sakshi Dhall, Gautam Srivastava ve Saibal K. Pal. "Ölçeklenebilir bir blok zinciri dağıtılmış ağı kullanarak sahte haberlerin kaynağını izleme." 2020 IEEE 17. uluslararası mobil ad hoc ve sensör sistemleri konferansı (MASS) içinde, s. 38-43. IEEE, 2020.

40 Akbar, Nur Arifin, Amgad Muneer, Narmine ElHakim ve Suliman Mohamed Fati. "Proof-of-work ve proof-of-stake blockchain konsensüsleri için dağıtık hibrit çift harcama saldırısı önleme mekanizması." Future Internet 13, no. 11 (2021): 285.

41 Sun, Xiaoqiang, F. Richard Yu, Peng Zhang, Zhiwei Sun, Weixin Xie ve Xiang Peng. "Blok zincirinde sıfır bilgi kanıtı üzerine bir araştırma." IEEE ağı 35, no. 4 (2021): 198-205.

doğruluk kontrol girişimlerinin güvenilirliğini, şeffaflığını ve verimliliğini artırabilirken, büyük ölçekli doğruluk kontrol projeleri için kullanılırken blok zinciri sistemlerinin ölçeklenebilirlik ve enerji tüketimi endişelerini akılda tutmak çok önemlidir. Sonuç olarak, İşbirlikçi Platformlar Arası İzleme, doğruluk kontrolü ve bilgi doğrulamada dönüştürücü bir çağın habercisidir. Yeni teknolojilerin ve tekniklerin kullanılması, doğruluk kontrolörlerinin geniş veri setlerini taramasına, yanlış bilgi eğilimlerini tespit etmesine ve çeşitli çevrimiçi platformlarda koordineli dezenformasyon kampanyalarını ortaya çıkarmasına olanak tanıyor. Teknolojik gelişmelerle desteklenen bu platformlar arası iş birliği, yanlış bilgiyle mücadelede birleşik bir cephe oluşturarak daha bilinçli ve dirençli bir bilgi ekosistemini teşvik etmektedir. Zorluklar devam etse de, İşbirliğine Dayalı Çapraz Platform İzlemenin yanlış bilgiyle mücadelede sunduğu potansiyel faydalar, bu yöntemi doğruluk kontrolörleri ve bilgi doğrulama girişimleri için cazip bir strateji haline getirmektedir.

AÇIK KAYNAK SORUŞTURMASI/ANALİZİ (OSINT)

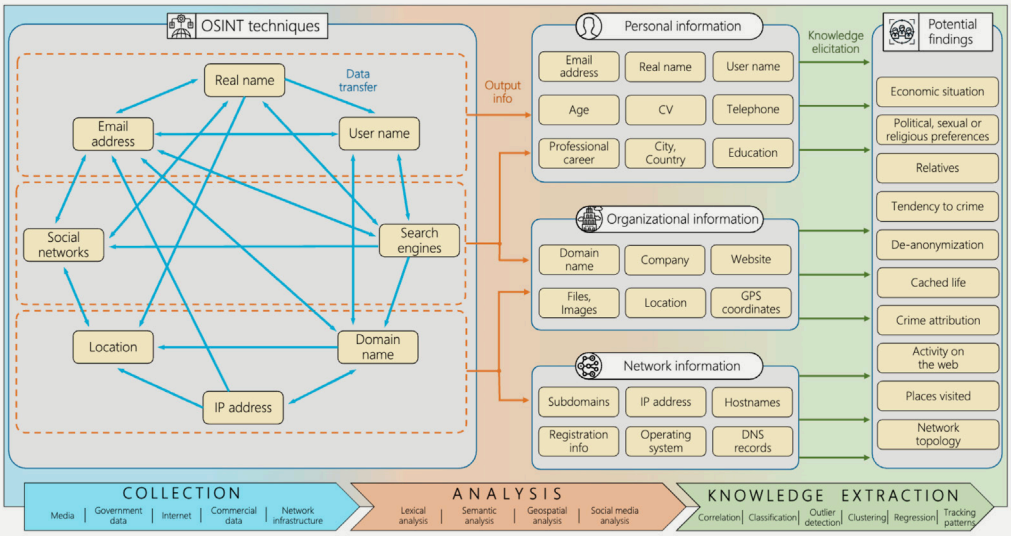
Fact-checker'lar iddiaları doğrulamak ve yanlış bilgileri çürütmek için açık kaynaklı istihbarattan ve kamuya açık verilerden yararlanabilir. Kitle kaynak kullanımı ve iş birliğine dayalı platformlardan yararlanmak, doğruluk kontrolörlerinin çok çeşitli uzmanlık ve kaynaklara erişmesini sağlayabilir. Açık Kaynak Araştırması (OSINT), iddiaları doğrulamak, yanlış bilgileri tespit etmek ve gizli gerçekleri ortaya çıkarmak için çeşitli kaynaklardan kamuya açık bilgilerden yararlanarak doğruluk kontrolü ve bilgi doğrulamada güçlü bir araç olarak ortaya çıkmıştır. OSINT ile ilişkili yeni teknolojilerin ve tekniklerin teknik özelliklerine odaklanan bu genişletilmiş analiz, OSINT'in doğruluk kontrolü ve bilgi doğrulama üzerindeki etkisini araştırmaktadır.

1. Dijital Forensik ve Görüntü Doğrulama:

Açık kaynak analizi (OSINT), çevrimiçi olarak paylaşılan görüntü ve videoların doğrulanması için büyük ölçüde dijital adli tıp tekniklerine dayanan kritik bir araçtır. Görsel içeriği doğrulamak ve manipüle edilmiş ya da yanıltıcı görüntüleri tespit etmek için, doğruluk kontrolörleri görüntü ters arama, meta veri analizi ve bağlamsal ipuçlarının araştırılması dahil olmak üzere bir dizi araç ve teknik kullanmaktadır. Mevcut dijital çağda, bu araçlar ve teknikler bilginin bütünlüğünün sağlanmasında ve kamuoyunun aldatıcı uygulamalardan korunmasında önemli bir rol oynamaktadır. Özellikle de yabancı bilgi manipülasyonu ve müdahalesine karşı mücadelede çok önemlidirler.

Bu gelişmelere öncülük eden derin öğrenme alanı, özellikle deepfake'ler şeklinde tahrif edilmiş veya manipüle edilmiş medya içeriğini ayırt etmek için benimsenmiştir. Generative Adversarial Networks (GANs) dahil olmak üzere bu yapay zeka modelleri sadece deepfake'leri yaratmak için değil aynı zamanda tespit etmek

için de kullanılmaktadır.[42] GAN'lar esasen iki yapay zeka sistemi arasındaki bir yarışmadır - biri deepfake'i yaratır ve diğeri onu tespit etmeye çalışır, böylece tespit yeteneğini geliştirir. Medya adli biliminin önemli bir parçası, görüntülerdeki değişiklikleri tanımlamayı veya bir görüntünün orijinal kaynağını ortaya çıkarmayı içeren görüntü adli bilimidir. Hata Seviyesi Analizi (ELA) ve JPEG Hayalet Tespiti gibi teknikler bu konuda sıklıkla kullanılmaktadır.[43] ELA, bir görüntünün farklı sıkıştırma seviyelerindeki alanlarını tanımlayarak olası değişiklik alanlarını vurgular. Öte yandan JPEG Hayalet Tespiti, JPEG sıkıştırma tutarsızlıklarına dayalı olarak görüntünün bir kısmının görüntünün geri kalanıyla karşılaştırılarak manipüle edilip edilmediğini tespit etmeye yardımcı olur. İşin video tarafında, video karelerinin piksel düzeyinde analizi yoluyla tespit edilebilen aydınlatma, gölgeler ve hatta kalp atış hızı veya nefes alma hızı gibi ince fizyolojik sinyaller gibi video öğelerindeki tutarsızlıkların incelenmesine dayanan kurcalama tespit teknikleri sıklıkla kullanılır. Bu arada, ses adli tıp alanı, ses dosyalarındaki manipülasyonları ortaya çıkarmak için spektrogram analizi, konuşmacının kimliğini doğrulamak için ses biyometrisi ve hatta yapay olarak sentezlenmiş konuşmayı tanımlamak için gelişmiş yapay zeka modelleri gibi tekniklerle doludur.



Kaynak: Pastor-Galindo, Javier, Pantaleone Nespoli, Félix Gómez Mármol, ve Gregorio Martínez Pérez. "OSINT'in henüz sömürülmemiş altın madeni: Fırsatlar, açık zorluklar ve gelecekteki eğilimler." IEEE Access 8 (2020): 10282-10304.

42 Parveen, Azra, Zishan Husain Khan ve Syed Naseem Ahmad. "Dijital adli tıp araçlarının sınıflandırılması ve değerlendirilmesi." TELKOMNIKA (Telekomünikasyon Hesaplama Elektronik ve Kontrol) 18, no. 6 (2020): 3096-3106.

43 Harish Kumar, J., ve T. Kirthiga Devi. "Meta Verilere ve İstatistiksel Analize Dayalı Görüntü Dosyalarının Parmak İzi." Uluslararası Derin Öğrenme, Hesaplama ve Zeka Konferansı Bildirileri: ICDCI 2021, s. 105-118. Singapur: Springer Nature Singapore, 2022.

2. Coğrafi Bilgi Sistemler (GIS) ve Haritalama:

Coğrafi konum belirleme ve haritalama araçları gibi açık kaynaklı analiz (OSINT) teknikleri, olayların konumunun ve bağlamının doğrulanması için çok önemlidir. Bu yöntemler doğruluk kontrolcülerinin uydu görüntülerini, coğrafi etiketli verileri ve konuma dayalı bilgileri kullanarak iddiaları doğrulamasına ve yanlış anlatıları çürütmesine olanak tanır. Bu yöntemlerin kullanımı özellikle yabancı bilgi manipülasyonu ve müdahalesiyle mücadelede önem taşımakta olup, bilginin kaynağının doğrulanmasına ve yabancı aktörler tarafından yürütülen potansiyel yanlış bilgilendirme kampanyalarının tespit edilmesine yardımcı olmaktadır.[44]

Örneğin, sosyal medya paylaşımlarının coğrafi konumu, bilgilerin gerçekliğini doğrulamak ve yabancı konumlardan kaynaklanan potansiyel yanlış bilgilendirme kampanyalarını tespit etmek için güçlü bir araç olarak kullanılabilir. Gelişmiş algoritmalar, bu paylaşımların içeriğini ve meta verilerini analiz ederek kaynak konumlarını belirleyebilir.[45] Bu teknik, Ukrayna'daki çatışma sırasında paylaşılanlar da dahil olmak üzere çeşitli gönderi ve görsellerin gerçek konumunun belirlenmesinde çok önemli olmuştur. Ayrıca, coğrafi konum belirleme, farklı coğrafi bölgelerdeki belirli bilgilere yönelik duyguları ve tepkileri ayırt eden bir süreç olan duygu analizi ile de birleştirilebilir. Bu yaklaşım, dünya çapında çeşitli siyasi olaylara verilen tepkilerde görüldüğü gibi, konuma bağlı olarak yanlış bilginin etkisindeki farklılıkları belirlemeye yardımcı olur.

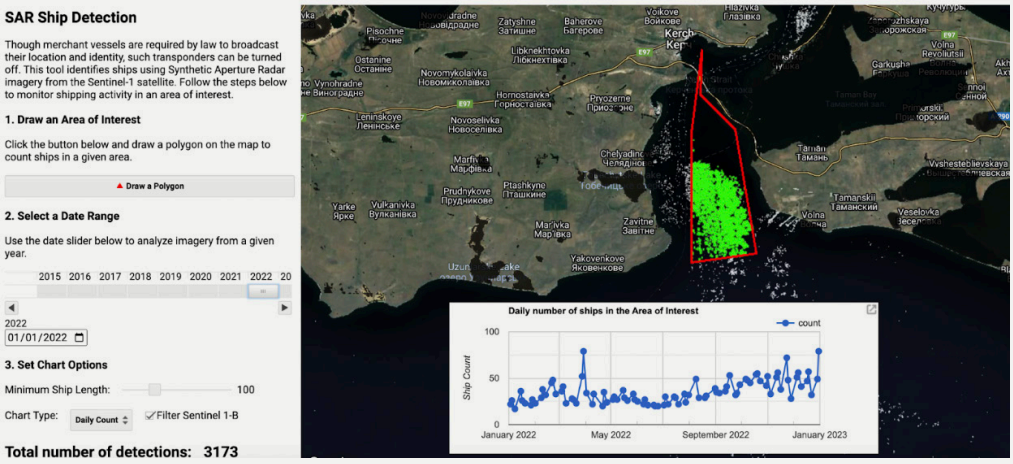
Yabancı aktörler tarafından bırakılan dijital ayak izinin analiz edilmesi, coğrafi konumlarının ve yanlış bilgilendirme kampanyalarına olası katılımlarının belirlenmesine yardımcı olabilir. Bu analiz, gelişmiş IP adresi izleme teknikleriyle birleştiğinde, çevrimiçi bilgilerin kaynağını belirleyebilir ve hatta verilerin kaynağını gizlemek için VPN kullanımını tespit edebilir. Bu taktikler, Rusya'nın 2016 ABD seçimlerine müdahalesine yönelik soruşturmalar sırasında çok önemli bir rol oynamıştır. Görselleştirmenin gücü de küçümsenmemelidir. Haritalama ve görselleştirme araçları bilginin coğrafi dağılımını etkili bir şekilde gösterebilir ve bölgeler arasındaki örüntülerin ve anormalliklerin belirlenmesine yardımcı olabilir. Uydu görüntüleri, Çin'in Sincan bölgesindeki insan hakları ihlallerinin doğrulanmasında görüldüğü gibi, olayların ve bilgilerin doğrulanması için güçlü bir araç olarak hizmet ederek belirli konumların gerçek dünya görünümünü sağlayabilir.

Yerel dillerde doğruluk kontrolü, belirli bölgeleri hedef alan yanlış bilgilendirme kampanyalarıyla mücadeleye yardımcı olur. Sosyal medya bağlantılarının ve yabancı aktörler ile yerel etkileyiciler arasındaki etkileşimlerin analiz edilmesi potansiyel manipülasyon çabalarını ortaya çıkarabilir. Fact-checker'lar ayrıca kaynakların gerçekliğini doğrulamak ve belirli olaylara veya olaylara yakınlıklarını teyit etmek için coğrafi konumu kullanabilir. Bilgilerin resmi devlet

44 Yadav, Ashok, Atul Kumar ve Vrijendra Singh. "Açık kaynak zekası: siber güvenlikte mevcut durum, uygulamalar ve gelecek perspektifleri üzerine kapsamlı bir inceleme." Yapay Zeka İncelemesi (2023): 1-32.

45 Evangelista, João Rafael Gonçalves, Renato José Sassi, Márcio Romero ve Domingos Napolitano. "Açık kaynak istihbaratının (OSINT) yapay zeka ile uygulanmasını araştırmak için sistematik literatür taraması." Uygulamalı Güvenlik Araştırmaları Dergisi 16, no. 3 (2021): 345-369.

verileriyle veya kamuya açık kaynaklarla çapraz referanslandırılması, iddiaların doğrulanmasında bir başka paha biçilmez tekniktir. Örneğin, COVID-19 salgını sırasında, doğruluk kontrolörleri vaka sayıları ve ölüm oranları hakkındaki iddiaları sağlık departmanlarından ve Dünya Sağlık Örgütü'nden alınan verilerle sık sık çapraz referanslandırmıştır.



Bellingcat'in Ukrayna açıklarında Rus silah ambargosunu delen gemileri tespit etmek için açık kaynak Sentetik Açıklıklı Radar kullanma çalışması. Kaynak: https://www.bellingcat.com/news/2023/05/11/grain-trail-tracking-russias-ghost-ships-with-satellite-imagery/

Fact-checker'lar ayrıca yerel uzmanlar ve kuruluşlarla iş birliği yaparak coğrafi konum verilerinin doğruluğunu artırabilir ve değerli bir bağlam sağlayabilir. Coğrafi etiketli görüntü ve videoların gerçekliğini doğrulama teknikleri, bunların manipüle edilmemesini veya yanıltıcı olmamasını sağlar. Bunlar, örneğin doğal afetler veya şiddet olaylarıyla ilgili görüntülerin doğrulanmasında faydalı olmuştur. Coğrafi konum belirleme yöntemleri aracılığıyla gerçek zamanlı konum doğrulama, doğruluk kontrolörlerinin ortaya çıkan yanlış bilgi kampanyalarına hızlı bir şekilde yanıt vermesini sağlar. Bu zamanında müdahale, yanlış bilginin gerçek dünyada anında etki yaratabileceği seçimler ve referandumlar gibi yüksek profilli etkinlikler sırasında çok önemliydi. Coğrafi konum belirleme ve haritalama tekniklerinin diğer doğruluk kontrol yöntemleriyle birleştirilmesi, yabancı bilgi manipülasyonu ve müdahalesiyle mücadele çabalarının etkinliğini artırmaktadır. Gelişmiş teknolojilerden yararlanmak ve yerel uzmanlarla iş birliği yapmak, doğruluk kontrolörlerinin yanlış bilginin kaynaklarını ve kökenlerini daha iyi tespit etmelerini sağlar, böylece doğru ve güvenilir bilginin halka ulaşmasını sağlar.

3. Karanlık Web İzleme:

Açık kaynak istihbaratı (OSINT) bazı durumlarda dezenformasyon kampanyalarını ve kötü niyetli aktörleri tespit etmek için karanlık web'in izlenmesini gerektirir. Karanlık web'de gezinmek çeşitli zorluklar yaratsa da, internetin bu gizli köşelerinden ortaya çıkarılabilecek bilgiler, yanlış bilgi yaymaya yönelik gizli çabaları ortaya

çıkarmak için paha biçilmezdir.[46] Doğruluk kontrolü ve yabancı bilgi manipülasyonu ve müdahalesiyle mücadelede karanlık web izlemenin önemi küçümsenemez. Genellikle yasadışı faaliyetlerle bağlantılı olan dark web, erişim için özel bir yazılım gerektiren deep web'in bir bileşenidir ve yanlış bilgi, dezenformasyon ve diğer zararlı içerik türlerinin yayılması için gelişen bir pazar haline gelmiştir. Karanlık ağın karmaşık ağında gezinmek için, derin ve karanlık web taraması gibi gelişmiş yöntemler ve teknikler kullanılmaktadır.[47] Doğruluk kontrolörleri, özel araçlar ve hizmetler kullanarak gizli web sitelerine ve forumlara erişebilir, daha geniş bir bilgi yelpazesinin toplanmasını sağlayabilir ve potansiyel yanlış bilgi kaynaklarını ortaya çıkarabilir. Bu tür araştırmaların bir örneği, kolluk kuvvetlerinin yasadışı uyuşturucu pazarlarını belirlemek ve izlemek için benzer teknikler kullandığı uyuşturucu kaçakçılığıyla mücadelede görülmüştür.

Bu çabada, karanlık web forumlarından, sohbet günlüklerinden ve tartışmalardan metin içeriğini analiz etmek için kullanılan NLP teknikleri de aynı derecede önemlidir. NLP aracılığıyla duygu analizi ve konu modellemesi, koordineli dezenformasyon kampanyalarının tespit edilmesine ve trend olan yanlış bilgilendirme konularının belirlenmesine yardımcı olabilir. Bu tür uygulamalardan biri, 2020 ABD Başkanlık Seçimleri sırasında NLP'nin kamuoyunu manipüle etmeyi amaçlayan dezenformasyon kampanyalarını ortaya çıkarmak için kullanılmasıydı.[48] Görsel içerik analizi, doğruluk kontrolörlerinin kullandığı bir başka tekniktir. Gelişmiş görüntü ve video analizi, dark web'de paylaşılan potansiyel olarak manipüle edilmiş veya yanıltıcı görsel içeriği tespit ve analiz edebilir; derin öğrenme algoritmaları, deepfake'leri tespit etmek ve görüntü ve videolardaki tutarsızlıkları belirlemek için araçlar olarak hizmet eder. Dark web'in belirgin özelliklerinden biri de kripto para birimlerinin yaygınlığıdır.

Kripto para işlemlerinin izlenmesi, yanlış bilgilendirme kampanyalarının finansman kaynaklarının izlenmesine ve aktörler arasındaki mali bağlantıların tespit edilmesine yardımcı olur. Bu ödemelerin izlenmesi, fidye yazılımı saldırılarına yönelik soruşturmalara benzer şekilde, yabancı müdahale çabalarıyla ilgili finansman modellerini ortaya çıkarabilir. Aynı şekilde, dark web sahte belgeler, hacklenmiş veriler ve diğer dezenformasyon türlerini satan ve dağıtan pazarlara da ev sahipliği yapmaktadır. Doğruluk kontrol uzmanları bu platformları izleyerek, tıpkı güvenlik araştırmacılarının siber suç trendlerini takip ederken yaptığı gibi, yanlış bilgiyle ilgili ürünlerin mevcudiyetini ve talebini ortaya çıkarabilir. Yapay zeka, dark web izlemede de ayrılmaz bir rol oynamaktadır.

Makine öğrenimi modelleri, yanlış bilgilendirme kampanyalarıyla ilişkili belirli anahtar kelimeleri veya ifadeleri tanımak için eğitilebilir ve yabancı bilgi manipülasyonu ve müdahalesine işaret edebilecek kalıpları ve anormallikleri belirlemede önemli

46 He, Siyu, Yongzhong He, ve Mingzhe Li. "Karanlık web üzerindeki yasadışı faaliyetlerin sınıflandırılması." İçinde 2. Uluslararası Bilgi Bilimi ve Sistemleri Konferansı Bildirileri, s. 73-78. 2019.

47 Rawat, Romil, Vinod Mahor, Sachin Chirgaiya ve Bhagwati Garg. "Karanlık web siber suçlusu tarafından teknolojik özellikli otomatik şehir altyapısının yapay siber casusluk tabanlı korunması." Nesnelerin Zekası: AI-IoT Tabanlı Kritik Uygulamalar ve Yenilikler (2021): 167-188.

48 Ibrishimova, Marina Danchovsky, ve Kin Fun Li. "Bilgi doğrulama ve doğal dil işleme kullanarak sahte haber tespitine yönelik bir makine öğrenimi yaklaşımı." Akıllı Ağ ve İşbirlikçi Sistemlerdeki Gelişmeler: 11. Uluslararası Akıllı Ağ ve İşbirlikçi Sistemler Konferansı (INCoS-2019), s. 223-234. Springer Uluslararası Yayıncılık, 2020.

yardım sağlar. Doğruluk kontrol kuruluşları, dark web verilerine erişmek ve analiz etmek ve potansiyel tehditleri belirlemek için genellikle kolluk kuvvetleri ve istihbarat kurumlarıyla iş birliği yapar. Terörizm ve siber suçlarla mücadele için kurulanlara benzer bu iş birliği ilişkileri, dark web izleme çabalarının etkinliğini artırır. Dark web verileriyle uğraşırken veri gizliliği ve güvenlik önlemleri çok önemlidir. Şifreleme ve güvenli veri depolama hassas bilgileri korur ve araştırmacıların ve doğruluk kontrolcülerinin kimliklerini güvence altına alır. Bu uygulamalar, düşmanca ortamlarda çalışan veya hassas bilgilerle uğraşan gazeteciler ve aktivistler tarafından kullanılanlardan farklı değildir. Yanlış bilgi dil sınırı tanımadığından, dark web içeriğinin çeşitli dillerde izlenmesi çok önemlidir. Gelişmiş dil işleme yetenekleri, güvenlik araştırmacılarının uluslararası siber tehditleri izlemelerine benzer şekilde, farklı dillerdeki içeriklerin analiz edilmesine yardımcı olur.

Karanlık web istihbarat platformları, karanlık web faaliyetlerini izlemek ve analiz etmek için özel araçlar ve veri akışları sunar. Bu kaynaklar, siber güvenlik firmalarının tehditleri izlemek için kullandıklarına benzer şekilde, yanlış bilgileri tespit etme ve bunlarla mücadele etme konusunda doğruluk kontrol kuruluşlarına önemli destek sağlar. Dark web faaliyetlerinin gerçek zamanlı olarak izlenmesi, ortaya çıkan tehditlere ve yanlış bilgilendirme kampanyalarına zamanında yanıt verilmesini sağlar. Bu anında tespit, gerçek zamanlı siber tehditlerle başa çıkmada siber güvenlik olay müdahalesine benzer şekilde, hızlı doğruluk kontrolü ve çürütme sağlar. Dark web izleme faaliyetinde bulunmak siber güvenlik, veri analizi ve yasal hususların anlaşılması konularında güçlü bir temel gerektirir. Sorumlu ve etkili bir izleme sağlamak için ilgili makamlar ve uzmanlarla iş birliği yapmak gibi etik hususlar da son derece önemlidir. Doğruluk kontrol kuruluşları bu gelişmiş yöntem ve teknikleri kullanarak yabancı bilgi manipülasyonu ve müdahalesini proaktif bir şekilde tespit edebilir ve bunlara karşı koyabilir, böylece kamusal söylemi koruyabilir ve doğru ve güvenilir bilgiyi teşvik edebilir.

4. Etik Hususlar ve Kaynak Doğrulama:

Doğruluk kontrol uzmanları, açık kaynaklı istihbarat (OSINT) aracılığıyla sunulan geniş ve karmaşık bilgi ağında gezinirken, sıkı etik kurallara uymaları hayati önem taşımaktadır. Sorumlulukları sadece kaynakları doğrulamanın ötesinde, gizliliğe saygı göstermeyi ve araştırmalarında hassas bilgileri sorumlu bir şekilde ele almayı da kapsar. Doğruluk kontrolünde temel etik hususlardan biri şeffaflık ve ifşadır.

Doğruluk kontrol kuruluşlarının metodolojileri, finansman kaynakları ve ortaya çıkabilecek potansiyel çıkar çatışmaları hakkında açıklık sağlamaları beklenir. Bu hususları açıkça ifşa ederek, izleyicileri nezdinde güven oluşturabilir ve doğruluk kontrol sonuçlarının bütünlüğünü koruyabilirler. Örneğin, İngiltere merkezli önde gelen bir doğruluk kontrol kuruluşu olan Full Fact, kamu güvenini korumak için finansman kaynaklarını açıkça paylaşmaktadır. Benzer şekilde, tarafsızlık ve

önyargısızlık da etik doğruluk kontrolünün bir diğer kritik unsurunu oluşturur. Siyasi veya ideolojik bağlam ne olursa olsun, doğruluk kontrolü iddiaları değerlendirmek için objektif kriterler kullanılarak tarafsız bir şekilde yapılmalıdır. Tarafsızlığa olan bu bağlılık, üretilen değerlendirmelerin adil ve doğru olmasını sağlar. Bununla birlikte, doğruluk kontrolünde hakikat arayışı, bireyler ve toplumlar üzerindeki potansiyel etkiyle de dengelenmelidir.

Fact-checker'lar bulgularının potansiyel yansımalarını göz önünde bulundurmalı ve yanlış bilgileri çürütürken zararlı içeriği güçlendirmemeye dikkat etmelidir. Bu durum, yanlış bilginin yayılmasının gerçek dünyada zarara yol açabileceği seçim dönemleri veya toplumsal huzursuzluk zamanları gibi hassas durumlarda görülebilir. Gizlilik korumasını sürdürmek, kullanıcı verilerini ve kişisel bilgileri ele almanın çok önemli bir parçasıdır ve doğruluk kontrolörleri veri gizliliği düzenlemelerine uymalıdır. Verilerin anonimleştirilmesi ve kullanıcılardan açık rıza alınması gibi teknikler bu açıdan son derece önemlidir. Yanlış bilgileri çürütme görevlerini yerine getirirken, doğruluk kontrolörleri de yanlış bilgileri istemeden daha fazla yaymamak için dikkatli olmalıdır. Bu, yanlış iddiaları aşırı görünürlük kazandırmadan çürüttükleri nüanslı bir yaklaşım gerektirir. Bu ilke, COVID-19 salgını sırasında sağlık kuruluşlarının ve doğruluk kontrolörlerinin potansiyel olarak zararlı yanlış bilgileri güçlendirmeden doğru bilgi sağlamayı dikkatlice dengelemeleri gerektiğinde örneklenmiştir. Yapay zeka ve otomasyonun doğruluk kontrolüne giderek daha fazla dahil edilmesiyle birlikte, bu teknolojilerin sorumlu kullanımı daha da önemli hale gelmektedir.

Bu teknolojilerin sunduğu faydalara rağmen, doğruluk kontrol uzmanları, olası önyargı ve hatalardan kaçınmak için uygulamalarının etik hususlar tarafından yönlendirildiğinden emin olmalıdır. Doğruluk kontrolü çalışmalarında eğitim de önemli bir rol oynamaktadır. Doğruluk kontrolörleri, halkı eleştirel düşünme ve medya okuryazarlığı konusunda eğitmeye odaklanarak, izleyicileri yanlış bilgileri bağımsız olarak tespit etme konusunda güçlendirebilir. Bu tür bir sosyal yardım, örneğin Birleşmiş Milletler'in küresel medya okuryazarlığını geliştirme kampanyasında olduğu gibi, dünyanın dört bir yanındaki medya kuruluşları tarafından etkili bir şekilde kullanılmıştır. Son olarak, doğruluk kontrolörlerinin kötü niyetli aktörlerin olası misillemelerinin farkında olmaları gerekir. Tıpkı gazetecilerin bazen çalışmaları nedeniyle tehditlerle karşı karşıya kalmaları gibi, doğruluk kontrolcüleri de güvenliklerini korumak için önlemler almalıdır. Bu tür önlemlerin uygulanması, doğruluk kontrolörlerinin potansiyel misilleme karşısında hayati önem taşıyan çalışmalarına devam edebilmelerini sağlar ve yanlış bilgiye karşı devam eden mücadeleyi destekler.

Sonuç olarak, Açık Kaynak Araştırması doğruluk kontrolünü ve bilgi doğrulamayı önemli ölçüde etkilemiştir. Doğruluk kontrol uzmanları dijital adli tıp, sosyal medya analizi, veri madenciliği ve diğer OSINT tekniklerinden yararlanarak iddiaları doğrulayabilir, yanlış bilgileri tespit edebilir ve kamuoyuna doğru bilgi sağlayabilir. Yeni teknolojiler ve otomasyon, OSINT çabalarının verimliliğini artırarak, doğruluk

kontrolörlerinin büyük miktarda veriyi incelemesine ve potansiyel yanlış bilgileri hızla tespit etmesine olanak tanır. İşbirliği, çok dilli yetenekler ve etik hususlar, OSINT'in faydalarını en üst düzeye çıkarmak ve doğruluk kontrolü ve bilgi doğrulama girişimlerinin inandırıcılığını ve güvenilirliğini sağlamak için gereklidir. Teknoloji ilerlemeye devam ettikçe, OSINT yanlış bilgiyle mücadelede ve daha bilinçli ve hesap verebilir bir bilgi ekosistemini teşvik etmede çok önemli bir araç olmaya devam edecektir.

İÇERİK SINIFLANDIRMASI İÇİN MAKİNE ÖĞRENİMİ

Makine öğrenimi modelleri, içeriği potansiyel olarak yanıltıcı veya gerçeklere dayalı olarak sınıflandırmak üzere eğitilebilir, böylece doğruluk kontrolörlerinin çabalarını önceliklendirmelerine ve en kritik bilgi manipülasyonu vakalarına odaklanmalarına yardımcı olur. İçerik Sınıflandırma teknikleri için Makine Öğrenimi, yanlış bilgileri belirleme, içeriği kategorize etme ve doğruluk kontrol çabalarına öncelik verme sürecini otomatikleştirerek doğruluk kontrolü ve bilgi doğrulamada devrim yaratmıştır. Bu genişletilmiş analiz, İçerik Sınıflandırması için Makine Öğrenimi ile ilişkili yeni teknolojilerin ve tekniklerin teknik özelliklerine odaklanarak, doğruluk kontrolü ve bilgi doğrulama üzerindeki etkisini araştırmaktadır.

1. Aktif Öğrenme ve Döngüde İnsan (HITL) Yaklaşımı:

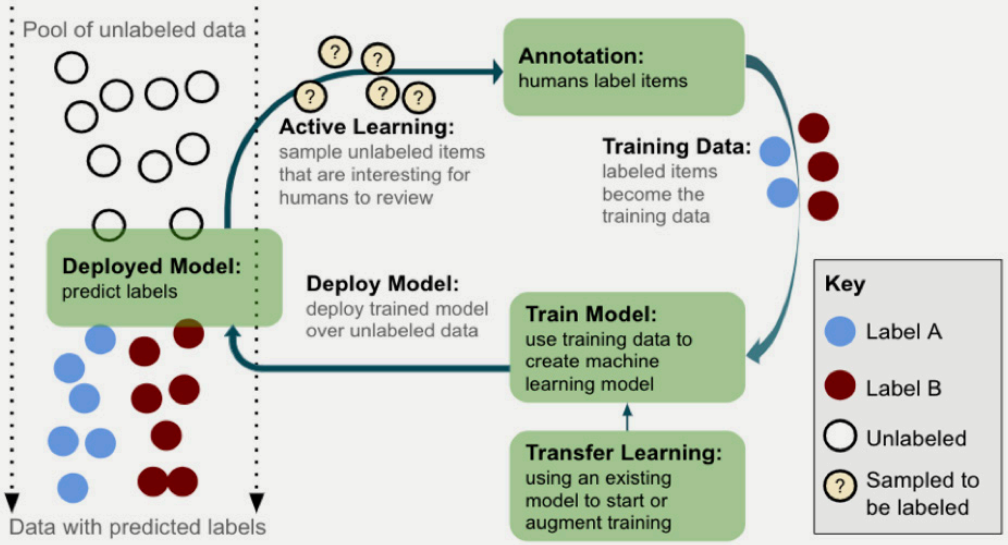
Aktif Öğrenme ve Döngüde İnsan (HITL) Yaklaşımının kombinasyonu, doğruluk kontrol yöntemlerini önemli ölçüde geliştirmek ve yabancı bilgi manipülasyonu ve müdahalesiyle mücadele etmek için güçlü bir araç seti sağlar.[49] Makine öğrenimi algoritmalarının gücünü insan uzmanlığıyla birleştiren bu teknikler, doğruluk kontrol sürecini optimize eder, doğruluğu artırır ve yanlış bilgi kampanyalarını etkili bir şekilde hedef alır.

Bir makine öğrenimi yaklaşımı olan aktif öğrenme, doğruluk kontrolünün verimliliğini artırmada önemli bir rol oynamaktadır.[50] İnsan açıklaması için en bilgilendirici veri noktalarını seçerek çalışır ve hangi iddiaların veya bilgi parçalarının en acil insan doğrulaması gerektirdiğine öncelik verilmesine yardımcı olur. Uygulamada, makine öğrenimi modeli başlangıçta küçük bir etiketli veri kümesi üzerinde eğitilir. Daha sonra belirsizlik ya da zorluk arz eden veri noktaları belirlenir ve bunlar daha sonra doğruluk kontrol uzmanları tarafından açıklanır. Model bu güncellenmiş veri kümesi üzerinde yeniden eğitilir ve böylece zaman içinde performansı yinelemeli olarak iyileştirilir. Bu süreç, ihtiyaç duyulan manuel açıklama sayısını azaltır ve böylece daha verimli bir şekilde yüksek doğruluk elde edilir.

49 Liu, Zimo, Jingya Wang, Shaogang Gong, Huchuan Lu ve Dacheng Tao. "Döngü içinde insan yeniden tanımlama için derin takviye aktif öğrenme." IEEE/CVF uluslararası bilgisayarla görme konferansı bildirilerinde, s. 6122-6131. 2019.

50

Aktif öğrenmeyi tamamlayan HITL yaklaşımı, karmaşık görevleri yerine getirmek için insan uzmanlığını makine öğrenimi algoritmalarıyla birleştirir. Doğruluk kontrolü alanında bu yaklaşım, doğruluk kontrolörlerine yapay zeka sistemlerinin kararlarını doğrulama ve yönlendirme yetkisi vererek insan yargısını ve muhakemesini sürece entegre eder. Makine öğrenimi modelleri öneriler veya tahminler sunar, bunlar daha sonra sonuçlar yayınlanmadan önce doğruluk kontrolörleri tarafından gözden geçirilir ve doğrulanır. Bu işbirlikçi süreç, doğruluk kontrol çabalarının doğru ve güvenilir olmasını sağlarken aynı zamanda makine öğreniminin hızından ve ölçeklenebilirliğinden de yararlanır. Bu çerçeveler içinde, belirsizlik örneklemesi ve güvene dayalı etiketleme gibi belirli stratejiler çok önemli roller oynamaktadır. Belirsizlik örneklemesinde model, insan incelemesi için emin olmadığı veri noktalarını seçer. Bu, doğruluk kontrol uzmanlarının modelin en az emin olduğu iddialara odaklanmasını, olası hataları düzeltmesini ve modelin performansını artırmasını sağlar.



Aktif Öğrenme ve Transfer Öğrenme yaklaşımları arasında bir karşılaştırma.
Kaynak: https://livebook.manning.com/book/human-in-the-loop-machine-learning/chapter-1/v-11/16

Güvene dayalı etiketleme de benzer şekilde çalışır ve model her tahmine bir güven puanı atar. Düşük güven puanına sahip iddialar manuel doğrulama için doğruluk kontrolörlerine yönlendirilir, bu da yanlış bilginin yayılma riskini azaltır. Doğruluk kontrolörleri ayrıca, eğitim veri setinin çeşitli konuları ve bakış açılarını temsil eden çok çeşitli iddiaları ve bilgileri kapsamasını sağlamak için veri çeşitliliği örneklemesini kullanır. Bu geniş temsil, doğruluk kontrol çabalarının kapsamlı bir şekilde ele alınmasına yardımcı olur ve çeşitli alanlardaki yanlış bilgilere karşı koyar. Ayrıca, doğruluk kontrolü yapanlardan gelen geri bildirimler ve güncellemeler, makine öğrenimi modelini geliştirmek için sürekli olarak dahil edilmektedir. Bu sistem yeni verilerden ve insan kararlarından adapte olur ve öğrenir, böylece daha etkili bir süreç ortaya çıkar. HITL yaklaşımı ayrıca farklı insan perspektiflerini ve yargılarını entegre ederek yapay zeka modellerindeki potansiyel önyargıları azaltmaya yardımcı olur. Doğruluk kontrol uzmanları bu önyargıları tespit edip

düzeltebilir, adil ve objektif bir doğruluk kontrol sürecini koruyabilir. Bu işbirlikçi yaklaşım aynı zamanda doğruluk kontrol sürecinin ölçeklenebilirliğini ve verimliliğini de artırır. Aktif öğrenme, en ilgili veri noktalarına odaklanarak manuel ek açıklama yükünü azaltır.

Dahası, HITL yaklaşımı gerçek zamanlı doğruluk kontrolüne olanak tanıyarak ortaya çıkan yanlış bilgi kampanyalarına ve yabancı bilgi manipülasyonuna hızlı yanıt verilmesini kolaylaştırır. Sonuç olarak, doğruluk kontrol kuruluşları aktif öğrenme ve HITL yaklaşımı gibi gelişmiş yöntemleri benimseyerek yabancı bilgi manipülasyonu ve müdahalesiyle mücadele çabalarını önemli ölçüde geliştirebilir. Yapay zeka modelleri ve insan doğruluk kontrolörleri arasındaki bu iş birliği süreci, doğruluk ve doğru raporlama ilkelerini destekleyen sağlam ve etik bir doğruluk kontrol ekosistemi sağlar.

2. Az Atışlı Doğruluk Kontrolü için Transfer Öğrenimi:

Yanlış bilginin benzeri görülmemiş bir hızla yayıldığı bir çağda, transfer öğrenme ile desteklenen az sayıda doğruluk kontrolü, bu sorunla etkin bir şekilde mücadele eden güçlü ve yenilikçi bir yaklaşımdır. Bu teknik, doğru doğruluk kontrolü gerçekleştirmek için sınırlı bir etiketli veri kümesinin yanı sıra önceden eğitilmiş modeller kullanır. Bu, özellikle etiketli verilerin az olduğu senaryolarda avantajlıdır. Böyle bir yaklaşım, yabancı bilgi manipülasyonu ve müdahalesiyle mücadelede önemli bir potansiyele sahiptir ve doğruluk kontrolörlerinin ortaya çıkan yanlış bilgi kampanyalarına hızla yanıt vermesine ve sınırlı kaynaklarla iddiaları etkili bir şekilde doğrulamasına olanak tanır.

Az sayıda doğruluk kontrolünde, BERT (Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri), GPT (Üretken Önceden Eğitilmiş Transformatör) ve RoBERTa (Sağlam Bir Şekilde Optimize Edilmiş BERT Ön Eğitim Yaklaşımı) gibi önceden eğitilmiş dil modelleri etkili olmaktadır.[51] Büyük miktarda metin verisi üzerinde eğitilen bu modeller, karmaşık dilsel kalıpları, bağlamı ve anlambilimi anlayarak dilin zengin temsillerini yakalar. Temel olarak hareket ederek, doğruluk kontrolü alanında metni anlamak ve analiz etmek için sağlam bir temel sağlarlar. Bu temel üzerine inşa edilen önceden eğitilmiş model, ince ayar olarak bilinen bir süreçten geçer. Burada model, tipik olarak etiketlenmiş iddialardan ve bunların doğruluklarından (doğru veya yanlış) oluşan daha küçük bir doğruluk kontrol örnekleri veri kümesi üzerinde eğitilir.

Bu özel eğitim, modelin bilgisini doğruluk kontrolü bağlamına uyarlamasına ve anlayışını eldeki özel görevle uyumlu hale getirmesine olanak tanır. Aktarımlı öğrenmeyi tamamlayan az sayıda örnekle öğrenme paradigması kullanılmaktadır. Bu yaklaşım, modelleri sınıf başına minimum sayıda örneğe (bu durumda, doğruluk

51 Lee, Nayeon, Yejin Bang, Andrea Madotto, Madian Khabsa ve Pascale Fung. "Towards few-shot fact-checking via perplexity." arXiv ön baskı arXiv:2103.09535 (2021).

kontrol etiketleri) dayalı olarak doğru tahminler yapmak üzere eğitir. Sonuç olarak model, az sayıda etiketli örnekten elde ettiği bilgiyi etkili bir şekilde genelleştirir. Bu genelleme kavramı, eğitim verilerini her sınıf için prototiplere dönüştüren az sayıda öğrenme yöntemi olan prototipik ağların kullanılmasıyla daha da geliştirilmiştir. Doğruluk kontrolü bağlamında, bu prototipler doğru ve yanlış iddiaları temsil eder ve modelin sınırlı etiketli verilerle yeni iddiaların doğruluğunu tahmin etmesine olanak tanır. Bu süreç, meta-öğrenme ve metrik öğrenme yoluyla daha da geliştirilmiştir. Bu yaklaşımlar, modelleri minimum sayıda örneğe dayanarak yeni görevlere hızla adapte olacak şekilde eğitir ve veri noktaları arasındaki benzerliği ölçen bir mesafe metriği öğrenir.

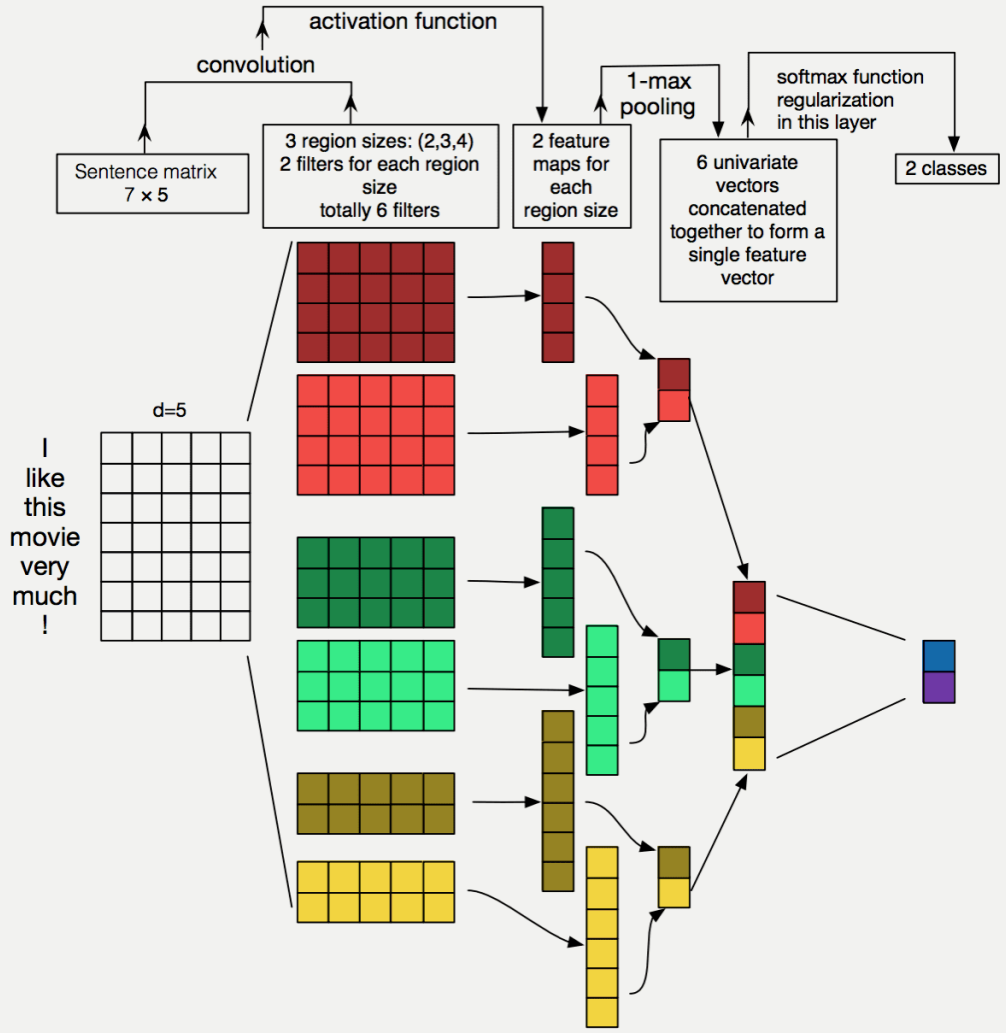
Bu yöntemler, modelin sınırlı etiketli verilerle yeni iddialara hızlı bir şekilde uyum sağlama yeteneğinin çok önemli olduğu az sayıda doğruluk kontrolünde hayati bir rol oynamaktadır. Transfer öğrenmenin esnekliği, alanlar arası bağlamlara kadar uzanır. Burada model, bilgiyi haber makaleleri gibi bir alandan sosyal medya gönderileri gibi başka bir alana aktarır. Bu esneklik, bilgi manipülasyonunun meydana gelebileceği çeşitli platformlarda doğruluk kontrolünde değerli olduğunu kanıtlamaktadır. Daha önce tartışılan aktif öğrenme ve Döngüde İnsan (HITL) stratejileri de az atımlı öğrenmeyi güçlendirir. Aktif öğrenme, insan açıklaması için en bilgilendirici veri noktalarına odaklanarak, insan gözden geçirenleri doğruluk kontrol sürecine entegre eden ve böylece doğruluğu sağlamak için modelin tahminlerini doğrulayan ve yönlendiren HITL ile el ele çalışır. Az vuruşlu öğrenme ve transfer öğrenmenin birleşimi, gerçek zamanlı doğruluk kontrolü ve hızlı yanıt yeteneklerinin önünü açar. Modelin yeni iddialara hızlı bir şekilde adapte olmasını sağlar, bu da ortaya çıkan yanlış bilgilendirme kampanyalarını ve yabancı bilgi manipülasyonunu engellemek için çok önemlidir.

Ayrıca, transfer öğrenmenin az-atışlı öğrenme ile kullanılmasının bir sonucu olarak kapsamlı etiketli verilere olan ihtiyacın azalması, doğruluk kontrol çabalarının ölçeklenebilirliğini ve verimliliğini artırmaktadır. Özetle, az sayıda doğruluk kontrolü için transfer öğrenmede gelişmiş yöntemlerin benimsenmesi, doğruluk kontrol kuruluşlarının sınırlı etiketli veriye sahip senaryolarda bile yabancı bilgi manipülasyonu ve müdahalesine etkili bir şekilde karşı koymasını sağlar. Önceden eğitilmiş dil modellerini, az-atışlı öğrenmeyi ve insan doğrulamasını bir döngü içinde birleştirerek, doğrulanmış bilgilerin halka hızlı bir şekilde yayılmasını teşvik eden sağlam ve doğru bir doğruluk kontrol ekosistemi geliştirilir.

3. Geliştirilmiş Doğruluk için Topluluk Yöntemleri:

Doğruluk kontrolünde kullanılan sofistike teknikler olan topluluk yöntemleri, yabancı bilgi manipülasyonu ve müdahalesine karşı sağlam bir önlem olarak hizmet eder. Bu teknikler, kolektif tahminler yapmak için çeşitli bireysel modellerin veya algoritmaların çıktılarını benzersiz bir şekilde harmanlayarak doğruluk kontrol

sürecinin doğruluğunu ve güvenilirliğini artırır.[52] Bu yaklaşım sayesinde, topluluk yöntemleri önyargıyı etkili bir şekilde azaltır, genellemeyi iyileştirir ve doğru ve sağlam doğruluk kontrol sonuçları sunar. Topluluk yöntemlerinin önemli bir gücü, model mimarisinin çeşitliliğinde yatmaktadır. Önceden eğitilmiş dil modelleri, grafik tabanlı modeller ve derin sinir ağları gibi çeşitli mimariler kullanarak, topluluğa çeşitlilik katarlar. Verilerin farklı yönleri üzerinde eğitilen her model, benzersiz kalıpları ve ilişkileri yakalar ve kodlar. Örneğin, önceden eğitilmiş dil modelleri bir iddianın dilsel nüanslarını anlamada üstün olabilirken, grafik tabanlı modeller iddiada yer alan varlıklar arasındaki ilişkileri analiz etmede daha etkili olabilir.



Cümle Sınıflandırması için Evrişimsel Sinir Ağı Mimarisi. Kaynak: Zhang, Ye, ve Byron Wallace. "Cümle Sınıflandırması için Evrişimsel Sinir Ağlarının Duyarlılık Analizi (ve Uygulayıcılar Kılavuzu)." ArXiv, (2015). Erişim tarihi: 3 Ağustos 2023. /abs/1510.03820.

Topluluk yöntemlerindeki temel teknikler arasında Bagging (Bootstrap Aggregating) ve Boosting yer alır.[53] Torbalama, bootstrap örnekleri olarak bilinen eğitim verilerinin farklı alt kümeleri üzerinde aynı modelin birden fazla örneğinin eğitilmesini içerir. Bireysel modellerin tahminleri daha sonra oylama veya ortalama alma gibi

52 Ahmad, Iftikhar, Muhammad Yousaf, Suhail Yousaf ve Muhammad Ovais Ahmad. "Makine öğrenimi topluluk yöntemleri kullanılarak sahte haber tespiti." Karmaşıklık 2020 (2020): 1-11.

53 Gupta, Surbhi, ve Munish Kumar. "Boosting ve bagging metodolojilerini kullanan adli belge inceleme sistemi." Soft Computing 24 (2020): 5409-5426.

yöntemlerle birleştirilerek aykırı tahminlerin etkisi azaltılır. Öte yandan Boosting, zayıf öğrenicilerin (mütevazı performans gösteren modeller) doğruluğunu artırmayı amaçlar. Bu, yanlış sınıflandırılan örneklere daha fazla ağırlık atayarak elde edilir. Model ağırlıklarını önceki iterasyonların performansına göre güncelleyen iteratif güçlendirme sayesinde genel doğruluk iyileştirilir. Topluluk yöntemleri ayrıca Yığınlama ve Rastgele Orman gibi sofistike tekniklerden de yararlanır.[54] Yığınlamada, çeşitli modellerden gelen tahminler bir meta-model veya daha üst düzey bir model aracılığıyla birleştirilir. Burada, temel modellerin tahminleri meta-model için özellik olarak kullanılır ve topluluğun birden fazla bilgi kaynağından yararlanmasına olanak tanır. Torbalamaya dayalı bir topluluk yöntemi olan Rastgele Orman yöntemi, tahminler yapmak için birden fazla karar ağacı kullanır, bu da aşırı uyumu azaltır ve bireysel ağaçlardan gelen tahminlerin ortalamasını alarak doğruluğu artırır.

Gradient Boosting Machines (GBM) ve Adaboost, topluluk yöntemlerinde daha ileri teknikleri temsil etmektedir.[55] GBM, önceki modeller tarafından yapılan hataları düzeltmek için yeni modellerin eğitildiği ve böylece birden fazla zayıf öğrenicinin tahminlerini birleştirerek doğruluğu yinelemeli olarak geliştiren bir güçlendirme tekniğidir. Popüler bir boosting algoritması olan Adaboost, eğitim örneklerine ağırlıklar atar ve yanlış sınıflandırılmış örnekleri vurgulamak için modelleri yinelemeli olarak eğitir, böylece sınıflandırılması zor örneklere odaklanarak doğruluğu artırır. Gradyan artırmanın bir başka uygulaması da hız ve performans için optimize edilmiş olan XGBoost'tur (Extreme Gradient Boosting) ve düzenli hale getirme ve paralel işlemeyi içerir, bu da onu topluluk yöntemleri içinde oldukça etkili kılar. Özellikle, topluluk yöntemleri, doğruluğu ve sağlamlığı artırmak için metin, görüntü ve meta veri gibi farklı veri türleri üzerinde eğitilen modellerden gelen tahminleri birleştirerek veri temsilinde çeşitliliğe de izin verir. Ayrıca tahminlerin belirsizliğini tahmin edebilir ve doğruluk kontrolü sonuçları için güven puanları sağlayabilirler. Bu yaklaşım, tahminlerde güven sunmanın izleyicilerde güven oluşturmaya yardımcı olabileceği doğruluk kontrolü alanında özellikle değerlidir.

Çevrimiçi topluluk öğrenimi, yeni veriler elde edildikçe toplulukta sürekli güncellemeleri ve iyileştirmeleri kolaylaştırarak bu yöntemlerin yeteneklerini daha da genişletir. Bu gerçek zamanlı ayarlama ve öğrenme, gerçek zamanlı doğruluk kontrolüne ve ortaya çıkan yanlış bilgilere hızlı bir şekilde yanıt verilmesine olanak tanır. Özetle, topluluk yöntemleri, doğruluk kontrolünün doğruluğunu artırmak için güçlü bir yaklaşım sunar. Doğruluk kontrol kuruluşları bu gelişmiş yöntem ve tekniklerden yararlanarak yabancı bilgi manipülasyonu ve müdahalesine karşı mücadelelerinde güvenilir ve kapsamlı sonuçlar elde edebilirler. Birden fazla modelin ortak zekâsından yararlanan topluluk yöntemleri, doğruluk kontrol çabalarının dayanıklılığını artırır ve kamuoyunun iyi bilgilendirilmesini teşvik eder. Bu yöntemlerin etkinliği, farklı doğruluk kontrol kuruluşları arasında iş birliği çabaları, veri kaynaklarının çeşitlendirilmesi ve model perspektifleri yoluyla daha da artırılabilir.

54 AG, Priya Varshini, Anitha Kumari K, ve Vijayakumar Varadarajan. "Rastgele orman tabanlı yığılmış topluluk yaklaşımı kullanarak yazılım geliştirme çabalarını tahmin etme." Elektronik 10, no. 10 (2021): 1195.

55 Verma, Pawan Kumar, Prateek Agrawal, Vishu Madaan ve Radu Prodan. "MCred: BERT ve CNN kullanarak sahte haber tespiti için çok modlu mesaj güvenilirliği." Journal of Ambient Intelligence and Humanized Computing (2022): 1-13.

4. Bağlamsal Anlama için Sinirsel Dil Modelleri:

Doğruluk kontrolü ve yabancı bilgi manipülasyonu ve müdahalesine karşı koyma konusundaki ilerlemelerin ön saflarında sinirsel dil modellerini buluyoruz. Bu modeller, derin öğrenme tekniklerini ve çok miktarda metin verisi üzerinde büyük ölçekli ön eğitimleri bir araya getirerek dil bağlamı ve semantiği hakkında derin bir anlayış oluşturur. Bu gelişmiş bağlamsal anlayışla, sinirsel dil modelleri karmaşık iddiaları doğru bir şekilde anlamak, incelikli yanlış bilgileri tespit etmek ve bilgilerin doğruluğunu etkili bir şekilde doğrulamak için donatılmıştır.

En gelişmiş nöral dil modelleri arasında yer alan BERT ve GPT gibi Transformatör Tabanlı Modeller, nöral dil anlayışında önemli adımlar atmaktadır. Bu modeller, kendi kendine dikkat mekanizmalarını kullanarak, bir cümledeki kelimeler arasındaki ilişkileri yakalama yeteneğine sahiptir ve uzun menzilli bağımlılıkları anlamalarını kolaylaştırır. Bu modelleri geliştirme süreci genellikle iki kritik adımı içerir: ön eğitim ve ince ayar. Başlangıçta, bu modeller büyük ölçekli metin verisi derlemleri üzerinde ön eğitime tabi tutulur. Bu sayede zengin dil temsillerini ve bağlamsal yerleştirmeleri öğrenirler. Bunu takiben, belirli doğruluk kontrol verileri üzerinde eğitildikleri ve böylece dil anlama yeteneklerini iddiaları doğrulama görevi için uyarladıkları ince ayar işleminden geçerler.

Sinirsel dil modelleri, bağlamsal kelime katıştırmaları oluşturarak bağlamı üstün bir şekilde anlamayı başarır. Bu katıştırmalar, kelimelerin anlamını çevrelerindeki bağlama göre yakalar. Bu, modellerin çok anlamlı kelimeleri ve kelime öbeklerini anlamasını sağlayarak daha iyi doğruluk kontrolü performansı sağlar. Bazı gelişmiş nöral modeller metin, görüntü ve meta veri gibi çok modlu girdileri entegre edebilmektedir. Bu çok modlu bağlamsal anlayış, daha doğru ve bütünsel doğruluk kontrolü için çeşitli veri türlerini birleştirerek iddiaların kapsamlı bir şekilde yorumlanmasını sağlar. Bu modeller tarafından kullanılan dikkat mekanizmaları, bilgileri işlerken bir cümlenin veya belgenin en alakalı kısımlarına odaklanmalarını sağlar. Bu, doğruluk kontrolü gerektiren iddialardaki kilit unsurların belirlenmesine yardımcı olur ve daha iyi bağlamsal anlayış sağlar. Ayrıca, bu modeller dil oluşturma teknikleri ve karşıt eğitim ile donatılmıştır. Dil üretimi, potansiyel yanlış bilgilendirme kampanyalarını simüle edip tespit edebilir veya yanlış iddialara karşı argümanlar oluşturabilir.

Öte yandan karşıt eğitim, modeli karşıt saldırılara ve yanlış bilgilendirme girişimlerine karşı dayanıklı hale getirebilir.[56] Transfer öğrenmeden yararlanan sinirsel dil modelleri, haber makaleleri gibi bir alandaki bilgiyi sosyal medya gönderileri gibi başka bir alana uygulayabilir. Bu çapraz alan anlayışı, yabancı bilgi manipülasyonuyla mücadele etmek için çeşitli bilgi kaynaklarının doğruluğunu kontrol etmek için hayati önem taşır. Diğer yetenekler arasında, nöral modellerin iddialardaki kilit varlıkları tanımladığı ve bu varlıklarla ilgili bilgileri doğruladığı

55 Verma, Pawan Kumar, Prateek Agrawal, Vishu Madaan ve Radu Prodan. "MCred: BERT ve CNN kullanarak sahte haber tespiti için çok modlu mesaj güvenilirliği." Journal of Ambient Intelligence and Humanized Computing (2022): 1-13.

56 Tariq, Abdullah, Abid Mehmood, Mourad Elhadef ve Muhammad Usman Ghani Khan. "Sahte Haber Sınıflandırması için Çekişmeli Eğitim." IEEE Access 10 (2022): 82706-82715.

doğruluk değerlendirmesi için varlık tanıma yer alır. Sıfır atışlı öğrenme de modellerin, tam olarak bu iddialar üzerinde açıkça eğitilmemiş olsalar bile, görülmemiş veya yeni iddialar için tahminlerde bulunmalarını sağladığı için önemli bir rol oynamaktadır. Ortaya çıkan yanlış bilgilere verilen bu hızlı yanıt, doğruluk kontrolörleri için çok değerlidir. Tek bir dilin ötesine geçen bu modeller, birden fazla dil için ince ayar yapılarak diller arası doğruluk kontrolünü kolaylaştırabilir.

Son olarak, zaman içinde performansı artırmak için bilgilerini yeni verilerle ve doğruluk kontrol sonuçlarıyla güncelleyerek sürekli öğrenmeden de geçebilirler. Bu, gerçek zamanlı doğruluk kontrolünü ve ortaya çıkan yanlış bilgi kampanyalarına yanıt vermeyi destekler. Sinirsel dil modellerinin bağlamsal anlama yeteneklerinden yararlanarak, doğruluk kontrol kuruluşları yabancı bilgi manipülasyonu ve müdahalesiyle mücadele çabalarını önemli ölçüde artırabilir. Modellerin bağlamı, semantiği ve dildeki nüansları işleme yeteneği, doğruluk kontrolünün doğruluğunu ve verimliliğini artırarak güvenilir bilginin halka zamanında yayılmasını sağlar. Bu modellerin sağlam ve etik doğruluk kontrolü uygulamaları için hassas bir şekilde ayarlanmasını ve optimize edilmesini sağlamak, araştırmacılar, dil uzmanları ve doğruluk kontrolörleri arasında iş birliğine dayalı çabalar gerektirir.

Sonuç olarak, İçerik Sınıflandırma için Makine Öğrenimi, yanlış bilgilerin tanımlanmasını otomatikleştirerek, içeriği kategorize ederek ve doğruluk kontrol çabalarına öncelik vererek doğruluk kontrolü ve bilgi doğrulamayı önemli ölçüde etkilemiştir. NLP, denetimli öğrenme, aktif öğrenme ve transfer öğrenme, bu tekniklerin başarısını sağlayan temel teknik unsurlar arasındadır. Doğruluk kontrolörleri, gelişmiş Makine Öğrenimi modellerinden yararlanarak büyük miktarda içeriği hızlı ve doğru bir şekilde işleyebilir ve doğruluk kontrol girişimlerinin verimliliğini ve etkinliğini artırabilir. Bu teknolojilerin sürekli gelişimi, yanlış bilgiyle mücadelede ve daha bilinçli ve dirençli bir bilgi ekosisteminin teşvik edilmesinde önemli bir rol oynamaya devam edecektir.

DEEPFAKE TESPİT TEKNOLOJİLERİ

FIMI kampanyalarında deepfake'lerin yükselişi göz önüne alındığında, fact-checker'lar manipüle edilmiş ses ve video içeriğini doğru bir şekilde tespit etmek için gelişmiş deepfake tespit teknolojilerini benimseyebilir. Deepfake Tespit Teknolojileri, özellikle deepfake teknolojisinin karmaşıklığı ilerlemeye devam ettikçe, yanlış bilgiyle mücadelede çok önemli hale gelmiştir. Deepfake Tespiti ile ilgili yeni teknolojilerin ve tekniklerin teknik özelliklerine odaklanan bu genişletilmiş analiz, bunların doğruluk kontrolü ve bilgi doğrulama üzerindeki etkilerini araştırmaktadır.

1. Yüz Tanıma ve Yer İşareti Tespiti:

Yüz tanıma ve yer işareti tespiti, deepfake içerik ve yabancı bilgi manipülasyonu ve müdahalesine karşı mücadelede kritik unsurlar olarak ortaya çıkmıştır. Bu teknikler, yüz özelliklerini doğru bir şekilde tanımlamak ve analiz etmek için bilgisayarla görme ve derin öğrenmenin gücünden yararlanarak deepfake içeriklerdeki manipüle edilmiş yüzleri tespit etmeyi kolaylaştırır. Derin öğrenme alanında, Evrişimsel Sinir Ağları (CNN'ler) yüz tanıma görevlerinde bir devrim yaratmıştır. Bu sofistike modeller, doğru yüz eşleştirme ve tanımlamayı kolaylaştıran ayırt edici yüz özelliklerini ve yerleştirmelerini öğrenme kapasitesine sahiptir.

Bir adım daha ileri giderek, 3D yüz tanıma teknikleri sağlam yüz temsilleri oluşturmak için derinlik bilgisini kullanır. Bu yaklaşım, zorlu aydınlatma koşulları ve pozlar altında performansı önemli ölçüde artırarak deepfake tespitinde değerli bir araç haline getirir. Manipüle edilmiş yüzlerin tanımlanmasında doğruluğu ve sağlamlığı artırmak için, birden fazla yüz tanıma modeli topluluk yöntemleri aracılığıyla birleştirilebilir. Bu topluluk modelleri, manipüle edilmiş yüzleri birden fazla referans veritabanıyla karşılaştırarak tespit edebilir. Ayrıca, önceden eğitilmiş yüz tanıma modelleri, manipüle edilmiş yüzlerin tanımlanmasında uzmanlaşarak derin sahte algılama veri kümeleri üzerinde ince ayar yapılabilir. Transfer öğrenimi olarak bilinen bu uygulama, sınırlı etiketli deepfake verisiyle bile verimli bir eğitim sağlar.

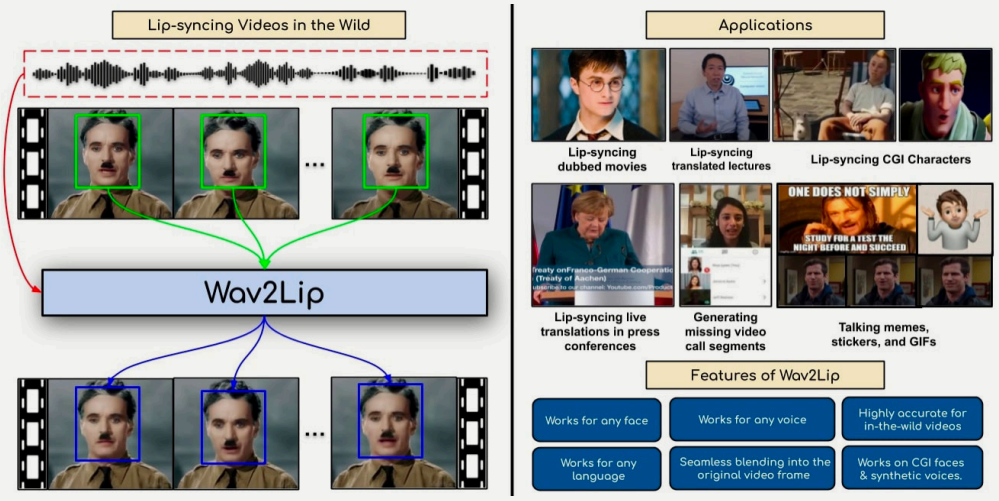
Dikkat mekanizmaları da bu bağlamda güçlü bir araçtır. Gözler, burun ve ağız gibi yüz işaretlerini vurgulamak için kullanılabilirler, bu da doğru yüz analizine ve işaret tespitine yardımcı olur. Kesin yer işareti konumlandırması için kademeli regresyon modelleri özellikle etkilidir. Yüz işaretlerinin konumlarını yinelemeli olarak hassaslaştırarak tespitin doğruluğunu artırırlar. Ayrıca, videolarda gerçek zamanlı yüz işareti tespiti, yüz ifadelerinin ve hareketlerinin sürekli olarak izlenmesini ve analiz edilmesini sağlar. Benzer şekilde, yüz derinlik verilerini analiz eden 3D yüz işaret algılama teknikleri, işaretleri doğru bir şekilde bulabilir ve 2D görüntü manipülasyonlarına karşı sağlamlık sunar. Yüzler arasında doğrudan karşılaştırma yapmak için yüz işaretlerini hizalama teknikleri kullanılabilir. Yüz özelliklerini standart bir konuma getirirler ve bu da derin sahteciliklerin tespit edilmesine önemli ölçüde yardımcı olur.

Tespitten kaçmak için tasarlanmış düşmanca saldırılara karşı sağlamlığı artırmak için, yüz tanıma modelleri düşmanca eğitimle donatılabilir. Bu eğitim şekli, yüz tanıma sistemlerinin düşmana karşı dayanıklılığını büyük ölçüde artırır. Ortaya çıkan deepfake içeriğe gerçek zamanlı olarak hızlı yanıt vermek için, hızlı yüz tanıma ve yer işareti tespiti sağlayan gerçek zamanlı çıkarım için modelleri optimize etmek çok önemlidir. Ayrıca, çapraz alan teknikleri bu modellerin farklı görüntü ve video kaynakları arasında genelleştirilmesine olanak tanıyarak farklı platformlarda güvenilir deepfake tespiti sağlar. Son olarak, mahremiyet konusunda giderek daha fazla endişe duyulan bir dünyada, yüz tanıma ve işaret tespiti süreçleri sırasında bireylerin kimliklerini korumak için mahremiyeti koruyan teknikler dahil edilebilir.

Doğruluk kontrol kuruluşları, bu gelişmiş yüz tanıma ve yer işareti tespit yöntem ve tekniklerinden yararlanarak deepfake tespit yeteneklerini büyük ölçüde geliştirebilir. Manipüle edilmiş yüzleri doğru bir şekilde tanımlama ve analiz etme becerisi, doğruluk kontrolörlerinin deepfake içeriği hızla tanımlamasına ve yaygın bir şekilde yayılmasını önlemesine olanak tanır. Deepfake tespiti ve manipülasyonunda ortaya çıkan zorluklara uyum sağlamak için araştırmacılar, bilgisayarla görme uzmanları ve doğruluk kontrolörleri arasında iş birliğini teşvik etmek çok önemlidir.

2. Dudak Senkronizasyonu ve Konuşma Analizi:

Dudak senkronizasyonu ve konuşma analizi, deepfake tespiti alanında ve yabancı bilgi manipülasyonu ve müdahalesine karşı mücadelede kilit bileşenlerdir. Bu durum özellikle manipüle edilmiş görsel-işitsel içeriğin yanlış bilgi yaymak için kullanıldığı durumlarda geçerlidir. Derin öğrenme, doğal dil işleme ve görsel-işitsel analiz, dudak hareketleri ve ses arasındaki tutarsızlıkları saptamak için gelişmiş yöntem ve tekniklerde kullanılmaktadır. Bu, deepfake videoların ve seslerin ortaya çıkarılmasına ve tespit edilmesine yardımcı olur.



Yakın zamanda yapılan bir çalışma, GAN sesini %99 başarı oranıyla artırmıştır. Kaynak: K R Prajwal, Rudrabha Mukhopadhyay, Vinay P. Namboodiri ve C.V. Jawahar. 2020. Doğada Dudaktan Konuşma Üretimi için Tek İhtiyacınız Olan Bir Dudak Senkronizasyonu Uzmanı. İçinde 28. ACM Uluslararası Multimedya Konferansı Bildirileri (MM '20). Association for Computing Machinery, New York, NY, ABD, 484-492. https://doi.org/10.1145/3394171.3413532

Bir videodaki ses parçası ile konuşmacının dudak hareketleri arasındaki senkronizasyonun analizini içeren dudak senkronizasyonu tespiti, Evrişimli Sinir Ağları (CNN'ler) veya Tekrarlayan Sinir Ağları (RNN'ler) gibi gelişmiş derin öğrenme modellerini kullanır. Bu modeller zamansal kalıpları yakalamak ve olası dudak senkronizasyonu hatalarını belirlemek için tasarlanmıştır, böylece manipüle edilmiş içeriği işaretlemek için bir araç sağlar. Bu süreci daha da geliştirerek, tespit için daha kapsamlı bir veri kümesi sağlamak üzere dudak senkronizasyonu analizinde ses ve görsel özellikler birleştirilir. Erken füzyon veya geç füzyon gibi teknikler, modelin ses ve görsel verileri birlikte işlemesini sağlar. Bu sadece dudak senkronizasyonu tespitinin doğruluğunu güçlendirmekle kalmaz, aynı zamanda deepfake içeriğe karşı daha sağlam bir savunma sağlar.

Konuşma analizi cephesinde, konuşma tanıma modelleri sesi metne dönüştürmek için kullanılırken, Metinden Konuşmaya (TTS) modelleri metni sentezlenmiş konuşmaya dönüştürür. Tanınan konuşma ile sentezlenen konuşma karşılaştırılarak, tutarsızlıklar ve potansiyel deepfake manipülasyonları daha verimli bir şekilde tespit edilebilir. Konuşmacı doğrulama ve tanımlama teknikleri deepfake tespit sürecini daha da destekler. Doğrulama, sesteki konuşmacının kimliğini doğrularken, tanımlama yöntemleri bilinen konuşmacıları belirler. Her iki teknik de deepfake tespit senaryolarında sesin gerçekliğini değerlendirmede etkilidir. Bu yöntemleri tamamlayan ses izi analizi, bireyler için benzersiz ses izleri oluşturarak potansiyel taklitlerin veya ses manipülasyonlarının tanımlanmasını sağlar. Ayrıca, konuşmadaki prozodi ve tonlama kalıplarının analizi, deepfake manipülasyonuna işaret edebilecek doğal olmayan dalgalanmaları veya anormallikleri ortaya çıkarabilir.

Ayrıca, modlar arası hizalama, ses ve görsel verileri zaman içinde hizalayarak dudak hareketlerinin ve konuşma içeriğinin doğrudan karşılaştırılmasına olanak tanır. Bu hizalama, doğru dudak senkronizasyonu analizini daha da kolaylaştırır. Deepfake içeriğin hem görsel hem de işitsel bileşenlerine uygulanan pertürbasyon teknikleri, sağlamlığı ve gerçekliği değerlendirir. Bu, deepfake manipülasyonlarını ortaya çıkarabilir ve yanlış bilgilere karşı ek güvenlik katmanları sağlayabilir. Dudak senkronizasyonu ve konuşma analizinde topluluk modellerinin kullanılması, birden fazla modeli birleştirerek deepfake tespitinin güvenilirliğini ve etkinliğini artırır. Bu, tespit yöntemlerinin sağlamlığını artırır ve sonuçların doğruluğunda genel bir artışa katkıda bulunur. Daha da önemlisi, bu modeller gerçek zamanlı çıkarım için optimize edilmiş olup dudak senkronizasyonu ve konuşma modellerinin hızlı bir şekilde analiz edilmesini sağlar. Bu, ortaya çıkan deepfake içeriğe hızlı yanıt verilmesini sağlar ve yanlış bilginin yayılmasını önlemeye yardımcı olur.

Gelişmiş teknikler aynı zamanda birden fazla dilde içeriğin analiz edilmesini sağlayarak diller arası deepfake tespitine olanak tanır ve yabancı bilgi manipülasyonuyla mücadeleye yardımcı olur. Son olarak, konuşma analizindeki düşmanca sağlamlık, konuşma analizi modellerinin düşmanca eğitim teknikleriyle donatılmasıyla desteklenmektedir. Bu, deepfake'lerde kullanılabilecek düşmanca ses saldırılarına karşı sağlamlığı artırır. Dudak senkronizasyonu ve konuşma

analizindeki bu gelişmiş yöntem ve teknikler, doğruluk kontrol kuruluşlarına yabancı bilgi manipülasyonu ve müdahalesiyle mücadele etmek için güçlü araçlar sağlar.

Dudak senkronizasyonu tutarsızlıklarının ve konuşma anormalliklerinin doğru bir şekilde tespit edilmesi, doğruluk kontrolörlerinin manipüle edilmiş içeriği hızlı bir şekilde tespit etmesini ve yanlış anlatıların yayılmasını önlemesini sağlar. Araştırmacılar, NLP uzmanları, görsel-işitsel analistler ve doğruluk kontrol uzmanları arasındaki iş birliğinin, deepfake tespiti ve manipülasyonundaki gelişen zorlukları ele almak için bu teknikleri sürekli olarak iyileştirmek ve uyarlamak için devam etmesi çok önemlidir.

3. Davranış ve Hareket Analizi:

Deepfake Tespit Teknolojileri, videolardaki doğal olmayan hareketleri ve eylemleri tespit etmek için davranış analizinden önemli ölçüde yararlanır. Bu tespit biçimi, yabancı bilgi manipülasyonu ve müdahalesiyle mücadele ederken, özellikle de manipüle edilmiş videolar bireylerin davranışlarını veya hareketlerini değiştirerek aldatmayı amaçladığında çok önemlidir. Bilgisayar görüşü, makine öğrenimi ve davranış analizinden yararlanan bu gelişmiş teknikler, bir videodaki hareketleri gerçek hayat senaryolarında beklenen kalıplarla karşılaştırır. Bu yaklaşım, videolardaki anormalliklerin, tutarsızlıkların ve deepfake manipülasyonlarının tespit edilmesine yardımcı olarak gerçek ve deepfake içeriği birbirinden ayırır.

Bu tekniklerin kritik bir bileşeni poz tahmini ve izlemedir. Gelişmiş poz tahmin modelleri, bir video boyunca bir kişinin vücudundaki eklemler ve uzuvlar gibi önemli noktaları doğru bir şekilde tespit edip izleyebilir. Bu izleme süreci, hareket kalıplarının dikkatli bir şekilde analiz edilmesini ve böylece vücut duruşu ve davranışındaki potansiyel manipülasyonların belirlenmesini sağlar. Bunun yanı sıra, bir bireyin yürüyüş şeklini ve ritmini incelemek için yürüyüş analizi kullanılır. Bu süreçte kullanılan gelişmiş yöntemler, deepfake manipülasyonunun veya taklitçiliğin göstergesi olabilecek yürüyüş farklılıklarını tespit edebilir. Deepfake tespit araçlarının cephaneliğine yüz mikro ifadeleri analizi de ekleniyor. İnce duyguları ve tepkileri ortaya çıkarabilen kısa duygusal ifadeler olan yüz mikro ifadeleri dikkatle incelenir. Gelişmiş yöntemler bu mikro ifadelerdeki deepfake manipülasyonuna veya duygusal tepkilerdeki tutarsızlıklara işaret edebilecek değişiklikleri belirleyebilir.

Algılama teknolojileri ayrıca etkinlik tanıma modellerini de içerir. Bu modeller videolardaki konuşma, koşma veya el kol hareketleri gibi belirli eylemleri veya davranışları tanımlayabilir. Tanınan bu faaliyetlerin analizi, deepfake içerikte bulunabilecek anormalliklerin tespit edilmesinde çok önemli bir rol oynar. Bu analiz, duygu tanıma teknikleri ile desteklenmektedir. Bu teknikler, videolardaki bireylerin duygusal durumlarını tanımlayarak ve bu duyguları videonun bağlamıyla karşılaştırarak, duygusal ifadelerin gerçekliğini değerlendirmeye yardımcı olur. Anomali tespit modelleri, deepfake'lere karşı mücadelede bir diğer hayati

araçtır. Bu modeller, beklenen kalıplardan sapan nadir veya anormal davranış ve hareketleri tespit ederek potansiyel deepfake manipülasyonlarının belirlenmesine yardımcı olur. Ayrıca, sahnenin gerçekçiliğini ve tutarlılığını değerlendirmek için çevre ve diğer bireylerle etkileşimler gibi videonun bağlamsal bir analizi kullanılır. Bu, bireyler için benzersiz davranış profilleri oluşturan ve sırasıyla davranış özelliklerine dayalı olarak bireylerin kimliğini doğrulayan davranış profili oluşturma ve biyometri ile birlikte, potansiyel taklitleri veya manipülasyonları tanımlamak için çok önemlidir.

Tespit sürecini daha da güçlendiren, faaliyet ve davranış aktarımı tespit teknikleridir. Bu teknikler, manipüle edilmiş bir videoda bir kişinin eylemlerinin başka bir kişiye aktarıldığı durumları tespit edebilir. Bu teknolojiler aynı zamanda birden fazla davranış ve hareket analizi modelini bir araya getiren topluluk modellerinden de faydalanmaktadır. Bu kombinasyon, deepfake tespitinde doğruluğu ve sağlamlığı artırarak manipüle edilmiş içeriğin daha güvenilir bir şekilde tanımlanmasını sağlar. Ortaya çıkan deepfake içeriğe hızlı yanıt verebilmek için bu modeller gerçek zamanlı çıkarım için optimize edilmiştir. Bu, davranış ve hareket kalıplarının hızlı bir şekilde analiz edilmesini sağlayarak yanlış bilginin zarar vermeden önce yayılmasını önler. Video içeriğinin kapsamlı bir şekilde anlaşılmasını sağlamak için multimodal davranış analizi kullanılır. Bu, davranış analizinde video, ses ve meta veriler gibi birden fazla modalitenin entegrasyonunu içerir. Son olarak, davranış analizinde düşmana karşı sağlamlığı sağlamak için modeller düşmana karşı eğitim teknikleriyle donatılabilir. Bu, derin sahtekarlıklarda kullanılabilecek düşman saldırılarına karşı sağlamlıklarını artırır.

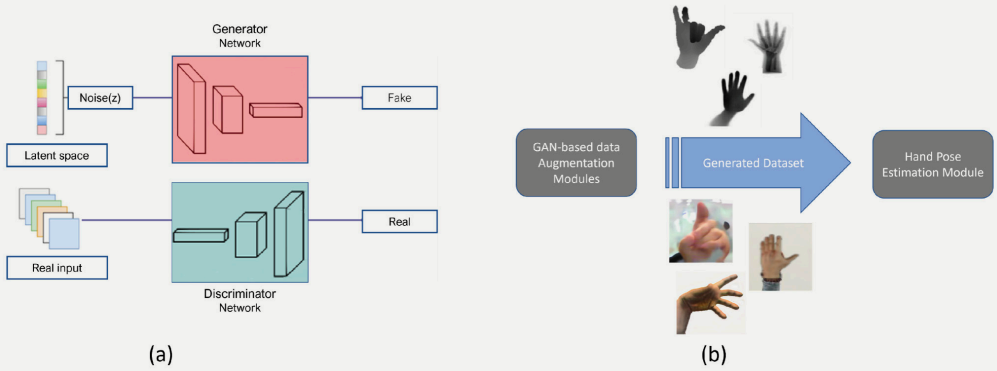
Doğruluk kontrol kuruluşları, davranış ve hareket analizindeki bu gelişmiş yöntem ve tekniklerden yararlanarak deepfake tespit yeteneklerini önemli ölçüde geliştirebilir ve yabancı bilgi manipülasyonu ve müdahalesiyle daha etkili bir şekilde mücadele edebilir. Davranış ve hareketlerdeki anormalliklerin ve tutarsızlıkların doğru bir şekilde tanımlanması, doğruluk kontrolörlerinin manipüle edilmiş içeriği hızlı bir şekilde tespit etmesini sağlayarak yanlış anlatıların yayılmasına karşı koruma sağlar. Bu tekniklerin deepfake tespiti ve manipülasyonun yarattığı zorluklarla birlikte gelişmeye devam etmesini sağlamak için araştırmacıların, bilgisayarla görme uzmanlarının, davranış analistlerinin ve doğruluk kontrol uzmanlarının sürekli olarak iş birliği yapması ve bu teknikleri geliştirmesi çok önemlidir.

4. GAN Tespit Teknikleri:

Deepfake tespiti, özellikle de Generative Adversarial Network (GAN) tarafından üretilen medyanın tanımlanması, yabancı bilgi manipülasyonu ve müdahalesiyle mücadelede önemli bir unsurdur. GAN'lar hiper-gerçekçi deepfake içerik oluşturmada yaygın olarak kullanıldığından, bu tür GAN tarafından oluşturulan medyanın varlığını ayırt edebilen gelişmiş yöntem ve tekniklere olan ihtiyaç giderek artmaktadır. Bu zorluğun üstesinden gelmek için, GAN'a özgü ince yapaylıkları ve tutarsızlıkları tespit etmek üzere derin öğrenme, istatistiksel analiz ve adli incelemenin bir kombinasyonu kullanılmaktadır.

Çekişmeli özellik öğrenimi bu tekniklerin temel taşıdır. Bu süreçte, GAN tarafından üretilen ve gerçek medya arasında ayrım yapmak için ayrı bir sınıflandırıcı eğitilir.[57] Bu sınıflandırıcı, belirli GAN eserlerini ve özelliklerini tanımayı öğrenerek GAN tarafından oluşturulan içeriği doğru bir şekilde tespit edebilmektedir. Bu süreci tamamlayan, görüntüleri veya videoları daha küçük yamalara ayıran yama tabanlı analiz tekniğidir. GAN tarafından oluşturulan yamalar genellikle gerçek medyada bulunanlardan farklı olarak benzersiz yapaylıklar gösterdiğinden, her yama GAN'a özgü desenler için ayrı ayrı analiz edilir. Tespit sürecini daha da iyileştirmek için, piksel desenlerinde GAN tarafından oluşturulan içeriğin göstergesi olan düzensizlikleri ve korelasyonları tanımlamak için Yüksek Sıralı İstatistikler (HOS) gibi gelişmiş istatistiksel yöntemler kullanılır.[58] Bu, sentez işlemi sırasında ortaya çıkmış olabilecek belirli gürültü modellerine odaklanan gürültü analizi ile desteklenmektedir. Bu gürültü modellerindeki tutarsızlıkların belirlenmesi, GAN tarafından üretilen içeriğin gerçek medyadan ayırt edilmesine yardımcı olur. Frekans alanı analizi ve renk uzayı analizi gibi ek yöntemler bu tespit sürecine başka bir katman ekler.

GAN tarafından üretilen görüntüler ve videolar Fourier alanında olağandışı frekans dağılımları sergileyebilir ve bunlar frekans alanı analizi ile tespit edilebilir. Benzer şekilde, renk uzayları analizi, GAN oluşturma işlemi sırasında kullanılan farklı renk uzaylarından kaynaklanabilecek belirli renk dağılım modellerini tespit eder. Ek olarak, GAN'lar tarafından bırakılan gradyanlar ve ağ aktivasyonları gibi benzersiz izleri tanımlamak için teknikler geliştirilmiştir. Bu izlerin analiz edilmesi, GAN tarafından oluşturulan içeriğin varlığına ilişkin değerli bilgiler sağlayabilir. Meta verilerin ve sıkıştırma eserlerinin incelenmesi de video dosyalarında GAN tarafından oluşturulan içeriğin tespit edilmesinde önemli bir rol oynamaktadır.



GAN videolarını ve görüntülerini üretici ve ayırt edici ağlar aracılığıyla otomatik olarak tespit etmeyi amaçlayan yeni popüler bir çalışma. Kaynak Farahanipad F, Rezaei M, Nasr MS, Kamangar F, Athitsos V. El Pozu Tahmini Problemi için GAN Tabanlı Veri Artırımı Üzerine Bir Araştırma. Teknolojiler. 2022; 10(2):43. https://doi.org/10.3390/technologies10020043

57 Wang, Can, Shangfei Wang ve Guang Liang. "Çekişmeli özellik öğrenme yoluyla kimlik ve poz açısından sağlam yüz ifadesi tanıma." İçinde 27. ACM uluslararası multimedya konferansı bildirileri, s. 238-246. 2019.

58 Lee, Ji-Yeoun. "Saarbruecken ses veritabanını kullanan akıllı bir patolojik ses algılama sistemi için derin öğrenme yöntemlerinin deneysel değerlendirmesi." Uygulamalı Bilimler 11, no. 15 (2021): 7149.

Sağlam ve doğru tespit için GAN sınıflandırma toplulukları kullanılır. Bu topluluk yöntemleri, her biri GAN'a özgü artefaktların farklı yönlerine odaklanan birden fazla GAN sınıflandırıcısını birleştirir. Ayrıca, gerçek veriler üzerinde eğitilen modellerin GAN örneklerine önceden maruz kalmadan görünmeyen GAN tarafından oluşturulmuş içeriği tanımlamasına olanak tanıyan sıfır atış GAN tespiti gibi teknikler, bu tespit yöntemlerinin karmaşıklığına katkıda bulunur. Deepfake tespitinin talepleri gerçek zamanlı yanıtlar gerektirmektedir, bu nedenle GAN tespit modelleri gerçek zamanlı çıkarım için optimize edilmiştir. Bu, medya içeriğinin hızlı bir şekilde analiz edilmesine ve böylece ortaya çıkan deepfake tehditlerine hızlı yanıtlar verilmesine olanak tanır. Kapsamlı tespit yetenekleri sağlamak için, görsel-işitsel içeriği ve meta verileri analiz edebilen çapraz modal GAN tespit teknikleri geliştirilmiştir. Video tabanlı deepfake'ler göz önüne alındığında, gelişmiş teknikler GAN tarafından oluşturulan dizileri tanımlamak için videolardaki kareler arasındaki zamansal tutarlılığa odaklanmaktadır.

Bu yaklaşım, GAN tespitinin kapsamını hareketsiz görüntülerin ötesine genişletmektedir. Doğruluk kontrol kuruluşları, GAN tespitinde bu gelişmiş yöntem ve teknikleri kullanarak deepfake tespit yeteneklerini önemli ölçüde geliştirmektedir. GAN tarafından üretilen içeriğin doğru bir şekilde tanımlanması, doğruluk kontrolörlerinin manipüle edilmiş medyayı hızlı bir şekilde tespit etmesini sağlayarak dezenformasyonun yayılmasını önler ve bilgi ekosisteminin bütünlüğünü korur. GAN tabanlı manipülasyon ve aldatmacanın yarattığı zorlukların üstesinden gelmeye devam etmek için araştırmacılar, derin öğrenme uzmanları ve doğruluk kontrolörleri arasında iş birliği şarttır.

Sonuç olarak, Deepfake Tespit Teknolojileri, deepfake'lerin giderek daha sofistike hale gelmesiyle birlikte artık doğruluk kontrolü ve bilgi doğrulamanın önemli bir parçası haline gelmiştir. Gelişmiş görüntü ve video analizi, yüz tanıma, dudak senkronizasyonu analizi ve davranış analizi kullanan bu teknolojiler, sentetik medyanın doğru bir şekilde tanımlanmasını sağlamaktadır. Kıyaslama veri kümelerinin ve çok modlu yaklaşımların kullanımı deepfake tespitinin etkinliğini daha da artırırken, açıklanabilirlik ve gerçek zamanlı tespit yetenekleri de kritik unsurlar olarak öne çıkıyor. Deepfake teknolojisi gelişmeye devam ettikçe, Deepfake Tespitinde devam eden araştırma ve ilerlemeler, bilginin bütünlüğünü korumak ve güvenilir ve güvenilir bir dijital ekosistemi teşvik etmek için çok önemli olacaktır.

TÜRKİYE'DE DOĞRULAMA KURULUŞLARI ve VERİ DOĞRULAMA

Türkiye'deki doğruluk kontrol ekosistemine ilişkin en geniş ve en kapsamlı genel değerlendirme daha önce EDAM tarafından yayımlanmıştır.[59] Bu bölüm, o kapsamlı raporla aynı düzeyde ayrıntıya girmeyecek, ancak bu rapordaki tartışmaları Türkiye'deki daha geniş doğruluk kontrol ortamına bağlayacaktır. Türkiye, farklı kitlelere hitap eden çok sayıda yazılı, görsel ve dijital yayın kuruluşu ile canlı ve dinamik bir medya ortamına sahiptir. Ancak bu çeşitlilik aynı zamanda yanlış bilgi ve propaganda için de bir zemin oluşturmaktadır. Siyasi kutuplaşma, sansür ve hükümetin medya üzerindeki kontrolü, doğru bilginin yayılmasını daha da karmaşık hale getirmiştir. Bu bağlamda, doğruluk kontrolü, gerçeği kurgudan ayırmak ve kamuya mal olmuş kişi ve kurumları açıklamalarından sorumlu tutmak için giderek daha önemli hale gelmiştir.

Türkiye'de doğruluk kontrolüne olan talep ilk olarak 2000'lerin sonundaki H1N1 virüsü salgını sırasında tıbbi sahtekarlıkların artması ve ardından internet penetrasyonunun artmasıyla tetiklenmiştir. Daha sonra, siyasi çalkantılar ve 2014'ten sonraki bir dizi seçim nedeniyle, ülkenin siyasi sistemi daha otoriter hale geldikçe siyasi doğrulamaya odaklanan ikinci bir doğruluk kontrolcü dalgası ortaya çıktı. Daha sonra, 2018'den sonra giderek hükümet karşıtı bir doğruluk kontrol ekosistemi olarak algıladıkları şeye karşı bir denge sağlamayı amaçlayan hükümet yanlısı üçüncü bir doğruluk kontrol dalgası devreye girdi. Sadece iki Türk doğruluk kontrol platformu, Doğruluk Payı ve Teyit.org, şeffaflık ve editoryal kurallara uymalarını ve düzenli dış denetimlerden geçmelerini gerektiren Uluslararası Doğruluk Kontrol Ağı'nın (IFCN) üyesidir.

Evrim Ağacı ve Yalan Savar gibi diğer gruplar da IFCN kriterlerinin çoğunu karşılamaktadır ancak henüz üye değildir. Doğruluk kontrol platformlarının yükselişine rağmen, halkın önemli bir kısmı hala haber almak için interneti kullanmıyor veya internette karşılaştıkları iddiaları kontrol etmiyor. Bunu yapanlar ise genellikle arkadaşlarına ve ailelerine danışmak veya diğer haber kaynaklarıyla çapraz kontrol yapmak gibi geleneksel doğrulama yöntemlerine güvenmektedir. Ancak, doğruluk kontrolü kavramı Türkiye için hala nispeten yeni olduğundan bu durum değişebilir. Doğruluk kontrolü ile dezenformasyon arasındaki ilişki kesin olmamakla birlikte, bazı kanıtlar doğruluk kontrolünün yüksek profilli dezenformasyon vakalarına karşı koyabileceğini göstermektedir.

Doğruluk Payı, Evrim Ağacı ve Teyit.org gibi platformlar, daha karmaşık ve zamana duyarlı dezenformasyon biçimleriyle mücadele etseler de, doğruluk kontrol yükünün çoğunu omuzlamaktadır. Bununla birlikte, Türkiye'deki doğruluk kontrol ekosistemi iki umut verici eğilim göstermektedir. Birincisi, doğruluk kontrol kuruluşları, olumsuz koşullar altında bile hayatta kalabileceklerini, yenilik yapabileceklerini ve halka hizmet edebileceklerini kanıtlayarak objektif bilgi için bir talep yaratmışlardır. İkincisi, siyasi ve yasal baskılara rağmen başarılı ve siyasi olarak sürdürülebilir

59 Ünver, Hamid Akın, Fact-Checkers and Fact-Checking In Turkey (2020). EDAM Araştırma Raporları, 2020, SSRN: https://ssrn.com/abstract=3644285 adresinden erişilebilir.

olmuşlar, özellikle demokratik gerileme yaşayan ülkelerde diğerlerine ilham kaynağı olabilecek yeni doğrulama protokolleri ve katılım modelleri yaratmışlardır. Bununla birlikte, hayatta kalmak için uluslararası finansmana bağımlılık, bir platformun dezenformasyona etkili bir şekilde karşı koyma yeteneğini doğrudan etkilediği için bir sorun olmaya devam etmektedir. Bu nedenle, Türkiye'deki doğruluk kontrol ekosisteminin gelecekteki seyri uluslararası gözlemciler için önemini korumaktadır ve bu platformların ilerlemeleri, başarıları ve başarısızlıkları karşılaştırmalı siyasal iletişim, iletişim sosyolojisi ve teknoloji-enformasyon gibi alanlarla ilgili olmaya devam edecektir.

Türkiye'de bilgi manipülasyonu genellikle siyaset, din, sağlık ve ulusal güvenlik gibi hassas konular etrafında dönmektedir. Yanlış anlatılar ve söylentiler sosyal medya platformlarında hızla yayılarak toplumdaki bölünmeleri daha da derinleştirebilmekte ve halkın medyaya olan güvenini sarsabilmektedir. Ayrıca, seçim dönemlerinde yanlış bilgilerin yayılması seçmen algılarını ve demokratik süreçleri önemli ölçüde etkileyebilir. Türkiye'deki fact-checking kuruluşları, kanıta dayalı analizler sunarak ve yanlış iddiaları çürüterek bu zorlukların ele alınmasında kritik bir rol oynamaktadır. Vatandaşları doğru bilgilerle donatmayı ve eleştirel düşünmeyi teşvik etmeyi, böylece daha bilgili ve katılımcı bir toplumu teşvik etmeyi amaçlamaktadırlar.

Teyit.org, 2016 yılında kurulan Türkiye'nin öncü doğruluk kontrol kuruluşlarından biridir. Türkçe'de "doğrulama" anlamına gelen "teyit" kelimesi, kuruluşun temel misyonunu yansıtmaktadır. Teyit.org'un gazeteci ve araştırmacılardan oluşan ekibi, medyada ve sosyal platformlarda dolaşan yanlış bilgileri, söylentileri ve yanıltıcı iddiaları çürütmeye kendini adamıştır. Metodoloji: Teyit.org, doğruluk kontrolü için kamusal etki ve güvenilirliğe dayalı iddia seçimini içeren sistematik bir yaklaşım izlemektedir. İddiaları desteklemek veya çürütmek için kaynakları doğrular ve ilgili kanıtları toplar. Kuruluş, ifadelerin doğruluğunu kategorize etmek için bir derecelendirme sistemi kullanır ve bulgularının açık ve kolay anlaşılır özetlerini sağlar. Doğruluk kontrol raporları şeffaftır, metodolojileri açıklar ve sonuçlar için kanıt sağlar. Teyit.org, siyasi açıklamalar ve viral sosyal medya paylaşımlarından sağlıkla ilgili iddialara ve bilimsel yanlış anlamalara kadar çok çeşitli konuları ele almaktadır. Yanlış bilgi kalıplarını belirlemek ve dezenformasyon kampanyalarını tespit etmek için açık kaynaklı araçlar ve veri analizi teknikleri kullanmaktadır. Teyit.org'un kayda değer vakalarından biri, pandemi sırasında COVID-19 ile ilgili yanlış bilgilerin çürütülmesini içeriyordu. Kuruluş, potansiyel tedaviler, önleyici tedbirler ve virüsün kökenleri hakkındaki iddiaları kontrol etti. Teyit.org, doğru bilgiler sağlayarak kritik bir dönemde halk sağlığı bilincine ve yanlış bilgilerle mücadeleye katkıda bulundu.

Doğruluk Payı: 2014 yılında kurulan ve "Doğruluk Payı" anlamına gelen Doğruluk Payı, Türkiye'nin bir diğer önde gelen doğruluk kontrol kuruluşudur. Ekibi, politikacıların, kamuya mal olmuş kişilerin ve medya kuruluşlarının açıklamalarını doğrulamaya

adanmıştır. Metodoloji: Doğruluk Payı'nın doğruluk kontrol süreci, titiz kaynak doğrulama ve kanıt toplamayı içerir. Doğruluğu sağlamak için karmaşık vakalarda konu uzmanlarına danışırlar. Kuruluş, ifadeleri doğruluklarına göre kategorize etmek için "Doğru" ile "Yanlış" arasında bir derecelendirme sistemi kullanmaktadır. Doğruluk Payı, özellikle seçimler sırasında siyasi iddiaların doğruluğunun kontrol edilmesinde aktif olarak yer almıştır. Çeşitli siyasi aktörlerden gelen baskılara rağmen tarafsızlığını ve bağımsızlığını korumaya çalışmaktadır. Doğruluk Payı'nın ele aldığı özel bir vaka, ulusal bir seçim sırasında rakip siyasi partiler tarafından ortaya atılan iddiaların doğruluğunu kontrol etmeyi içeriyordu. Kuruluş, kampanya vaatlerini ve politika beyanlarını titizlikle analiz ederek seçmenlere bu iddiaların doğruluğuna ilişkin tarafsız bir değerlendirme sunmuştur. Böylece Doğruluk Payı, seçmenlerin güvenilir bilgiye dayalı bilinçli kararlar vermelerini sağlamıştır.

Hem Teyit.org hem de Doğruluk Payı, doğruluk kontrol kuruluşlarının küresel bir ittifakı olan Uluslararası Doğruluk Kontrol Ağı'nın (IFCN) imzacılarıdır. Poynter Enstitüsü'nün bir projesi olan IFCN, doğruluk kontrolünde mükemmelliği teşvik etmeyi ve dünya çapında doğruluk kontrolcüleri arasındaki iş birliğini güçlendirmeyi amaçlamaktadır. Teyit.org ve Doğruluk Payı, IFCN'ye üyelikleri sayesinde kaynaklara, en iyi uygulamalara ve uluslararası bir doğruluk kontrolcüleri topluluğunun desteğine erişim kazanmaktadır. Bu iş birliği, güvenilirliklerini artırmakta ve doğruluk kontrol metodolojilerinin küresel standartlarla uyumlu olmasını sağlamaktadır. Ayrıca, IFCN'nin Facebook ile ortaklığı, doğruluk kontrolörlerinin sosyal medya platformundaki potansiyel olarak yanlış bilgileri incelemesini ve çürütmesini ve böylece daha geniş bir kitleye ulaşmasını sağlar. IFCN, düzenli sanal toplantılar, web seminerleri ve çalıştaylar aracılığıyla üye kuruluşları arasında iş birliğini teşvik etmektedir. Farklı ülkelerden gelen doğruluk kontrol uzmanları bilgi ve deneyimlerini paylaşmakta, yanlış bilgi konusunda ortaya çıkan eğilimleri tartışmakta ve yenilikçi doğruluk kontrol araçlarını ve teknolojilerini keşfetmektedir.

Fact-Checking Europe, aralarında Türkiye'nin de bulunduğu farklı Avrupa ülkelerinden çeşitli doğruluk kontrol kuruluşlarının yer aldığı ortak bir projedir. Teyit.org bu ağın aktif bir katılımcısıdır ve sınır ötesi doğruluk kontrolü çabalarına ve bilgi paylaşımına katkıda bulunmaktadır. İşbirliğine Dayalı Girişimler: Fact-Checking Europe'un girişimleri arasında, seçimler ve önemli etkinlikler sırasında birden fazla ülkeden doğruluk kontrol uzmanlarının iddiaları doğrulamak ve yanlış anlatılara karşı koymak için birlikte çalıştığı ortak doğruluk kontrol projeleri yer almaktadır. Teyit.org, diğer Avrupa ülkelerinden ortaklarla iş birliği yaparak doğruluk kontrol kapasitesini güçlendirmekte ve erişimini daha geniş bir kitleye yaymaktadır. Bu iş birliği aynı zamanda doğruluk kontrol kuruluşları arasında veri ve araştırma bulgularının paylaşımını da kolaylaştırmaktadır. Üye kuruluşlar birlikte çalışarak birbirlerinin uzmanlıklarından faydalanabilir ve doğruluk kontrolü çabalarının genel etkinliğini artırabilirler.

Türkiye'de doğruluk kontrol kuruluşlarının karşılaştığı en önemli zorluklardan biri hükümet yetkililerinden gelen siyasi baskıdır. Hükümetin medya üzerindeki sıkı

kontrolü, doğruluk kontrolcülerinin güçlü figürler veya iktidar partisi tarafından ortaya atılan iddiaları çürütürken gözdağı veya tehditlerle karşılaşabileceği bir ortam yaratmıştır. Örneğin Teyit.org, seçim dönemlerinde üst düzey hükümet yetkilileri tarafından yapılan açıklamaları çürüttüğünde çok sayıda tepkiyle karşılaştı. Kuruluşun dezenformasyonu ifşa etmesi, yalanlanan yetkililerin destekçilerinin sosyal medyada Teyit.org'a karşı çeşitli karalama kampanyaları başlatmasına, kuruluşu önyargılı olmakla ve muhalefet partilerinin çıkarlarına hizmet etmekle suçlamasına yol açmıştır.

Türkçe Doğruluk Kontrolü Yapanların Kullanabileceği NLP Araçları:

Türkçe dilinde doğruluk kontrolünde NLP araştırmalarının sınırlarını oluşturan yeni Türkçe kütüphaneler ve sınıflandırıcılar bulunmaktadır.[60] Film Duygu Veri Seti,[61] İTÜ NLP Türkçe Duygu Analizi Veri Seti,[62] ve Türkçe Ürün Duygu Analizi Veri Seti,[63] dahil olmak üzere bu duygu, ruh hali ve sınıflandırıcı kütüphaneleri, doğruluk kontrolörlerinin kullanımına sunulan en yeni metin tabanlı araçlardan sadece birkaçıdır. Bu veri kümeleri, makine öğrenimi modellerinin Türkçe metinlerde ifade edilen duyguları daha iyi anlamasına ve yorumlamasına yardımcı olabilir. Ayrıca, en son geliştirilenlerden bazıları, BOUN Adlandırılmış Varlık Tanıma Veri Kümesi,[64] Türkçe Wikipedia Dökümü[65] ve Türkçe NLP kütüphaneleri için bir kıyaslama platformu olan 'Mukayese'[66] gibi dil modelleme ve metin sınıflandırma gibi görevler için yararlı olabilecek değerli Adlandırılmış Varlık Tanıma veri kümelerine sahiptir. NLP görevlerinde metin ön işleme için faydalı olabilecek kapsamlı ve yeni durak sözcük setleri de geliştirilmektedir. Genellikle düşük değerli kelimeler olarak kabul edilen durak kelimeler, NLP süreçlerinde genellikle filtrelenir. Bu havuz, farklı doğruluk kontrolü ve otomatik doğrulama biçimleri için faydalı olabilecek iki ayrı Türkçe durak sözcük listesi sunmaktadır.

Araçlar ve kütüphaneler açısından, Türkçe'de NLP görevlerini yerine getirmek için çeşitli kaynaklar bulunmaktadır. Örneğin 'Zemberek',[67], Türkçe için özel olarak geliştirilmiş açık kaynaklı bir NLP kütüphanesidir. Tokenleştirme, morfoloji, yazım denetimi ve daha fazlası gibi çok sayıda yardımcı program sağlar. 'Turkish Deasciifier'[68], sadece ASCII karakterleri kullanılarak yazılmış bir Türkçe metni, uygun Türkçe karakterleri kullanarak doğru yazılmış bir versiyona dönüştürebilen bir araç da bahsetmeye değer. Kelime Gömme ile ilgilenen doğruluk kontrolcüleri için 'Turkish Word2Vec'[69] Türkçe için önceden eğitilmiş kelime vektörleri sunmaktadır.

60 Türkçe NLP araçlarının sık güncellenen bir listesi Gökçe Merdun'un ve Turkish NLP Suite GitHub sayfasında bulunabilir: https://github.com/agmmnn/turkish-nlp-resources ve https://github.com/turkish-nlp-suite (buradan itibaren tüm bağlantılara en son 3 Ağustos 2023 tarihinde erişilmiştir)
61 Turkish Sentiment Analysis Dataset. http://humirapps.cs.hacettepe.edu.tr/tsad.aspx
62 ITU Türkçe Doğal Dil İşleme Yazılım Zinciri. http://tools.nlp.itu.edu.tr/api_usage.jsp
63 Vitamins and Supplements NER and Span Dataset. https://github.com/turkish-nlp-suite/Vitamins-Supplements-NER-dataset
64 UD Turkish BOUN. https://universaldependencies.org/treebanks/tr_boun/index.html
65 Turkish Wikipedia Dump. https://www.kaggle.com/datasets/mustfkeskin/turkish-wikipedia-dump
66 https://github.com/alisafaya/mukayese
67 https://github.com/Loodos/zemberek-python
68 https://github.com/emres/turkish-deasciifier
69 https://github.com/akoksal/Turkish-Word2Vec

```
text
```

```
"'Pek ala, Samara'da 6000 desyatin topragin var ve de 300 atın; e ne olmuş?' Bu soru bni
tamamen ele geçirdi ve başka ne düşüneceğimi bilemiyrdum. (Tolstoy)"
```

```
normalized_text = str(normalizer.normalize(JString(text)))
normalized_text
```

```
"' pek ala , samarada 6000 deşyatın toprağın var ve de 300 atın ; e ne olmuş ? ' bu soru
beni tamamen ele geçirdi ve başka ne düşüneceğimi bilemiyordum . ( tolstoy )"
```

```
punctuation_free = "".join([i for i in normalized_text if i not in string.punctuation])
punctuation_free
```

```
' pek ala  samarada 6000 deşyatın toprağın var ve de 300 atın  e ne olmuş  bu soru beni
tamamen ele geçirdi ve başka ne düşüneceğimi bilemiyordum  tolstoy '
```

```
digit_free = ''.join([i for i in punctuation_free if not i.isdigit()])
digit_free
```

```
' pek ala  samarada  deşyatın toprağın var ve de  atın  e ne olmuş  bu soru beni tamamen
ele geçirdi ve başka ne düşüneceğimi bilemiyordum  tolstoy '
```

Zemberek için Python kod örneği. Ekran görüntüsü: https://medium.com/technology-hits/turk-ish-text-preprocessing-with-zemberek-in-python-35930cc79afa Kaynak: Akın, Ahmet Afşin, ve Mehmet Dündar Akın. "Zemberek, Türk dilleri için açık kaynaklı bir NLP çerçevesi." Yapı 10, no. 2007 (2007): 1-5.

NLP araştırmaları genişlemeye devam ettikçe, bu tür kaynakların Türkçe dil işlemede yeni keşiflere ve ilerlemelere önemli ölçüde katkıda bulunma potansiyeli de artmaktadır.

SONUÇ

Dijital ortamın hızlı evrimi, doğru ve güvenilir bilginin önemini artırmıştır. Gerçekten de, yanlış bilgi ve dezenformasyonun yaygınlığı, toplumun tüm sektörlerinde yankı bulan ve küresel ölçekte siyasi, sosyal ve ekonomik istikrarı etkileyen destansı boyutlarda bir sorun haline gelmiştir. Teknolojideki gelişmeler, özellikle de yapay zeka ve makine öğrenimi, yanlış bilginin yayılmasına karşı mücadelede güçlü araçlar olarak ortaya çıkmıştır. Bu araçlar, büyük miktarda veriyi hızlı bir şekilde işleme, kalıpları ortaya çıkarma ve tahminlerde bulunma kabiliyeti sunmaktadır. Doğruluk kontrolü bağlamında, bu teknolojiler bilginin gerçekliğini ve güvenilirliğini doğrulamak için tasarlanmış araçlar ve sistemler için omurga sağlamaktadır.

Ancak, bu teknolojilerin gücü iki ucu keskin bir kılıçtır, çünkü aynı zamanda deepfakes gibi son derece sofistike ve ikna edici yanlış içerik oluşturmak için de kullanılırlar. Örneğin yapay sinir ağları, çok sayıda gelişmiş doğruluk kontrol modelinin temelini oluşturmaktadır. Bu modeller muazzam miktarda metinsel veriyi eleyebilir, iddiaları belirleyebilir ve doğruluklarını değerlendirebilir. Görsel medya dünyasında, konvolüsyonel sinir ağları (CNN'ler) ve tekrarlayan sinir ağları (RNN'ler) manipüle edilmiş içeriği tespit etmek için etkili bir şekilde kullanılmaktadır. Aynı zamanda, görüntü ve videolardaki kişilerin gerçekliğini doğrulamak ve taklitlere

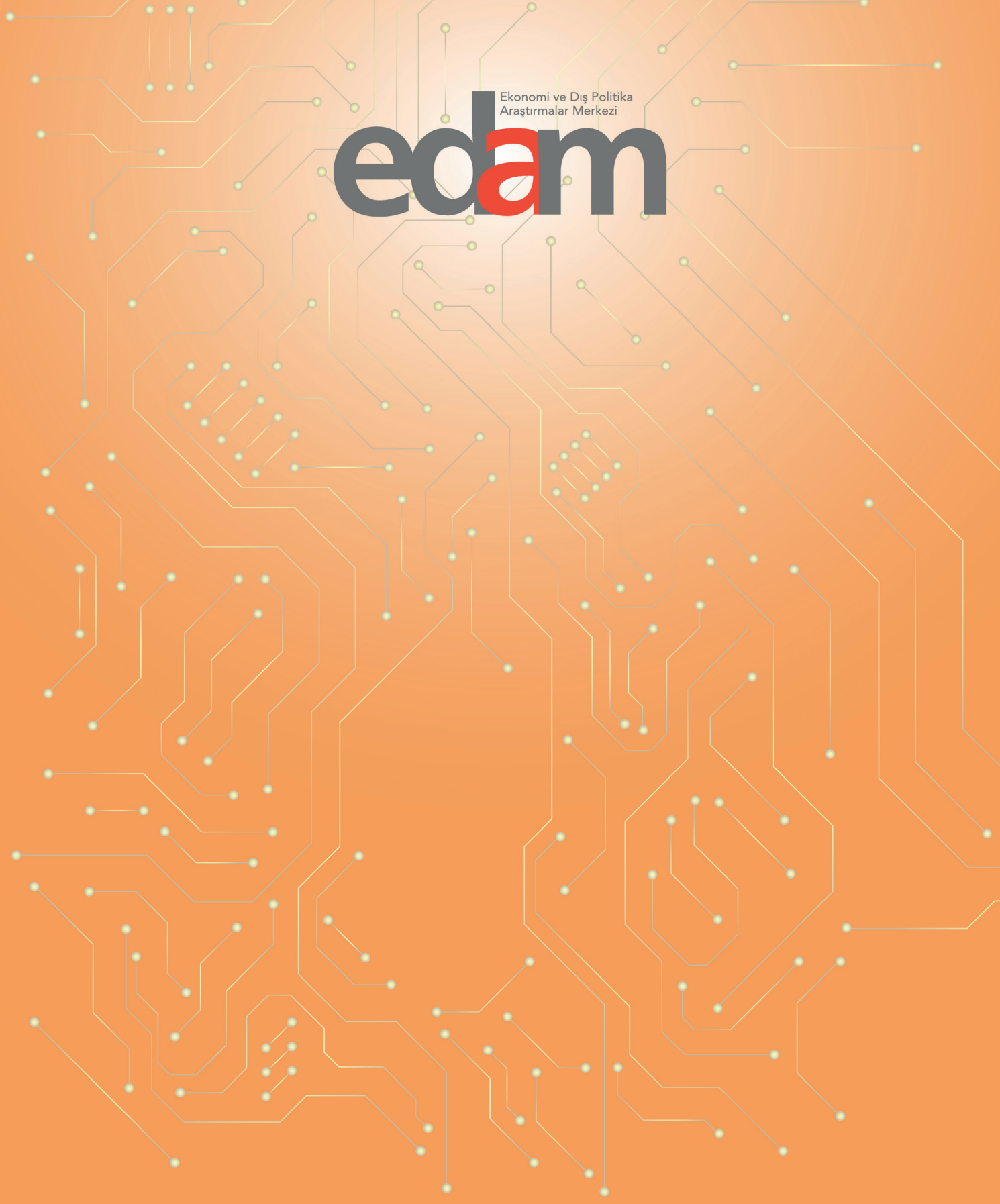
karşı koruma sağlamak için gelişmiş yüz ve ses kimlik doğrulama yöntemleri geliştirilmektedir. Paralel doğruluk kontrolü veri setleri ve çok dilli modeller, bu doğrulama tekniklerinin farklı dillere ve kültürel bağlamlara yayılmasını sağlayarak evrensel doğruluk kontrolü araçlarının geliştirilmesine olanak tanımıştır. Bu, Kanada'nın seçimler sırasında iki dilli doğruluk kontrolünü ele alışında görüldüğü gibi, çok dilli doğruluk kontrolüne yeni bir karmaşıklık düzeyi getirmiştir. Yapay zeka tarafından desteklenen gerçek zamanlı uyarılar ve erken uyarı sistemlerinin ortaya çıkışı da doğruluk kontrol girişimlerinin etkinliğini artırmıştır. Bu sistemler, potansiyel olarak yanıltıcı veya yanlış bilgileri anında işaretleyerek, doğruluk kontrolörlerinin hızlı bir şekilde yanıt vermesini ve yanlış bilgilerin yayılmasını engellemesini sağlar. Bu teknikler, son teknoloji ürünü medya adli tıp yöntemleriyle birlikte, bilgi doğrulamanın doğruluğunu ve hızını daha da geliştirerek, üzerinde oynanmış görüntüleri veya deepfake videoları hızlı bir şekilde çürütmeyi mümkün kılmıştır.

Kitle kaynak platformları ve halkın doğruluk kontrolüne katılımı da çok önemli bir rol oynamıştır. Teknolojinin yardımıyla bu platformlar, doğrulama sürecinde kolektif zekayı teşvik ederek geniş bir kullanıcı topluluğunun katılımını sağlayabilir. Avrupa çapındaki doğruluk kontrol platformu "FactCheckEU" bu işbirlikçi yaklaşımı örneklemektedir. Doğruluk kontrol kuruluşları iş birliğinin gücünün farkına varmış ve sosyal medya platformlarıyla yakın çalışmaya başlamıştır. Bu iş birliği, yanlış bilgilerin anında işaretlenmesine ve çürütülmesine olanak sağlayarak yayılmasını ve etkisini azaltmıştır. Facebook'un üçüncü taraf doğruluk kontrol kuruluşlarıyla ortaklığı ve Twitter'ın Birdwatch'ı bu stratejinin başlıca örnekleridir.

Yanlış bilginin jeopolitiği küçümsenemez. Dijital alan bilginin yayılması için birincil araç haline geldikçe, yanlış bilgi yayma becerisi silah haline geldi ve kamuoyunun manipüle edilmesine, demokratik süreçlere müdahale edilmesine ve hatta sosyal huzursuzluğun kışkırtılmasına yol açtı. Yanlış bilgi çatışmaları ateşleyip körükleyebilir, ulusları istikrarsızlaştırabilir ve jeopolitik güç dinamiklerini değiştirebilir. Kurumlara ve demokrasinin kendisine olan güveni sarsabilir. Doğruluk kontrolü ve bilgi doğrulama teknolojilerindeki ilerlemeler bu savaş alanında kilit öneme sahiptir. Bu teknolojiler doğru bilginin dolaşımını teşvik ederek demokrasinin korunmasına katkıda bulunur, çatışmaların kışkırtılmasını önler ve bilgi manipülasyonuna karşı toplumsal direnci artırır. Ancak bu teknolojilerin sürekli olarak iyileştirilmesi büyük önem taşımaktadır. Yanlış bilgilendirme teknikleri geliştikçe, bunlarla mücadele etmek için kullanılan teknolojiler de gelişmelidir. Bu da doğruluk kontrolü ve doğrulama teknolojilerinde sürekli araştırma, geliştirme ve yatırım yapılmasını gerektirmektedir.

Bilginin kötüye kullanımı ile etkin bir şekilde mücadele edebilmek için hükümetler, kuruluşlar ve teknoloji şirketleri yakın iş birliği içinde olmalı, kaynak ve bilgi paylaşımında bulunmalıdır. Ayrıca, bu teknolojilerin etik ve sorumlu bir şekilde kullanılmasını sağlayacak, bir yandan ifade özgürlüğünü korurken diğer yandan da olası kötüye kullanımlara karşı koruma sağlayacak düzenlemeler yapılmalıdır. Doğruluk kontrolü ve bilgi doğrulama teknolojilerinin geliştirilmesi ve yaygınlaştırılması şeffaf ve hesap verebilir olmalıdır. Sonuç olarak, doğruluk

kontrolü ve bilgi doğrulama teknolojilerindeki ilerlemeler jeopolitik üzerinde derin etkiler yaratmakta ve küresel ölçekte bilgi akışını ve güvenilirliğini etkilemektedir. Bu teknolojilerin hakikatin korunması, bilgilendirilmiş kamusal söylemin teşvik edilmesi ve demokrasinin korunmasındaki rolü büyük önem taşımaktadır.



Adres : Hare Sokak NO:16 AKATLAR 34335 İstanbul/Türkiye
Telefon : +90 212 352 18 54
Faks : +90 212 351 54 65
E-Posta : info@edam.org.tr